

*ENGINEERING DNA-BASED SELF-ASSEMBLY
SYSTEMS TO PRODUCE NANOSTRUCTURES AND
CHEMICAL PATTERNS*

by
John Zenk

Department of Chemical and Biomolecular Engineering
Johns Hopkins University

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the
degree of Doctor of Philosophy

Baltimore, Maryland

December 2016

© 2016 John Zenk
All Rights Reserved

Abstract

While commonly known as a material that stores biological information essential for life, few realize that deoxyribonucleic acid (DNA) is also a wonderful building (*i.e.* physical structures) and computing material. The field of DNA nanotechnology aims to use DNA primarily to build and control matter on the nanoscale. In 2006, a technique known as DNA origami was developed, which allows for the formation of about any shape on the nanoscale. Such DNA origami have been used in many applications: nanodevices, nanotubes, nanoreactors. However, the small surface area of the origami often limits its usefulness. One promising method for building large (micron-sized) DNA origami structures is to self-assemble multiple origami components into well-defined structures. To date, however, such structures suffer low yields, long reaction times and require experimental optimization with no guiding principles. One primary reason is that a governing theory and experimental measurements behind such a self-assembly process are lacking.

In this work, we develop coarse-grained computational simulations to describe and understand the self-assembly of finite-sized, multicomponent complexes (*e.g.*, nine different DNA-origami components that form a square grid complex). To help inform the model, we experimentally investigate how various interface architectures between two self-assembling DNA origami components affect the reaction kinetics and thermodynamics. We further develop the accuracy of our simulations by incorporating these measurements and other thermodynamic measurements from our group and implement a computational algorithm that optimizes the interaction strengths between self-assembling components for reaction efficiency (*i.e.*, speed and yield of the complex). With these experimentally-informed simulations, we suggest design improvements and provide yield predictions to an experimentally demonstrated tetrameric complex.

Finally, with the overarching idea of using DNA-based components to self-assemble to produce ordered structures and patterns, we build a reaction-diffusion system whose reactions are programmed using DNA strand displacement and diffusion which occurs in a hydrogel, wherein patterns develop, and liquid reservoirs, which are used to supply the high energy components. With this reaction-diffusion system we create stable (*i.e.*, unchanging in space and time) one and two-dimensional patterns of DNA molecules with millimeter-scale features.

Thesis Advisor

Dr. Rebecca Schulman

Thesis Committee Members

Dr. Michael A. Bevan

Dr. Jeffrey J. Gray

Dr. Kalina Hristova

Dr. Margaret E. Johnson

What I cannot create, I do not understand.

Richard Feynman

Acknowledgements

After the first week in graduate school, I knew exactly who I wanted to learn from for the next several years. Professor Rebecca Schulman, aside from her amazing productivity and academic endeavors (which speak for themselves), has the qualities I've always wanted in an advisor: super smart, passionate, driven and with a knack for cranking out cutting-edge scientific ideas, but also someone who is extraordinarily kind, relatable, motivational, and generous. Her great sense of humor is just a bonus. She helped stretch me intellectually beyond what I thought I was capable of. It goes without saying that without her, none of this would be possible for me. In 40 years, when I reflect on the choices I've made in my life, I am sure choosing Professor Schulman as my Ph.D. thesis advisor will rank as one of the best.

Next, I must say thank you to my lab mates. My lab mates remind me of a quote attributed to Albert Einstein: "Use for yourself little, but give to others much." My lab mates have given to me much. Namely, an amazing environment in which we could discuss research ideas and future goals, speculate about what's going to be the next big thing, debate the implications of Supplementary Figure 7, and decide where to play laser tag. Though I am not sure how little they've used for themselves, given that the long-time lab décor is multiple towering (and growing) yellow and orange plastic pillars made of only used pipette tip holders. These people are amazing: not only do they possess excellent academic rigor, a quality that one hopes to find in lab mates, but they also possess a wonderful sense of team mindedness and unity. I could not think of any better environment in which to work for several years. I very much appreciate all of the members of Schulman Lab with whom I've interacted: Deepak Agarwal, Phil Dorsey, Josh Fern, Tyler Jorgenson, Jen Lu, Seth Reinhart, Dominic Scalise, Sam Schaffter, and Ankur Verma. Thank you all. Angelo Cangialosi and Abdul Majeed Mohammed deserve a special thank you for being central to my Ph.D. experience and for helping me in lab and in life.

Furthermore, I'd like to recognize the amazing undergraduate students at Johns Hopkins University who worked with me over the years. I was truly amazed by how quickly these students grasped the concepts and protocols to become productive lab members and teammates. Chanon Tuntivate was an absolute joy to work with. His work-horse work ethic mixed with his innate curiosity led him to be an extremely productive experimentalist. Chanon often thought of clever solutions to complex problems when other people (including myself) could not. Kaiyuan Wang, driven by an intense desire to perform science, has a wonderful ability to dissect experiments and improve their protocols. Kai is always thinking about the next experiment, asking "what should we investigate next." Matthew Billups became the jack of all trades, performing experiments and computer simulations. I am grateful for his help on both types of projects.

I would be remiss if I didn't harken back to the origin of my interest in academic research. Thank you to Professor Christine Hrenya, my undergraduate mentor, and my graduate student mentor while in the Hrenya lab, Dr. Peter Mitrano who stoked my desire to pursue an advanced degree. I had such a marvelous time working within the group to uncover how granular materials behave. If scientific research could be so rewarding (and fun!), why wouldn't I want to spend the next handful of years doing the same?

Outside the scientific community, I must send an enormous and special THANK YOU to the 410 CrossFit community and especially its founders, Gary and Marcie Ostroski. 410 gave me an active, rich life outside of the lab. I was daily inspired and encouraged by all at 410, including, but not limited to: Ryan Carver, Connor Cook, Mike Dominelli, Jeff Nelson, Tanvi Shroff, and Nitesh Turaga. Doing burpees and squats every day provided perspective and gave me a way to constructively expel my frustration from bad simulations and failed experiments!

Next, I'd like to thank my family for supporting me in my pursuit of a Ph.D., with an especially huge thanks to my Mom, who has always supported my academic endeavors and first

encouraged me many years ago to pursue an engineering degree. My Dad and brother Brian also deserve a huge thank you for their constant love, support and good sense of humor.

Finally, I'd like to say a much deserved thank you to Katie Steefel. There are no words to describe how grateful I am for her. Without her constant love, support and encouragement, I know I wouldn't be where I am or who I am today.

Table of Contents

CHAPTER 1	Introduction	1
1.1	Contributions	9
CHAPTER 2	An Assembly Funnel Makes Biomolecular Complex Assembly Efficient	10
2.1	INTRODUCTION	10
2.2	RESULTS	12
2.3	DISCUSSION	22
2.4	METHODS	24
2.4.1	Stochastic kinetic simulations	24
2.4.2	Rate constants and physical parameters	24
2.5	SUPPORTING INFORMATION	25
2.5.1	Text S2.1 Supporting Methods	25
2.5.2	Text S2.2 Computational Specifics	31
2.5.3	Text S2.3 Assembly Regime Criteria	32
2.5.4	Text S2.4 Assembly Distribution Selection	34
2.5.5	Text S2.5 Further Explanation of Bond Coupling Effect on Yield (from Figure 2.5)	35
2.5.6	Text S2.6 Computing Thermodynamic Equilibrium of Large Complexes	37
2.5.7	Supporting Figures	38
2.6	ACKNOWLEDGEMENTS	51
CHAPTER 3	Kinetics and Thermodynamics of Watson-Crick Base Pairing-Driven DNA Origami Dimerization	52
3.1	INTRODUCTION	53
3.2	RESULTS AND DISCUSSION	55
3.3	CONCLUSION	69
3.4	MATERIALS AND METHODS	71
3.4.1	Self-assembly of DNA nanostructures	71
3.4.2	AFM yield measurements	71
3.4.3	Kinetics and Thermodynamics Measurements	71
3.4.4	2D Infinite Lattice Formation	72
3.5	SUPPORTING INFORMATION	73
3.5.1	Supporting Note 3.1: Nanostructure specifics	73
3.5.2	Supporting Note 3.2: Fluorescence Specifics	74
3.5.3	Supporting Note 3.3: Thermodynamic measurements	78
3.5.4	Supporting Note 3.4: Kinetic measurements	91
3.5.5	Supporting Note 3.5: Two-dimensional Lattice with 6bp Floppy Linkers	107
3.5.6	Supporting Note 3.6: Design of Origami Edge Staples and Linker Sequences	108
3.6	ACKNOWLEDGMENTS	109
CHAPTER 4	Optimizing Component-Component Interaction Energies in the Self-Assembly of Finite, Multicomponent Structures	110
4.1	INTRODUCTION	111
4.2	MODEL AND METHODS	113
4.2.1	Directed Evolution of Interfaces	116
4.2.2	Directed Evolution of Interfaces of a DNA Origami-Based Complex with Experimentally Measured Kinetic and Thermodynamic Parameters	118
4.2.3	Stoichiometric Imbalance	118
4.3	RESULTS AND DISCUSSION	119
4.4	CONCLUSION	129
4.5	SUPPORTING INFORMATION	131
4.5.1	Supporting Information Note 1: Directed Evolution of Interfaces Simulations	131
4.5.2	Supporting Information Note 2: Stoichiometric Imbalances Simulations	136

4.5.3	Supporting Information Note 3: Simulation Parameters from Experimental Measurements of the DNA origami-based complex.....	146
4.6	ACKNOWLEDGEMENTS.....	153
CHAPTER 5	Stable DNA-based Reaction-Diffusion Patterns.....	154
5.1	INTRODUCTION.....	154
5.2	RESULTS AND DISCUSSION.....	156
5.3	CONCLUSION.....	172
5.4	MATERIALS AND METHODS.....	173
5.4.1	DNA complex preparation.....	173
5.4.2	Hydrogel preparation.....	174
5.4.3	Reaction Monitoring.....	175
5.4.4	Liquid Reservoir Exchange.....	175
5.4.5	Simulations.....	176
5.5	SUPPORTING INFORMATION.....	177
5.5.1	Supplementary Note S1: Supplementary Data.....	177
5.5.2	Supplementary Note S2: Image Processing.....	183
5.5.3	Supplementary Note S3: Experimental Notes.....	184
5.5.4	Supplementary Note S4: Leak Reactions and Recapture Kinetics.....	186
5.5.5	Supplementary Note S5: Reaction-Diffusion Device Fabrication.....	190
5.5.6	Supplementary Note S6: DNA Sequences.....	191
5.6	ACKNOWLEDGEMENTS.....	193
CHAPTER 6	Conclusions and Future Directions.....	194
	References.....	196
	Curriculum Vitae.....	213

List of Figures

Figure 1.1.	Sequencing and synthesis costs are decreasing over time.....	3
Figure 2.1.	Self-assembly model for a 4-component square grid (“2x2”) complex.....	13
Figure 2.2.	Thermodynamic equilibrium is a good predictor of yield for isothermal assembly after long assembly times for 1-dimensional complexes, but not 2- or 3-dimensional complexes.	15
Figure 2.3.	An assembly funnel means that complex assembly occurs via a small number of pathways.....	17
Figure 2.4.	Complexes form rapidly in the assembly funnel regime.....	19
Figure 2.5.	The amount of bond coupling, or additivity of bond energies during cooperative binding steps does not significantly affect assembly yields above a small threshold.....	20
Figure 2.6.	Design of components so that particular assembly pathways are favored can drastically increase assembly yields.	21
Figure 3.1.	Schematic of the DNA origami tile dimer system.....	56
Figure 3.2.	Measurement of dimerization yield and thermodynamic parameters.	60
Figure 3.3.	Reaction rate constants for multivalent DNA origami dimerization.....	66
Figure 3.4.	Two-dimensional origami tile lattice with two tiles labeled “a” and “b” designed with four 6bp SEs per interface.....	68
Figure 4.1.	Simulated multicomponent self-assembly and component optimization shown with a 2x2 grid system.	115
Figure 4.2.	A mixture of strong and weak interfaces between components can enhance the fitness of 2D and 3D structures.	120

Figure 4.3. Structures with high fitness assemble via annealing and produce high yield.	123
Figure 4.4. Fitness with variable stoichiometry (σ) as a function of fitness with no stoichiometric imbalance for 3x3 square grid system.....	125
Figure 4.5. Optimizing the self-assembly system where four different DNA-origami components assemble into a tetramer complex.	127
Figure 5.1. Schematic of DNA strand displacement reactions and the setup of reaction diffusion systems.	158
Figure 5.2. Diffusion coefficient measurement for ssDNA and dsDNA in 1% agarose.	162
Figure 5.3. Designed processes of pattern generation in a RD system using DNA strand displacement reactions.	165
Figure 5.4. Stable two-dimensional gradient.	167
Figure 5.5. Two stable gradients of Output species in one hydrogel with two non-interacting (orthogonal) sets of reactants.	170

List of Supplementary Figures

Figure S 2.1 Valid and invalid species in the 2x2 grid complex.	26
Figure S 2.2 Valid and invalid reactions for the 2x2 grid complex.....	27
Figure S 2.3 Yields of 1x3 to 1x9 line complexes at various isothermal conditions.....	38
Figure S 2.4 Yield of 1x9 line complex at various reaction times, τ , subject to different isothermal assembly conditions.	39
Figure S 2.5 Assembly size distribution at different isothermal assembly conditions after $\tau = 1000$	39
Figure S 2.6 Yields of 2x2, 3x3 and 4x4 square grid complexes at different isothermal assembly conditions and bond coupling constants (a_0).....	40
Figure S 2.7 Yields of 2x2, 3x3 and 4x4 square grid complexes as a function of bond coupling constant, a_0 , at various isothermal conditions (solid lines) and anneal (dash-dot line).....	41
Figure S 2.8 Yield of 3x3 square grid complex for many isothermal conditions, from $\eta = -6$ to $\eta = 6$ in increments $\eta = 0.2$	41
Figure S 2.9 Reducing the number of components in the simulation does not significantly affect yield predictions.	42
Figure S 2.10 Yield for 2x2 and 3x3 square grid complexes at various isothermal conditions, including yield predictions after long reaction times, $\tau = 10000$	42
Figure S 2.11 Yields of 2x2, 3x3, 4x4 and 5x5 square grid complexes at various reaction times, τ , subject to different assembly conditions.	43
Figure S 2.12 Yield of 2x2, 3x3 and 4x4 spiral complexes at various reaction times, τ , subject to different assembly conditions.	43
Figure S 2.13 Assembly size distributions (in # of components) for 2x2, 3x3 and 4x4 square grid complexes at various isothermal conditions and bond coupling constants.	44
Figure S 2.14 Timescales of nucleation and rearrangement together determine the rate of complex formation.	44
Figure S 2.15 Size distribution of intermediates for various 2D complexes.	45
Figure S 2.16 During an anneal, most complexes are produced during the phase of the anneal that passes through the assembly funnel.	46
Figure S 2.17 During very long anneals, component depletion can increase the amount of time that the system effectively stays within the assembly funnel regime.	47
Figure S 2.19 Deterministic and stochastic solutions are almost identical.....	49

Figure S 2.20 Yields of 2x2x2 cube complexes as a function of bond coupling constants a_0 at various isothermal conditions.	49
Figure S 2.21 Yield of 2x2x2 cube complexes as a function of bond coupling constant, a_0 at various isothermal conditions (in terms of η).	50
Figure S 2.22 Yield of 2x2x2 cube complex at various reaction times, τ , subject to different isothermal assembly conditions.	50
Figure S 2.23 Assembly size distributions for 2x2x2 cube complex at various isothermal conditions and bond coupling constants.	51
Figure S 3.1 CanDO and CaDNAo models of DNA origami tiles used in this work.	73
Figure S 3.2 DNA origami components melt above 55°C.	74
Figure S 3.3 T1 – T2 binding does not reduce fluorophore activity when no quencher is present.	76
Figure S 3.4 Free linkers in solution (unbound to origami) do not bind to their complementary linker for 5bp SEs.	77
Figure S 3.5 Free 6bp linkers in solution (unbound to origami) can bind to their complementary linker for linkers with 6bp SEs.	78
Figure S 3.6 Fluorescence quenching data as a proxy for AFM yield.	80
Figure S 3.7 Thermodynamic measurements for interfaces with 5bp SE linkers.	81
Figure S 3.8 Thermodynamic measurements for interfaces with 5bp SE linkers (continued).	82
Figure S 3.9 Thermodynamic measurements for interfaces with 6bp SE linkers.	83
Figure S 3.10 Thermodynamic measurements for interfaces with 6bp SE linkers (continued). ...	84
Figure S 3.11 Thermodynamic measurements for interfaces with 6bp SE floppy linkers.	85
Figure S 3.12 Thermodynamic measurements for interfaces with 6bp SE floppy linkers (continued).	86
Figure S 3.13 Dimerization achieves equilibrium in 15 minutes for 4 linkers with 5bp SEs.	87
Figure S 3.14 Dimerization achieves equilibrium in 60 minutes for 6 linkers with 6bp SEs.	87
Figure S 3.15 Measured values of standard Gibbs free energy at 298K, ΔG_{2980} , for all interface designs.	90
Figure S 3.16 Measured values of standard enthalpy, $-\Delta H_{2980}$, and entropy $-298 \cdot \Delta S_{2980}$ for all interfaces.	90
Figure S 3.17 Kinetics measurements: forward reaction rate constants for interfaces with 5bp SE linkers.	93
Figure S 3.18 Kinetics measurements: forward reaction rate constants for interfaces with 5bp SE linkers.	94
Figure S 3.19 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE linkers.	95
Figure S 3.20 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE linkers.	96
Figure S 3.21 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE floppy linkers.	97
Figure S 3.22 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE floppy linkers (continued).	98
Figure S 3.23 Kinetics measurements: reverse reaction rate constants for interfaces with 5bp SE linkers.	99
Figure S 3.24 Kinetics measurements: reverse reaction rate constants for interfaces with 5bp SE linkers (continued).	100
Figure S 3.25 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE linkers.	101

Figure S 3.26 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE linkers (continued).	102
Figure S 3.27 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE floppy linkers.	103
Figure S 3.28 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE floppy linkers (continued).	104
Figure S 3.29 Equilibrium constants (K_{eq}) of dimerization.	105
Figure S 3.30 Less than ~5% of the edge linkers dissociate from origami, even at high temperatures.	106
Figure S 3.31 Two-dimensional origami tile lattice with two tiles labeled “a” and “b” designed with four 6bp floppy SEs per interface.	107
Figure S 3.32 Designed strength of SE, ΔG_{2980} , as calculated by the nearest neighbor model	109
Figure S 4.1 Efficient assembly is achieved through strong interaction strengths in 1D assembly.	131
Figure S 4.2 Directed evolution of interfaces in simulation increases the fitness of all structures.	131
Figure S 4.3 The rate of structure formation of optimized structures is faster than or just as fast as the rate of formation of initial structures.	132
Figure S 4.4 Multiple different optimized structures can be found.	133
Figure S 4.5 On a population scale, structure formation rates higher, especially at short assembly times, or just as high via isothermal assembly versus via annealing at ideal isothermal conditions.	134
Figure S 4.6 Structure formation rates higher, especially at short assembly times, or just as high via isothermal assembly versus via annealing at ideal isothermal conditions.	135
Figure S 4.7 Stoichiometric imbalances limit yield in the 1x5 line system.	136
Figure S 4.8 Stoichiometric imbalances limit yield for a fixed assembly time in the 1x5 line system.	137
Figure S 4.9 Stoichiometric imbalances limit yield in the 2x2 grid system.	138
Figure S 4.10 Stoichiometric imbalances limit yield for a fixed assembly time in the 2x2 grid system.	139
Figure S 4.11 Stoichiometric imbalances limit yield in the 3x3 grid system.	140
Figure S 4.12 Stoichiometric imbalances limit yield for a fixed assembly time in the 3x3 grid system.	141
Figure S 4.13 Stoichiometric imbalances limit yield in the 4x4 grid system.	142
Figure S 4.14 Stoichiometric imbalances limit yield for a fixed assembly time in the 4x4 grid system.	143
Figure S 4.15 Stoichiometric imbalances limit yield in the 2x2x2 cube system.	144
Figure S 4.16 Stoichiometric imbalances limit yield for a fixed assembly time in the 2x2x2 cube system.	145
Figure S 4.17 Optimizing the self-assembly system where nine different DNA-origami components assemble into a complex.	153
Figure S 5.1 Simulations of reaction and diffusion processes with different rate constants for the release reaction.	177
Figure S 5.2 Simulated reaction and diffusion of the release-recapture processes in 1D.	178
Figure S 5.3 A linear gradient forms when Output concentration is initially high within the hydrogel substrate.	179
Figure S 5.4 Stable gradients form consistently.	180
Figure S 5.5 Growing gradient in two-dimensions with large dimensions (24 x 24 x 8 mm).	181

Figure S 5.6 Growing gradient in two-dimensions with small dimensions (16 x 16 x 8 mm)....	182
Figure S 5.7 Dark frame correction to reduce edge effects in image montages.....	183
Figure S 5.8 Calibrating counts from calibration wells to concentration of Output species.	184
Figure S 5.9 Evaporation in calibration wells mitigated by using Scotch Tape.....	185
Figure S 5.10 Potential leak reactions considered in the model.	188

List of Supplementary Tables

Table S 2.1 Complex specifics and parameter space explored in this work.....	31
Table S 2.2 Criteria for labeling assembly regimes.....	33
Table S 3.1 Measured Thermodynamic Parameters for 5bp SE linkers (kcal/mol)*	88
Table S 3.2 Measured Thermodynamic Parameters for 6bp SE linkers (kcal/mol)	88
Table S 3.3 Measured Thermodynamic Parameters for 6bp floppy SE linkers (kcal/mol).....	89
Table S 3.4 Concentration and Annealing Protocol for Infinite Lattice*	108
Table S 4.1 Thermodynamic parameters of Interfaces from Ref. ¹¹⁵ taken at 25°C.....	147
Table S 4.2 Forward Reaction Rate Constants, k_{on} ($\text{M}^{-1}\text{s}^{-1}$), taken from <i>Chapter 3</i> (Ref. ¹⁴⁴).....	149
Table S 4.3 Reverse Reaction Rate Constants, k_{off} (s^{-1}), taken from <i>Chapter 3</i> (Ref. ¹⁴⁴).....	150
Table S 5.1 List of reactions modeled in RD system.....	189

CHAPTER 1 INTRODUCTION

Popular fascination with deoxyribose nucleic acid (DNA) did not truly begin until James Watson and Francis Crick, using Rosalind Franklin's X-ray crystallography data, discovered the three-dimensional structure of DNA. Now known as B-DNA, the structure consists of a double stranded helix with a major and minor groove whose constituent two strands run anti-parallel to one another and are "held" together by base-pairing between specific molecules within the two strands that pair with their complement molecule based on the number of hydrogen bonds that can form. All of these details the authors describe in their 1953 *Nature* article, which contained a mere six references, one figure, and less than 1000 words¹. In their seminal paper, they also correctly predict that Adenosine hydrogen bonds to Thymine (A-T) while Guanine hydrogen bonds with Cytosine (G-C); hence these binding pairs are now known as Watson-Crick complements. Indeed, Watson and Crick even correctly predicted the distance that separates adjacent base pairs in a B-form double-stranded DNA (dsDNA), 3.4 angstroms, which is a fundamental quantity commonly used in the DNA nanotechnology field (the field of this thesis) today.

Fast forward thirty years. While the vast majority of DNA-centered studies and applications up to this point were biological in nature, in the early 1980s, a researcher named Nadrian Seeman started thinking about DNA differently. He conceived of and created an entirely new field by using DNA for structural purposes, instead of biological, by exploiting DNA's Watson-Crick complementarity and the fact that double stranded DNA (dsDNA) is much stiffer than single stranded DNA (ssDNA). In 1982 and '83, he described building junctions and lattices from many ssDNA strands that could partially hybridize with other ssDNA strands, such that the equilibrium state of the mixture would form a well-defined structure^{2,3}. Biological systems exhibit such structural organization when they form Holiday Junctions, he noted. Such lattice structures

could be useful to hold a protein in place to more easily determine the protein's crystal structure using x-ray crystallography, for example. Shortly thereafter his group experimentally demonstrated the formation of a 4-way junction from four ssDNA strands, marking the first human-engineered DNA-based nanoscale structure⁴. The field of structural DNA nanotechnology was born.

Fast forward another 30 years. The study and engineering of DNA-based systems, both biological and non-biological, has taken massive leaps as new techniques and tools have made the technology increasingly accessible, which has been in large part driven by the healthcare implications of DNA. Costs over time of DNA sequencing and synthesis follow highly nonlinear trends, similar to Moore's law in the semiconductor industry: Intel co-founder Gordon Moore accurately predicted that transistor density on an integrated chip would double approximately every two years for at least a decade⁵, which has held true for the past four decades. Similarly, costs of synthesis for short DNA oligonucleotides have decreased more than an of magnitude in the past two decades while during that time DNA sequencing has decreased by more than six orders of magnitude (Figure 1.1)⁶. In fact, the trend of sequencing costs is notably more nonlinear than even Moore's law, with approximate cost per megabase (1 million bases) sequencing decreasing by a factor of more than 10^5 over the course of ~13 years. While the initial sequencing effort was fueled by private and government-backed multibillion dollar investments to decode the human genome, DNA sequencing is now a massive business success, with sequencing companies like Illumina, Inc., less than 20 years old, boasting over \$25 billion in market capitalization (at the time of this writing). Such increasing accessibility has fueled innovation of DNA.

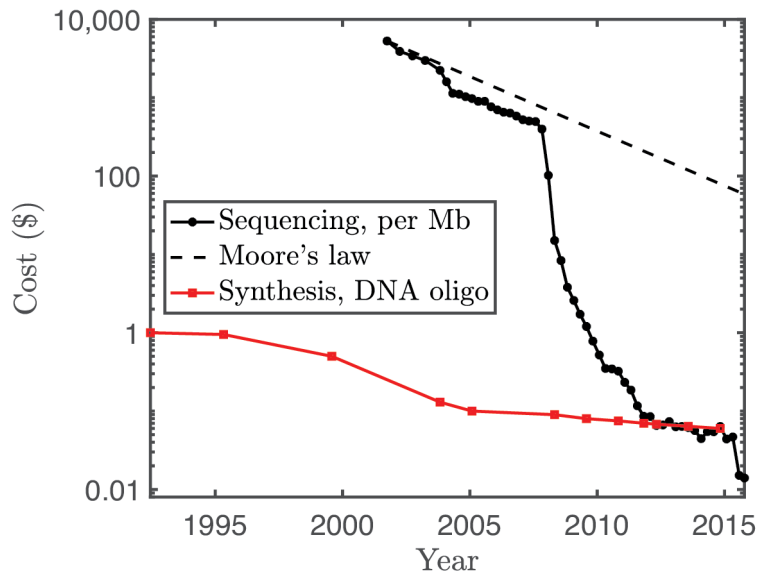


Figure 1.1. Sequencing and synthesis costs are decreasing over time. The cost to sequence a megabase (Mb), or one million bases, of DNA has non-linearly decreased over time, deviating from even the steeply declining Moore's law, which serves as a common benchmark for rapidly advancing technologies and depicts the cost per Mb halving every two years. Synthesis of short DNA oligo costs have decreased over time, although not as significantly as sequencing costs. Sequencing data taken from Ref. ⁶ and synthesis data taken from Refs. ^{7,8}.

One potentially high-value non-biological application for DNA is as a material for information storage, where information is “written” by synthesizing and “read” by sequencing ⁹. The theoretical maximum information density in DNA is extreme: 455 exabytes, or about 455 million terabytes per gram of ssDNA ⁹ (in comparison, as of this writing the entirety of Facebook's datacenters house ~400,000 terabytes of data), which is likely the reason why biological systems have evolved to store their information into DNA sequences. One more such non-biological application, whose origins predate the prior example and forms the foundation of this thesis involves using DNA as a material in order to create nanoscale structures and patterns from the bottom-up with specific shapes and functions ^{2,4}.

Along with decreasing costs of DNA synthesis came increasing complexity of DNA-based nanostructures. A major inflection point for the field of structural DNA nanotechnology occurred in 2006 ¹⁰, when a technique known as DNA origami, which allows for the creation of nanoscale

structures with almost any shape and is a technique discussed in *Chapters 2-4* and used in *Chapter 3*, was introduced. In this technique, a long, circular scaffold strand (7249 bases) is “folded” by a molar excess of many shorter (~30 bases long) “staple” ssDNA strands, where the staple sequences and crossover patterns determine the final shape of the DNA origami. Aside from the ability to be folded into a variety of nanoscale shapes, the main quality that differentiates DNA origami from other nanoscale fabrication techniques is that an origami surface can serve as a high-resolution nanobreadboard: a wide variety of biomolecular components (*e.g.*, aptamers, proteins, gold nanoparticles) can be precisely placed within about 6 nanometers of one another. Because of these unique characteristics, DNA origami enabled the creation of many nanodevices: nanoscale reactors^{11, 12, 13}, and 3D structures with various logic functions^{14, 15, 16}, biophysical measurement tools¹⁷, molecular robots^{18, 19}, actuators and sensors^{20, 21, 22}, devices to control self-assembly^{23, 24, 25}, novel drug delivery systems^{26, 27, 28}, and biologically-inspired structures^{29, 30, 31}, to name a few. Fabrication of origami is simple and robust: design³² and structure prediction software³³ allow for rapid prototyping of structures and optimized experimental protocols enable rapid (several minutes) folding³⁴. The time from design conception to final folded structure could take as little as one week.

While such DNA origami structures have been used to template the assembly of electronic devices, light harvesting structures, chemical assembly lines and templates for molecular walkers, their limited surface area is often a limiting factor for nanoengineers. For example, a common origami design is a two-dimensional, flat rectangular structure which has a surface area of ~6000 nm², enough for about a dozen molecular components. There has therefore been significant interest in building larger templates with the same ability to specifically position molecules with almost nanoscale resolution, but over larger surface area. Larger surface areas would enable a new generation of origami-based devices, for example, an artificial “leaf,” where components whose reaction cascades mimic photosynthesis³⁵. In fact, first steps to achieve such a goal have been

accomplished: studies have shown that significantly increased activity can be achieved by using DNA origami to decrease the physical distance between two enzymes that share an intermediate ¹¹ and that protein complexes can be assembled on an origami surface ³⁶. One route to this method is to hierarchically self-assemble individual nanostructures into a larger, multicomponent structure. This approach has proven difficult, and one fundamental challenge is that we know little about how molecular structures assemble specifically at the size scale of tens to a few hundred nanometers: prior to the studies conducted, described in this thesis, there were no kinetic and thermodynamic measurements for the self-assembly of structures with DNA-origami components.

Thus, most current methods for designing self-assembly processes to produce multi-component origami structures are of the “guess and check” variety, i.e., design a DNA origami component and its interface then determine if it self-assembles and if it does, determine if it will assemble with other components. These methods lack guiding principles and rules for which to build larger origami structures. Protein engineering, for example, has extensive experiment-informed computational tools such as Rosetta ³⁷, in order to help design protein assembly processes, which has led to the creation of protein-only 2D infinite lattices ³⁸ and finite complexes ³⁹. Although recent studies have been conducted to understand the kinetics and thermodynamics of smaller DNA-based structures, to our knowledge no systematic study has been performed on larger and more useful DNA origami components. Many nanostructures spawned from the field of DNA nanotechnology lack sophistication and scalability largely due to the lack of experimental backbone in the form of thermodynamic and kinetic information for origami components.

The final idea from the field of DNA nanotechnology that this thesis draws on is DNA strand displacement (DSD) reactions, which are commonly used in the subfield of DNA computing ^{40, 41}. The main idea behind DNA computing is that billions (or many more) of chemical reactions (computations) could occur in parallel, counteracting the slow speed at which each reaction

progresses relative to its silicon counterpart. Chemical computations are governed by specific rules, such as Watson-Crick complementary base pairing. Such chemical programs have been shown to perform various computations, such computing the square root of a four bit number⁴². The computations occur *via* DSD reactions which are be easily “programmed” into DNA by manipulating its sequence and the sequence of those molecules with which it will interact. In a simple DSD reaction, for example, one ssDNA molecule might displace another in a partially double-stranded complex (otherwise known as a DNA logic gate) because the ssDNA is fully complementary to a strand in the complex. Such a reaction is driven by the free energy decrease that arises from DNA hybridization. DSD has many advantages: specificity⁴³, scalability⁴² and kinetic tunability^{44,45}. To control the rates and equilibrium of strand displacement reactions, often a “toehold” is introduced to the complex. A toehold-mediated strand displacement reaction occurs when a short region of ssDNA (the “toehold” region) on a partially double-stranded DNA complex assists an input strand of DNA to bind to and displace an incumbent strand. The steps of such a reaction are typically designed to traverse the following: (1) hybridization of the input strand to the toehold, (2) branch migration of the input strand toward the incumbent strand and (3) complete dissociation of the incumbent strand. The forward reaction rate constant, k_{on} , can be tuned across about six orders of magnitude by changing the length and sequence of the toehold: each nucleotide added to the length of the toehold (up to ~7 bp) increases the k_{on} by about one order of magnitude⁴⁵. Branch migration involves a random walk *via* reversible association and dissociation steps of a single base pair of two strands that compete to bind to a shared complementary base on a third strand. Dissociation of the incumbent strand is usually the goal, which can then participate in downstream reactions as an input strand. Such reactions are used in dynamic DNA-based devices, from the first demonstration, DNA tweezers,⁴³ where toehold mediated DSD reactions were used to open and close the device. Software has been developed to design reactions and structures,

including secondary structure and compute the equilibrium state of the strands at various environmental conditions, such as salt concentrations and temperatures⁴⁶.

This thesis is centered around one general area of DNA nanotechnology: the design and implementation of DNA-based self-assembly systems to form structures and patterns using non-covalent interactions. In *Chapter 2* we develop a coarse-grained model to describe the self-assembly of finite-sized structures comprised of multiple different components and use computer simulations in order to: (1) understand the most effective methods to assemble such structures, (2) elucidate phenomena which might lead to poor assembly and (3) suggest a simple, general method for designing components which enhance assembly efficiency. In *Chapter 3* we investigate how to tune interaction strength between two self-assembling components by systematically altering the linking architecture between two DNA origami components, and measuring the resulting kinetics (*i.e.*, forward and reverse reaction rate constants) and thermodynamics (*i.e.*, yield, equilibrium constant) of the reversible dimerization reaction at a range of temperatures. These rate parameters were then incorporated into our coarse-grained model. *Chapter 4* builds on our simulations in *Chapter 2* by introducing a simple optimization algorithm whose goal is to find the set of components that assembles most efficiently across a broad range of reaction conditions. Coupling the optimization algorithm and kinetic and thermodynamic measurements of interfaces, measured in *Chapter 3*, our simulations provide a simple method to optimize the self-assembly of such structures and suggest an improved design of an experimentally reported DNA origami-based tetrameric complex and provide yield predictions for the improved structure. In keeping with the overarching theme of using self-assembling DNA to create organized material from the bottom-up, *Chapter 5* shows that by using diffusing ssDNA and dsDNA components programmed to undergo specific strand displacement reactions (*i.e.*, a reaction-diffusion system), we can produce spatiotemporally stable (*i.e.*, unchanging) concentration profiles of a DNA oligonucleotide species

in a hydrogel. Generally, this chapter provides an experimental framework for creating more complex DNA-based patterns in a hydrogel from the bottom-up using a reaction-diffusion system. This thesis concludes with a brief discussion in *Chapter 6* of the general implications and future areas of study that might arise from this work.

1.1 Contributions

Chapter 2

A version of *Chapter 2* was published:

John Zenk and Rebecca Schulman. An Assembly Funnel Makes Biomolecular Assembly Efficient, *PLoS ONE*, **9**, e111233 (2014).

John and Rebecca developed the model, performed simulations and wrote the paper.

Chapter 3

A version of *Chapter 3* was published:

Reproduced with permission from J. Zenk, C. Tuntivate, R. Schulman, Kinetics and Thermodynamics of Watson-Crick Base Pairing Driven DNA Origami Dimerization, *J. Am. Chem. Soc.*, 138, 3346-3354 (2016). Copyright 2016 American Chemical Society.

John, Chanon and Rebecca designed the research; John and Chanon performed the research; John, Chanon and Rebecca analyzed data; John and Rebecca wrote the paper.

Chapter 4

A version of *Chapter 4* is in preparation for submission:

John Zenk, Matthew Billups and Rebecca Schulman, Optimizing Component-Component Interaction Energies in the Self-Assembly of Finite, Multicomponent Structures, *in preparation*.

John and Rebecca designed the research; John and Matthew performed the simulations; John, Matthew and Rebecca analyzed the data; John and Rebecca wrote the paper.

Chapter 5

A version of *Chapter 5* is in preparation for submission.

John Zenk, Dominic Scalise, Kaiyuan Wang, Phillip Dorsey, Joshua Fern, Ariana Cruz and Rebecca Schulman, Stable DNA-Based Reaction-Diffusion Patterns, *in preparation*.

John, Dominic and Rebecca designed the research; Dominic developed the idea for stable patterns and designed the molecules; John, Kaiyuan, Phillip and Ariana performed the experiments; John performed the simulations and analysis; Joshua designed the sequences for the orthogonal system; John, Dominic and Rebecca wrote the paper.

CHAPTER 2 AN ASSEMBLY FUNNEL MAKES BIOMOLECULAR COMPLEX ASSEMBLY EFFICIENT

SUMMARY

Like protein folding and crystallization, the self-assembly of complexes is a fundamental form of biomolecular organization. While the number of methods for creating synthetic complexes is growing rapidly, most require empirical tuning of assembly conditions and/or produce low yields. We use coarse-grained simulations of the assembly kinetics of complexes to identify generic limitations on yields that arise because of the many simultaneous interactions allowed between the components and intermediates of a complex. Efficient assembly occurs when nucleation is fast and growth pathways are few, *i.e.* when there is an assembly "funnel." For typical complexes, an assembly funnel occurs in a narrow window of conditions whose location is highly complex specific. However, by redesigning the components this window can be drastically broadened, so that complexes can form quickly across many conditions. The generality of this approach suggests assembly funnel design as a foundational strategy for robust biomolecular complex synthesis.

2.1 INTRODUCTION

Within cells, bottom-up phenomena organize biomolecules into structures with sizes ranging from angstroms to microns. Precise control over structure at the angstrom and nanometer scales is important for optimizing catalysis^{47,48}, the action of molecular machines⁴⁹ or molecular recognition⁵⁰. Larger biomolecular structures orchestrate processes such as translation, adhesion, or controlled transport. One goal of chemistry and molecular engineering is therefore to develop analogous bottom-up methods for controlling biomolecular structure across the same range of dimensions^{51,52}.

Different physical processes are responsible for the *in vivo* formation of structure across these length scales. Stable nanometer- or angstrom-scale structures generally form as the result of folding a protein or RNA chain with a particular sequence⁴⁸. Folding larger structures from a single chain is difficult because synthesizing long, sequence-specific polymers without errors is a challenge⁵³ and the potential for a folding process to become frustrated increases quickly with polymer length^{54, 55}. Larger structures instead form through a hierarchical assembly process in which folded components self-assemble together into a larger complex. Examples of such complexes include the ribosome, proteasome and antibodies. Some complexes, including the nuclear pore complex⁵⁶, cell adhesions⁵⁷ or the kinetochore⁵⁸ can contain hundreds of components and reach sizes of more than a micron. Complex formation is ubiquitous: in *Escherichia coli*, for example, more than 20% of known polypeptides become reported members of protein complexes⁵⁹.

While the development of strategies for the design of synthetic self-assembling complexes have long lagged behind the design of folding processes, recently, a wealth of designed complexes assembled from proteins³⁹, nucleic acids³⁵ and other components^{60, 61} has spurred interest in developing rules and general strategies for designing complexes^{2, 62, 63}. Generally, design methods attempt to maximize complex yield by maximizing the thermodynamic stability of the complex or the free energy difference between the complex and other potential structures with the inherent assumption that thermodynamic equilibrium will be achieved^{2, 64, 65, 66}. Yet, in practice, complexes that are thermodynamically stable often assemble with low yields or may take as long as weeks to assemble properly^{67, 68, 69}. While unaccounted-for experimental effects such as stoichiometric imbalances between components¹⁰ might explain lower yields or slower than expected assembly times, kinetic factors that could limit yield are rarely investigated. To improve yields and dynamics,

there are currently few strategies other than complex redesign³⁴ when thermodynamic design considerations fail.

Here, we test the assumption that self-assembly processes for biomolecular complexes generally reach a high-yield equilibrium state by simulating the kinetics of a variety of generic, idealized assembly reactions. We find that for typical biomolecular complex self-assembly reaction rates and component concentrations^{70, 71, 72, 73}, it may take days or weeks to reach a state close to equilibrium, even when equilibrium yields are low. Thus, design processes that rely solely on thermodynamics to predict yields may meet with mixed success because yield is limited kinetically rather than thermodynamically. Our simulations also identify two key reasons why some self-assembly processes can be slow. First, near the melting temperature of the complex, low nucleation rates limit the rate of formation of complexes. Second, far below the melting temperature, assembly may occur rapidly through many different pathways, combinatorially trapping intermediate assembly products. Once assembly reaches this trapped state, complexes can form only after intermediates disassemble, which can be very slow. Avoiding both of these regimes is required to achieve high yield. For many common complexes, this requirement means that complex formation happens efficiently only under a narrow range of physical conditions. We show that designing components that skirt such kinetic pitfalls can significantly speed up assembly and enhance yields.

2.2 RESULTS

A simple model of multicomponent biomolecular complex self-assembly. To characterize the kinetics of complex assembly, we use a simple model of assembly in which rigid components of a generic complex bind to one another via orientation-specific pairs of complementary interfaces (Fig. 2.1). We assume that all components have identical interaction energy at each interface and the same initial concentration. Interaction between non-matching interfaces, or crosstalk, is neglected, reflecting rapid advances in the design of specific

biomolecular interfaces^{39, 67, 68, 69, 74, 75}. Multiple different rigid components and their unique interfaces could be easily fabricated from DNA, for example, using existing techniques such as DNA origami¹⁰ or DNA bricks⁶⁸. In our model, we consider all binary reactions that produce a complex or any connected subset of components, which we call intermediates (see Section 2.5 Supporting Methods).

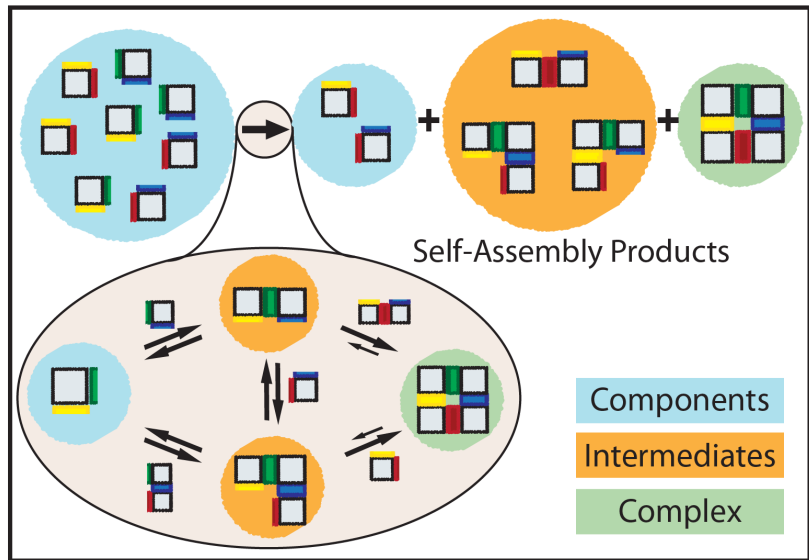


Figure 2.1. Self-assembly model for a 4-component square grid (“2x2”) complex. The square, rigid components have specific binding rules on each edge denoted by edge colors. Like colored edges interact, whereas edges with different colors and black edges do not interact. An initial, fixed number of components is depleted during self-assembly. At the end of the process, the solution contains a mixture of components, intermediates and complexes.

Our simulations use on and off rates similar to those measured for oligonucleotide⁷³, protein⁷¹, DNA tile⁷², and ribosomal subunit-RNA⁷⁰ reactions. Simulated assembly protocols are simple and modeled after those in broad experimental use^{10,34,69}. Assembly timescales are realistic, ranging from $\tau = 1$, or about 30 minutes to $\tau = 1000$, or about 2.7 weeks for 1 nM of components (*i.e.*, concentrations typical for large (megadalton) DNA nanostructures¹⁰), or 30 minutes for 1 μ M of components (see Equation 2.1). To model the interplay of changes in bond energy that could result from multi-bond reactions (*e.g.*, from entropic or allosteric effects⁷⁶), we introduce a dimensionless bond coupling term α that determines how the free energy of interaction scales with

the number of bonds formed (see Equation 2.2). We use the dimensionless parameter $\eta \equiv \log_{10}\left(\frac{k_{on}}{k_{off,1}}[X]_0\right)$ as an analog for inverse temperature (*e.g.*, high values of η correspond to low temperatures and strong interactions and vice versa) and define yield as the fraction of total material in complete complexes (see Equation 2.3).

The goal of our study is to understand how yields of self-assembled biomolecular complexes vary with complex size (in terms of number of components), geometry and reaction parameters (*e.g.*, $k_{off,b}$, a_o) by using kinetic simulations and as a result, learn how to design complexes and assembly protocols to increase yields. In order to elucidate general principles, we focus on a set of generic complexes: 1-dimensional “line” complexes of different lengths, 2-dimensional square “grid” complexes with different numbers of components on a side and a 3-dimensional “cube” complex.

Estimating the yield of a complex by considering its free energy relative to the free energies of other potential products is a standard method of estimating the yield of a self-assembly reaction⁷⁷, but such estimates are relevant only when assembly reactions are close to equilibrium. To determine whether typical reactions approach equilibrium, we modeled the kinetics of assembly using component concentrations, reaction times and rates typical of experimental self-assembly reactions^{70, 71, 72, 73}. To understand the effects of temperature, we initially studied reactions that take place at a single temperature (a single value of η). Isothermal assembly of 1D line complexes quickly achieved yields near those predicted at thermodynamic equilibrium for all interaction energies considered (Fig. 2.2a and Figs. S2.3 and S2.4). The system as a whole also approached equilibrium, as demonstrated by the concentrations of both complexes and intermediates (Fig. S2.5). Yields of line complexes were highest when the interactions between components were strongest, in agreement with both thermodynamic predictions and similar studies of self-assembly kinetics⁷⁸.

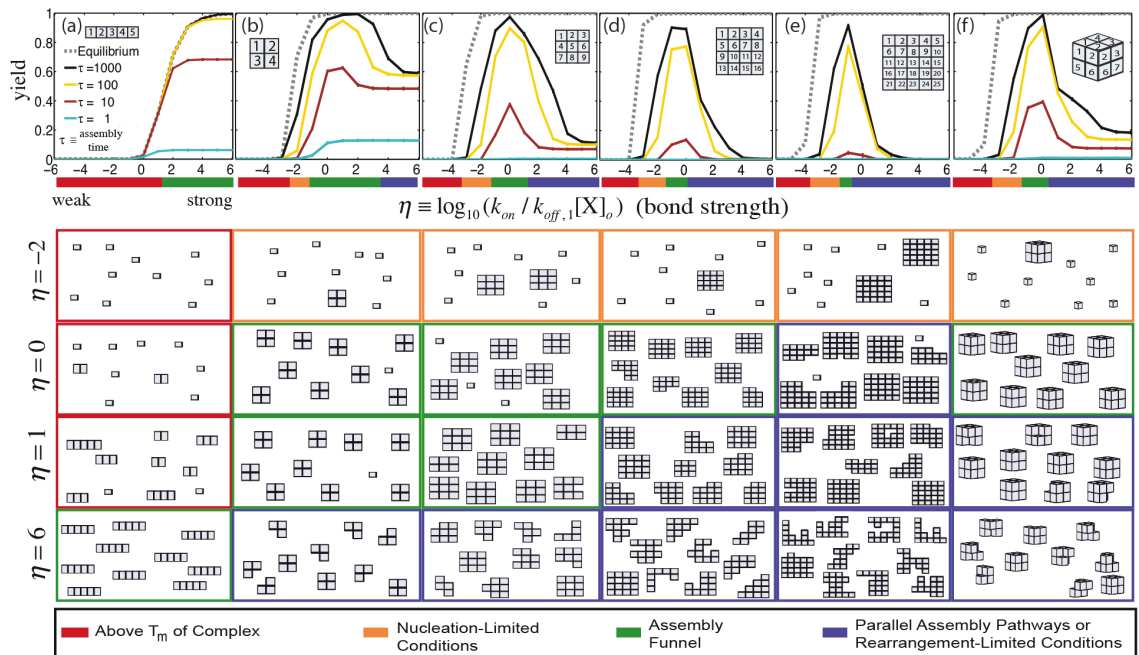


Figure 2.2. Thermodynamic equilibrium is a good predictor of yield for isothermal assembly after long assembly times for 1-dimensional complexes, but not 2- or 3-dimensional complexes. Assembly yields for a (a) 1x5 line complex, (b) 2x2, (c) 3x3, (d) 4x4 and (e) 5x5 square grid complex and (f) 2x2x2 cube complex as a function of the dimensionless temperature parameter, η . Inset diagram depicts the complex. Numbers on the components in the complex indicate component identity (e.g. component “1” is different than component “2”). The dashed line indicates thermodynamic equilibrium. Dimensionless reaction time is defined as $\tau = k_f[X]_0 t$ where k_f is the macroscopic forward reaction rate constant and $[X]_0$ is the initial concentration of components. Colored bars and boxes below figures represent the four different assembly regimes (Note S2 in Section 2.5.2). The assembly funnel regime is considered to be where the complex is thermodynamically favored (*i.e.*, $yield_{eq} > 0.5$) and assembly is rapid such that $yield_{\tau=1000} \geq 0.8 yield_{eq}$. Assembly “snapshots” (below graphs) are taken at $\tau = 1000$ and $\eta = -2$ (top row), $\eta = 0$, $\eta = 1$, and $\eta = 6$ (bottom row) and comprised of ten random species drawn from the reaction mixture, weighted by concentration (Note S3 in Section 2.5.3). Error bars indicate the standard deviation of the reported quantity after 10 simulations and where omitted, are $<1\%$. Here and elsewhere unless otherwise noted, there is no bond coupling ($a_o = 1$).

Strong interactions maximize yield for 1-dimensional systems only. While strong interactions maximize the yield of line complexes, strong interactions in even small grid or cube complexes with no bond coupling ($a_o = 1$) produced yields far lower than yields expected at equilibrium for simulated reaction times as long as weeks ($\tau = 1000$, ~ 2.7 weeks for $[X]_0 = 1\text{nM}$) (see Figs. 2.2b-f and Figs. S2.8 and S2.9). Further, after a certain point, increasing reaction time

only marginally increases yield (Figs. S2.11, S2.22). For example, increasing the assembly time from $\tau = 100$ to $\tau = 1000$ increased the yield of 3x3 grid complexes by at most ~10%. Similarly marginal increases in yield were observed when assembly times were increased further to $\tau = 10000$ (Fig. S2.10). These results suggest that these self-assembly processes rarely approach the equilibrium state in practice.

Slow nucleation and molecular rearrangement rates can limit yield. To understand why grid and cube complexes assembled so slowly, we investigated the composition of the simulated solution after the completion of the reaction ($\tau = 1000$) under many isothermal assembly conditions (Fig. 2.2, and Figs. S2.13, S2.15 and S2.23). Above the melting temperature of a given complex, no complexes form. Just below the melting temperature, the most abundant species aside from complexes were components, suggesting that yield under these highly reversible conditions is limited by the long times required to nucleate intermediates. Under effectively irreversible conditions (*i.e.*, high values of η), intermediates that cannot interact with one another to form complexes were the most common species, including the four 3-component intermediates in the 2x2 square grid complex and the 5 to 8-component intermediates in the 3x3 square grid complex (Fig. S2.14). Under these conditions, components or smaller intermediates must detach from a larger intermediate and attach to another intermediate, or “rearrange”, in order to complete a complex, which is an energetically unfavorable and therefore slow process. This rearrangement-limited regime is present for the assembly of grid and cube complexes but not line complexes because the intermediates to line complexes never need to rearrange to produce complexes. These results are corroborated by studies of viral capsid assembly⁷⁹ as well as homomeric⁸⁰ and ring-like protein complex assembly⁸¹, where nucleation and rearrangement rates were found to influence assembly efficiency and fidelity.

A high-yield assembly funnel regime occurs at medium-strength component interactions. The results thus far indicate that the self-assembly of grid and cube complexes could occur with high yields when bond strengths are neither too weak for fast nucleation nor too strong to prevent components in intermediates from rearranging. Indeed, our simulations show that there is a small window of medium component-component interaction strength where complexes are stable and assemble with high yields without requiring infeasibly long assembly times. We called this regime the “assembly funnel” regime, because in this regime the energy landscape contains a small number of smooth downward paths to complete complex formation, similar to a protein folding funnel⁸² or a protein binding funnel⁸³. This regime for grid and cube complexes is generally near $k_{on}[X]_0 = k_{off,1}$ or $\eta = 0$. In our simulations of 2x2 to 5x5 square grid complexes, we found that increasing complex size shrinks the size of the already small assembly funnel regime by disfavoring forward conditions (*i.e.*, where $\eta > 0$). Increasing complex size increases the number of ways components can become “stuck” in incompatible intermediates, so completing a larger complex requires more molecular rearrangement on average than completing a smaller one.

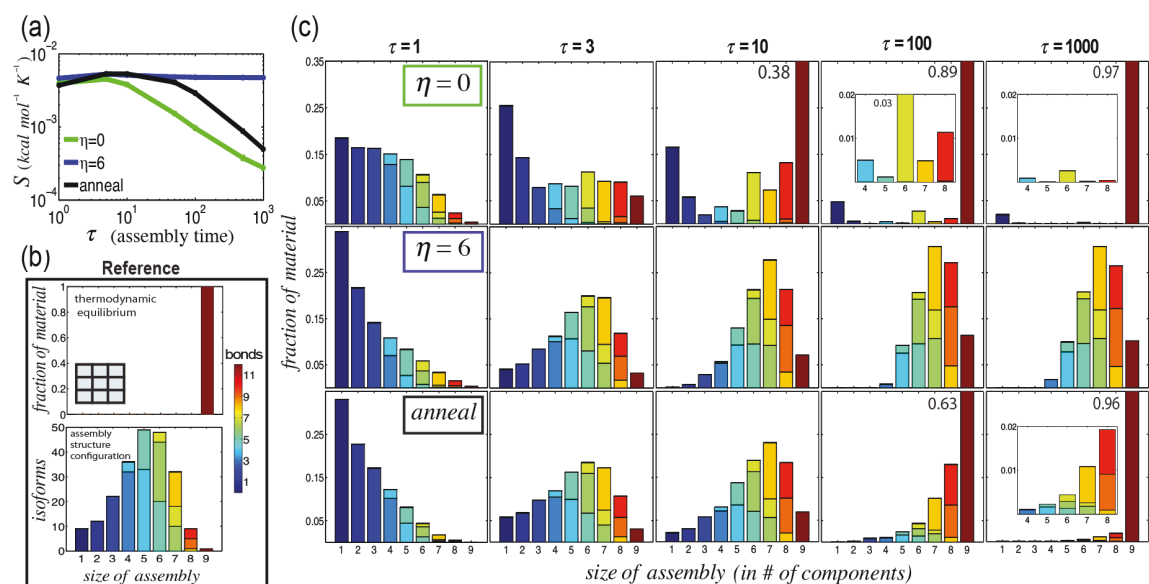


Figure 2.3. An assembly funnel means that complex assembly occurs via a small number of pathways. The possible set of reaction pathways govern assembly outcome under rearrangement-limited

conditions, whereas thermodynamically favorable pathways govern assembly outcome in the assembly funnel regime. (a) Conformational entropy (S) of the system under different assembly conditions as a function of assembly time, τ . (b) Reference energy distributions of a 3x3 square grid complex based on thermodynamics and assembly configuration. Color spectrum indicates the number of bonds in an assembly. (c) Partition of energies at different times during self-assembly in the assembly funnel regime at $\eta = 0$ (green box), rearrangement-limited conditions at $\eta = 6$ (blue box), and during an anneal (black box). Over the course of an anneal, η transitions from -6 to 6 , spending $\tau/100$ at 100 different linearly decreasing isothermal conditions. Values at the top right are complex yields. Inset plots show detail. Error bars $<1\%$.

Reaction conditions determine the set of possible assembly pathways. To further understand the influence of pathways on complex formation, we examined the kinds of intermediates that tend to arise and persist by measuring the conformational entropy, or distribution of species sizes and free energies, of the system. The conformational entropy is given by $S = -R \sum_j \sum_i f_{ij} \ln(f_{ij})$ where f_{ij} is the fraction of species with energy i and j components. Higher values of conformational entropy correspond to broader distributions of assembly sizes and free energies. Under rearrangement-limited conditions, conformational entropy initially increases as many different intermediates form, and then plateaus (Fig. 2.3a). The species that form and remain are those that are most easily accessible via reaction pathways rather than those that are energetically favorable (Fig. 2.3b,c). In contrast, assembly in the assembly funnel regime favors the production of a relatively small number of intermediates, those lowest in free energy, so conformational entropy decreases with time as these low-energy intermediates and complexes form. Because complex size and geometry determine the possible reaction pathways and the types of assembly intermediates that can form⁸⁴, they also control the propensity of an assembly process to become “stuck” under a given set of reaction conditions.

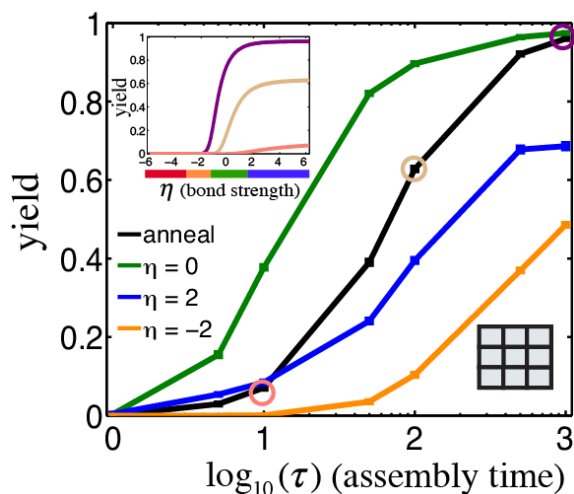


Figure 2.4. Complexes form rapidly in the assembly funnel regime. Yield of 3x3 square grid complex as a function of reaction time by assembling via annealing and at various isothermal assembly conditions: $\eta = -2$ (orange, nucleation-limited), $\eta = 0$ (green, assembly funnel) $\eta = 2$ (blue, parallel assembly pathways and rearrangement-limited). Inset plot (top left) depicts yield during an anneal as a function of interaction strength for different reaction times: $\tau = 10$ (salmon), $\tau = 100$ (beige), and $\tau = 1000$ (purple). Inset diagram (bottom right) depicts the complex.

The time spent in the assembly funnel regime determines the yield. While complexes form quickly in the assembly funnel regime, the specific reaction conditions that generate an assembly funnel depend on the set of possible reaction pathways as well as kinetic and thermodynamic parameters that are generally unknown and difficult to estimate. One solution to this problem is to assemble via annealing. A typical annealing protocol begins at a temperature above the melting temperature of the complex, which is then gradually decreased until effectively irreversible conditions are achieved. To determine how yields using this protocol compare to those during isothermal assembly, we simulated annealing for square grid complexes. We found that yields during an anneal are predominately determined by the amount of the time spent in the assembly funnel regime. As the temperature decreases, few complexes form before the assembly funnel regime is reached. Within the funnel regime, complexes form rapidly, primarily through thermodynamic pathways (Figs. 2.3, 2.4 and Figs. S2.16-S2.18). After the annealing moves out of the assembly funnel regime, complexes are stabilized, but relatively few new complexes form.

Thus, assembly via annealing is relatively efficient even when it is not known which conditions that generate an assembly funnel, which is in agreement to recent computational findings on DNA brick self-assembly⁸⁵. However, to produce high yields, an anneal must be slower than a comparable isothermal assembly process in the assembly funnel regime because complex formation is slow for the majority of the anneal. This effect becomes more pronounced as complex size increases because the range of reaction conditions that produce an assembly funnel decreases. Thus, for very large complexes, it may be important to find ideal isothermal conditions, even when annealing is a practical option for assembly³⁴.

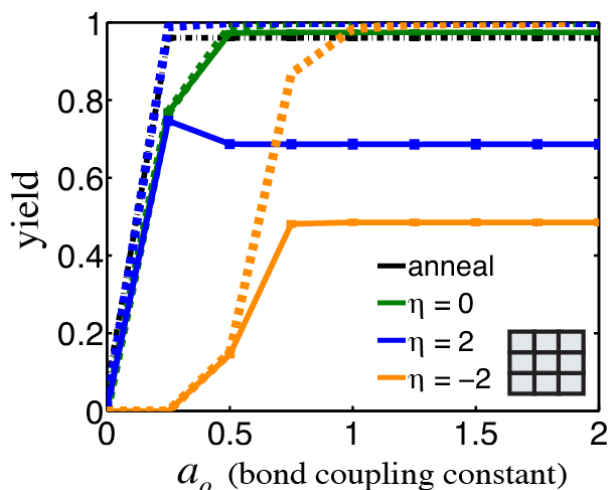


Figure 2.5. The amount of bond coupling, or additivity of bond energies during cooperative binding steps does not significantly affect assembly yields above a small threshold. Yield of a 3x3 square grid complex as a function of the bond coupling constant, a_o under many isothermal assembly conditions (solid lines, color) and after an anneal (black) for reaction time $\tau = 1000$. Dashed lines show yields at thermodynamic equilibrium for isothermal conditions with the same color. Error bars $<1\%$.

Just a small amount of bond coupling between components is needed for high yield.

2- and 3-dimensional complexes are generally stabilized by the interactions of multiple bonds between components, and the specific free energy changes that result from multi-bond interactions also shape the energy landscape for assembly⁸⁶. To determine how the free energy of multi-bond interactions influences yield, we characterized changes in yield as we altered the coupling between

multiple interfaces on a component. Surprisingly, we found that bond coupling was not an important determinant of assembly yield (see Fig. 2.5 and Figs. S2.7, S2.21). Although positive coupling ($a_o > 1$) slightly broadens the set of conditions where complex yields are high at thermodynamic equilibrium (Figs. S2.6, S2.20), it leads neither to increased nucleation rates nor component rearrangement rates and thus does not increase yields in practice. Negative coupling ($a_o < 1$) does not always reduce yields in the assembly funnel regime and can even marginally enhance yields under rearrangement-limited isothermal conditions by destabilizing some intermediates (Note S4 in Section 2.5.4). Thus, high-yield assembly can be obtained under the proper assembly conditions for a wide range of bond coupling values, as any coupling value is subject to equal pressures on nucleation and rearrangement rates.

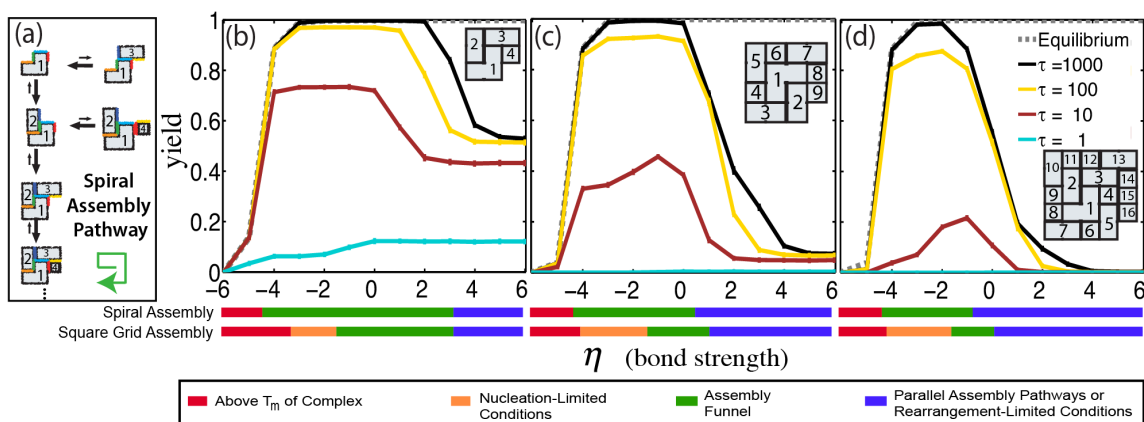


Figure 2.6. Design of components so that particular assembly pathways are favored can drastically increase assembly yields. (a) Schematic of spiral complex assembly via the favored assembly pathway. On the favored assembly pathway, assembly begins with the “L” shaped component, labeled “1”. At each assembly step, a component attaches through two interfaces (following the green arrow). Other components can only attach through one. Lengths of reaction arrows indicate propensities in the assembly funnel regime. Assembly yields for a (b) 2x2 (4 component), (c) 3x3 (9 component) and (d) 4x4 (16 component) spiral complex as a function of a dimensionless temperature parameter, η . Inset diagram depicts the complex and numbers on the components in the complex indicate component identity. Colored bars below the figure represent the four different assembly regimes for spiral complexes and grid complexes containing the same number of components. Error bars <1%.

Components can be designed to assemble efficiently because they assemble *via* an assembly funnel under most conditions. While it is challenging to optimize reaction conditions

to produce high yields, might it be possible to create components that broaden the assembly funnel regime and thus self-assemble a desired complex more efficiently? To address this question, we designed components for a 2-dimensional target structure that were expected to have a smaller barrier to nucleation than the components of the grid complex we studied above. In a “spiral complex,” a spiral-shaped growth pathway allows all components to attach to the growing assembly via multiple bonds, so that there is no nucleation barrier to assembly. Because all other growth pathways require that components interact with one another via a single bond, the single spiral-shaped growth pathway is favored (Fig. 2.6a). Compared to square grid complex counterparts, the 4-, 9- and 16-component spiral complexes assemble faster and even achieve thermodynamic equilibrium in nucleation-limited regimes, broadening the reaction conditions that generate an assembly funnel (Fig. 2.6b-d). As a result, an anneal produces complexes more quickly, by almost an order of magnitude (Figs. S2.11 and S2.12). While the spiral scheme does not improve yield in the rearrangement-limited regime, this exercise suggests that effective self-assembly design strategies will likely promote rapid, high-yield complex formation by considering reaction pathways as well as nucleation and rearrangement rates.

2.3 DISCUSSION

Most existing strategies for the design and analysis of self-assembly processes use the thermodynamics of a complex as a starting point for predicting structure and yield. This strategy has been successful for understanding the assembly process of homogeneous or periodic crystals and superlattices⁸⁷. While in principle, these strategies can be extended to guide the design of finite, heterogeneous complexes, we find that for a large class of multicomponent assembly processes, these strategies are insufficient because assembly is kinetically limited. Our results are echoed by experimental studies in which complex yields are low even when the desired product is strongly

thermodynamically favored^{10,68}, and in which assembly can be made efficient by assembling at a constant temperature at which assembly is optimal³⁴, in what we term the assembly funnel regime, if such a regime can be found. In fact, the assembly funnel assembly strategy has been used in the self-assembly homogenous multicompartment micelles⁸⁸.

Although optimizing assembly conditions appears difficult, this work suggests that it may be much more productive to design components such that they assemble efficiently through one or a small number of reaction pathways. This strategy of designing components that assemble efficiently appears to be important *in vivo*, as the components of protein complexes are under evolutionary pressure to assemble via ordered pathways⁸⁹.

One major assumption in this work is perfectly formed components: we do not address the challenge to form the components in the first place. In successfully forming biomolecular complexes, components must first be properly synthesized and folded or fabricated before they can associate to form a complex. Components that misfold or degrade can alter the assembly landscape by allowing the possibility of nonspecific interactions (e.g., resulting in aggregated products, as clearly evidenced by diseases such as amyloidosis), which provides another, perhaps even larger, challenge in understanding complex assembly.

While this work will need to be extended to take into account artifacts of assembly such as component defects and differences in component stoichiometry and bond energies, this work adds to growing evidence that the physics of assembly of multicomponent, aperiodic structures is not simply an extension of principles for assembling homogeneous or periodic structures⁵². Assembly of multicomponent lattices and crystals also appear to occur far from equilibrium in general^{90,91} even when component depletion is offset by continued production of new components, as happens in *in vivo* systems. Specific attention to effects that arise in multicomponent systems, such as the

interplay between combinatoric and thermodynamic factors explored here, are likely to be important in developing the capacity to self-assemble larger, more intricate structures robustly.

2.4 METHODS

2.4.1 Stochastic kinetic simulations.

The dynamics of the reactions to form a complex are determined using Gillespie sampling of stochastic chemical kinetics⁹². While typically stochastic fluctuations are not important to assembly results, the Gillespie algorithm makes it possible to statistically sample kinetic trajectories that would otherwise be inaccessible because the numerical integration of the coupled set of ODEs for mass action kinetics is intractable for most of the complexes we study (Table S2.1 and Note S1 in Section 2.5.1). For small complexes where comparison is possible, stochastic kinetic simulations and mass action kinetics produce nearly identical results (Fig. S2.21).

2.4.2 Rate constants and physical parameters.

For all reactions, the macroscopic on rate constant is assumed to be constant, $k_{on} = k_f = 6 \times 10^5 (M * sec)^{-1}$, reflecting experimental data for DNA and proteins^{70, 71, 72, 73}, which additionally simplifies analysis by providing an energy landscape for assembly. Because in practice intermediates and complexes may diffuse more slowly than components due to their increased size, this assumption likely underestimates assembly times. We define dimensionless time:

$$\tau = k_f [X]_0 t, \tag{2.1}$$

where k_f is the macroscopic on rate constant, $[X]_0 = 0.1$ nM is the initial component concentration and t is dimensional reaction time in seconds.

To model the interplay of changes in bond energy that could result from multi-bond reactions, we introduce a dimensionless bond coupling term α that determines how the free energy of interaction scales with the number of bonds formed. This bond coupling term is given by:

$$\alpha(b) = (a_o - 1)(b - 1), \quad (2.2)$$

where a_o is a dimensionless coupling constant and b is the number of bonds formed in the reaction. Interfaces are energetically independent in the case of zero ($a_o = 1$) bond coupling. Negative coupling ($0 \leq a_o < 1$) means that the interaction of multiple bonds is less favorable than the sum of the individual bond energies whereas positive coupling ($a_o > 1$) means the same interaction is more favorable than the sum of the individual bond energies. The coupling term appears in the macroscopic off rate equation:

$$k_{off,b} = k_f \exp\left(\frac{(b + \alpha)\Delta G^o}{RT}\right), \quad (2.3)$$

where ΔG^o is the change in standard Gibbs free energy for a component-component interaction through a single bond, T is absolute temperature and R is the universal gas constant.

For detailed information on species and reaction enumeration algorithms, as well as kinetic simulation specifics, see Supporting Methods, Section 2.5.1.

2.5 SUPPORTING INFORMATION

2.5.1 Text S2.1 Supporting Methods

Enumeration Algorithms

We model the self-assembly of 1D “line” complexes, 2D “square grid” complexes, 2D “spiral” complexes, and 3D “cube” complexes. In each of the complexes we model, every component in each complex is distinct. For a given complex, we assume that any connected set of

components smaller than the full complex is a valid intermediate and any reaction involving two such species (either components or intermediates) that forms another valid intermediate is allowed. To ensure that all possible assembly and disassembly reactions could occur during our kinetic simulations, we built a full list of what intermediate assemblies could form while assembling a complex and what reactions are allowed. We describe our methods enumerating complex intermediates and assembly reactions below.

Intermediate Enumeration

To enumerate all valid intermediate assemblies (*i.e.*, species with more than one component) for a given complex, we iterate through all possible subsets of components in the complex. A valid intermediate has full connectivity (*i.e.*, every component in the intermediate shares at least one edge with another component of the subset, see Figure S2.1). This definition means that intermediates of 2D and 3D complexes may have voids. We chose to include these intermediates because there is experimental evidence that they are present as assembly intermediates¹⁰.

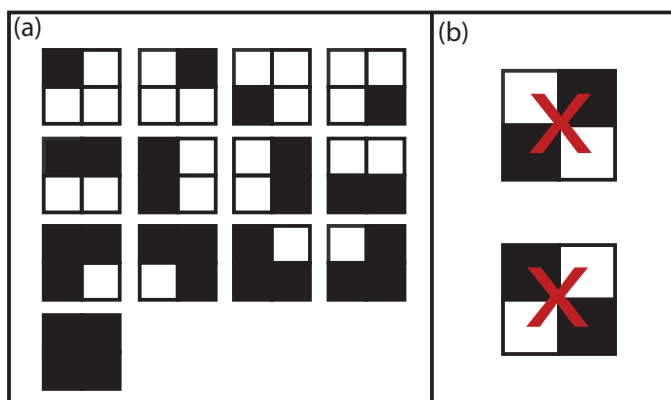


Figure S 2.1 Valid and invalid species in the 2x2 grid complex. (a) Valid species are components, full complexes and multiple component configurations, or intermediates, where all components comprising an intermediate have at least one bond (shared edge) with another component. A black box represents an occupied site whereas a white box represents unoccupied sites on the lattice. (b)

Invalid intermediate assemblies are denoted by red “X”s and are lattice configurations that are not connected (do not share an edge) and are not included in our model.

Reaction Enumeration

After enumerating all valid intermediate assemblies, we determine which species can interact to form larger assembly products. We assume all reactions are reversible and binary (*i.e.*, every forward reaction has two reactants and one product). Valid reactions are those in which two valid species interact given (1) they share no component and (2) the product of the reaction is also a valid species (Figure S2.2). If both of these criteria are satisfied the reaction is included in the list of possible reactions. We assume aggregation reactions (*i.e.*, those involving two valid intermediate assemblies such as in Figure S2.2b) are valid because experimental evidence suggests nature often assembles structures via parallel pathways^{93, 94, 95, 96}, which has also been demonstrated as an effective *in vitro* and synthetic assembly strategy^{88, 97, 98}.

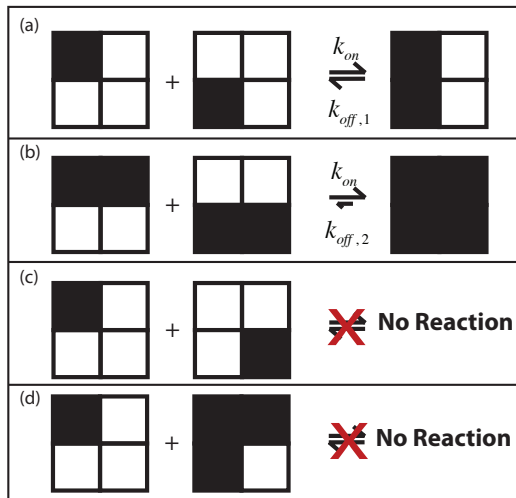


Figure S 2.2 Valid and invalid reactions for the 2x2 grid complex. Examples of valid reactions in a 2x2 grid complex in which (a) one bond, or (b) two bonds are formed. The reverse reaction rate (indicated roughly as arrow length) will change with reaction conditions and bond coupling. Reactions such as in (c) and (d) are not included in our model. In (c), the components do not interact at any edges and would not produce a valid species as a product, and in (d) the reactants

share components in the same position, which would in practice block that reaction from happening.

Kinetic Simulations

The macroscopic forward and reverse reaction rate constants are defined respectively by:

$$k_{on} = k_f, \quad (2.4)$$

$$k_{off,b_i} = k_f \exp\left(\frac{(\alpha + b_i)\Delta G^o}{RT}\right), \quad (2.5)$$

where R is the universal gas constant, T is absolute temperature, and α is a bond coupling term. Bond coupling is a function of both the number of bonds formed in the i^{th} reaction, b_i , and a constant parameter, a_0 . Coupling is defined as:

$$\alpha(b_i) = (a_0 - 1)(b_i - 1), \quad (2.6)$$

where $\alpha = 0$ for all single bond interactions. For 1D complexes, all reactions happen via a single bond so the value of the constant parameter, a_0 , does not affect the reaction. In 2D and 3D simulations, we vary the parameter a_0 from $0 \leq a_0 \leq 2$ in order to model a zero, negative and positive values of bond coupling.

For the reaction of $A + B \rightleftharpoons C$ through b_i bonds, the macroscopic reaction rates are:

$$r_{on} = k_{on}[A][B], \quad (2.7)$$

$$r_{off,b_i} = k_{off,b_i}[C], \quad (2.8)$$

where $[A]$, $[B]$ and $[C]$ are the respective concentrations of species A , B and C .

We use stochastic kinetic simulations of the dynamics of self-assembly in which we sample assembly trajectories of the complex. This method enables us to obtain accurate results without requiring tremendous computational resources, as would be the case when integrating many (tens to millions, see Table S2.1) highly coupled ordinary differential equations of the mass action kinetics model.

For a stochastic kinetic simulation, macroscopic rates must be converted into reaction propensities. The microscopic forward reaction propensity constant is defined as $k_f = \frac{k_f}{N_A V} \frac{1}{\text{#molecules} \cdot \text{second}}$ such that the microscopic forward and reverse reaction propensities are defined respectively:

$$a_{on} = k_f A_{\#} B_{\#}, \quad (2.9)$$

$$a_{off,b_i} = k_f \exp\left(\frac{(\alpha + b_i)\Delta G^o}{RT}\right) C_{\#}, \quad (2.10)$$

where N_A is Avogadro's number, V is the volume of the reaction vessel and $A_{\#}$, $B_{\#}$ and $C_{\#}$ are number of molecules of A, B and C, respectively. For a given simulation, the initial number of all components is the same. The volume of the reaction vessel is fixed throughout the reaction.

Reaction time in seconds, t , is nondimensionalized using parameter τ such that for the starting component concentration, $[X]_0$ and the macroscopic forward reaction rate constant, k_f is given by:

$$\tau = k_f [X]_0 t. \quad (2.11)$$

Likewise, reaction temperature is nondimensionalized by the on and off reaction rates through a parameter, η , where:

$$\eta \equiv \log_{10}\left(\frac{k_{on}}{k_{off,1}} [X]_0\right). \quad (2.12)$$

High values of η correspond to strong interaction strengths (low temperatures) and low values of η correspond to weak interaction strengths (high temperatures). Together, η and τ define the reaction conditions for the self-assembly of a given complex. Specifically, the initial concentration of components, reaction time and temperature together correspond to a particular η and τ .

We use Gillespie’s algorithm ⁹² to sample trajectories of the stochastic kinetics of a given reaction. Because the trajectories are all different, we perform multiple simulations for each complex to obtain a sample of the space of trajectories. Ten simulations are performed for every set of reaction conditions in order to obtain assembly yield average and error bars. Error bars indicate the standard deviation of the reported quantity. For all complexes except the 5x5 square grid, error bars are <1% of the mean.

To report specific assembly times and temperatures, we chose specific reaction rates and concentrations that are inspired by published self-assembly experiments using biomolecular components. We use a diffusion-limited forward reaction rate constant of $k_f = 6 \times 10^5 \frac{1}{M \text{ sec}}$ for all reactions, a value close to those measured for oligonucleotide ⁷³, protein ⁷¹, DNA tile ⁷², and ribosomal subunit-RNA ⁷⁰ reactions. In practice, the rate constants of reactions involving large assemblies are likely to be slower, reflecting slower diffusion rates. As such, this assumption likely overestimates yields under some conditions.

We chose the standard energetic parameters for two 5 base-pair DNA-DNA “sticky end” hybridization reactions in our definition of $\Delta G^o = \Delta H^o - T\Delta S^o$ where $\Delta H^o = -102.4 \text{ kcal/mol}$, $\Delta S^o = -0.3 \text{ kcal/mol/K}$ which are typical free energy parameters as predicted by the nearest-neighbor model ⁹⁹. These values are also similar to a wide variety of protein-ligand complexes ¹⁰⁰ and protein-protein interactions in complex formation ¹⁰¹.

In this work, the conformational entropy of the system, similar to conformational entropy in protein folding ¹⁰², is a measure of the distribution of occupied energy states in a system and is given by the Boltzmann sampling over all states:

$$S = -R \sum_j \sum_i f_{i,j} \ln(f_{i,j}) \quad (2.13)$$

where R is the universal gas constant and $f_{i,j}$ is the fraction of species with energy i and j number of components. Like conformational entropy in protein folding, the conformational entropy in Equation S2.13 does not take into account other entropic effects, such as molecular vibration within a species. It is therefore not a direct measure of total system entropy.

Table S 2.1 Complex specifics and parameter space explored in this work

Complex	Number of Components	Number of Valid Species	Number of Reactions	Number of each Starting Component	Iterations
1x3	3	6	4	10000	10
1x4	4	10	10	10000	10
1x5	5	15	20	10000	10
1x6	6	21	35	10000	10
1x7	7	28	56	10000	10
1x8	8	36	84	10000	10
1x9	9	45	120	10000	10
2x2*	4	13	18	10000	10
3x3	9	218	1381	10000	10
4x4	16	11506	224305	10000	10
5x5	25	2301877	115804894	1000	10
2x2x2	8	167	1079	10000	10
Spiral 2x2	4	14	21	10000	10
Spiral 3x3	9	281	1941	10000	10
Spiral 4x4	16	16898	357022	10000	10

*This complex is also enumerated in ⁶⁵ and has an identical list of possible species and reactions to the work presented here

2.5.2 Text S2.2 Computational Specifics

All enumeration algorithms and stochastic simulations were performed using custom code written in C. Within the C code, multiple optimization techniques were employed in order to decrease computation time. The code used to simulate assembly of the 5x5 square grid complex especially required optimization because more than 30 million possible species can be formed during assembly and roughly 100 million highly coupled reactions are possible (see Table S2.1).

The complexity of this reaction process made the reaction selection step in simple Gillespie algorithm excruciatingly slow. All simulations ran on an in-house server with 8 cores (2 x Quad Core Intel® i7 960) and 24 Gb RAM. Thermodynamic equilibrium was calculated using a Gauss-Newton algorithm implemented in MATLAB for smaller complexes (less than 16 components). For larger complexes, we implemented an iterative algorithm in C that converged to thermodynamic equilibrium by choosing a reaction at random (with equal probability to choose any reaction) then zeroing the net reaction rate by changing the species concentrations. The algorithm converged to a final solution when the net reaction rate for all reactions was zero to machine precision (see Note S2.5 in Section 2.5.5 for specifics). All code is available upon request.

2.5.3 Text S2.3 Assembly Regime Criteria

We demarcate various assembly regimes by using the following values: the yield of the complex at thermodynamic equilibrium, or y_{eq} , the yield of the complex after $\tau = 1000$, or $y_{\tau=1000}$, and the mean size (in number of components) of the intermediates after $\tau = 1000$ and number of components in a complex, or $\bar{N}_{int,\tau=1000}$ and N_{cplx} , respectively. Table S2.2 shows the specific criteria for demarcating assembly regimes.

Table S 2.2 Criteria for labeling assembly regimes

Color	Regime	Definition
	Above T_m of complex	$y_{eq} < 50\%$
	Nucleation-limited conditions	$y_{eq} \geq 50\%$, $y_{\tau=1000} < 0.8y_{eq}$, $\bar{N}_{int,\tau=1000} \leq 0.5N_{cplx}$
	Assembly funnel	$y_{eq} \geq 50\%$, $y_{\tau=1000} \geq 0.8y_{eq}$
	Parallel assembly pathway and rearrangement-limited conditions	$y_{eq} \geq 50\%$, $y_{\tau=1000} < 0.8y_{eq}$, $\bar{N}_{int,\tau=1000} > 0.5N_{cplx}$

Above T_m of complex

Molecular engineers can effectively design complexes (usually the target end product of self-assembly) using thermodynamic principles to have a low free energy and thus achieve high yields after very long assembly times ($\tau \rightarrow \infty$). So designing complexes that are not stable or subjecting components to assembly conditions that give $y_{eq} < 50\%$ is unfavorable from an engineering perspective and hence assigned a red label in our plots.

Nucleation-limited conditions

Nucleation-limited conditions thermodynamically favor complex formation ($y_{eq} \geq 50\%$) but weak component-component interactions limit yields ($y_{\tau=1000} < 0.8y_{eq}$). Under these conditions the mean intermediate size (in number of components) is less than or equal to half of

the size of the complex after $\tau = 1000$ (or $\bar{N}_{int,\tau=1000} \leq 0.5N_{cplx}$, see Figure S15), suggesting that nucleation is the primary growth mechanism of complex assembly.

Assembly funnel

The regime where an assembly funnel is present (the assembly funnel regime) is the most favorable regime of assembly, where complexes are highly thermodynamically favored ($y_{eq} \geq 50\%$) and complexes form rapidly enough to achieve high-yield in finite times ($y_{\tau=1000} \geq 0.8y_{eq}$).

Parallel assembly pathway and rearrangement-limited conditions

The parallel assembly pathway and rearrangement-limited regime occurs when complexes are highly thermodynamically favored ($y_{eq} \geq 50\%$) but the dynamics of assembly are slow ($y_{\tau=1000} < 0.8y_{eq}$). Additionally, under these conditions the mean intermediate size (in number of components) is greater than half of the size of the complex after $\tau = 1000$ (or $\bar{N}_{int,\tau=1000} > 0.5N_{cplx}$, see Figure S2.15), suggesting that the rearrangement of components between intermediates is necessary to form complexes.

2.5.4 Text S2.4 Assembly Distribution Selection

In Figure 2.2 of the main text, ten species are drawn at random from the reaction mixture. The probability of selecting species j , P_j , for a given sample species is proportional to the mean concentration of j , \bar{C}_j , calculated from the results of ten simulations at a given isothermal assembly condition (η) after $\tau = 1000$ such that:

$$P_j = \frac{\bar{C}_j}{\sum_i \bar{C}_i}, \quad (2.14)$$

where \bar{C}_i is the mean concentration of species i after the same simulations.

2.5.5 Text S2.5 Further Explanation of Bond Coupling Effect on Yield (from Figure 2.5)

To determine how energetic changes from multi-bond interactions influences yield, we altered the bond coupling between multiple interfaces on a component. As expected, when the bond coupling constant is very small (*e.g.*, $a_o = 0$), complexes and intermediates are unstable and thus yield is low under all assembly conditions. But when the bond coupling constant is large ($a_o \geq 1$), complexes and intermediates are very stable and yield is essentially independent of the bond coupling constant for all assembly conditions above a relatively low threshold. This suggests that large values of bond coupling constant are not required for high-yield self-assembly. However, when $0 < a_o < 1$, how bond coupling affects yield depends on the assembly conditions.

Nucleation-limited conditions

In the nucleation-limited regime, while thermodynamic analysis would predict that a larger degree of coupling would increase yield (see Figure S2.7) we found that for a 3x3 square grid complex, some coupling (*e.g.*, $a_o \approx 0.25$) is important in this regime to stabilize intermediate assemblies. However, higher values of the bond coupling constant do not improve yield. The results suggests that in this regime, where the dynamics of assembly are governed by the nucleation rate, increasing intermediate stability past a point of sufficient stability (the specific value depends on complex size and geometry) does not lead to higher nucleation rates.

Assembly funnel

Assembly in the assembly funnel regime achieves yields close to thermodynamic equilibrium for all values of bond coupling. Increasing the bond coupling constant from $a_o = 0$ increases the stability of intermediates and complexes in which some components are bound to at least three neighbors, resulting in higher yield at thermodynamic equilibrium and after finite times. However, above some small value of the bond coupling constant that is dependent on the complex size and geometry, further increasing the constant beyond this value does not increase the yield, or the size of the assembly funnel. For example, the value of bond coupling constant where this occurs for a 3x3 square grid complex is $a_o \approx 0.5$, and yield is bounded by $y_{\tau=1000} \approx 95\%$. Likewise, the assembly funnel regime for the 3x3 square grid complex includes roughly the same range of η values for all bond coupling constants larger than this value.

Parallel assembly pathway and rearrangement-limited conditions

In a moderately rearrangement-limited regime ($\eta \approx 2$), small values of the bond coupling constant can actually improve assembly dynamics by enhancing rearrangement rates (*e.g.*, $a_o \approx 0.25$ for a 3x3 square grid complex). Here, intermediates are less stable, which results in increased rates of disassembly and hence higher yields than assembly under the same conditions but with a higher bond coupling constant. While enhanced yield is possible under moderately rearrangement-limited regimes, under effectively irreversible conditions ($\eta \approx 6$) yield is independent of bond coupling because single component-component interactions are so strong that all disassembly is unlikely.

2.5.6 Text S2.6 Computing Thermodynamic Equilibrium of Large Complexes

This section describes the algorithm used to compute the concentrations of components, intermediates and complexes at equilibrium for larger (16 and 25 component) complexes. We start with an initial guess of component and complex concentrations that preserves mass balance among reaction components with respect to the initial conditions of the reaction. The algorithm then chooses a reaction at random (with equal probability to choose any reaction in the list of possible reactions), and zeros the net reaction rate by computing the change in the species concentrations that would give the concentration of species at equilibrium.

For a given reaction $A + B \rightleftharpoons C$, the equilibrium constant, K_{eq} , is given by:

$$K_{eq} = \frac{k_{on}}{k_{off,b_i}} = \frac{[C]_{eq}}{[A]_{eq}[B]_{eq}}. \quad (2.15)$$

The equilibrium concentrations of the species will be $[A]_{eq} = [A] - \delta$, $[B]_{eq} = [B] - \delta$ and $[C]_{eq} = [C] + \delta$, where δ is some change in concentration and, $[A]$, $[B]$ and $[C]$ are current species concentrations. One can solve for δ using the quadratic formula. We update the species concentrations if δ is both larger than a small threshold value and as a fraction of current species concentration δ is larger than a small tolerance value. If neither is the case, we deem the reaction to be at equilibrium.

This process of randomly selecting reactions and determining whether the species have achieved their equilibrium concentration continues until enough consecutive sampled reactions (roughly n^2 for n reactions) are within the tolerance value of the equilibrium point. This ensures that all reactions are equilibrated with high probability. At this point the algorithm terminates.

2.5.7 Supporting Figures

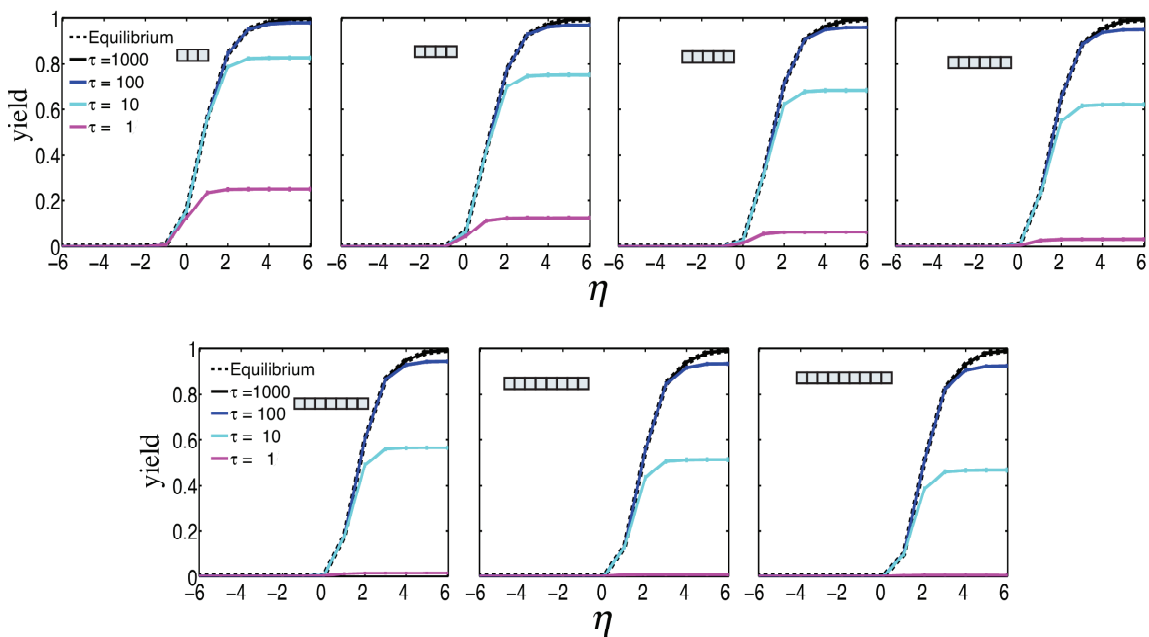


Figure S 2.3 Yields of 1x3 to 1x9 line complexes at various isothermal conditions. Dashed lines indicate thermodynamic equilibrium. Inset diagrams depict the complexes. Here, as in the main text, $\eta \equiv \log_{10}(k_{on}/k_{off,1}[X]_0)$ and $\tau = k_f[X]_0 t$. For all figures in the Supporting Information, unless otherwise noted, there is no bond coupling ($a_o = 1$) and error bars are $<1\%$.

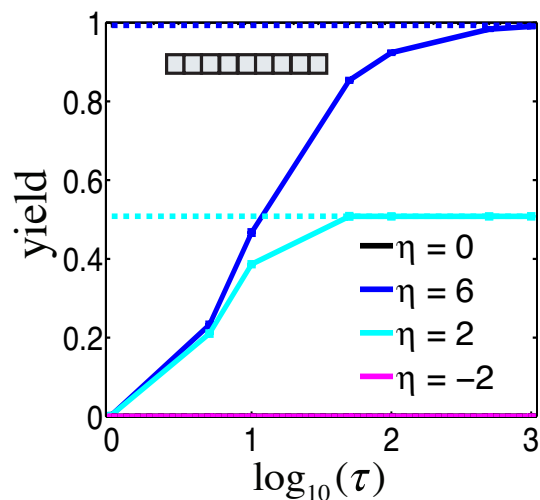


Figure S 2.4 Yield of 1x9 line complex at various reaction times, τ , subject to different isothermal assembly conditions. Dashed lines indicate equilibrium values at a given value of η . Inset diagram depicts the complex.

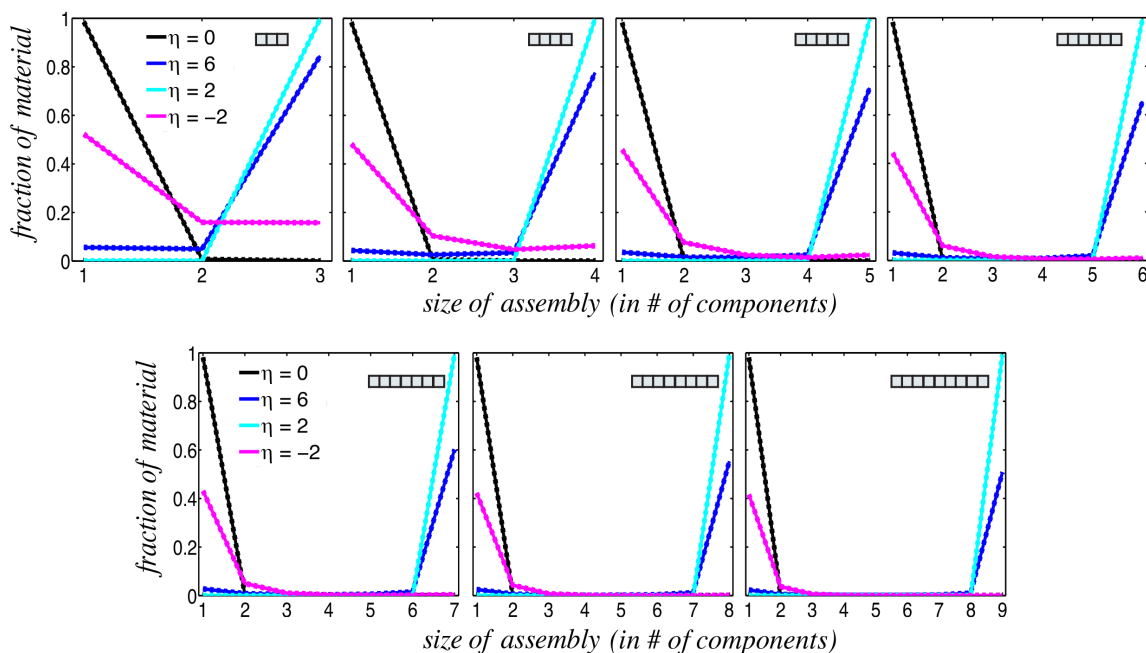


Figure S 2.5 Assembly size distribution at different isothermal assembly conditions after $\tau = 1000$. Thermodynamic equilibrium predictions are dashed lines and in all cases directly overlay the reported fractions. Inset diagrams depict the complexes.

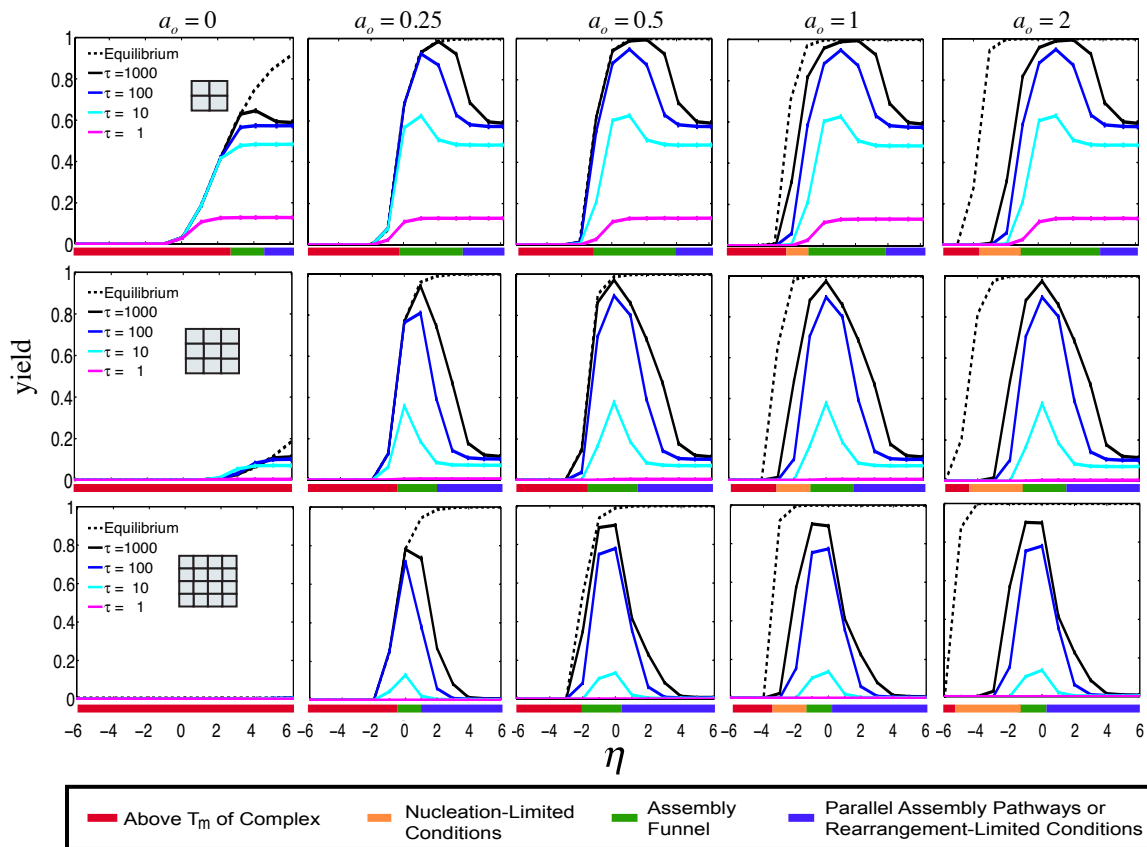


Figure S 2.6 Yields of 2x2, 3x3 and 4x4 square grid complexes at different isothermal assembly conditions and bond coupling constants (a_o). Dashed lines indicate yield at thermodynamic equilibrium. Inset diagrams depict the complexes. As bond coupling increases, intermediates and complexes become more stable (as seen by the increase in melting temperature at thermodynamic equilibrium) but nucleation rates remain approximately constant such that complex yields approach equilibrium for negative bond coupling under nucleation-limited conditions (e.g., $-3 \lesssim \eta \lesssim -1$) but remain far from equilibrium for positive bond coupling.

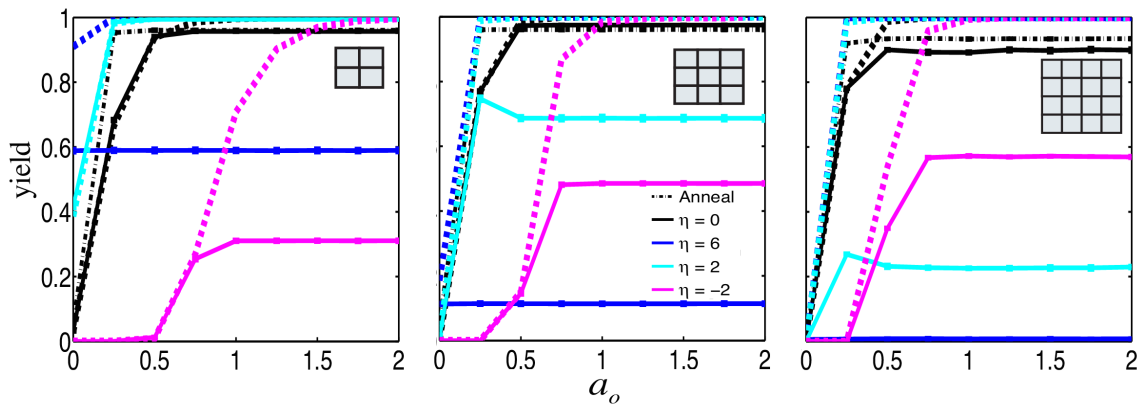


Figure S 2.7 Yields of 2x2, 3x3 and 4x4 square grid complexes as a function of bond coupling constant, a_o , at various isothermal conditions (solid lines) and anneal (dash-dot line). Dashed lines indicate equilibrium values at the given value of η . Inset diagrams depict the complexes. Above a relatively low threshold of bond coupling (whose exact value depends on assembly size and assembly conditions), assembly yields are largely insensitive to bond coupling values (see Note S4 for further explanation).

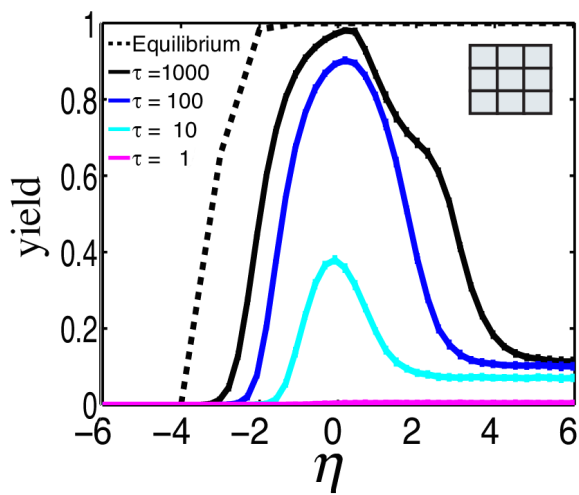


Figure S 2.8 Yield of 3x3 square grid complex for many isothermal conditions, from $\eta = -6$ to $\eta = 6$ in increments $\eta = 0.2$. Inset diagram depicts the complex.

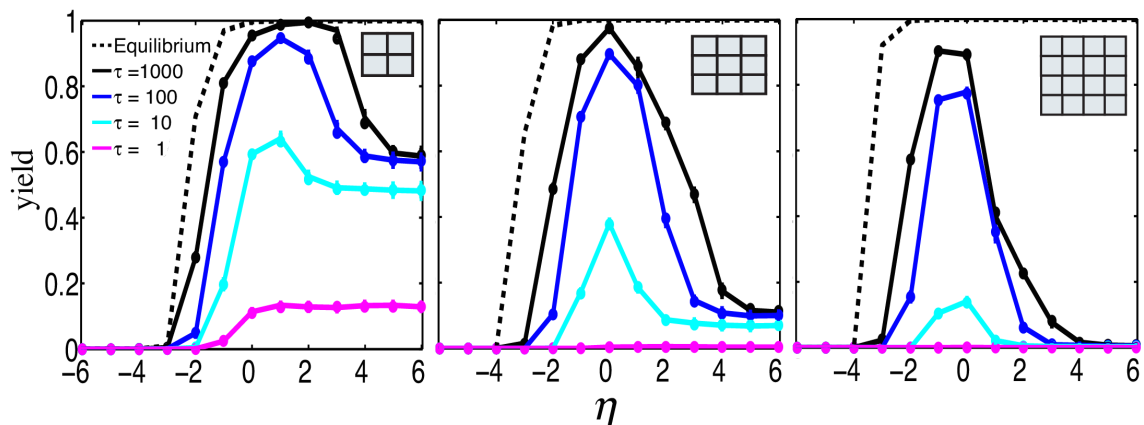


Figure S 2.9 Reducing the number of components in the simulation does not significantly affect yield predictions. Yield of 2x2, 3x3 and 4x4 square grid complexes at various isothermal conditions starting with 1000 (instead of 10000) of each component, with the simulated volume adjusted so that $[X]_o$ is unchanged. Dots indicate the yield of complexes at various isothermal conditions starting with 10000 of each component. Dashed line indicates yield at thermodynamic equilibrium. Inset diagrams depict the complexes.

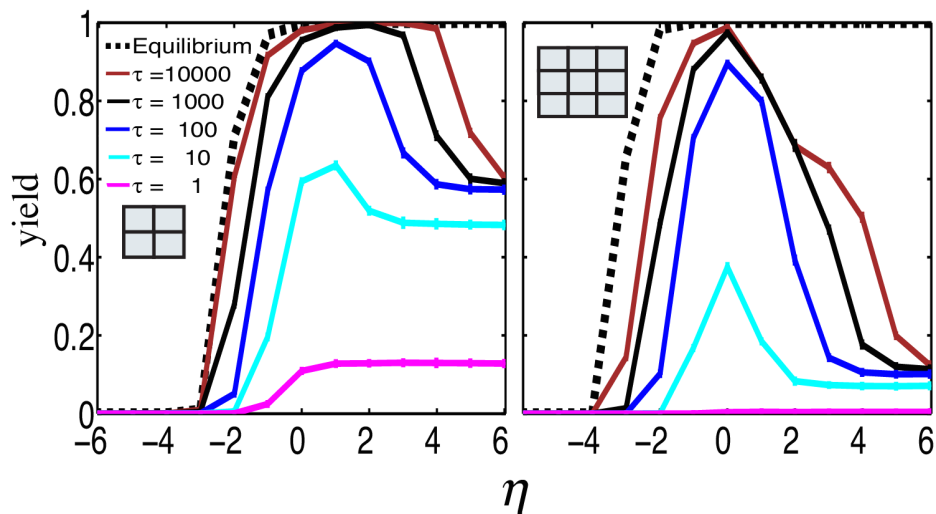


Figure S 2.10 Yield for 2x2 and 3x3 square grid complexes at various isothermal conditions, including yield predictions after long reaction times, $\tau = 10000$. Dashed line indicates the yield at thermodynamic equilibrium. Inset diagrams depict the complexes. These results suggest that further increasing assembly time beyond what we consider in the main text does not significantly increase yields under most conditions.

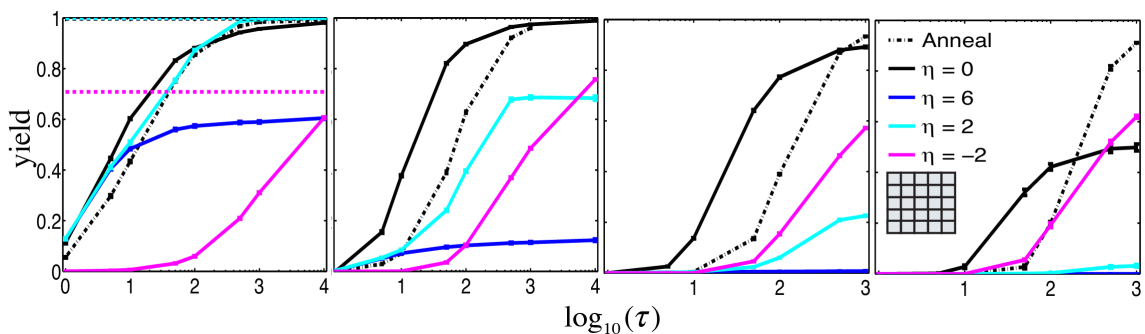


Figure S 2.11 Yields of 2x2, 3x3, 4x4 and 5x5 square grid complexes at various reaction times, τ , subject to different assembly conditions. Inset diagrams depict the complexes. Dashed lines correspond to thermodynamic equilibrium and color corresponds to the value of η . Dash-dot line connects complex yields of anneals with various reaction times, τ . For 2x2, 3x3 and 4x4 square grid complexes, $\eta = 0$ is within the assembly funnel regime, but for the 5x5 complex $\eta = 0$ is within the parallel pathways and rearrangement-limited regime.

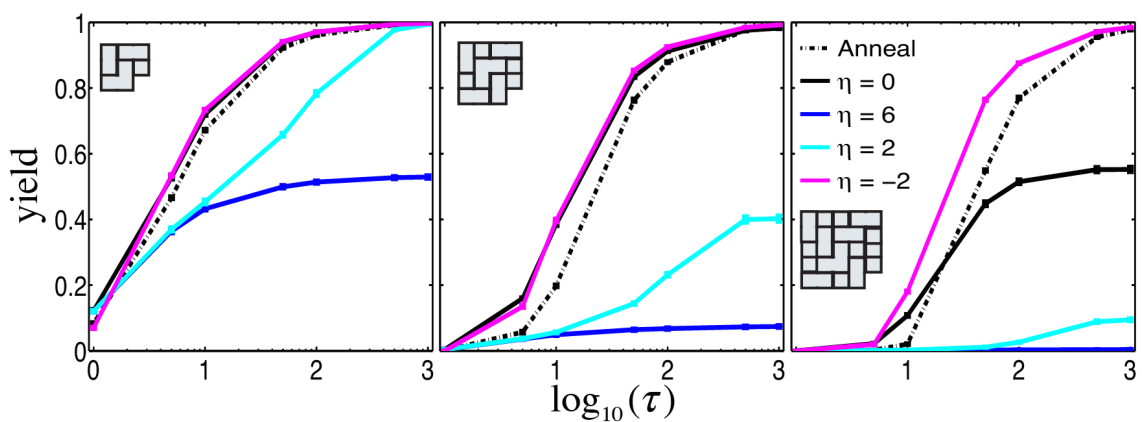


Figure S 2.12 Yield of 2x2, 3x3 and 4x4 spiral complexes at various reaction times, τ , subject to different assembly conditions. Inset diagrams depict the complexes. Dash-dot line connects complex yields after anneals with various reaction times, τ .

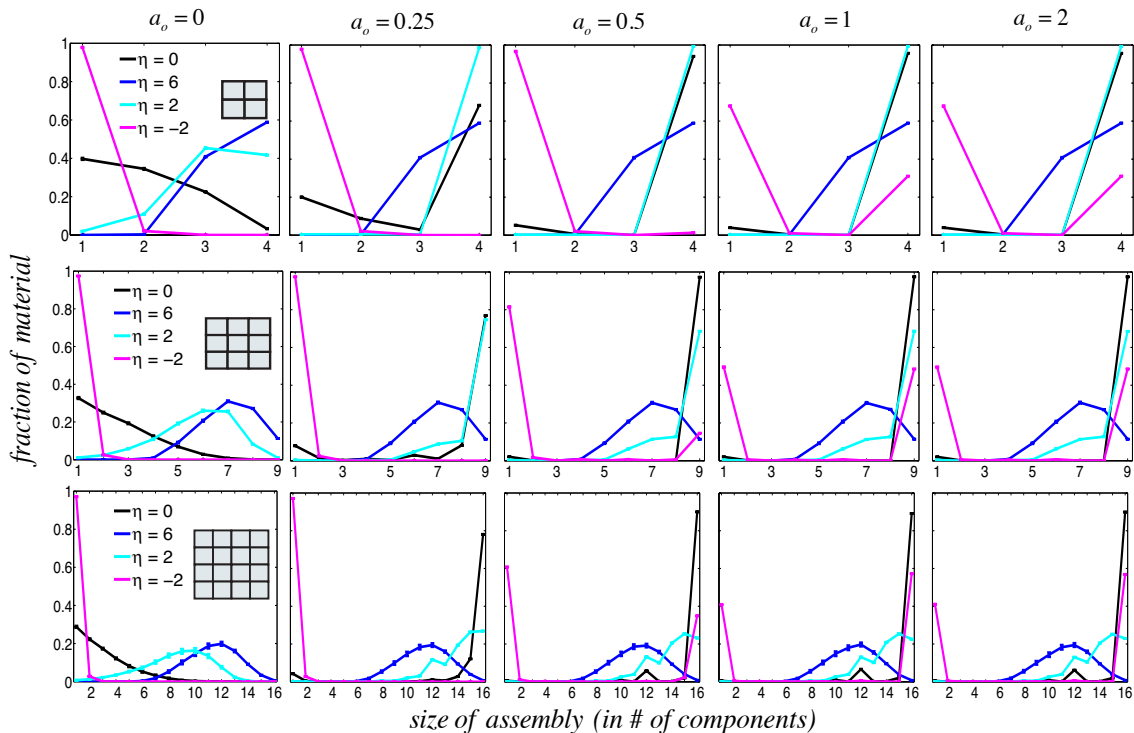


Figure S 2.13 Assembly size distributions (in # of components) for 2x2, 3x3 and 4x4 square grid complexes at various isothermal conditions and bond coupling constants. Inset diagrams depict the complexes. All plots are shown after $\tau = 1000$.

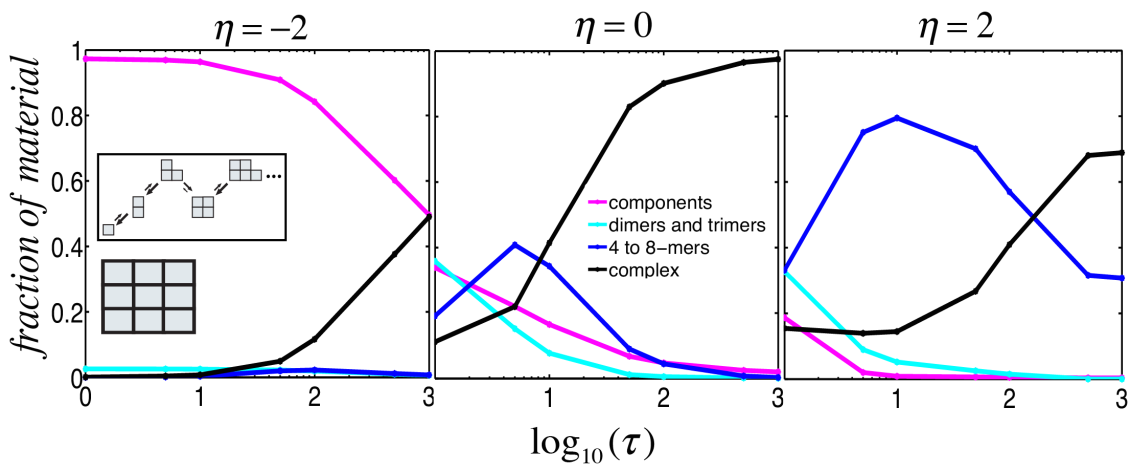


Figure S 2.14 Timescales of nucleation and rearrangement together determine the rate of complex formation. Both of these timescales are functions of complex size and geometry. The fraction of material in various species as a function of reaction time for 3x3 square grid assembly under different assembly regimes: nucleation-limited at $\eta = -2$, assembly funnel regime at $\eta = 0$

and rearrangement-limited at $\eta = 2$. Inset diagram depicts a possible reaction pathway for nucleation and arrow size indicates relative reaction propensities.

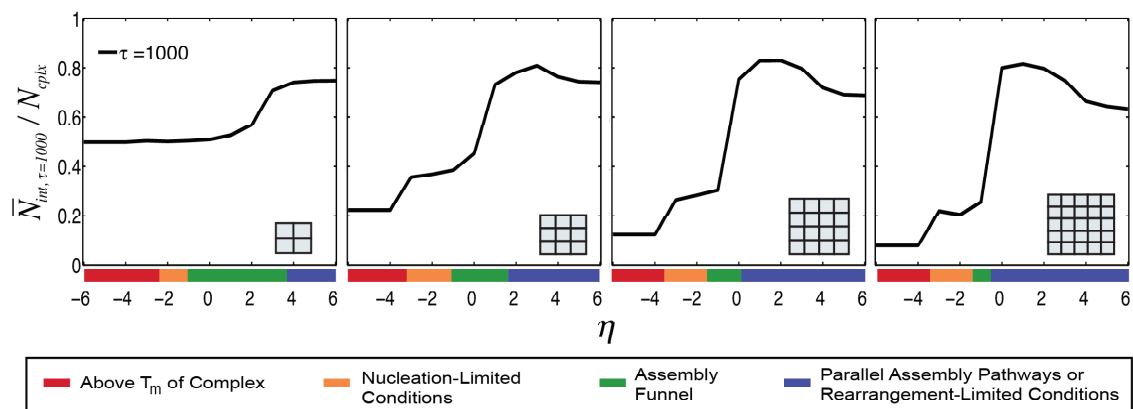


Figure S 2.15 Size distribution of intermediates for various 2D complexes. Mean size of intermediates (in number of components) after $\tau = 1000$, $\bar{N}_{int,\tau=1000}$, normalized by the number of components in the complex, N_{cplx} , for 2x2, 3x3, 4x4 and 5x5 square grid complexes at different isothermal assembly conditions. Inset diagrams depict the complexes. The mean intermediate size is defined as the mean size of the species in the system, not including complexes or components. Nucleation-limited conditions produce mean intermediate assembly sizes equal or less than half of the size of a complex whereas rearrangement-limited conditions allow intermediates to grow to be, on average, greater than half of the size of a complex.

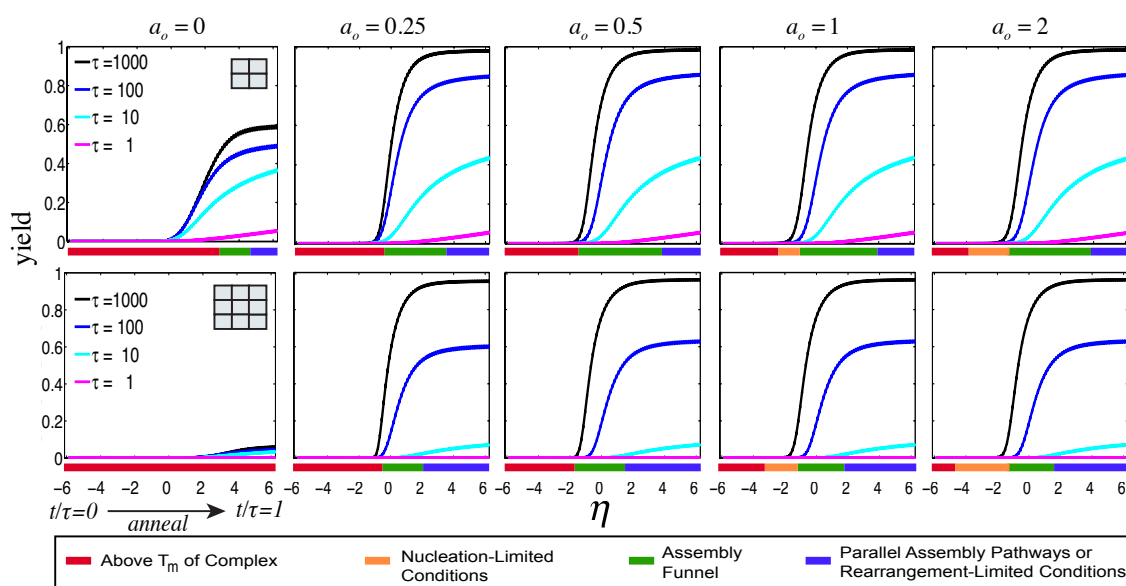


Figure S 2.16 During an anneal, most complexes are produced during the phase of the anneal that passes through the assembly funnel. Yield of 2x2 and 3x3 square grid complexes during the course of an anneal for various bond coupling constants. Inset diagrams depict the complexes being assembled. The anneal begins from left to right, with the total time of the anneal given as the τ value in the legend. The annealing process is simulated by changing the strength of component-component interactions as the reaction proceeds. At the start of the simulation ($t/\tau = 0$), $\eta = -6$ and over the course of the simulation the interaction strength is logarithmically increased 100 times, in equal reaction time intervals (i.e., $\tau/100$), to ultimately obtain $\eta = 6$ at the end of the simulation ($t/\tau = 1$). In practice, this annealing protocol corresponds to a linear decrease in temperature over time. Assembly regimes are determined by isothermal assembly (see Figure S2.6).

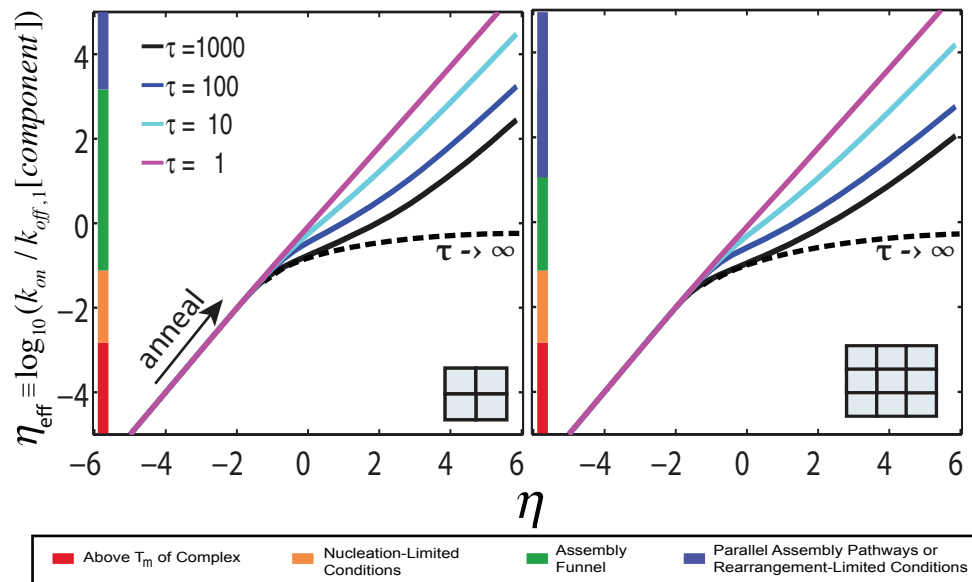


Figure S 2.17 During very long anneals, component depletion can increase the amount of time that the system effectively stays within the assembly funnel regime. Effective reaction propensity is given by $\eta_{eff} \equiv \log_{10}(k_{on}/k_{off,1}[\text{component}])$ where $[\text{component}]$ is the current average component concentration, for the 2x2 and 3x3 square grid complexes as a function of annealing conditions after various annealing times. Color bars on the left side of the figures correspond to different assembly regimes. Inset diagrams depict the complexes. Effective reaction propensities for slower anneals remain in the assembly funnel regime for longer periods of time, not only because of their increased time of anneal, but also because components are depleted during annealing. This decrease offsets the effect of the off rate ($k_{off,1}$) decreasing as the temperature decreases. As a result, during a slow anneal η_{eff} can be in the assembly funnel regime even as η drops into rearrangement-limited conditions. During fast anneals ($\tau = 1$), the off rate changes much faster than components deplete, accounting for the linear relationship between η and η_{eff} . Dashed line approximates the η_{eff} for an ideal anneal (where $\tau \rightarrow \infty$). In an ideal anneal, components would deplete in proportion to the decrease in the off rate and thus always remain in the assembly funnel regime after initially entering it.

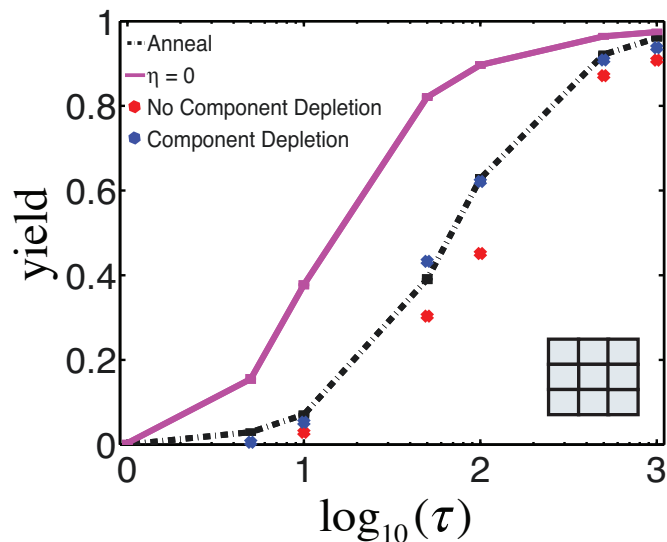


Figure S 2.18 The time spent in the assembly funnel regime can be used to predict the outcome of an anneal. Yield of 3x3 square grid complex as a function of reaction time for an isothermal assembly ($\eta = 0$) and for an anneal. Inset diagram depicts the complex. For a 3x3 square grid complex, the assembly funnel regime ranges from $-1 < \eta < 1$ (see Figure 2.2). The red and blue dots are estimated yields calculated by computing the time the anneal spends in the assembly funnel regime and, with this value, estimating yield by linear interpolation of an $\eta = 0$ isothermal assembly. With no component depletion effects (red), a given anneal of time τ_{anneal} , will spend $\tau_{assembly\ funnel} \approx \frac{\tau_{anneal}}{6}$ in the assembly funnel regime. With component depletion effects (blue, see Figure S16), the time spent in the assembly funnel regime will correspond to the time that the anneal remained $-1 < \eta_{eff} < 1$ so that the slower the anneal, the higher the fraction of total reaction time spent in the assembly funnel regime. For example, when $\tau_{anneal} = 1$, $\tau_{assembly\ funnel} = 0.18 \tau_{anneal}$ and when $\tau_{anneal} = 1000$, $\tau_{assembly\ funnel} = 0.34 \tau_{anneal}$. The method of estimating yield via annealing that includes component depletion effects more closely resembles the actual annealing yield, suggesting that component depletion effects, which serve to increase the time spent in the assembly funnel regime and in turn enhance yields, occurs during annealing.

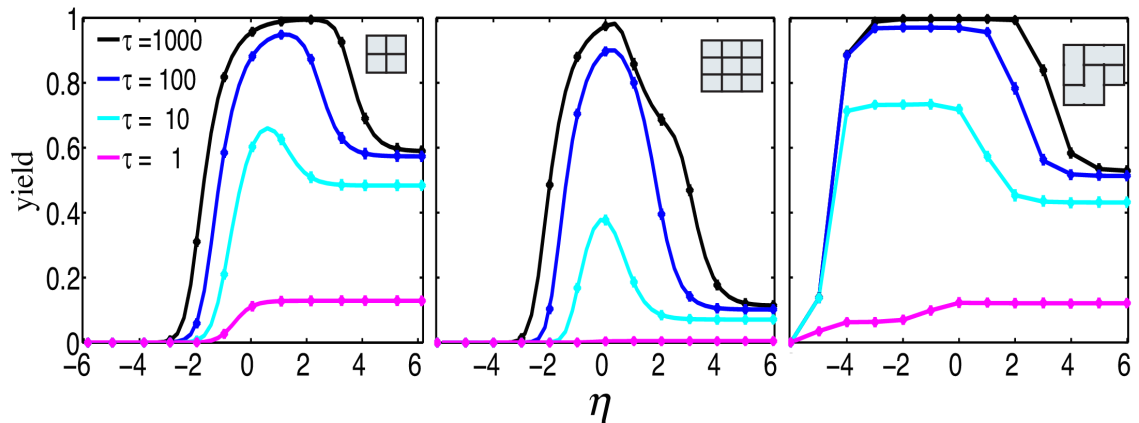


Figure S 2.19 Deterministic and stochastic solutions are almost identical. To test the similarity of the stochastic solution to the deterministic solution, we simulated the ODEs for the respective complexes using MATLAB's ode23s solver. Deterministic solution (solid lines) and overlaid stochastically sampled values (dots) of yield for 2x2 and 3x3 square grid and 2x2 spiral complexes at various isothermal conditions. Inset diagrams depict complexes.

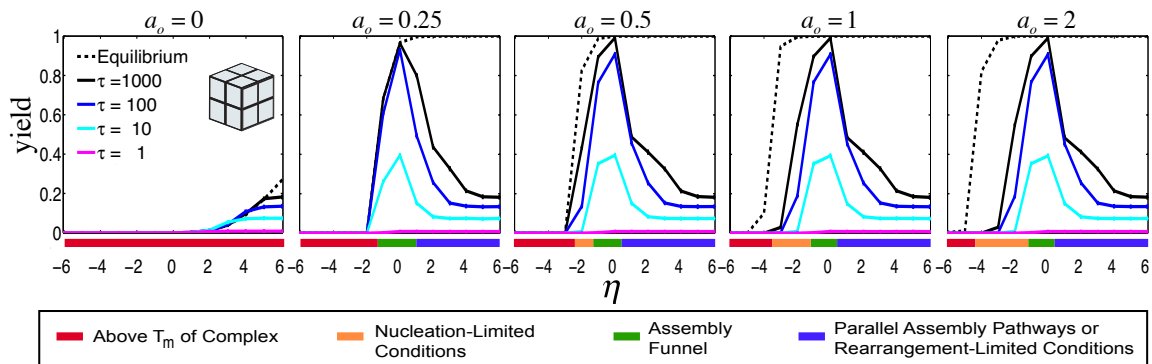


Figure S 2.20 Yields of 2x2x2 cube complexes as a function of bond coupling constants a_o at various isothermal conditions. Dashed line indicates complex yield at thermodynamic equilibrium. Inset diagram depicts the complex.

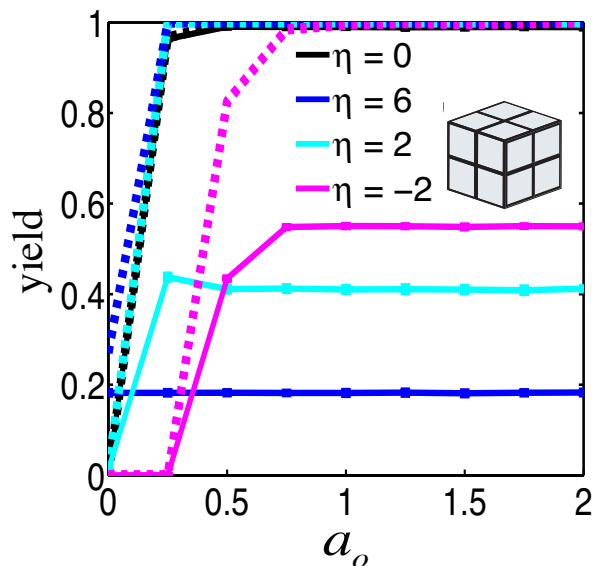


Figure S 2.21 Yield of 2x2x2 cube complexes as a function of bond coupling constant, a_o at various isothermal conditions (in terms of η). Dashed lines indicate equilibrium values at the given value of η . Inset diagram depicts the complex.

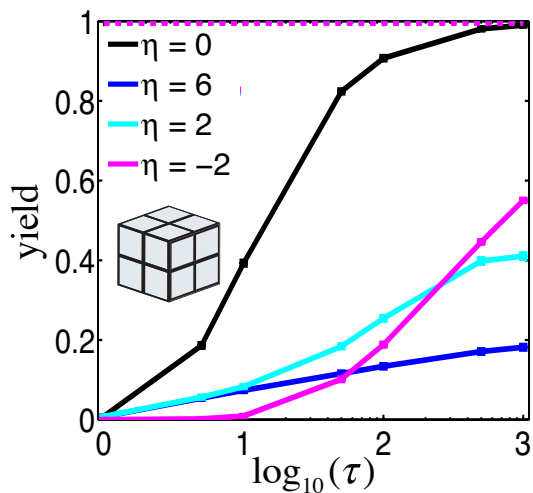


Figure S 2.22 Yield of 2x2x2 cube complex at various reaction times, τ , subject to different isothermal assembly conditions. Dashed lines indicate equilibrium values of yield at the given value of η (equilibrium yield is unity for all values of η shown). Inset diagram depicts the complex.

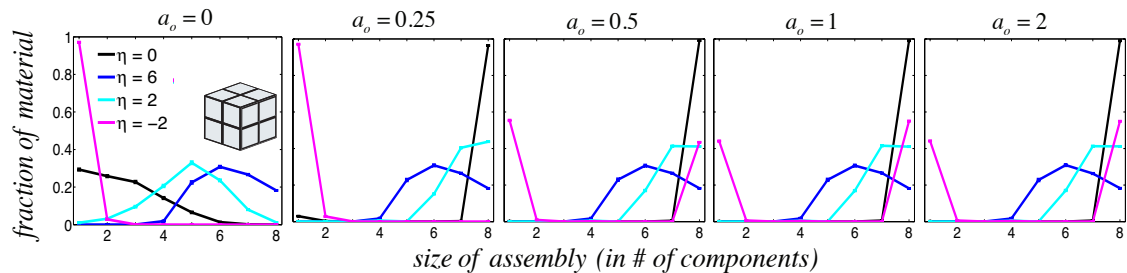


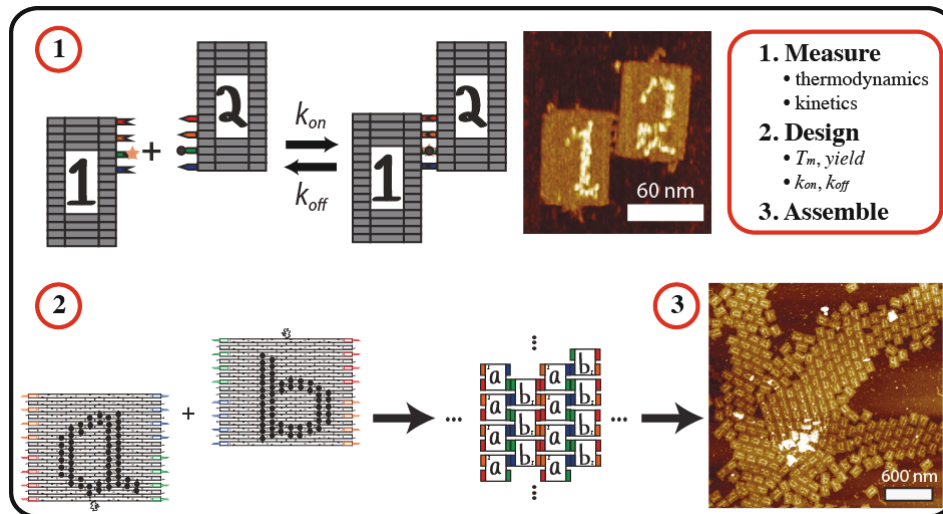
Figure S 2.23 Assembly size distributions for 2x2x2 cube complex at various isothermal conditions and bond coupling constants. All plots are shown after $\tau = 1000$. Inset diagram depicts the complex.

2.6 ACKNOWLEDGEMENTS

The authors thank Shourya Sonkar Roy Burman, Jeffrey Gray, Abdul Majeed Mohammed, Dominic Scalise and Josh Fern for helpful discussions and advice on the manuscript. This research was supported by grants from the National Science Foundation and a Turing Centenary Fellowship to R.S.

CHAPTER 3 KINETICS AND THERMODYNAMICS OF WATSON-CRICK BASE PAIRING-DRIVEN DNA ORIGAMI DIMERIZATION

SUMMARY



We investigate the kinetics and thermodynamics of DNA origami dimerization using flat rectangle origami components and different architectures of Watson-Crick complementary single-stranded DNA (“sticky end”) linking strategies. We systematically vary the number of linkers, the length of the sticky ends on the linker and linker architecture and measure the corresponding yields as well as forward and reverse reaction rate constants through fluorescence quenching assays. Yields were further verified using atomic force microscopy. We calculate values of H^0 and ΔS^0 for various interface designs and find non-linear van’t Hoff behavior, best described by two linear equations, suggesting distinct regimes of dimerization between those with and those without well-formed interfaces. We find that self-assembly reactions can be tuned by manipulating the interface architecture without suffering a loss in yield, even when yield is high, ~75-80%. We show that the second order forward reaction rate constant (k_{on}) depends on both linker architecture and number

of linkers used, with typical values on the order of $10^5 - 10^6 \frac{1}{M \cdot sec}$, which are similar to bimolecular association of small, complementary DNA strands. The k_{on} values are generally non-Arrhenius, tending to increase with decreasing temperature. Finally, we use kinetic and thermodynamic information about the optimal linking architecture to extend the system to an infinite, two-component repeating lattice system, and show that we can form micron-sized lattices, with well-formed structures up to $8 \mu m^2$.

3.1 INTRODUCTION

With the advent of DNA origami¹⁰ came the possibility of high resolution nanobreadboards, enabling unprecedented control of matter on the nanoscale³⁵. This feature has allowed the creation of DNA origami nanodevices that promise many potentially useful applications including nanoscale reactors^{11, 12, 13} and devices^{14, 15, 16}, molecular robots^{18, 19}, sensors and actuators^{20, 21, 22}, devices to control self-assembly^{23, 24}, novel drug delivery systems^{27, 28}, and molecular tools to probe or mimic biological components^{30, 31}.

One major limitation of these technologies is the amount of surface area of an individual origami onto which molecular components can be placed with nanoscale resolution, which is typically limited by the size of the DNA scaffold (usually $\sim 7kb$). Therefore the surface area of the origami only affords attachment of a small number of components, limiting the complexity and utility of the devices¹⁰³. Many methods have been proposed to overcome this problem: origami have been made from multiple scaffolds¹⁰⁴, individual origami have been assembled or organized into complexes⁶⁷, and surfaces have been used to confine the geometry in order to enhance self-assembly outcomes¹⁰⁵. The thermodynamics and kinetics of smaller DNA tiles have been well-studied^{106, 107, 108, 109}. Two- and three-dimensional DNA origami have been extensively characterized, computationally modeled and built with various inter-origami geometries^{32, 34, 110, 111}.

However, a systematic characterization of interfaces involving DNA origami components has yet to be accomplished.

Recent successes in the design of protein assemblies, including the successful design and assembly of protein-only 2D infinite lattices³⁸ and finite complexes³⁹, suggest that predicting structure and tuning interfacial energies using data-driven computational modeling software³⁷ is essential to optimize self-assembly outcomes. If DNA nanotechnologists are to reliably build larger, more complex architectures using a DNA origami breadboard, an understanding of origami interfaces, including kinetics and thermodynamics, is required³⁵.

Our goal in this paper is to understand how to control the strength and reversibility of interfacial reactions between origami components when assembly is driven by the Watson-Crick (WC) hybridization of multiple sets of short “sticky ends” (SEs). We chose this architecture because large libraries of such interfaces with low crosstalk can be designed¹¹², suggesting a way to rapidly scale the complexity of assembly reactions. We characterized the dimerization of origami tiles using a fluorescence quenching assay and correlated this assay to a robust but low-throughput measure of dimer yield, visual characterizations using an atomic force microscope (AFM). We measured multiple thermodynamic and kinetic parameters including yield, melting temperature, reaction rate constants, and ΔH^0 and ΔS^0 using the van't Hoff equation. We tested multiple different interface design strategies, including modifying the number of linkers per interface, their positional arrangement on the interface, SE length, and linker flexibility. We found that: (1) yield generally increases or remains constant with number of linkers (2) reaction outcomes can be tuned *via* interface design in order to change the dimer melting temperature while maintaining high yields, (3) dimerization occurs rapidly, with forward reaction rate constants comparable to smaller, DNA tile dimerization and complementary oligonucleotide hybridization reactions, which could suggest a unified design strategy for hierarchical DNA assembly processes, and (4) we can use the

information obtained about component interfaces in the dimerization assays to inform the design of a self-assembly system for 2D infinite lattices.

3.2 RESULTS AND DISCUSSION

As an example component for hierarchical assembly reactions involving origami components, we considered the dimerization of two 2D DNA origami rectangle tiles. The components were designed with caDNAno³² to be planar, with twist near to that of B-form DNA, *i.e.*, 10.44 bases per turn, similar to a previously described origami structure⁷⁴. CanDO^{33, 113} modeling software also predicted that the resulting origami structure would have little global twist (Supporting Information Figure S3.1 in Section 3.5.1). Hairpin staples assembled in specific positions on the two separate components allow the first component (labeled with a series of hairpins to depict the number “1”, called T1) to be distinguished from the second component (labeled with a series of hairpins to depict the number “2”, called T2) in AFM images (Figure 3.1A). The hairpins may induce unpredictable curvature to the origami components¹¹⁴. As homogeneous interfaces appear to bind to one another more effectively than heterogeneous interfaces¹¹⁵ the system was designed such that the set of scaffold sequences in the interfacial domain on T1 and T2 are the same. Therefore, upon dimerization T2 is rotated 180 degrees in plane relative to T1 in the dimer.

The interface for binding between the two components consisted of blocker edge staples, which help prevent non-specific interactions such as blunt-end stacking, and linker edge staples, which facilitate specific interactions between components *via* WC base pairing. Blocker edge staples contain a single-stranded poly-thymine (Thy4) domain on the 3' end, to mitigate blunt-end stacking¹¹⁶. Linker staples consist of five domains: a domain complementary to the origami scaffold, one or two Thy4 domains to allow linkers to traverse helices, two complementary intra-linker domains and a SE domain whose sequence is complementary to the opposite component.

The linkers were designed so that upon dimerization, the two origami components are separated by 21 bases, or approximately 2 full turns assuming B-form DNA¹¹⁷, so that the dimer would have an effectively planar structure (Figure 3.1B, C).

To understand how interface structure determines binding energy and the kinetics of component binding, we systematically altered three variables: (1) the number and (2) arrangement of the linkers at the interface and (3) the length of each SE. We considered the binding of components with interfaces containing between 3 and 7 linker pairs with either 5 or 6 base pair (bp) SEs. For interfaces with 3 and 5 linkers we tested multiple positional arrangements of the linkers.

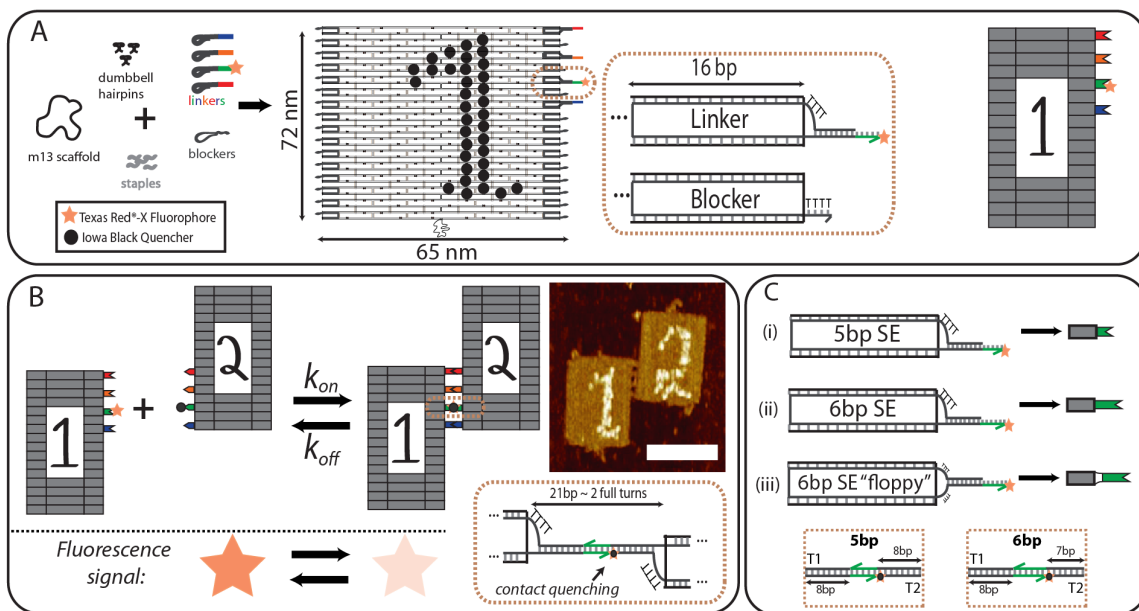


Figure 3.1. Schematic of the DNA origami tile dimer system. (A) DNA origami monomers (T1 shown) are assembled by annealing the scaffold with hairpins, blockers, linkers, and staples. Hairpins label tiles by providing height contrast on an AFM image. Blockers mitigate nonspecific interactions between origami such as blunt-end stacking. Linkers facilitate interaction specificity through a single-stranded “sticky end” (SE) domain. Far right, schematic of T1 and its interface. (B) Dimer formation occurs via SE hybridization. Complementary linkers in diagrams have the same color and complementary shapes. A fluorophore-quencher pair, drawn as an orange star and black circle, respectively, which is attached to the to the SE colored green is used to track the dimerization process. AFM image shows the assembled dimer. Scale bar is 60 nm. (C) The different linker architectures investigated in this work, including linkers with (i) 5 base pairs (bp) SEs (ii) 6bp SEs (iii) 6bp SEs floppy (two poly-Thymine regions after the scaffold-complementary region). Inset

diagrams show linking architecture between T1 (left) and T2 (right) for 5bp and 6bp linkers. For all interfaces, a 5' Iowa Black RQ on T2 quenches a 3' Texas Red®-X NHS Ester on T1.

We also tested the effects of linker flexibility by comparing linkers with and without a pair of Thy4 domains between the origami structure and the intra-linker complementary region: such linkers with a pair of Thy4 domains we call “floppy” linkers (Figure 3.1C). SE sequences were designed to reduce crosstalk and have approximately the same interaction energy per SE^{25, 99, 118} (Supporting Information Note 3.6 in Section 3.5.6).

To quantitatively characterize yield and rates of dimerization, we used a fluorescence quenching assay where the 3' end of one of the linking strands on T1 was modified with a Texas-Red ®-X (NHS Ester) fluorophore and the 5' end of the complementary T2 linker was modified with Iowa Black RQ quencher so that dimerization produces a decrease in fluorescence¹¹⁹ (Figure 3.1B). To ensure that the fluorophore-quencher interaction was solely responsible for changes in fluorescence upon binding, we performed control experiments where a fluorescently labeled T1 reacted with T2 with and without a quencher. We found that upon dimerization with a quencher, the reaction produced a dramatic decrease in fluorescence while dimerization without a quencher actually slightly enhanced fluorescence upon binding (Supporting Information Figure S3.3). To investigate whether the quenching effect was a result of origami binding or simply the complementary linkers binding in solution, we performed another control experiment where we mixed solutions of the T1 and T2 staples and linkers only, which produced no fluorescence change for 5bp linkers (Supporting Information Figure S3.4). About 5% of floppy linkers dimerize or bind to an origami interface at 25°C (Supporting Figure S3.5A). Binding between free linkers with 6bp SEs produced a significant fluorescence change below 35°C (Supporting Information Figure S3.5B) and indicated a ~20% yield at 25°C. The binding energy of origami dimerization for a given temperature can be measured by determining the yield of dimerization at equilibrium for known concentrations of monomer reactants. To find an assembly protocol that would stay close to

equilibrium as the temperature was changed, we mixed 5 nM of each of the two origami monomer types at equal concentrations (see Methods in Section 3.4) and measured the fluorescence at different temperatures for a set of different annealing and melting speeds. We found that for linkers with 5bp SE and floppy linkers, there was no significant hysteresis between 25°C and 55°C when the temperature decreased at 1°C per 15 minutes during annealing and increased at 1°C per 15 minutes during melting (Figure 3.2A and Supporting Information Figures S3.7-S3.8, S3.11-S3.12), suggesting that the system stayed close to equilibrium during both the heating and cooling processes. Decreasing the melting and cooling rates did not significantly change fluorescence values at any point during the process (Supporting Information Figure S3.13). The interfaces with 4-7 6bp SE linkers required a longer time at each temperature to achieve equilibrium, as hysteresis was observed when melted and annealed using the above protocol. Thus, samples with these interfaces were heated and cooled at a rate of 1°C per 60 min (Supporting Information Figures S3.9-S3.10, S3.14). In some cases the fluorescence signals from the first melt were noisy compared to the signal from subsequent anneal and melt cycles, which may be due to noise in the fluorescence signal (Supporting Information Note 3.2 in Section 3.5.2). To collect fluorescence data, we therefore melted and annealed each pair of components for at least 1.5 cycles (melt → anneal → melt) and excluded data from the first melt. In all anneal and melt experiments, the temperature never reached the melting temperature of the origami components (Supporting Information Figure S3.2).

Fluorescence-quenching assay can be used as a proxy for dimer yield with proper calibration. To verify that the fluorescence signal from these experiments could be used to determine dimerization yield, we also measured the yield of origami dimers at different temperatures using atomic force microscopy. We scanned random portions of a mica puck prepared after an assembly reaction proceeded in solution at a particular temperature (see Methods in Section

3.4) and using the labels on each component, classified the observed assemblies as either T1, T2, D (dimer) or a flipped dimer (Figure 3.2B). By combining the AFM data with the fluorescence-quenching assay data, we correlated the change in fluorescence to dimer yield and produced a calibration curve that indicated that the two measurements were linearly related (Figure 3.2C). To account for differences in fluorescence quenching efficiency for different linker architectures, we produced different calibration curves for each linker architecture using a distinct set of AFM measurements of yield (Supporting Information Figure S3.6) and used these calibration curves to convert fluorescence signals into a quantitative measure of yield. Notably, yield did not approach 1 at the coldest temperature studied in these curves but instead, yield varied widely at that temperature, depending on interface design. In producing the calibration curves, we assumed that quenching efficiency varies only with linker architecture (*i.e.* SE length and whether the linker contains a Thy4 insert) and not with number of SEs, as the linker architecture is the determining factor for the positions of the quencher and fluorophore, which ultimately determine fluorescence quenching efficiency.

Dimer yield at 25°C tends to increase with number of linkers per interface but saturates at ~80% yield. We found that in most cases, yield is strongly dependent on temperature as would be expected for a reaction driven by DNA hybridization¹²⁰. However, unlike the hybridization of two DNA strands, the melting transition was very broad, with yields between the minimum and maximum stretching over more than 20°C for most types of interfaces (Supporting Information Figures S3.7-S3.12). We found that for 5bp and floppy linkers, increasing the number of linkers, up to about 6 per interface (Figure 3.2D) tended to increase yield. Increasing the number of linkers to more than 6 did not necessarily produce a higher yield. This result echoes studies performed on the energetics of interaction between smaller DNA nanostructures, where increasing the number of linkers above a threshold did not increase the melting temperature¹⁰⁷. However,

linkers with 6bp SEs produced higher yields than floppy linkers at all temperatures tested with yields remaining high and approximately constant as the number of linkers per interface increased. The interface with three short 5bp SE produced the smallest yield of any interface, which is likely because the interactions between these sticky ends are not strong enough to maintain a bond between origami components. Interfaces with 6bp linkers produced the highest yields up to about 5 linkers per interface, thereafter the 5bp linkers and 6bp linkers had similar yields. Furthermore, while the arrangement of the linkers is important, there was not a clear pattern as to how linker arrangements affect interface energetics; one possibility is that arrangements of linkers that could form with relatively little structural distortion produced stronger interactions.

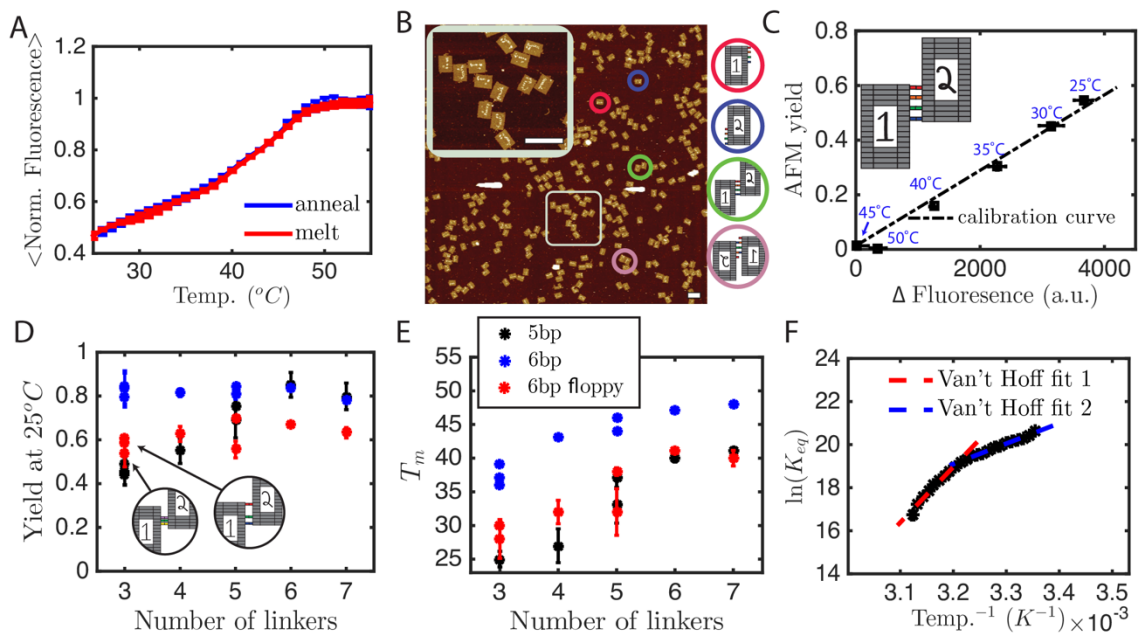


Figure 3.2. Measurement of dimerization yield and thermodynamic parameters. (A) Typical anneal / melt fluorescence curve depicting the averaged, normalized fluorescence of three mixtures of T1 and T2 with the same interface design that are cooled and then heated across a range of temperatures to produce a reversible binding transition (floppy 4 SE interface shown). (B) AFM scan of a T1-T2 mixture. Potential products include T1 alone (red) T2 alone (blue), the dimer D in correct orientation (green) and the dimer D in flipped confirmation (pink). Scale bars are 200nm. (C) Normalized fluorescence measurements of dimerization yield are linearly related to measurements of yield determined from AFM micrographs and were used to quantitatively measure origami yield (5bp 4 SE interface shown). All yield-fluorescence calibration measurements were made with the four-linker architecture (as seen in inset diagram). (D) Dimer yield and (E) melting temperature, T_m (defined as the temperature at which the yield is 50%), as a function of linker architecture and

number of linkers. Multiple data points at a given number of linkers indicate different linker arrangement on the origami, as seen in the inset of (D-E); color key is the same in both plots. (F) Equilibrium constants of dimer binding as a function of temperature, shown as a van't Hoff plot. The curve shape is typical of origami dimerization for the systems we studied, with two different linear regions. Error bars indicate one standard deviation of the reported quantity and are too small to be seen in (F). Bootstrapping is used to determine error bars for the AFM yield and are too small to see in (C).

In the AFM images, most of the SEs between dimers appeared to be bound, although this might not be the case in solution. One important result deduced from AFM imaging is that the maximum yield of dimers for the temperatures we studied is well below 100%. The maximum yield we obtain at 25°C is ~80% (Figure 3.2D). We hypothesized that this relatively low maximum yield could be caused by linker swapping, where linkers or edge staples with the same scaffold complementary could swap positions with a specific linker, reducing the strength of the interface. However, we found that less than ~5% of linker swap positions when heated to 55°C, suggesting that linker swapping is not the major mechanism limiting binding yield (Supporting Information Figure S3.30). We then hypothesized that yield limitations could be caused by the fact that linkers were purchased with standard desalting, and not purified, so that linkers that are defective or truncated due to synthesis errors could be incorporated into the origami, affecting the binding properties of the interface. To test this hypothesis, we obtained yield measurements *via* AFM for origami with PAGE purified linkers in the 6bp 4SE interface. When directly comparing the yields measured with AFM at 25°C, we that the yield achieved with purified linkers was 83%, or about 10% higher than the case of linkers which are synthesized with standard desalting (Supporting Information Figure S3.9). Although incremental improvements in yield can be achieved via linker purification, the modest yield enhancements might not be worth the typically high cost or laborious process of purification, especially if larger, more complex, multicomponent systems were to be created. It is not clear whether limitations on yield are due to imperfections in the structures being assembled (so that some structures never participate in a binding reaction) or whether origami interfaces reversibly achieve an equilibrium yield of significantly less than 1 over a broad range of

temperatures. Importantly, these results suggest that studies of DNA nanostructure hybridization should not assume maximal yield at low temperatures, but instead should calibrate fluorescence studies using a separate metric.

Melting temperature generally increases with the number of linkers per interface. We next wanted to determine how different interface designs affected the thermal stability of the dimer, so we calculated the melting temperature, T_m , defined as the temperature at which yield is 50%. For all linkers, the T_m increased on average as the number of linkers in an interface increased (Figure 3.2E), suggesting multivalent interactions provide thermal stability to the structure. This effect has been observed in the assembly of smaller DNA tiles¹⁰⁷ and also in DNA-mediated colloid assembly¹²¹, where increasing the surface density of SEs per colloid increases the T_m . Floppy linkers produced lower T_m values than non-floppy linkers for all of the number of linkers per interface studied, and had similar T_m values to interfaces with 5bp SEs. Floppy linkers have a larger entropic penalty for forming contacts due to the increased configurational space for linkers to explore, as the poly-T region adds additional translational and rotational degrees of freedom, which likely reduces the thermal stability of the dimer. In general, this method of adding more linkers to an interface can be used to thermally tune a self-assembly reaction. In fact, in the case of 6bp linkers, one could modulate the number of linkers per interface, thereby tuning the thermal properties of the reaction, without suffering a loss in yield.

Two linear van't Hoff fits most accurately describe multivalent DNA origami dimerization. Our next goal was to measure thermodynamic parameters, ΔH° and ΔS° , of dimerization. For each dimerization reaction, we plotted the log of the equilibrium constant, $\ln(K_{eq})$, for binding as a function of inverse temperature to create a van't Hoff plot. For a simple chemical reaction, a van't Hoff plot should show a linear relationship between inverse temperature, with the slope corresponding to $-\Delta H^\circ/R$ and the y-intercept corresponding to $\Delta S^\circ/R$, which is what

we observed for two of the 5bp 3SE interfaces. However, for all other interfaces, we observed a nonlinear relationship between $\ln(K_{eq})$ and inverse temperature. Specifically, we observed that two linear fits to the van't Hoff plot, one at high temperatures (~40-55°C, labeled “van't Hoff fit 1”) and another at low temperatures (~25-40°C, labeled “van't Hoff fit 2”), were appropriate (Supporting Information Figures S3.7-S3.12). At high temperatures, K_{eq} changed rapidly, likely a result of interactions occurring between well-formed interfaces (i.e., interfaces with defect-free linkers bound to defect-free components). At lower temperatures, K_{eq} changed less dramatically, possibly because: (1) some interfaces are defective (i.e. they contain less than their specified number of full-length linkers) due to truncation errors in DNA synthesis, (2) origami tiles are malformed, creating variable energetics between components and/or (3) for 6bp linkers the free linkers compete with origami to bind to an interface thus slowing or disallowing origami dimerization. This dimerization model also explains the broad (typically > 20°C) melting transition we observed: we are not simply observing one reaction happening between well-formed components with well-formed interfaces, but instead are observing multiple reactions occurring between a distribution of components and their interfaces, thus leading to a broad melting transition. Therefore, when we examine the ΔH° , ΔS° and ΔG° for each interface, we believe the van't Hoff at higher temperatures is likely to more accurately reflect the interaction energy of the well-formed interface (Supporting Information Tables S3.1-S3.3 and SI Figures S3.15-S3.16).

Next, we wanted to characterize the kinetics of origami dimerization as a function of interface design and solution temperature. To measure the forward (k_{on}) and reverse (k_{off}) reaction rates, we performed temperature jump experiments, a technique which has been widely used for studying interaction kinetics of biomolecular complexes, especially between complementary RNA or DNA strands^{122, 123, 124} and monitored reaction progress with a fluorescence quenching assay. In these experiments, we perturbed the system from equilibrium by either rapidly

cooling the solution from a higher temperature to a final lower temperature (to determine k_{on}) or heating the solution from a lower temperature to a final higher temperature (to determine k_{off}) (see Methods in Section 3.4 and Supporting Information Note 3.4 in Section 3.5.4). From fluorescence measurements, we obtained the dimer yield as a function of time and performed a least squares fit of k_{on} to the equation (Figure 3A):

$$\frac{d(\text{yield})}{dt} = k_{on} \left((1 - \text{yield})^2 C_0 - \frac{(\text{yield})}{K_{eq}} \right), \quad (3.1)$$

where C_0 is the initial concentration of an origami component and K_{eq} is the equilibrium constant, defined mathematically as $K_{eq} = \frac{[D]_e}{[T1]_e [T2]_e} = \frac{\text{yield}_e}{C_0 * (1 - \text{yield}_e)^2}$, where $[D]_e$, $[T1]_e$ and $[T2]_e$ are the concentrations of dimer, T1 and T2, respectively, at equilibrium and yield_e is the yield at equilibrium. Values of yield_e for all temperatures studied were experimentally determined from the anneal/melt experiments. Similarly, to determine the first order reverse rate constant, we fit k_{off} to the following equation (Figure 3.3C):

$$\frac{d(\text{yield})}{dt} = k_{off} (K_{eq} (1 - \text{yield})^2 C_0 - \text{yield}). \quad (3.2)$$

Interestingly, we found that forward reaction rates for all interface designs are in the range of $\sim 10^5 - 10^6 \frac{1}{M \cdot s}$ (Figure 3.3B, Supporting Information Figures S3.17-S3.22), similar in scale to bimolecular association rates for both smaller double-crossover DNA tiles, which range from $\sim 10^5$ to $10^6 \frac{1}{M \cdot s}$ ^{108, 109} and DNA oligonucleotides, which range from $\sim 10^6$ to $10 \frac{1}{M \cdot s}$ ^{45, 125, 126}. Furthermore, at the temperatures we tested (45°C to 25°C), the forward reaction rate constants do not monotonically increase with increasing temperature as would be expected for many bimolecular chemical reactions¹²⁷ or even dimerization of DNA nanostructures at lower temperatures¹⁰⁸. Such non-Arrhenius behavior has been previously reported in multiple studies of dsDNA duplex formation^{128, 129} and is explained by the prevailing model of successful DNA duplex formation^{128,}

¹³⁰. In this model, dimerization begins *via* a slow nucleation step involving two or three bases on either helix aligning to nucleate a metastable intermediate and then proceeds through a fast zipping-up mechanism to form the complex. In short, non-Arrhenius behavior arises from the fact that more bases are required to form a nucleus at higher temperatures, making the nucleation rate and thus the overall forward reaction rate, slower¹³⁰. Thus, we did not observe clear trends for k_{on} with temperature, as we would expect to see if we explored lower temperatures (~10-25°C) where the number of bases needed to nucleate the formation of a duplex is constant¹³⁰. The maximum value for k_{on} was found to be $5.7 \times 10^6 \frac{1}{M \cdot s}$ for multiple 6bp interfaces at 25°C, a similar order of magnitude for complementary DNA oligonucleotides, for components with multiple architectures of 6bp interfaces (Supporting Information Figures S3.19-S3.20). At the same temperature, the minimum value of k_{on} for interfaces that produced larger than a 50% yield was the floppy 3 SE interface with a value of $8.6 \times 10^5 \frac{1}{M \cdot s}$, which is more than 5X slower than the fastest reaction rate constant. Such a degree of difference in rates suggests that one could use this information to tune the rate of assembly as a method for controlling the assembly pathways.

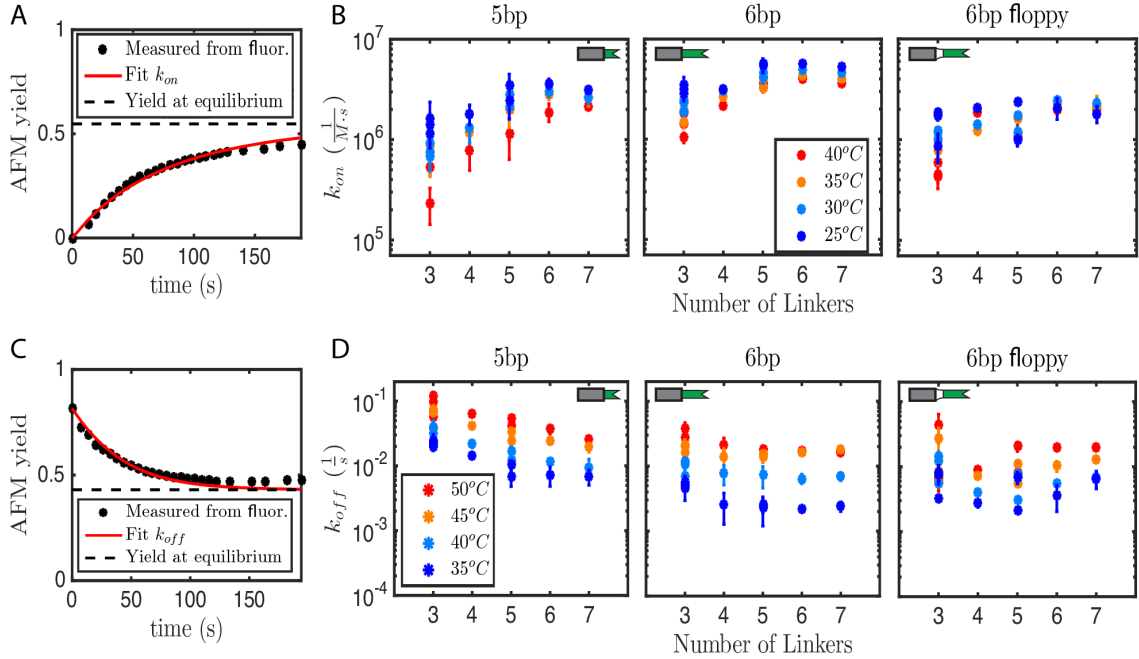


Figure 3.3. Reaction rate constants for multivalent DNA origami dimerization. (A) A temperature jump experiment used to measure k_{on} for one dimerization reaction (*e.g.*, from 50°C to 25°C). For each such experiment, the k_{on} is determined by fitting the data to a second order reaction with a known equilibrium constant. (B) k_{on} as a function of temperature, linker design and the number of linkers for the (left) 5bp SE linkers, (middle) 6bp SE linkers and (right) 6bp SE floppy linkers. (C) A temperature jump experiment (*e.g.*, from 25°C to 50°C) used to measure k_{off} for one dimerization reaction. (D) k_{off} as a function of temperature, linker design and number of linkers for the (left) 5bp SE linkers, (middle) 6bp SE linkers and (right) 6bp SE floppy linkers. Inset linker diagram depict type of linker. Error bars indicate one standard deviation of the reported quantity and in some cases are too small to see.

Increasing the number of linkers per interface tends to increase rates of association.

Additionally, multiple short SEs led to faster association rates than fewer SEs. This finding is consistent with findings in small DNA tile dimerization¹⁰⁸. Multivalent interactions increase the dimerization probability because they increase the frequency of nucleation: as the first SE partially hybridizes to its complementary SE, the effective concentration of complementary SEs increases. Intuitively, this makes sense because a greater number of linkers provide more opportunities for the SEs to nucleate an intermediate before zipping up. Floppy linkers have lower association rate constants than their non-floppy counterparts, possibly due to the fact that upon nucleation with one

or more complementary SEs, the other floppy linkers' SEs are likely misaligned, requiring more nucleation events to occur on average before a successful dimerization event.

Increasing the number of non-floppy linkers per interface tends to decrease rates of dissociation. As we expect for all interfaces, we found that k_{off} increased or stayed approximately the same with increasing temperature (Figure 3.3D, Supporting Information Figures S3.23-S3.28). In general, shorter SEs produce higher k_{off} values at a given temperature and number of linkers. Values of k_{off} for interfaces with 5bp SE tended to decrease logarithmically with the number of linkers at a given temperature. Interestingly, 6bp linkers generally had constant k_{off} values at a given temperature for interfaces with more than three linkers while floppy linkers had no clear trend for k_{off} with the number of linkers. We compared the methods for generating K_{eq} , either by fitting k_{on} and k_{off} or by measuring the component concentrations at equilibrium and found in most cases the separate, although not independent, measurements produced similar values (Supporting Information Figure S3.29).

Extending the two-component origami dimer system to a 2D “infinite” lattice enables growth of large structures, including tubes. Finally, we sought to use our findings of the thermodynamics and kinetics measurements of interfaces to optimize the assembly of a two-component, infinite lattice with multiple interfaces. As the interface with 6bp SE and four linkers produced high yields and a favorable T_m (*i.e.*, above the temperature at which non-specific aggregation and linker interference becomes prevalent and below the origami component melting temperature), we designed two origami tiles with four different sets of interfaces with 4 linkers and 6bp SEs which upon successful assembly would produce a diagonal two-dimensional cocrystal with a repeating pattern similar to a perpendicular striped design¹³¹. We labeled the tiles with hairpins “a” and “b” (Figure 3.4A) to distinguish them. Our strategy for 2D lattice assembly was to anneal in highly reversible regimes (*i.e.*, where the derivative of the melt/anneal curve is

maximum and just below; specifically, 40-50°C) with slow cooling, to promote growth of large, defect-free lattices¹³². In designing this annealing process, we assumed all of the interfaces have identical kinetic and thermodynamic constants and thus expect the extent of growth and growth rates for all interface directions to be equal, but might not be the case¹¹⁵.

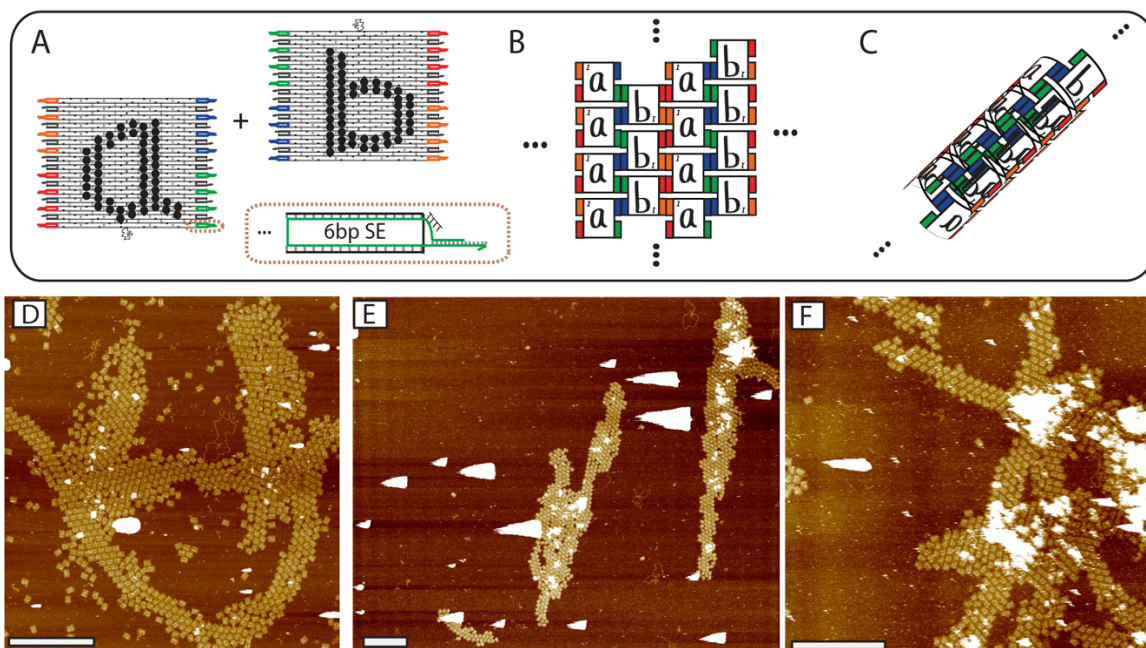


Figure 3.4. Two-dimensional origami tile lattice with two tiles labeled “a” and “b” designed with four 6bp SEs per interface. (A) Schematic of tiles “a” and “b” that form a lattice. These tiles are structurally identical to T1 and T2, except for the hairpin staples and linking schematic. For simplicity, linker edges are shown in a uniform color, although each SE sequence and scaffold-linker complementary region have unique sequences and are not self-complementary. (B) Schematic of 2D lattice (intended) and (C) tube (unintended) as possible confirmations of the “a”-“b” lattice. (D-F) AFM images of origami lattices annealed with different protocols: (D) 55 to 44°C at a rate of -0.5°C per hour; (E) 55 to 36°C at a rate of -0.5°C per hour; (F) 55 to 38°C at a rate of -1°C per hour. Scale bars located at the bottom left of the AFM images are 1 μm .

We expected that flat origami components would grow to produce a flat 2D lattice (Figure 3.4B) and found that indeed lattices grew and produced large structures, up to 8 μm^2 , with relatively few defects (Figures 3.4D–F, Supporting Information Figure S3.31). We also found that along with flat lattices, tubes form in solution (Figure 3.4C) and then unfurl on the mica surface, as evidenced on the AFM by a constant width lattice, with lengths approaching 10 μm (Figure 3.4E, Supporting

Information Figure S3.31). After finding evidence for large, tube-like structures forming in solution with the 6bp linkers, we hypothesized that the 6bp floppy linkers might be more conducive to forming extended lattice structures, as the flexibility in the linker might increase the entropic penalty for forming a tube. However, we also observed lattices consistent with nanotube formation when using the floppy linker, including some large 2D flat lattices with defects and long, constant width lattices suggesting tube-like structures in solution (Supporting Information Note 3.5 in Section 3.5.5). Furthermore, the floppy linkers produced lattices with a higher number of defects, as observed in the constant width structures on the AFM images, suggesting that non-floppy linkers provide more structural stability for large lattices than floppy linkers.

Other attempts to cocrystallize a planar infinite lattice from two rectangular origami components, such as Liu et al.¹³³, have produced structures with high aspect ratios, which are consistent with tube formation. Our findings here echo those observations and support the authors' subsequent strategy of designing components that allow helical axis-only growth by using components with orthogonal helical domains. Our findings reinforce this idea for rectangular, single layer origami components: growth along the transverse axis (*i.e.*, perpendicular to the helical axis) is likely to result in tubular structures.

3.3 CONCLUSION

Using fluorescence quenching assays and AFM calibration, we investigated how the structure of Watson-Crick base pair driven interfaces control the thermodynamics and kinetics of origami assembly. We found that equilibrium is rapidly achieved with interfaces composed of multiple short SE domains and that increasing the number of linkers per interface tended to increase the forward reaction rate and decrease the reverse reaction rate, suggesting multiple linkers work together both to speed up the reaction and to stabilize the dimer. We found that the kinetic and

thermodynamic properties (*e.g.*, T_m , k_{on}) of a self-assembly system can be tuned by manipulating the interface architecture (*e.g.*, number of linkers, length of SEs). Furthermore, some “floppiness” in linker design decreases the thermal stability, yield and reaction rates (*i.e.*, lower k_{on}) as compared to non-floppy linkers, likely as result of the increased entropic penalty associated with aligning the SEs. When comparing infinite lattices with non-floppy linkers to those with floppy linkers, we found the non-floppy linkers produced structures with fewer defects. Therefore, generally speaking, introducing floppiness into a linker is an unfavorable design strategy.

The ability to design biomolecular interfaces with tailored kinetic and thermodynamic properties is essential to reliably control self-assembly. The kind of analysis used in this study (*i.e.*, one that gives useful assembly parameters such as yield, k_{on} , k_{off}) will be helpful to being able to build and control assembly processes, as well as engineer the assembly pathways^{81, 134}. At the moment, we cannot predict thermodynamic and kinetic properties of an interface given the number of linkers and the length of their SEs, for example, but it will be important to develop theories or models that accurately describe these aspects of DNA nanostructure assembly. The experimental methods developed here will be important for providing input for developing these models and for testing them. Ultimately, by tuning the strengths of interfaces as well as their kinetic properties, finite structures and infinite lattices can be built reliably.

This work suggests that while proper interface structure can improve the yields of origami assembly, other effects, such as the fidelity of the interface and the structure of the origami components themselves are likely to be important in achieving reliable hierarchical assembly of origami structures. As evidenced by the tube formation in assembling “infinite” lattices, component curvature and flexibility play an important role in determining the final self-assembled structure.

3.4 MATERIALS AND METHODS

3.4.1 Self-assembly of DNA nanostructures.

We obtained m13mp18 ssDNA scaffold from Bayou Biolabs and all ssDNA staples from Integrated DNA Technologies (IDT) in RNase free water at a stock concentration of 100 μ M. All DNA strands were purchased with standard desalting, except for the fluorophore and quencher strands, which were HPLC purified, and the purified linkers were ordered PAGE purified. Unless otherwise noted, all experiments were performed at the following concentrations: 10X body staples, 5X blocker staples and 3X linker staples, where “X” indicates concentration relative to the scaffold. The scaffold concentration was 10 nM. All samples were prepared in TAE/12.5mM Magnesium Acetate Tetrahydrate (TAE Mg⁺⁺) buffer at a volume of 150 μ l.

3.4.2 AFM yield measurements.

AFM samples were handled under isothermal conditions, using a glove box with PID fan temperature control (Coy Labs). Buffers, pipettes, pipette tips, scotch tape and mica puck were left in the glove box for at least 30 minutes prior to sample prep in order to achieve thermal equilibrium. We then deposited 3 μ l of sample on the mica surface and performed three buffer washes with TAE Mg⁺⁺ in order to eliminate loosely adsorbed staple strands. To hinder origami dimerization on the mica surface, we then applied TAE Mg⁺⁺ also containing 5mM Nickel(II) Acetate Tetrahydrate in order to strongly adsorb the DNA nanostructures to the mica surface¹³⁵. The sample was imaged at room temperature.

3.4.3 Kinetics and Thermodynamics Measurements.

All fluorescence readings were performed using either a MX3005P or MX3000P Stratagene qPCR. The two origami tiles to undergo dimerization were mixed and placed in the real time thermal cycler at room temperature. Thermodynamic data was obtained by repeatedly (at least

twice, see Supporting Information Note 3.2-3.3 in Sections 3.5.2 and 3.5.3) melting the solution by heating it to 55°C and then annealing it to 25°C at a rate of +/- 1°C per 15 minutes, unless otherwise noted. To ensure that measurements were repeatable and to eliminate any initial noise associated with the fluorophore strand, which occurred on the first heating or cooling cycle (Supporting Information Note 3.2 in Section 3.5.2), measurements were collected over two heating and cooling cycles. Kinetics measurements were performed immediately after thermodynamic measurements and were obtained by rapidly heating (at a rate of +2.5°C/second) the sample from 25°C to a given temperature (in the case of measuring forward reaction rates) or by rapidly cooling (at a rate of -2.5°C/second) from 55°C (in the case of measuring reverse reaction rates) and monitoring the fluorescence in intervals of 2 to 6 seconds. All fluorescence measurements were performed at least in triplicate.

3.4.4 2D Infinite Lattice Formation.

All samples were plated and imaged using the same temperature-controlled protocol as the AFM dimer yield measurements.

3.5 SUPPORTING INFORMATION

3.5.1 Supporting Note 3.1: Nanostructure specifics

DNA Origami Specifics

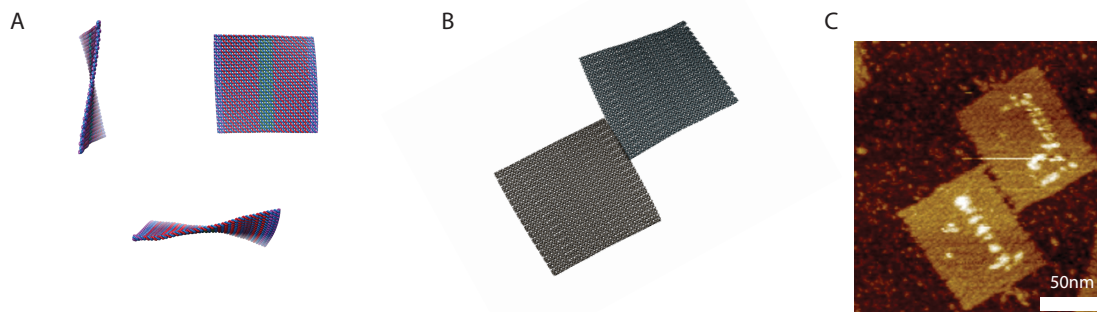


Figure S 3.1 CanDO and CaDNAno models of DNA origami tiles used in this work. The tile components were designed in caDNAno³² to be 32 helices tall with a helicity of 10.44 bases/turn, using 7008 bases of the ssDNA. T1 and T2 have the same origami design and staple sequences, except for the hairpins that distinguish the two tiles and the linker staples between the two tiles. For sequence specifics see Figure S3.33 and SI Note 3.7. (A) CanDO structure prediction of a single, tall, flat rectangle component³³. (B) Model of T1 and T2 binding generated from CanDO atomic model structure prediction without hairpins. However, CanDO structure prediction does not account for the curvature induced by hairpins on the origami surface, which has been shown to influence origami curvature¹¹⁴. The two tiles have many (>45) hairpins in order to ensure the differentiation of each origami component on the AFM, even if some hairpin staples did not properly incorporate into our system or were not distinguishable because of imaging artifacts: Due to the nature of the experiments described in this paper, it is important to be able to distinguish the identity of as many of the origami as possible. We rotated T2 relative to T1 by 180 degrees to facilitate the binding between homogeneous interfaces¹¹⁵. (C) AFM image of T1 and T2 binding with 4 linkers with 5bp SEs. Scale bar is 50 nm. As shown in Figure 1, the distance the linkers span between T1 and T2 is 21bp, or about 2 full turns of DNA.

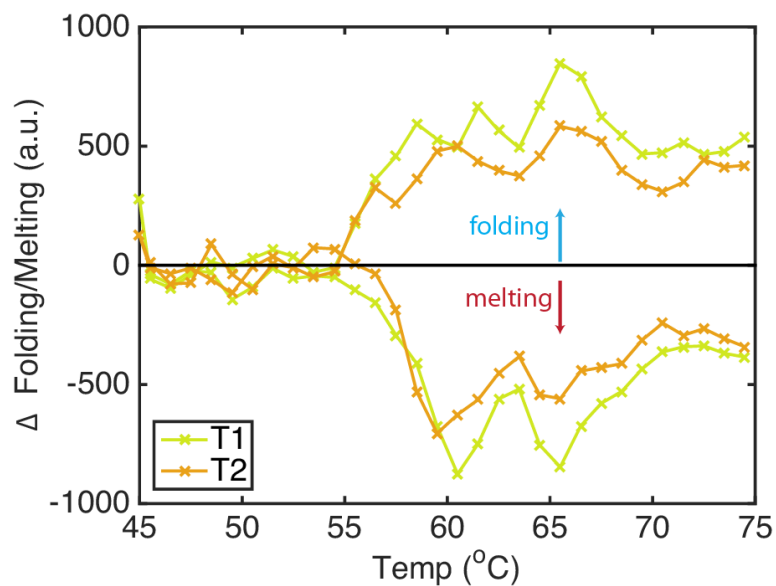


Figure S 3.2 DNA origami components melt above 55°C. Tile component melting and folding temperature as determined by DNA intercalation dye SYBR Green, similar to the protocol previously described in Ref. ³⁴. Rate of folding and melting of T1 and T2, as determined by the fluorescence derivative with respect to temperature of the full object reaction normalized by their respective staples-only reaction. The rate of cooling and heating is 1°C per hour. The melting of the DNA origami components takes place starting at 56°C, thus, all thermodynamic and kinetic protocols take place below this temperature.

3.5.2 Supporting Note 3.2: Fluorescence Specifics

Fluorescence Strategy

We chose to use a Texas Red®-X (NHS Ester) fluorophore, as its fluorescence has been shown to be relatively invariant to changes in temperature, pH, and nucleic acid sequence¹¹⁹. We used the Iowa Black® RQ dark quencher (IDT, Coralville, IA) to quench the Texas Red fluorophore. The fluorophore is located on a linker strand on Tile 1 (T1), on the end of a 3' sticky end (SE) and the quencher is on the 5' end of the Tile 2 (T2) linker strand (see Fig. 3.1 in main text). Contact quenching occurs when T1 and T2 bind via complementary linker SE sequences. Upon binding, the quencher contacts (within ~2 nm) the Texas Red fluorophore, which has been shown to produce high (> 95%) quenching efficiencies¹¹⁹. However, we found that the first anneal or melt cycle had

noisy fluorescence values. Data was reproducible only after the first cycle, after the fluorophore / quencher pair had been interrogated multiple times (at least 20). We speculate that this could be a fast photobleaching process, some non-uniformity in the Texas Red or Iowa Black Quencher synthesis, or some initial reaction of the dyes to the DNA or buffer environment.

Post Processing Fluorescence Measurements

To account for the changes in Texas Red fluorescence with temperature, the fluorescence reading for T1 alone was used as a baseline. The fluorescence of T1 alone was subtracted from the fluorescence readings of all T1+T2 reactions. Additionally, as well-to-well background fluorescence varies in a real time thermal cycler, all fluorescence baseline-adjusted measurements in a given interface design were linearly shifted in order to maximize the overlap between samples, using a least squares fitting algorithm, and the average of the samples was then linearly shifted to an initial, arbitrary starting value and then normalized to 1, in order to more easily and consistently interpret the change in fluorescence across all samples. The average and standard deviation of the fluorescence quenching were calculated using at least three replicates per linker architecture. Fluorescent data was smoothed *via* a moving average smoothing function built into MATLAB (*smooth* function).

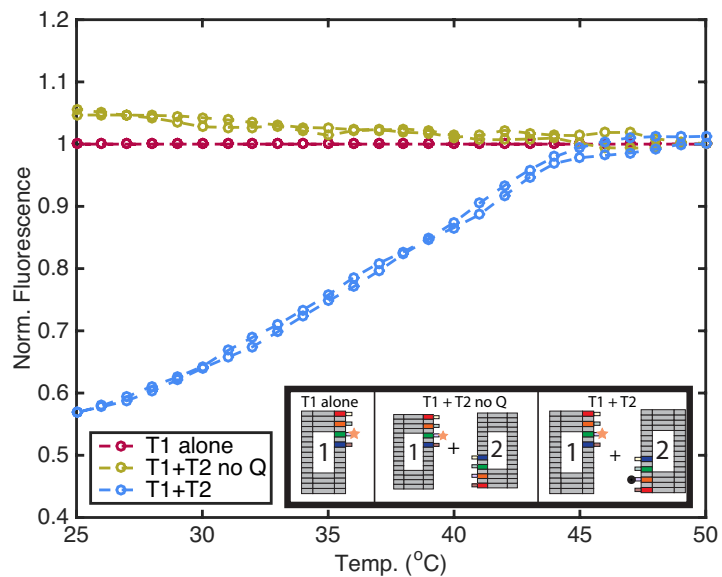


Figure S 3.3 T1 – T2 binding does not reduce fluorophore activity when no quencher is present. Normalized fluorescence (to T1 alone) change versus temperature for T1 alone, T1 and T2 together without a quencher and T1 and T2 together with a quencher, as shown in the schematic in inset. Dimers decrease fluorescence with decreasing temperature when a quencher is attached to T2 but increase slightly fluorescence if a quencher is not present. A similar increase has been previously reported in Ref. ¹⁰⁸ and is hypothesized to occur as a result of the fluorophore being “squeezed out” of interactions with DNA by the formation of dsDNA. In this hypothesis, the ability for the fluorophore to become excited could be enhanced upon dimerization because the fluorophore is no longer integrated into a stacking position within the ssDNA portion of the SE, allowing the dye to rotate more freely.

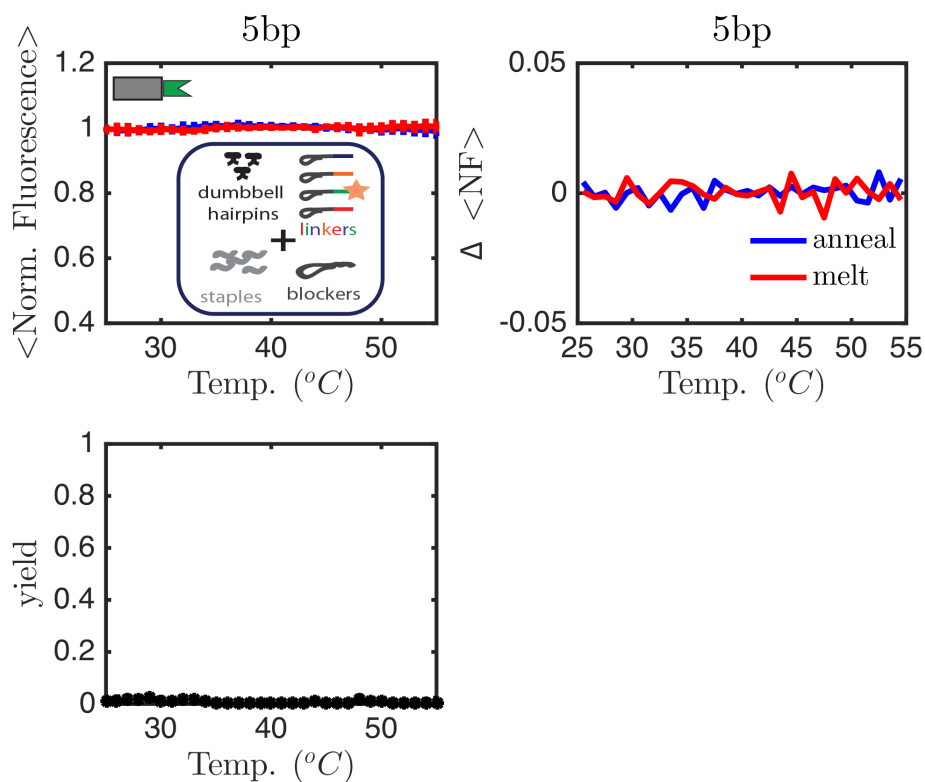


Figure S 3.4 Free linkers in solution (unbound to origami) do not bind to their complementary linker for 5bp SEs. Reaction mixtures for T1 and T2 were prepared without the scaffold (see inset diagram in top left plot for components included in the reaction) for 4SE systems with 5bp SEs. Samples were normalized with the scaffold-free reaction mixture of T1 alone. The plot in the top left shows the mean normalized fluorescence as a function of temperature. Error bars represent one standard deviation of the reported quantity. The plot in the top right shows the change in mean normalized fluorescence values as a function of temperature. The 5bp linkers do not show any change in fluorescence with temperature in the range of temperatures tested, indicating that free linkers in solution do not bind to their complement, as shown in the plot in the bottom left. Yield calculations assume the same fluorescence to yield calibration measurements obtained in Figure S3.6.

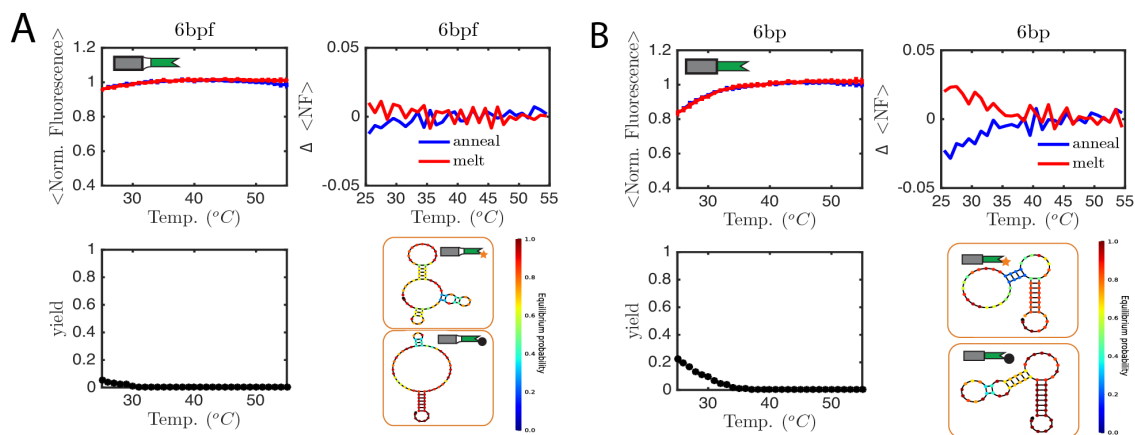


Figure S 3.5 Free 6bp linkers in solution (unbound to origami) can bind to their complementary linker for linkers with 6bp SEs. Reaction mixtures for T1 and T2 were prepared without the scaffold for 4SE systems with (A) floppy linkers and (B) non-floppy linkers with 6bp SEs. Samples were normalized with the scaffold-free reaction mixture of T1 alone. The plots in the top left show the mean normalized fluorescence as a function of temperature. Error bars represent one standard deviation of the reported quantity. The plots in the top right show the change in mean normalized fluorescence values as a function of temperature. The plots in the bottom left show and the diagrams in the bottom right show the secondary structure of the fluorophore and the linker according to the NUPACK⁴⁶ structure predictions at 35°C. For the system with non-floppy linkers, free linkers appear to bind to one another below ~35°C achieving ~20% yield at 25°C, which indicates that there is potentially competition between origami components and free linkers in binding to an origami interface. The floppy linkers, although their SE sequence is identical to the non-floppy linkers, do not show as much hybridization as the non-floppy linkers, yielding ~5% at 25°C, which is likely due to difference in the secondary structures between the floppy and non-floppy linkers. Specifically, the NUPACK prediction for the floppy linker of T1 (with the fluorophore) indicates secondary structure involving 3bp in its SE, most likely preventing or significantly slowing the reaction between the two free floppy linkers. Secondary structure is an important factor, along with SE length and sequence, in determining the extent to which free linkers bind in solution. Yield calculations assume the same fluorescence to yield calibration measurements obtained in SI Figure S6.

3.5.3 Supporting Note 3.3: Thermodynamic measurements

Fitting thermodynamic parameters

Finding thermodynamic parameters for the T1-T2 interaction was achieved using the definition for Gibbs free energy:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0, \quad (3.3)$$

and the van't Hoff Equation, assuming a two state transition between tiles:

$$\ln(K_{eq}) = -\frac{\Delta H^0}{RT} + \frac{S^0}{R}, \quad (3.4)$$

where ΔH^0 is the standard enthalpy change and ΔS^0 is the standard entropy change across a range of absolute temperatures, T and $R = 0.00198 \frac{kcal}{mol K}$ is the universal gas constant. K_{eq} is the equilibrium constant and defined as:

$$K_{eq} = \frac{yield_{eq}}{C_0 * (1 - yield_{eq})^2}, \quad (3.5)$$

where $yield_{eq}$ is defined as the yield of dimers at equilibrium and C_0 is the initial concentration of T1 and T2 (assuming negligible pipetting error so that the concentrations are the same). The $yield_{A M}$, or the fraction of material in dimers as determined by a series of AFM scans, is determined by the following formula:

$$yield_{A M} = \frac{2 * (N_D)}{(N_{T1}) + (N_{T2}) + 2 * (N_D)} \quad (3.6)$$

where N_D corresponds to the number of observed dimers, excluding those in the flipped conformation, (see Figure 3.2B), and N_{T1} and N_{T2} correspond to the counts of individual T1 and T2 components, respectively. Although flipped dimers would also result in a decrease in fluorescence, it was too difficult to determine with the AFM images if the flipped dimers were bound by linkers or happened to land next to one another in a flipped confirmation. For this reason, we excluded the flipped dimers in the yield calculation altogether. We expect this would not have a significant impact on the results, as likely some of the flipped dimers would indeed be in a quenched state, and others would be in a fluorescent state. Overall, the flipped dimers accounted for less than ~10% of all of the structures for all AFM scans. For the majority of interface designs and temperatures tested, flipped dimers accounted for < 3% of all structures. The T_m , or the

temperature at which the yield is 50% (see Figure 3.2F), was calculated using the yield curves (Figures S3.7-S3.14 bottom right plots).

Thermal Protocol for fluorescence quenching thermodynamics experiments

The annealing protocol to is as follows, with fluorescence readings taken at the end of each temperature hold at integral degrees:

- 1.) 25 to 55°C at +1°C per 15 minutes (melt 1)
- 2.) 55 to 25°C at -1°C per 15 minutes (anneal)
- 3.) 25 to 55°C at +1°C per 15 minutes (melt 2)

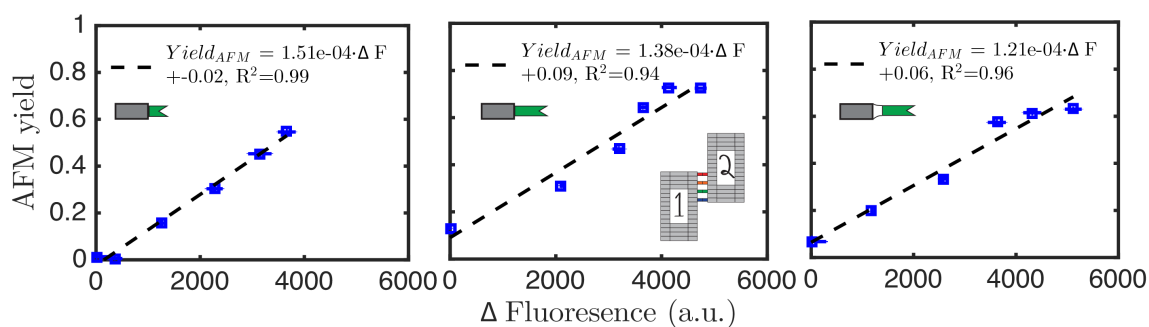


Figure S 3.6 Fluorescence quenching data as a proxy for AFM yield. Change in fluorescence versus AFM yield for 5bp SE (top left), 6bp SE (top right), and 6bp SE floppy (bottom left) SE (bottom right). Samples were annealed from 55°C to 25°C at a rate of 1°C per 15 minutes for the 5bp and 6bp floppy SEs systems and a rate of 1°C per hour for the 6bp SEs systems (see Methods). All systems used in this calibration process have four. All AFM yield data points have, at minimum, 300 counted objects. All fluorescence measurements were performed in at least triplicate. Error bars for the change in fluorescence represent one standard deviation; error bars for AFM yield generated by bootstrapping and indicate one standard deviation and are too small to see.

Thermodynamic Data

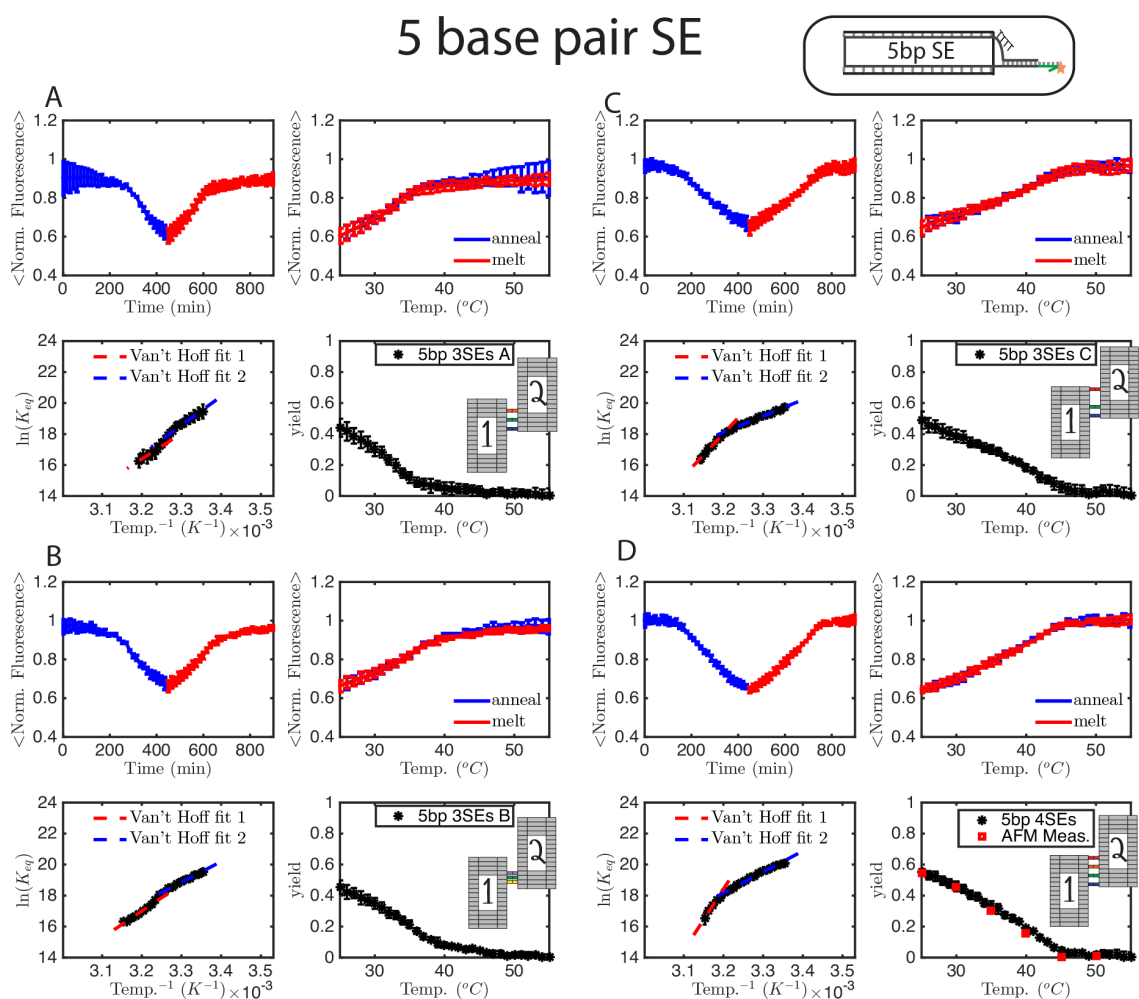


Figure S 3.7 Thermodynamic measurements for interfaces with 5bp SE linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Four plots per interface are (top left) fluorescence versus time, holding at 15 minutes per degree, (top right) fluorescence versus temperature, (bottom left) van’t Hoff plot and (bottom right) yield as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 5bp SE linker shown at the top of the figure. Error bars represent one standard deviation of the reported quantity.

5 base pair SE (Cont'd.)

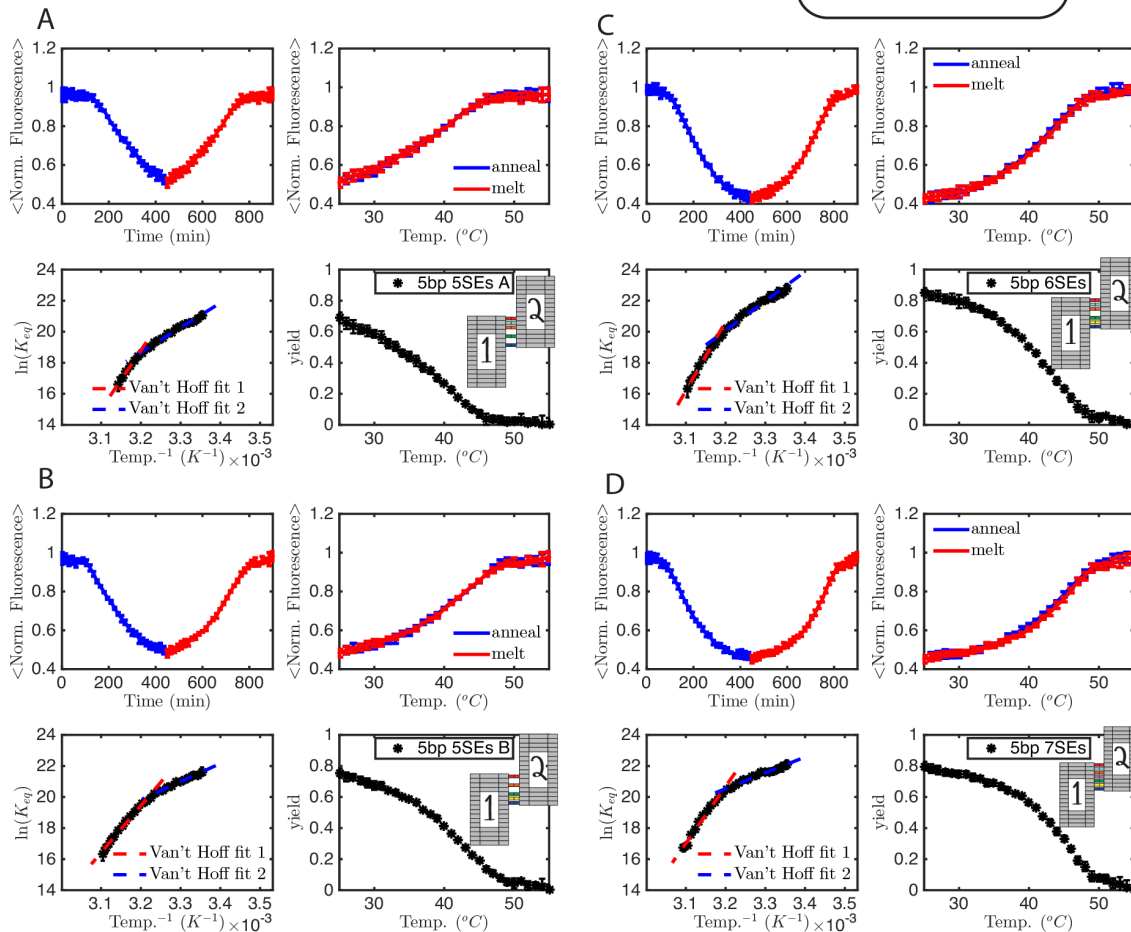
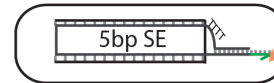


Figure S 3.8 Thermodynamic measurements for interfaces with 5bp SE linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Four plots per interface are (top left) fluorescence versus time, holding at 15 minutes per degree, (top right) fluorescence versus temperature, (bottom left) van’t Hoff plot and (bottom right) yield as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 5bp SE linker shown at the top of the figure. Error bars represent one standard deviation of the reported quantity.

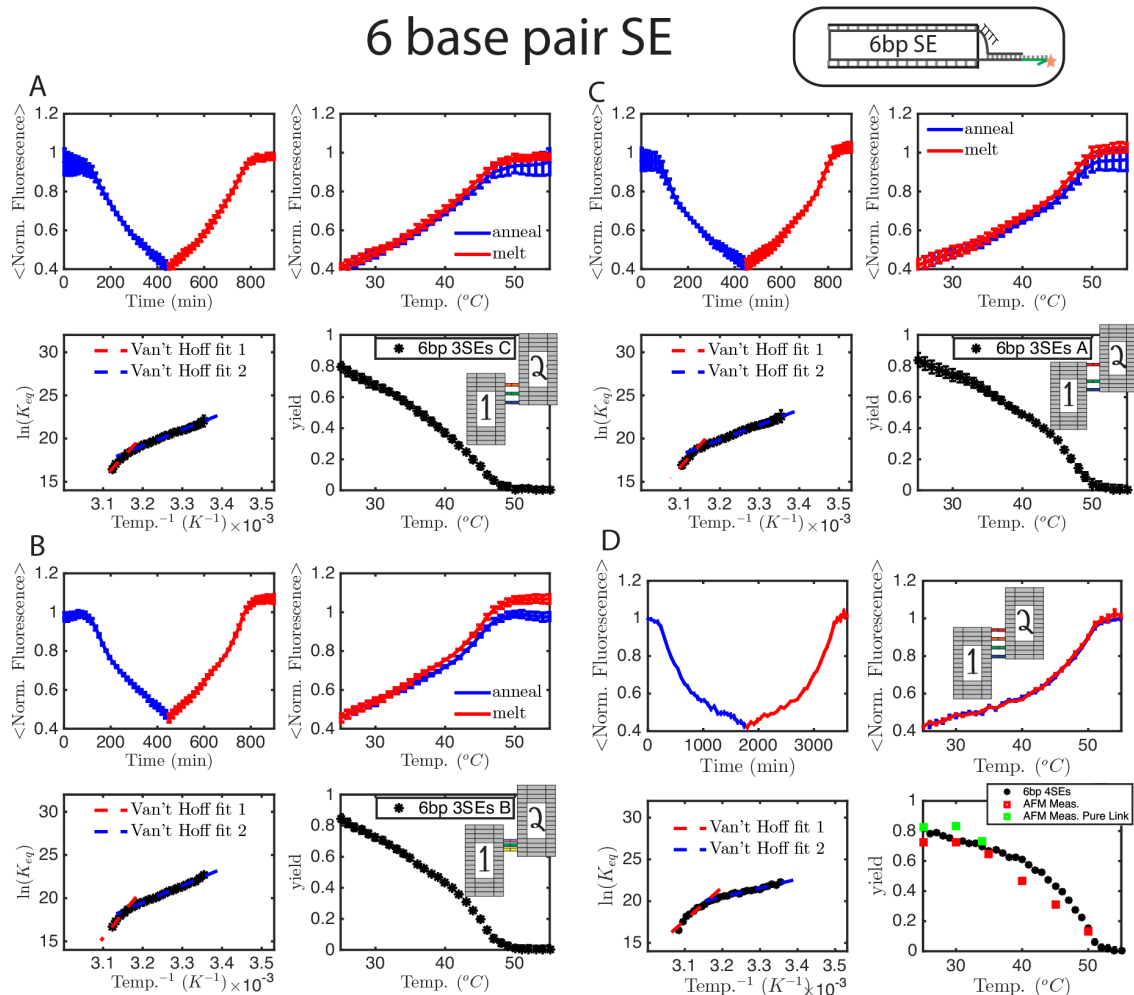


Figure S 3.9 Thermodynamic measurements for interfaces with 6bp SE linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Four plots per interface are (top left) fluorescence versus time, holding at 15 minutes per degree, (top right) fluorescence versus temperature, (bottom left) van’t Hoff plot and (bottom right) yield as a function of temperature. Note that (D) depicts yield as measured by AFM for linkers ordered PAGE purified (“AFM Meas. Pure Link”) and with standard desalting (“AFM Meas.”). Note the limits on the y-axis is larger than in the plots for other SE lengths for the van’t Hoff plots, reflecting the higher K_{eq} values. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE linker shown at the top of the figure. Error bars represent one standard deviation of the reported quantity.

6 base pair SE (Cont'd).

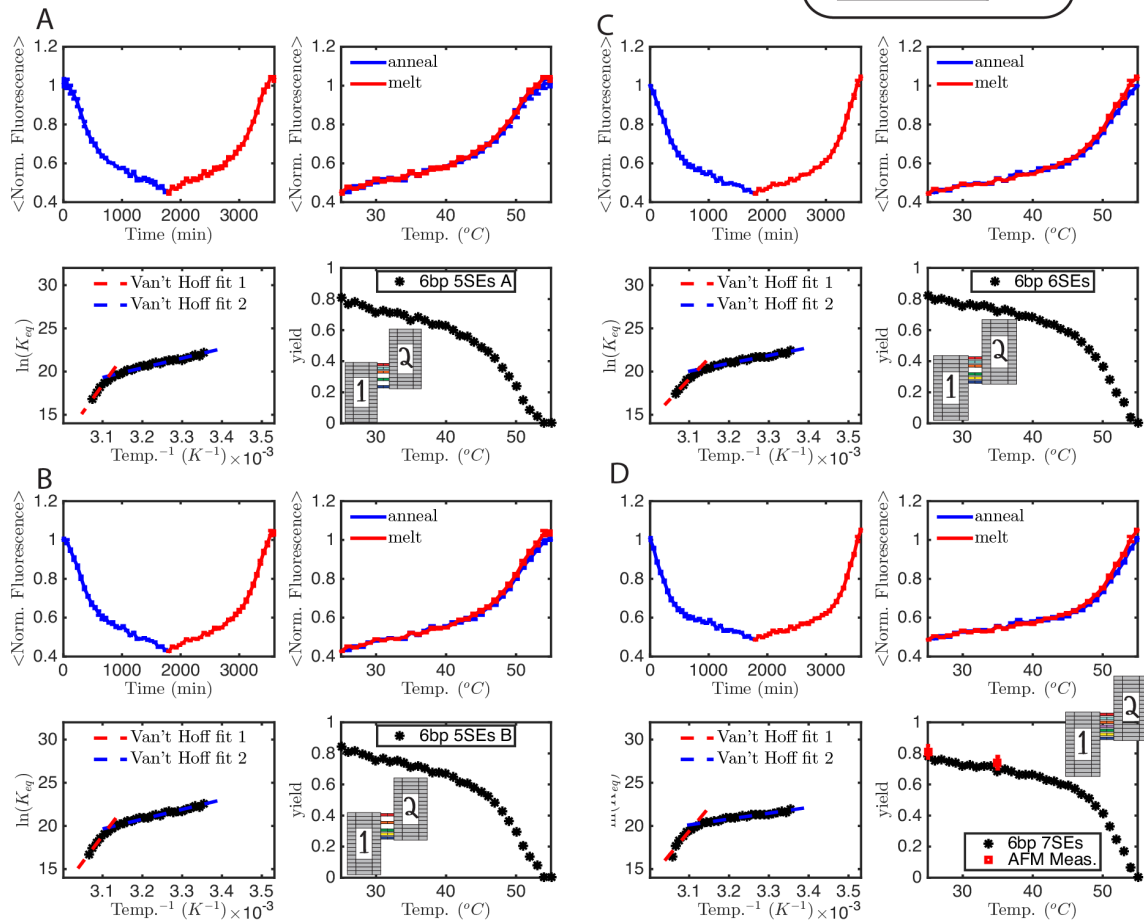


Figure S.3.10 Thermodynamic measurements for interfaces with 6bp SE linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Four plots per interface are (top left) fluorescence versus time, holding at 15 minutes per degree, (top right) fluorescence versus temperature, (bottom left) van’t Hoff plot and (bottom right) yield as a function of temperature. Note the limits on the y-axis is larger than in the plots for other SE lengths for the van’t Hoff plots, reflecting the higher K_{eq} values. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. To “spot check” other interface designs aside from the 4 linker, yield for 6bp 7SE at 25°C was verified on the AFM to be ~81% at 25°C, and ~74% at 35°C, suggesting the yield is indeed near 0 at 55°C. Cartoon schematic of an individual 6bp SE linker shown at the top of the figure. Error bars represent one standard deviation of the reported quantity.

6 base pair floppy SE

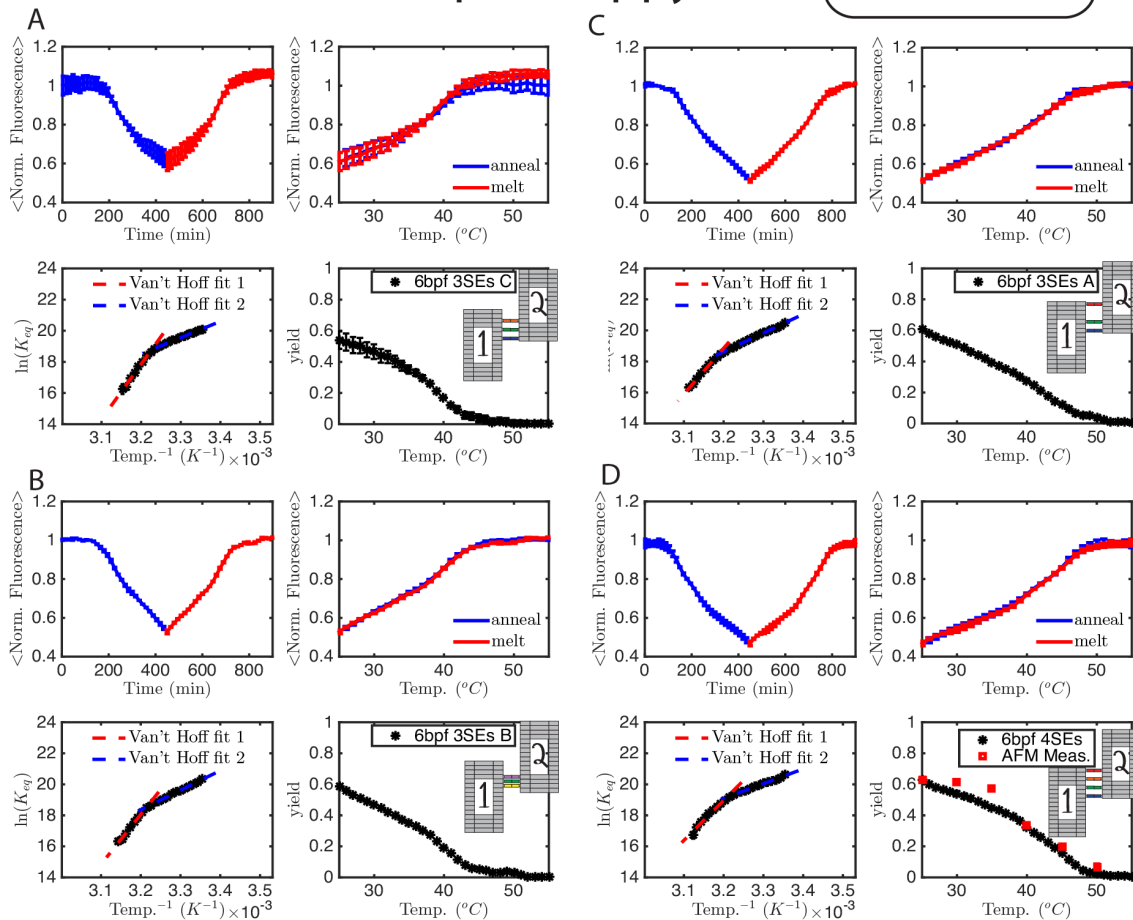


Figure S 3.11 Thermodynamic measurements for interfaces with 6bp SE floppy linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Four plots per interface are (top left) fluorescence versus time, holding at 15 minutes per degree, (top right) fluorescence versus temperature, (bottom left) van’t Hoff plot and (bottom right) yield as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE floppy linker shown at the top of the figure. Error bars represent one standard deviation of the reported quantity.

6 base pair floppy SE (Cont'd.)

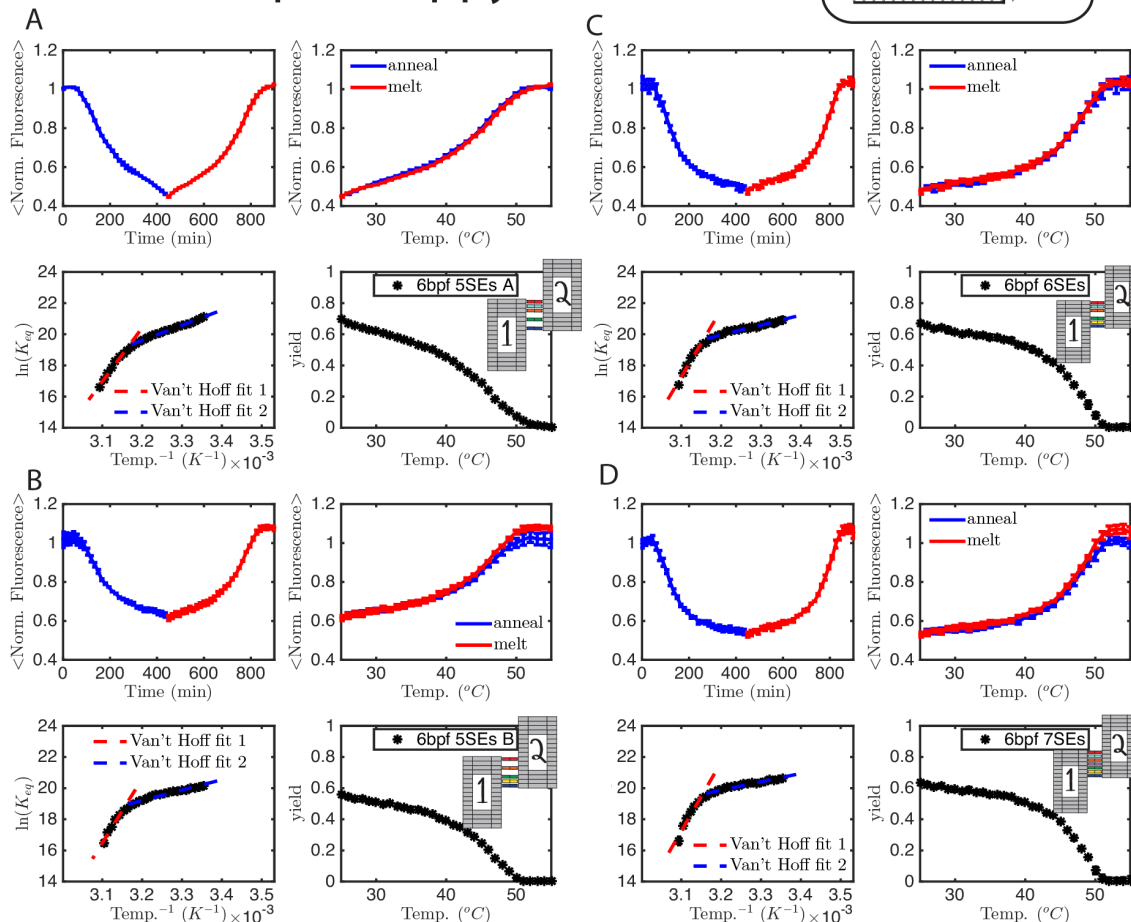


Figure S 3.12 Thermodynamic measurements for interfaces with 6bp SE floppy linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Four plots per interface are (top left) fluorescence versus time, holding at 15 minutes per degree, (top right) fluorescence versus temperature, (bottom left) van’t Hoff plot and (bottom right) yield as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE floppy linker shown at the top of the figure. Error bars represent one standard deviation of the reported quantity.

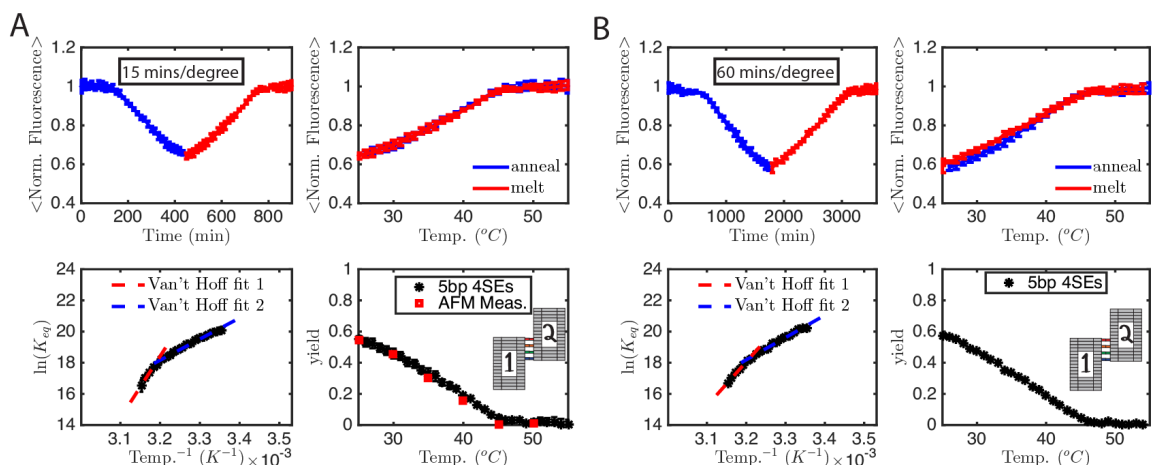


Figure S 3.13 Dimerization achieves equilibrium in 15 minutes for 4 linkers with 5bp SEs. Anneal/melt curves for (A) thermal protocols with heating/cooling rates of 1°C per 15 minutes and (B) 1°C per 60 minutes. Neither thermal protocol shows hysteresis in an anneal/melt cycle and both achieve approximate yields of $\sim 60\%$ at 25°C . We therefore believe reactions achieve equilibrium quickly, within 15 minutes for 5bp linkers. Inset diagrams depict the linking schematic.

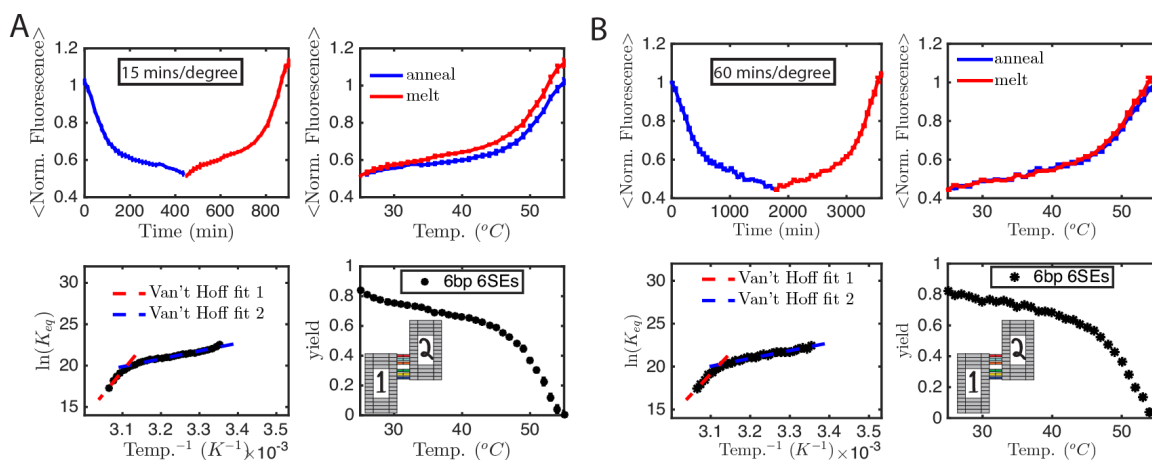


Figure S 3.14 Dimerization achieves equilibrium in 60 minutes for 6 linkers with 6bp SEs. Anneal/melt curves for (A) thermal protocols with heating / cooling rates of 1°C per 15 minutes and (B) 1°C per 60 minutes. The faster heating / cooling thermal protocol shows hysteresis in an anneal/melt cycle while the slower protocol does not. This is also the case for interfaces with 4 – 7 linkers and 6bp SE. Inset diagrams depict the linking schematic.

Measured Thermodynamic Parameters: ΔH^0 and ΔS^0

Table S 3.1 Measured Thermodynamic Parameters for 5bp SE linkers (kcal/mol)*

Interface	ΔH^0	ΔS^0	G^0	ΔH^0_2	ΔS^0_2	G^0_2
5bp 3SE A**	32.6±0.3	21.4±0.2	-11.2±0.0	36.8±0.9	25.2±0.9	-11.6±0.0
5bp 3SE B	34.2±3.3	22.6±3.1	-11.6±0.3	26.3±1.4	14.7±1.2	-11.6±0.2
5bp 3SE C	54.4±3.2	41.2±3.2	-13.2±0.2	20.8±2.7	9.2±2.6	-11.7±0.3
5bp 4SE	78.5±5.8	64.0±5.7	-14.5±0.2	26.9±3.1	14.9±3.1	-12.0±0.2
5bp 5SE A	75.2±6.0	60.6±5.9	-14.6±0.1	31.4±5.1	19.0±4.7	-12.5±0.4
5bp 5SE B	59.8±6.0	45.6±5.5	-14.2±0.5	23.2±8.2	10.4±7.8	-12.8±0.4
5bp 6SE	88.0±27.7	71.8±25.9	-16.3±1.8	37.5±14.2	23.9±13.6	-13.6±0.6
5bp 7SE	72.6±13.2	57.0±12.3	-15.6±0.9	20.2±5.0	7.2±4.7	-13.1±0.4

*VH1 is the fit for higher temperatures (~55°C to ~45°C); VH2 for lower temperatures (~45°C to 25°C). Error indicates one standard deviation of reported quantity and was calculated based on best fits to individual reaction samples.

**Nomenclature of interfaces (e.g. 5bp 3SE A) in this table is in the same order throughout SI. Please refer to Figures S3.7-S3.8 for interface cartoon schematic.

Table S 3.2 Measured Thermodynamic Parameters for 6bp SE linkers (kcal/mol)

Interface	ΔH^0	ΔS^0	G^0	ΔH^0_2	ΔS^0_2	G^0_2
6bp 3SE A	102.0±5.3	84.4±4.9	-17.6±0.4	34.3±7.4	21.0±6.8	-13.3±0.5
6bp 3SE B	112.7±9.1	95.0±8.6	-17.7±0.5	38.8±6.7	25.5±6.2	-13.3±0.4
6bp 3SE C	98.5±8.2	81.9±7.6	-16.6±0.6	36.7±3.5	23.7±3.3	-13.0±0.2
6bp 4SE	79.2±15.7	62.8±14.5	-16.5±1.2	21.2±2.6	8.1±2.5	-13.1±0.1
6bp 5SE A	126.9±3.2	106.3±3.0	-20.6±0.3	22.0±0.9	9.0±0.8	-13.1±0.1

6bp 5SE B	121.1±0.2	100.8±0.1	-20.4±0.0	22.3±1.1	9.0±1.0	-13.3±0.1
6bp 6SE	96.4±2.1	77.8±1.9	-18.6±0.2	18.0±0.3	4.8±0.3	-13.2±0.0
6bp 7SE	109.7±4.3	89.9±4.0	-19.8±0.3	13.4±0.3	0.5±0.2	-12.9±0.0

Table S 3.3 Measured Thermodynamic Parameters for 6bp floppy SE linkers (kcal/mol)

Interface	$-G^0$	$-2.8 S^0$	G^0	$-G^0_2$	$-2.8 S^0_2$	G^0_2
6bpf 3SE A	58.6±8.9	44.8±8.4	13.8±0.5	24.0±1.7	11.9±1.7	-12.1±0.1
6bpf 3SE B	63.6±3.4	50.0±3.3	13.6±0.1	25.0±1.2	13.0±1.2	-12.0±0.0
6bpf 3SE C	70.0±5.9	56.2±5.6	13.8±0.4	20.7±3.1	8.8±2.8	-11.9±0.3
6bpf 4SE	58.5±5.5	44.5±5.1	14.0±0.4	19.4±0.6	7.1±0.5	-12.2±0.1
6bpf 5SE A	68.3±1.9	53.0±1.7	15.2±0.2	19.3±0.6	6.8±0.6	-12.5±0.0
6bpf 5SE B	81.7±7.2	65.8±6.6	15.9±0.6	13.2±1.4	1.3±1.3	-11.9±0.2
6bpf 6SE	85.2±0.6	68.6±0.7	16.7±0.0	12.6±0.7	0.3±0.6	-12.4±0.1
6bpf 7SE	87.6±11.8	70.8±11.0	16.9±0.8	10.6±0.5	-1.6±0.6	-12.2±0.1

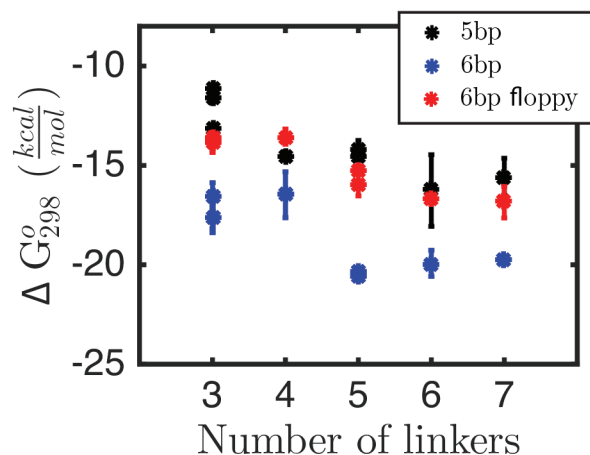


Figure S 3.15 Measured values of standard Gibbs free energy at 298K, ΔG_{298}^0 , for all interface designs. All values plotted are from van't Hoff fit 1 (higher temperatures), as this fit indicates reflects the thermodynamic properties of the fully-formed interface (see Results and Discussion section of main text for further discussion). Specific values can be found in Tables S3.1-S3.3. Error bars indicate one standard deviation of the ΔG_{298}^0 .

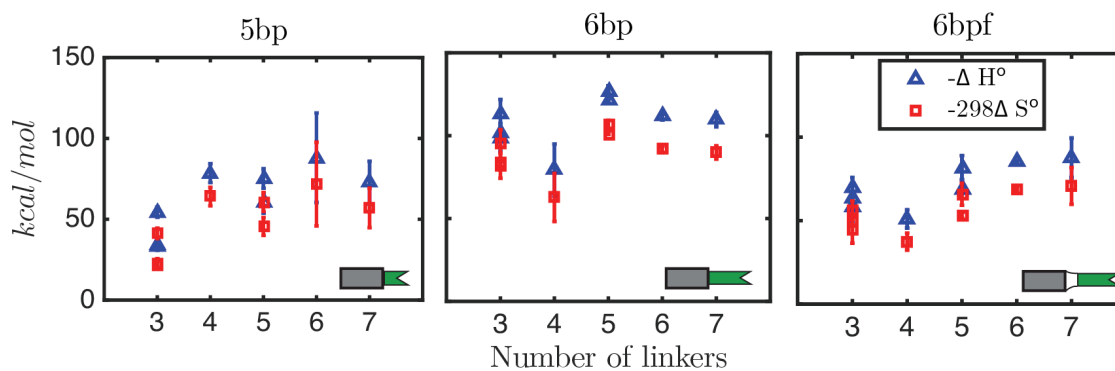


Figure S 3.16 Measured values of standard enthalpy, $-\Delta H_{298}^0$, and entropy $-298 * \Delta S_{298}^0$ for all interfaces. Plots are shown as follows: (left) 5bp SE, (middle) 6bp SE, and (right) 6bp floppy SEs. All values plotted are from van't Hoff fit 1 (higher temperatures), as this fit indicates reflects the thermodynamic properties of the fully-formed interface (see Results of main text for further

discussion). Specific values can be found in Tables S3.1-S3.4. Inset depicts cartoon schematic of linker design. Error bars indicate one standard deviation of the reported quantities.

3.5.4 Supporting Note 3.4: Kinetic measurements

The reaction between T1 and T2 produces the Tile 1-Tile 2 dimer, D according to the reaction:



where k_{on} is the the second order forward reaction rate constant and k_{off} is the first order reverse reaction rate constant. In our experiments, the initial concentration of components T1 and T2 were equal, at $C_0 = 5$ nM. The yield of the dimer is defined as the fraction of total material of one of the component types in the complex, *i.e.* $yield = \frac{[D]}{C_0}$. Since we know the equilibrium concentrations

for all of the species from our thermodynamics experiments, we know the equilibrium constant,

$$K_{eq} = \frac{k_{on}}{k_{off}} = \frac{[D]_e}{[T1]_e [T2]_e}$$

at all temperatures studied. With this information, we can fit a k_{on} for the

dimerization reaction:

$$\frac{d[]}{dt} = k_{on}[1][2] - k_{off}[], \quad (3.8)$$

which expands to:

$$\frac{d(yield)}{dt} = k_{on} \left((1 - yield)^2 C_0 - \frac{(yield)}{K_{eq}} \right). \quad (3.9)$$

In the same fashion, we obtain an expansion for Eq. 3.8 to fit the k_{off} for the dimerization reaction:

$$\frac{d(yield)}{dt} = k_{off} (K_{eq}(1 - yield)^2 C_0 - yield). \quad (3.10)$$

We used least squares fittings from the fluorescence data to Equations 3.9 and 3.10 to obtain the reaction rate constants, k_{on} (Figures S3.17-S3.22) and k_{off} (Figures S3.23-S3.28), respectively.

When fitting the reaction rate constants, we used the K_{eq} as obtained *via* the fluorescence measurements, where the yield was calibrated with the atomic force microscope (Supporting Note 3.3), in the thermodynamic protocol. Specifically, three sets of experiments are used to separately measure the three parameters we measured for each reaction: the equilibrium constant, the forward reaction rate constant, and the reverse reaction rate constant.

Thermal Protocol for fluorescence quenching kinetics experiments

Kinetic Measurements: Determining k_{on}

1. Hold 55°C for 15 minutes (take fluor. measurement every 5 minutes)
2. Hold 25°C for 15 minutes (take fluor. measurement every 6 seconds for 2 minutes, every 15 seconds for 8 minutes, every 30 seconds for 5 minutes)
3. Repeat steps 1 and 2, increasing the hold temperature of step 2 by 5°C every iteration until 50°C.

Kinetic Measurements: Determining k_{off}

1. Hold 25°C for 15 minutes (take fluor. measurement every 5 minutes)
2. Hold 55°C for 15 minutes (take fluor. measurement every 6 seconds for 2 minutes, every 15 seconds for 8 minutes, every 30 seconds for 5 minutes)
3. Repeat steps 1 and 2, decreasing the hold temperature of step 2 by 5°C every iteration until 30°C.

Kinetic Data: Forward Reaction Rate Constant

5 base pair SE

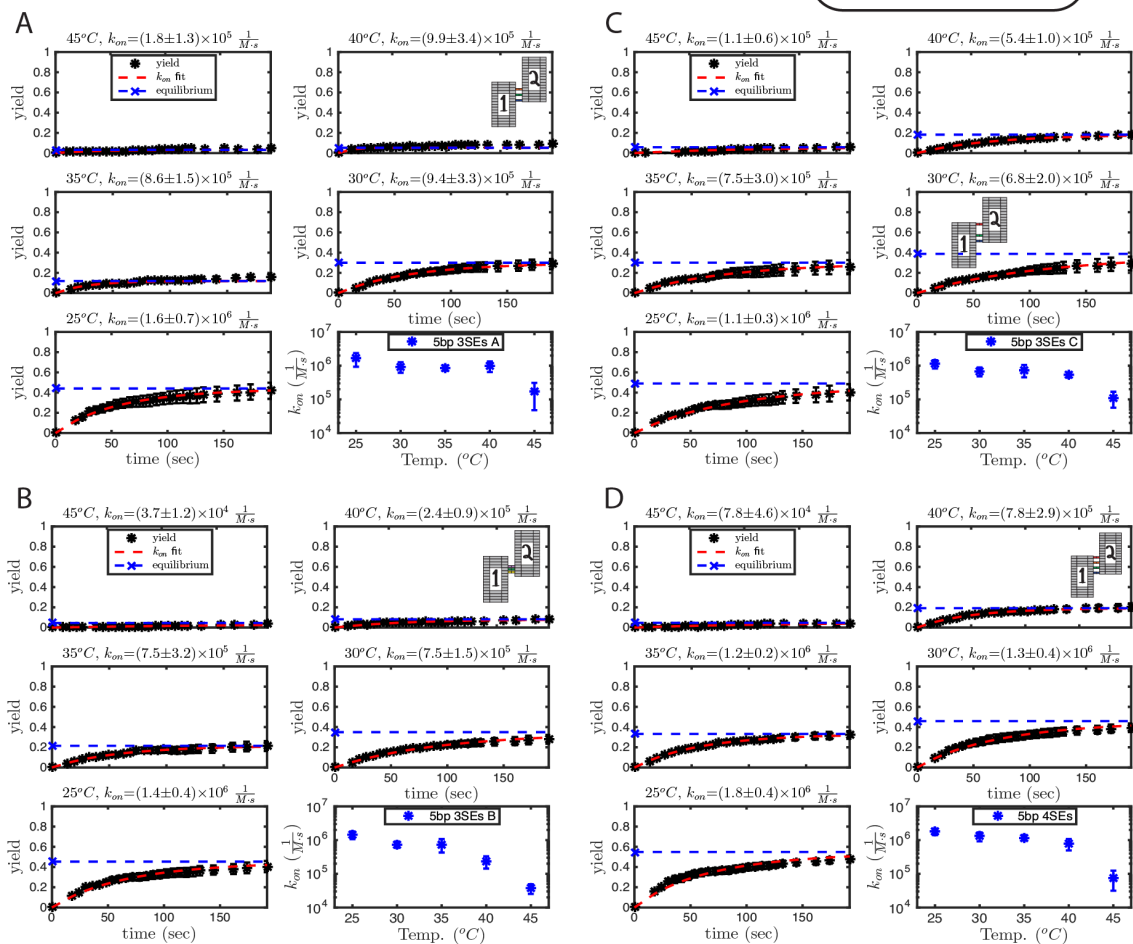


Figure S 3.17 Kinetics measurements: forward reaction rate constants for interfaces with 5bp SE linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Six plots per interface are fits for k_{on} for temperature jump experiments from 55°C to (top left) 45°C, (top right) 40°C, (middle left) 35°C, (middle right) 30°C, (bottom left) 25°C and (bottom right) k_{on} as a function of temperature. Values for k_{on} as a function of temperature are not reported when the error bar is equal to or greater than the value of k_{on} . Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 5bp SE linker shown at the top of the figure. Error bars indicate one standard deviation of the reported quantity.

5 base pair SE (Cont'd.)

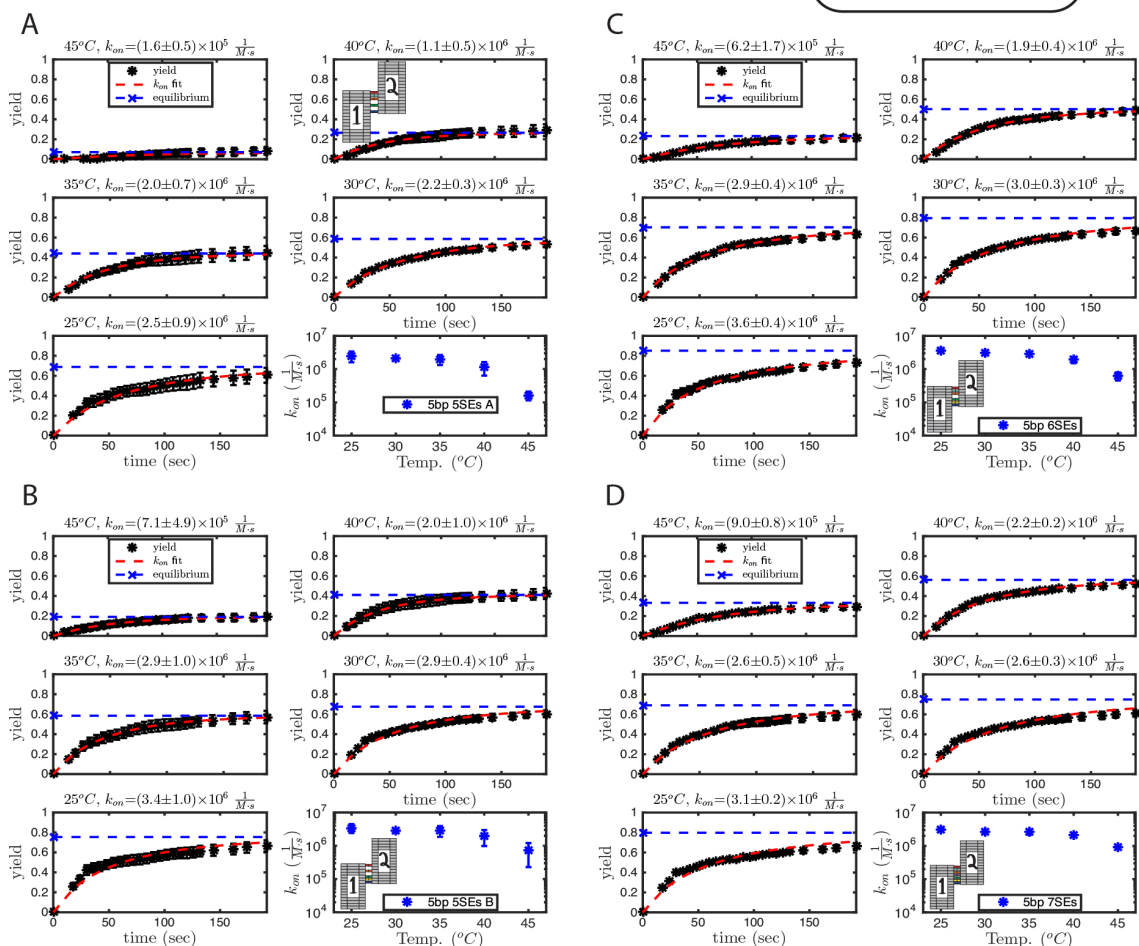


Figure S 3.18 Kinetics measurements: forward reaction rate constants for interfaces with 5bp SE linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Six plots per interface are fits for k_{on} for temperature jump experiments from 55°C to (top left) 45°C, (top right) 40°C, (middle left) 35°C, (middle right) 30°C, (bottom left) 25°C and (bottom right) k_{on} as a function of temperature. Values for k_{on} as a function of temperature are not reported when the error bar is equal to or greater than the value of k_{on} . Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 5bp SE linker shown at the top of the figure. Error bars indicate one standard deviation of the reported quantity.

6 base pair SE

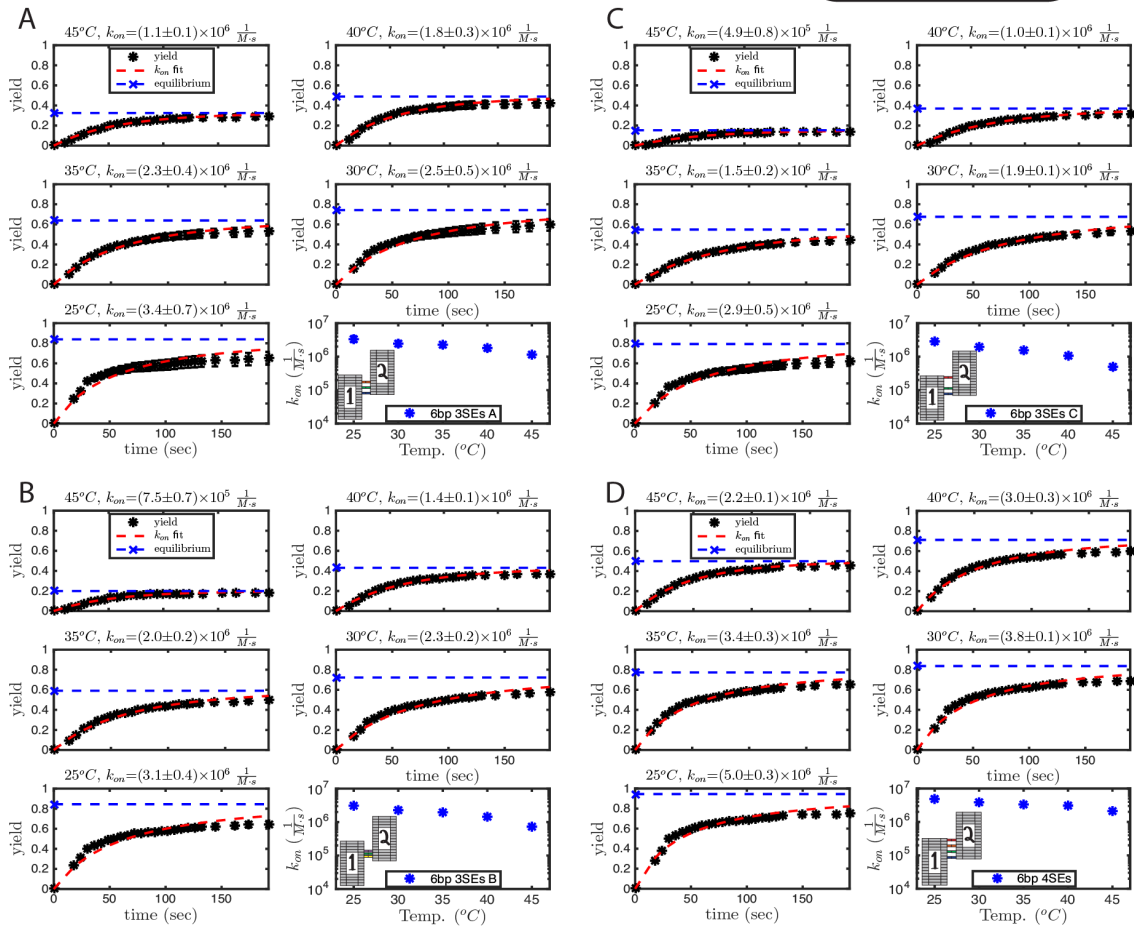


Figure S 3.19 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Six plots per interface are fits for k_{on} for temperature jump experiments from 55°C to (top left) 45°C, (top right) 40°C, (middle left) 35°C, (middle right) 30°C, (bottom left) 25°C and (bottom right) k_{on} as a function of temperature. Values for k_{on} as a function of temperature are not reported when the error bar is equal to or greater than the value of k_{on} . Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE linker shown at the top of the figure. Error bars indicate one standard deviation of the reported quantity.

6 base pair SE (Cont'd.)

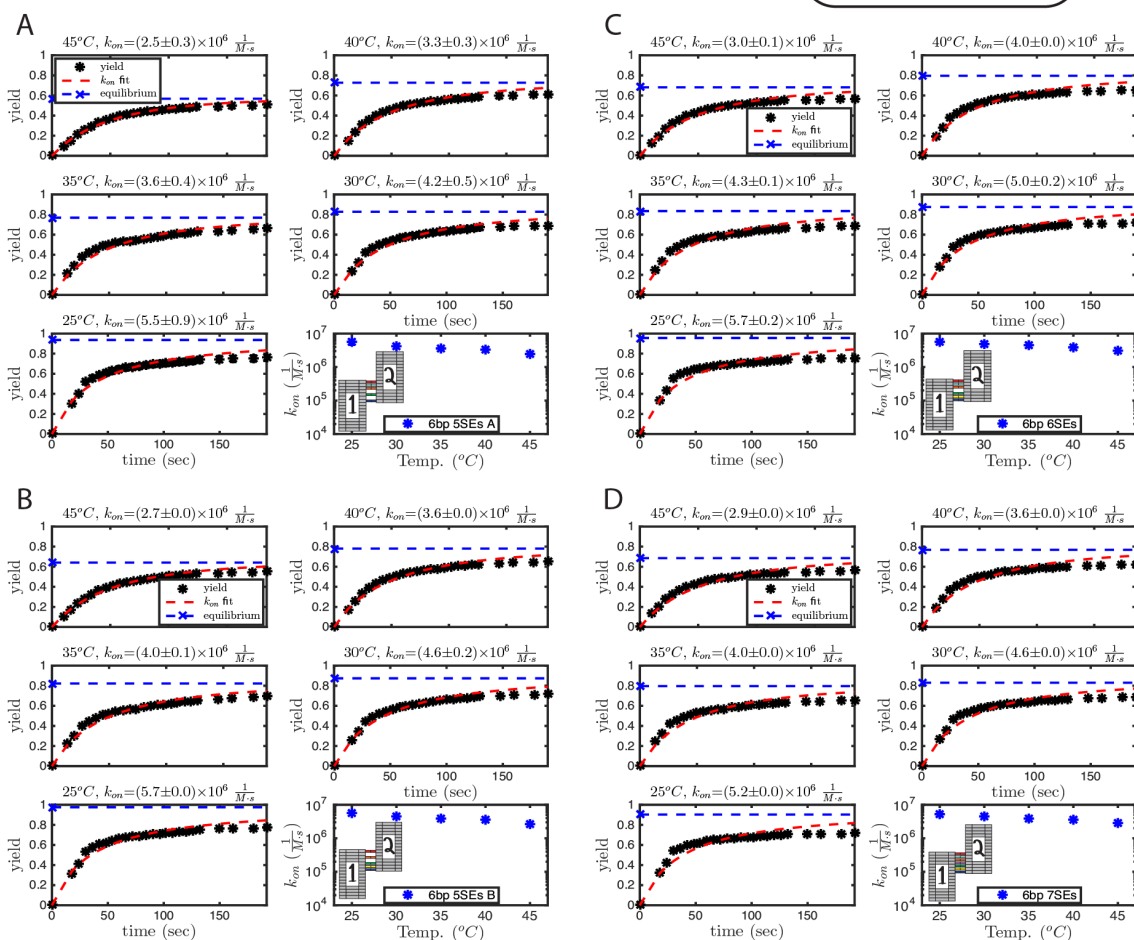


Figure S 3.20 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE linkers. Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Six plots per interface are fits for k_{on} for temperature jump experiments from 55°C to (top left) 45°C, (top right) 40°C, (middle left) 35°C, (middle right) 30°C, (bottom left) 25°C and (bottom right) k_{on} as a function of temperature. Values for k_{on} as a function of temperature are not reported when the error bar is equal to or greater than the value of k_{on} . Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE linker shown at the top of the figure. Error bars indicate one standard deviation of the reported quantity.

6 base pair floppy SE

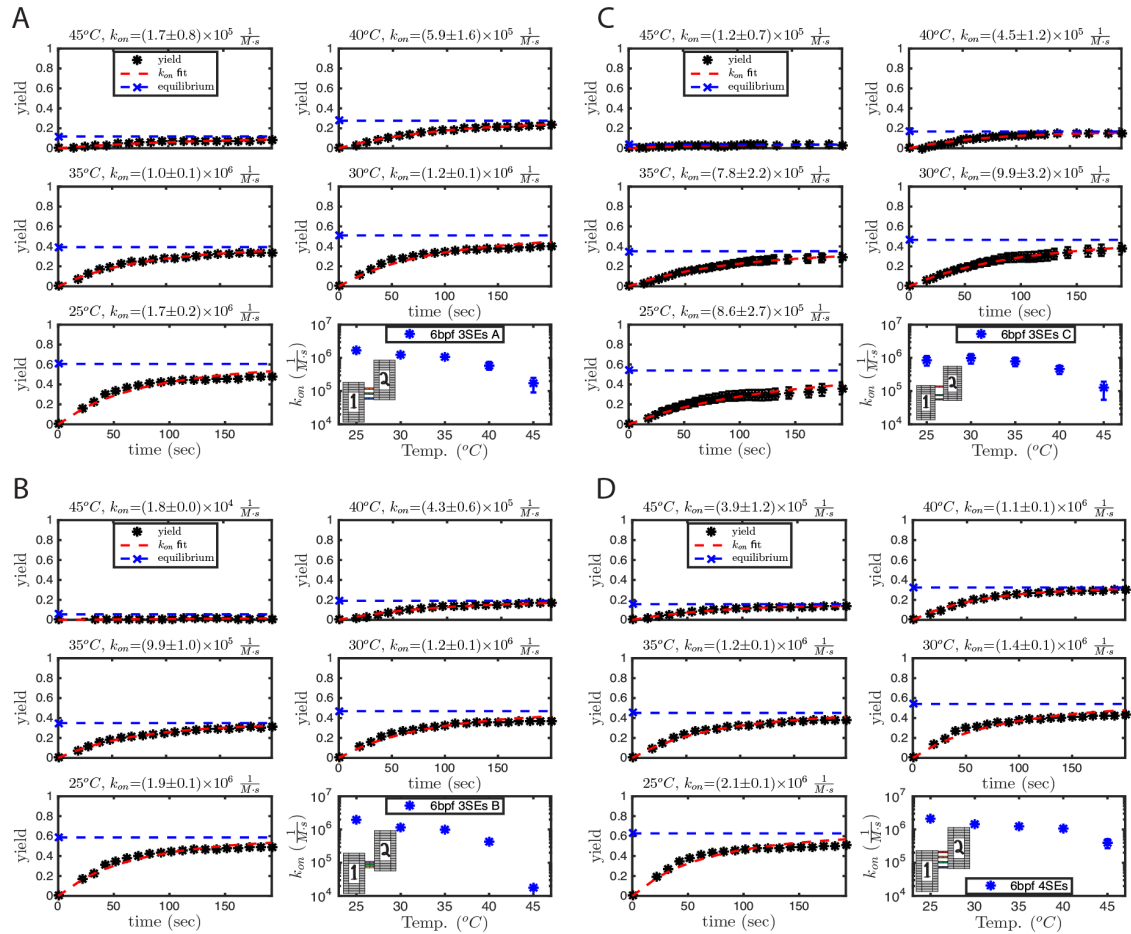
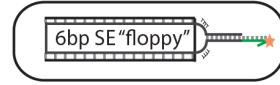


Figure S 3.21 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE floppy linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Six plots per interface are fits for k_{on} for temperature jump experiments from $55^{\circ}C$ to (top left) $45^{\circ}C$, (top right) $40^{\circ}C$, (middle left) $35^{\circ}C$, (middle right) $30^{\circ}C$, (bottom left) $25^{\circ}C$ and (bottom right) k_{on} as a function of temperature. Values for k_{on} as a function of temperature are not reported when the error bar is equal to or greater than the value of k_{on} . Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE floppy linker shown at the top of the figure. Error bars indicate one standard deviation of the reported quantity.

6 base pair floppy SE (Cont'd.)

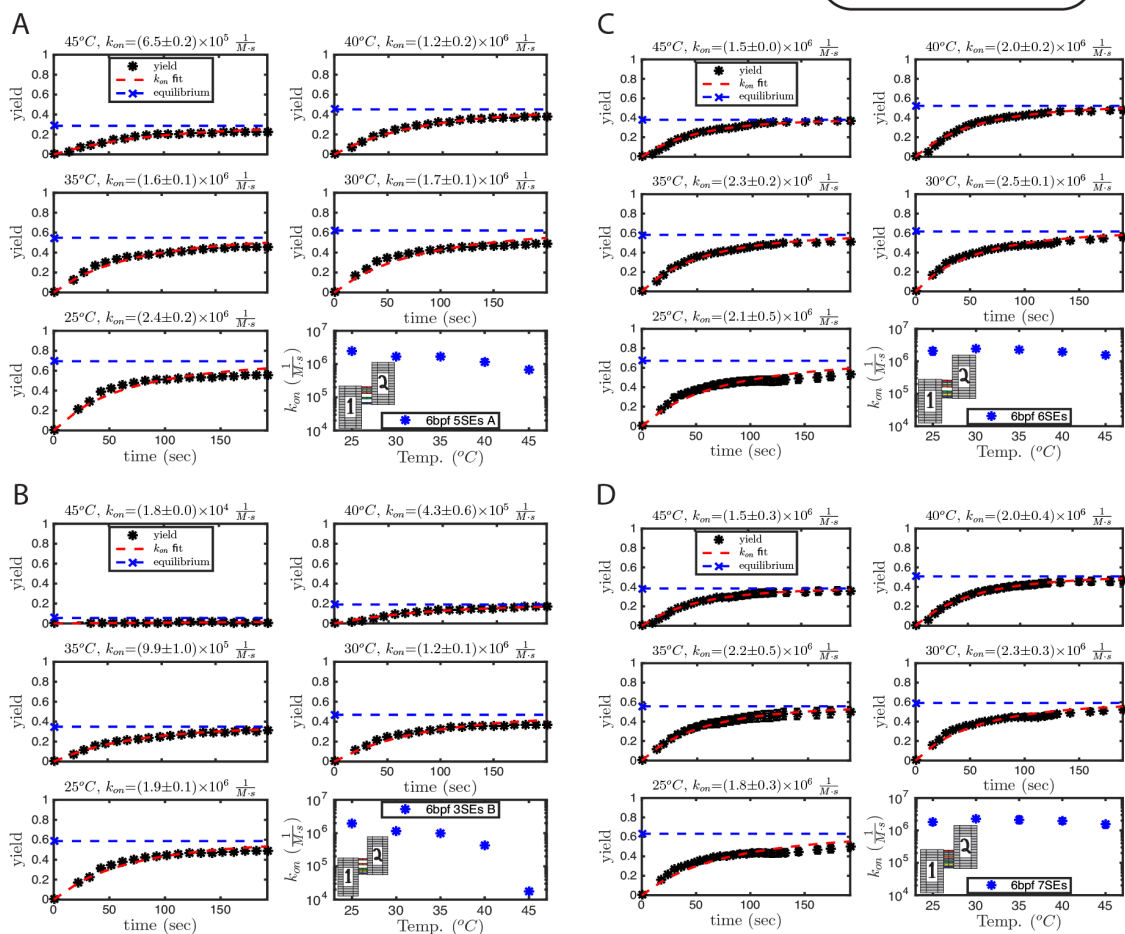


Figure S 3.2 Kinetics measurements: forward reaction rate constants for interfaces with 6bp SE floppy linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Six plots per interface are fits for k_{on} for temperature jump experiments from $55^{\circ}C$ to (top left) $45^{\circ}C$, (top right) $40^{\circ}C$, (middle left) $35^{\circ}C$, (middle right) $30^{\circ}C$, (bottom left) $25^{\circ}C$ and (bottom right) k_{on} as a function of temperature. Values for k_{on} as a function of temperature are not reported when the error bar is equal to or greater than the value of k_{on} . Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE floppy linker shown at the top of the figure. Error bars indicate one standard deviation of the reported quantity.

Kinetic Data: Reverse Reaction Rate Constant

5 base pair SE

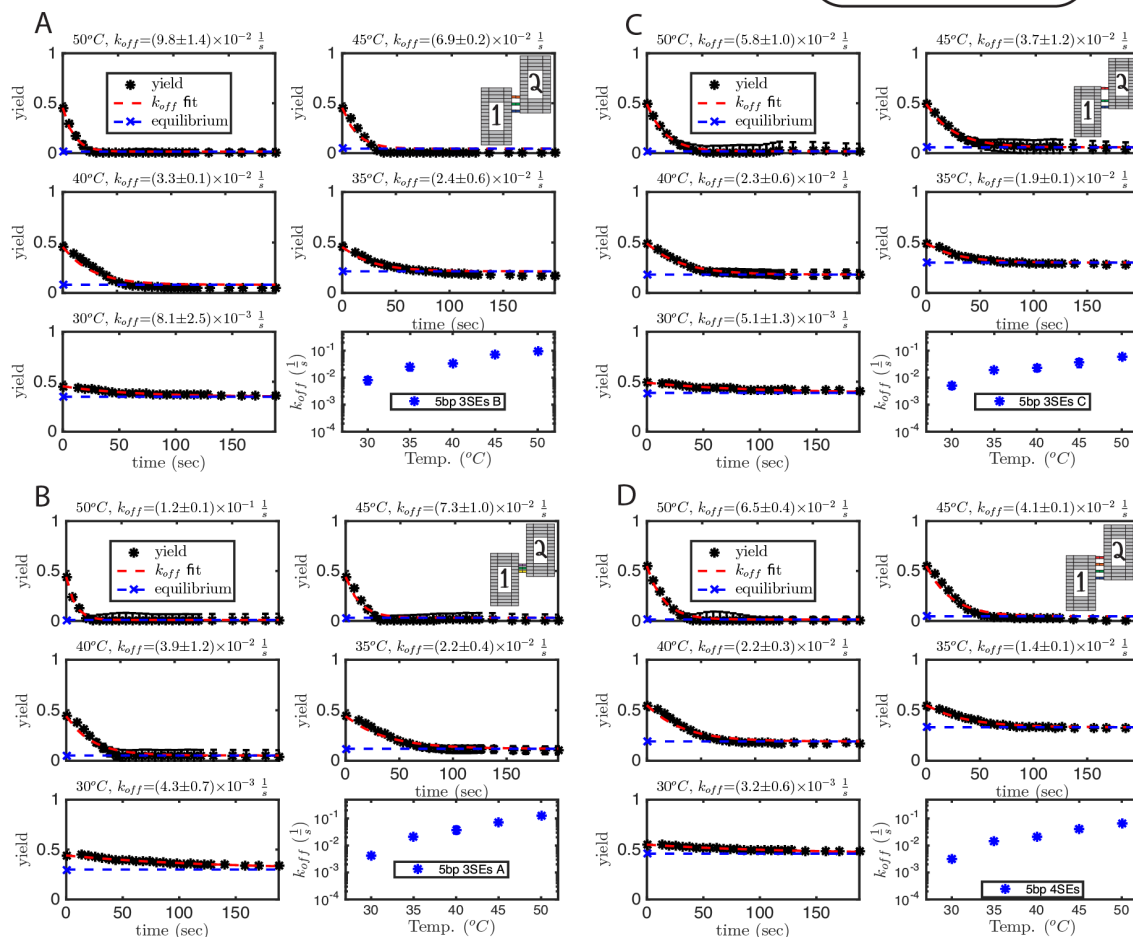


Figure S 3.23 Kinetics measurements: reverse reaction rate constants for interfaces with 5bp SE linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers, (E) 5 linkers in arrangement “A”, (F) 5 linkers in arrangement “B”, (G) 6 linkers and (H) 7 linkers. Six plots per interface are fits for k_{off} for temperature jump experiments from 25°C to (top left) 50°C, (top right) 45°C, (middle left) 40°C, (middle right) 35°C, (bottom left) 30°C and (bottom right) k_{on} as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 5bp SE linker shown at the top of the figure.

5 base pair SE (Cont'd.)

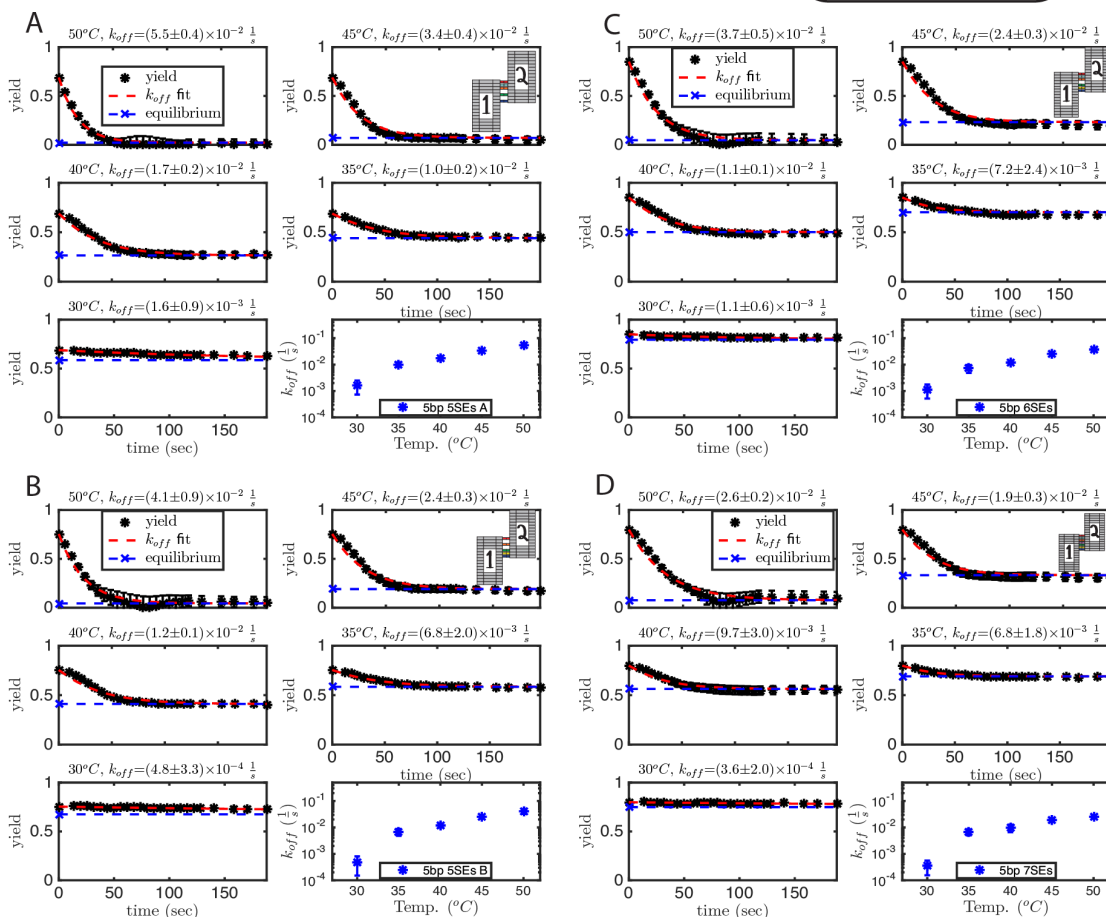


Figure S 3.24 Kinetics measurements: reverse reaction rate constants for interfaces with 5bp SE linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Six plots per interface are fits for k_{off} for temperature jump experiments from $25^{\circ}C$ to (top left) $50^{\circ}C$, (top right) $45^{\circ}C$, (middle left) $40^{\circ}C$, (middle right) $35^{\circ}C$, (bottom left) $30^{\circ}C$ and (bottom right) k_{on} as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 5bp SE linker shown at the top of the figure.

6 base pair SE

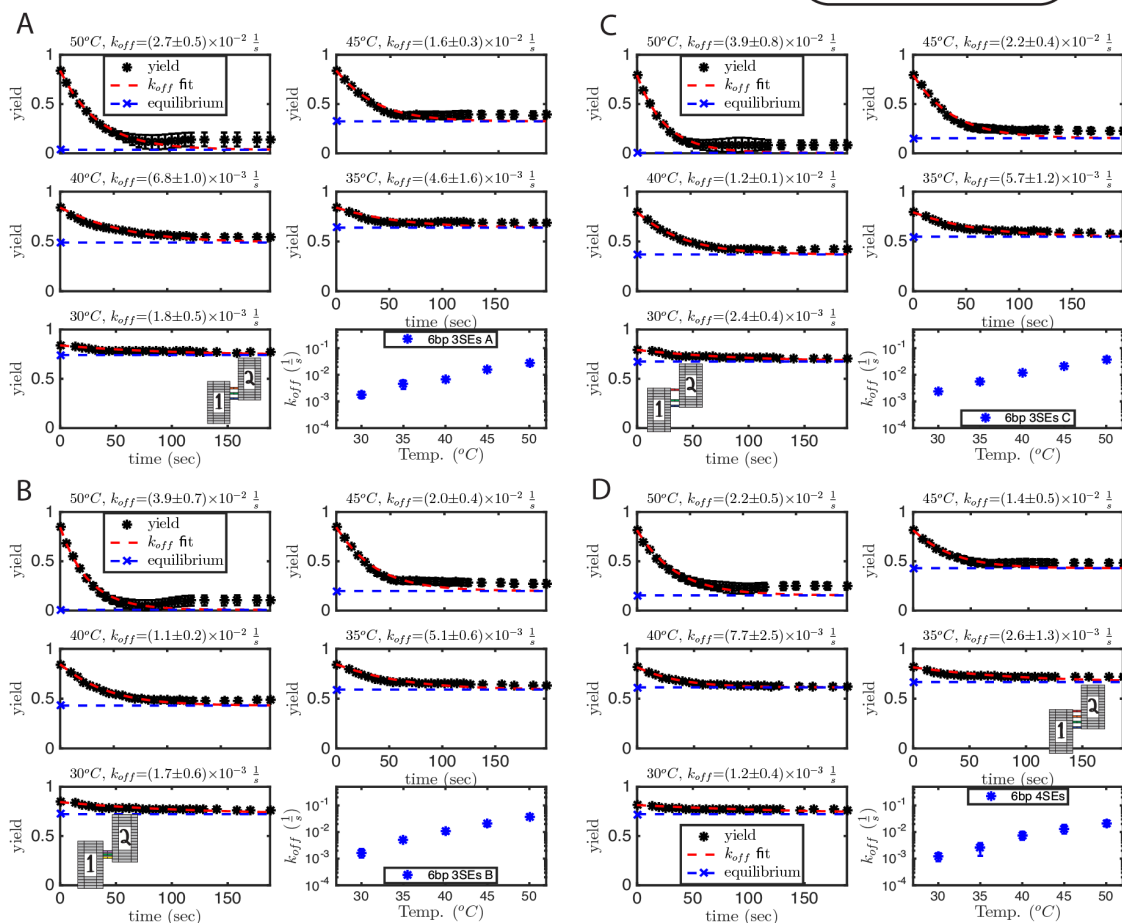


Figure S 3.25 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Six plots per interface are fits for k_{off} for temperature jump experiments from 25°C to (top left) 50°C, (top right) 45°C, (middle left) 40°C, (middle right) 35°C, (bottom left) 30°C and (bottom right) k_{on} as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE linker shown at the top of the figure.

6 base pair SE (Cont'd.)

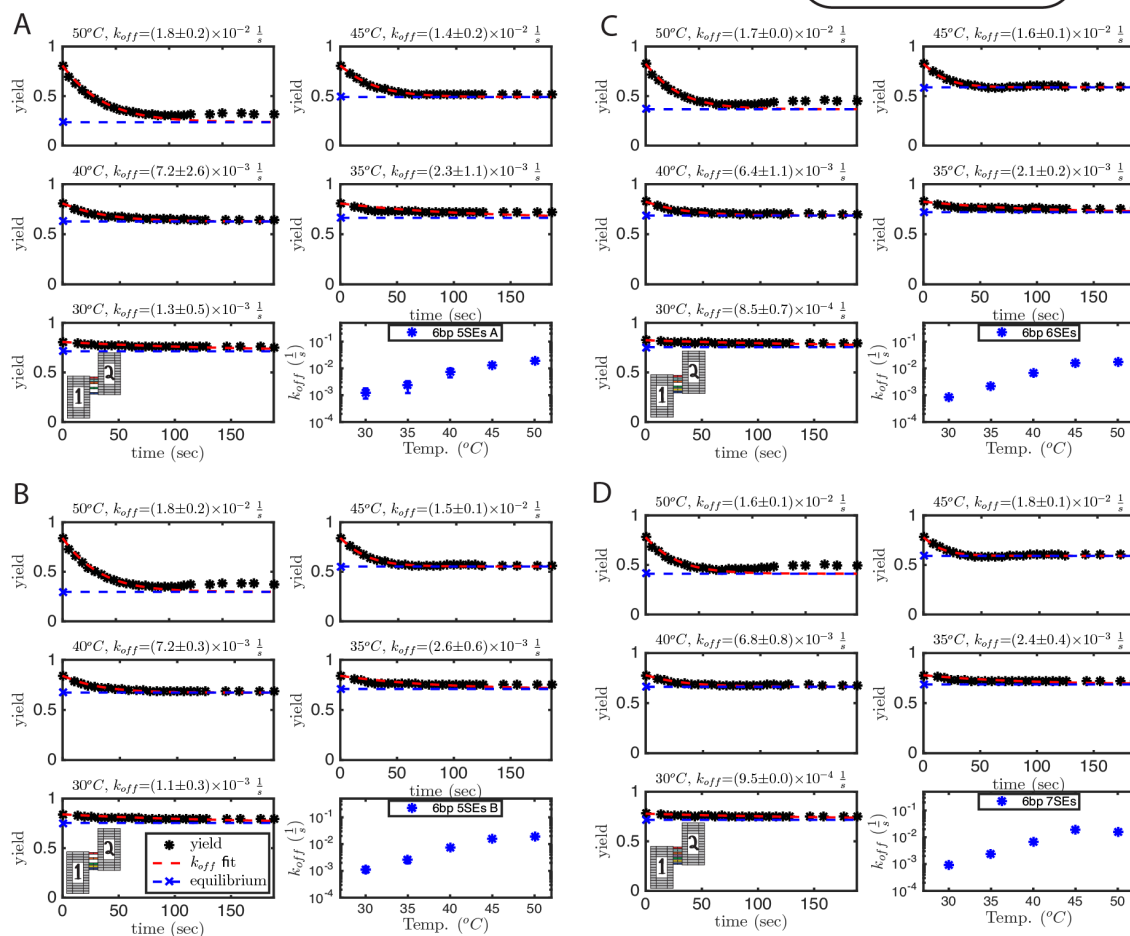


Figure S 3.26 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Six plots per interface are fits for k_{off} for temperature jump experiments from 25°C to (top left) 50°C, (top right) 45°C, (middle left) 40°C, (middle right) 35°C, (bottom left) 30°C and (bottom right) k_{on} as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE linker shown at the top of the figure.

6 base pair floppy SE

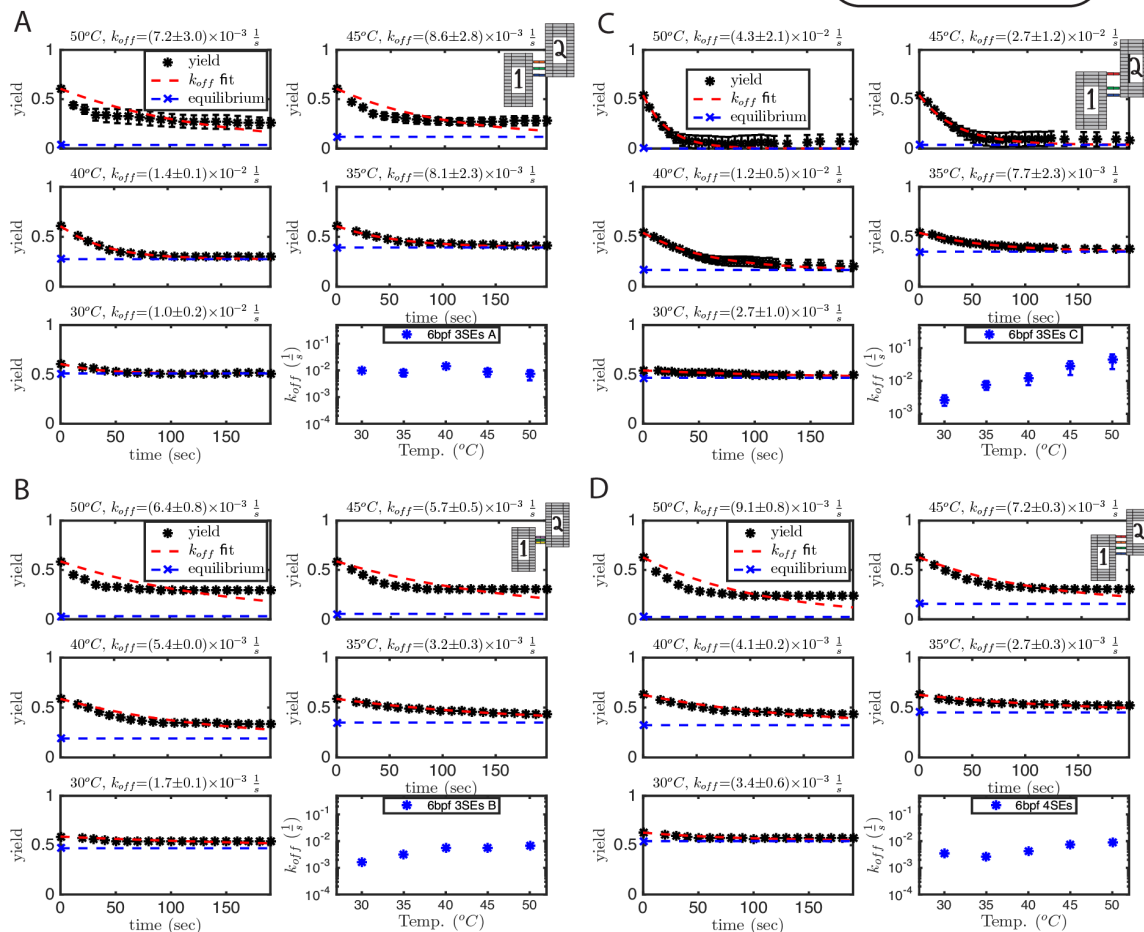


Figure S 3.27 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE floppy linkers. Interfaces are comprised of (A) 3 linkers in arrangement “A” (B) 3 linkers in arrangement “B”, (C) 3 linkers in arrangement “C”, (D) 4 linkers. Six plots per interface are fits for k_{off} for temperature jump experiments from 25°C to (top left) 50°C, (top right) 45°C, (middle left) 40°C, (middle right) 35°C, (bottom left) 30°C and (bottom right) k_{on} as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE floppy linker shown at the top of the figure.

6 base pair floppy SE (Cont'd.)

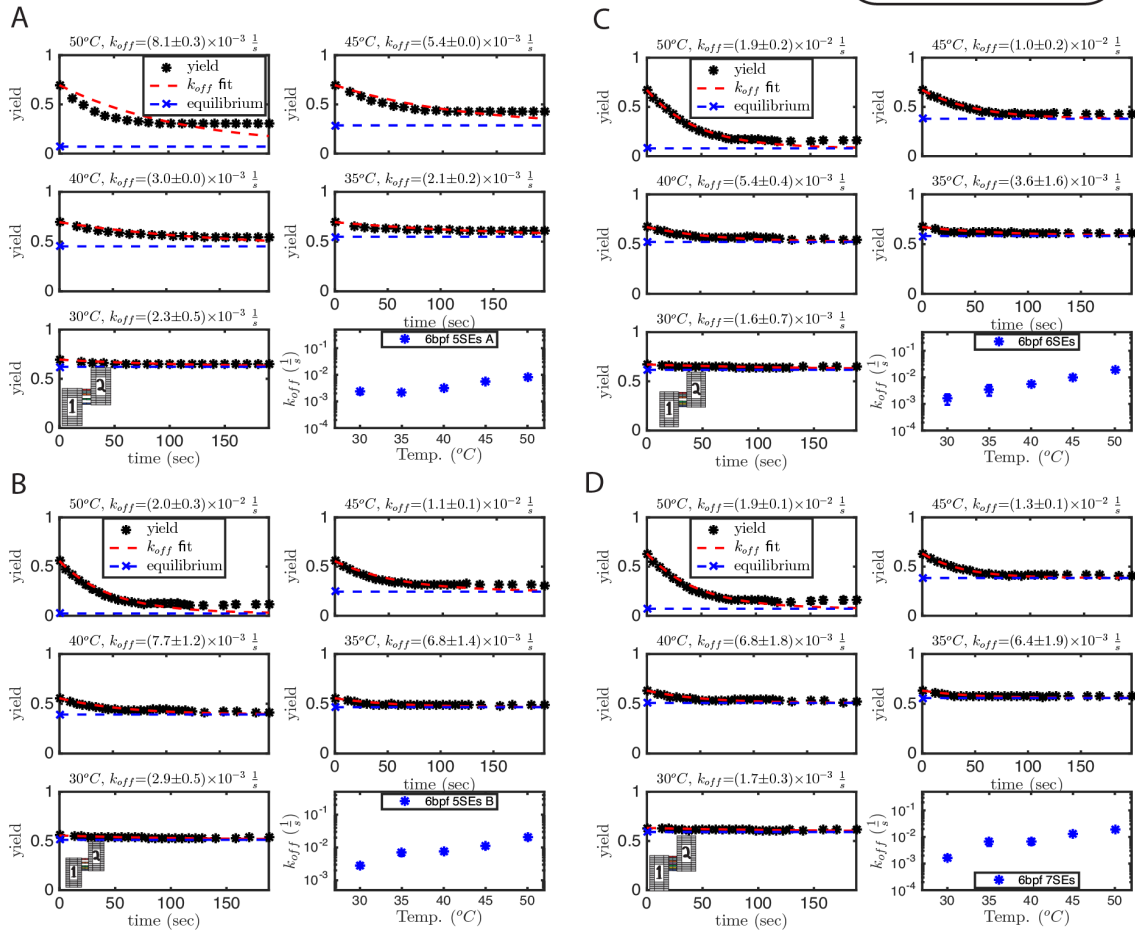
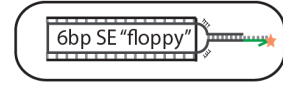


Figure S 3.28 Kinetics measurements: reverse reaction rate constants for interfaces with 6bp SE floppy linkers (continued). Interfaces are comprised of (A) 5 linkers in arrangement “A” (B) 5 linkers in arrangement “B”, (C) 6 linkers, (D) 7 linkers. Six plots per interface are fits for k_{off} for temperature jump experiments from 25°C to (top left) 50°C, (top right) 45°C, (middle left) 40°C, (middle right) 35°C, (bottom left) 30°C and (bottom right) k_{on} as a function of temperature. Cartoon schematic of T1 and T2 to the left of plots depicts interface arrangement. Cartoon schematic of an individual 6bp SE floppy linker shown at the top of the figure.

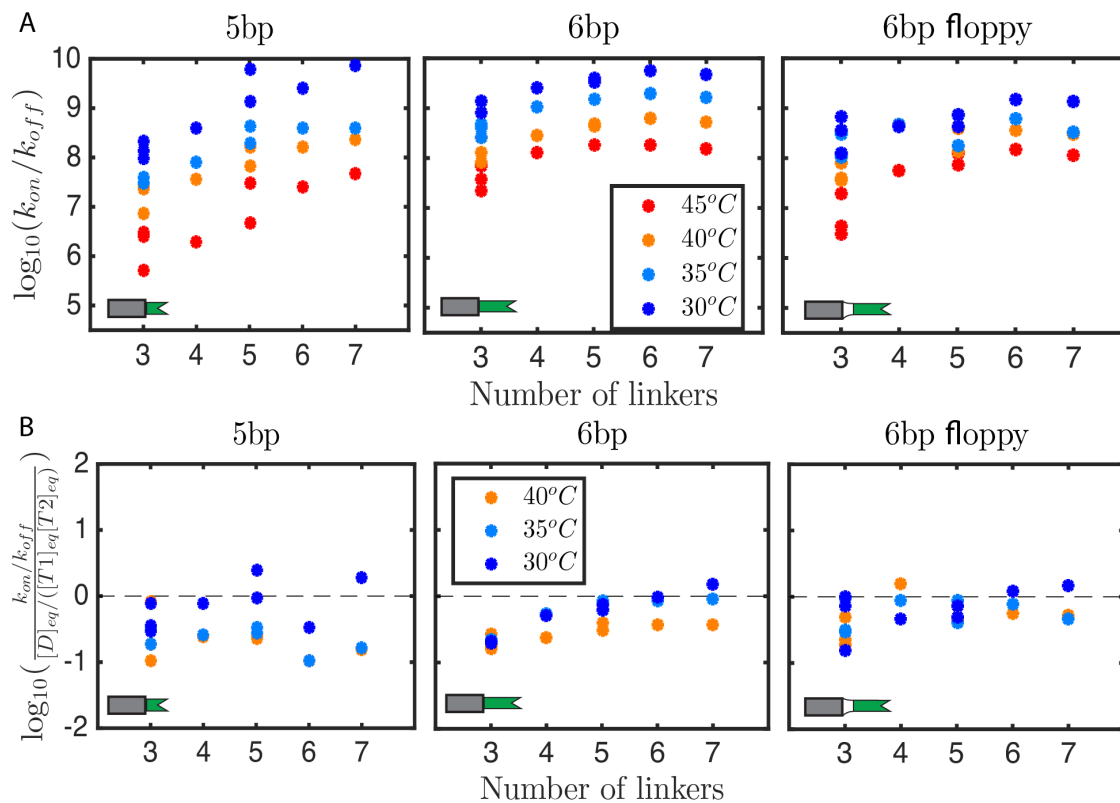


Figure S 3.29 Equilibrium constants (K_{eq}) of dimerization. (A) K_{eq} values calculated from the fit k_{on} and k_{off} reaction rate constants and as a function of temperature and interface design. As expected, higher temperatures generally give lower values of K_{eq} across all interface designs. (B) Ratio of K_{eq} values found from fit reaction rate constants and yields at equilibrium and used as a measure of error in our fits. Ideally, the fit kinetic parameters would exactly reflect the equilibrium yield and thus the ratio of the two would be unity. However, these measurements are not truly independent, as equilibrium yields are used to fit reaction rate constants so we expect some degree of dependence between the two calculations. (Left) 5bp SE linker design, (middle) and (right) 6bp SE floppy linker designs. Black dashed lines serve as guide for the eye, indicating equal values for the equilibrium constant for both methods.

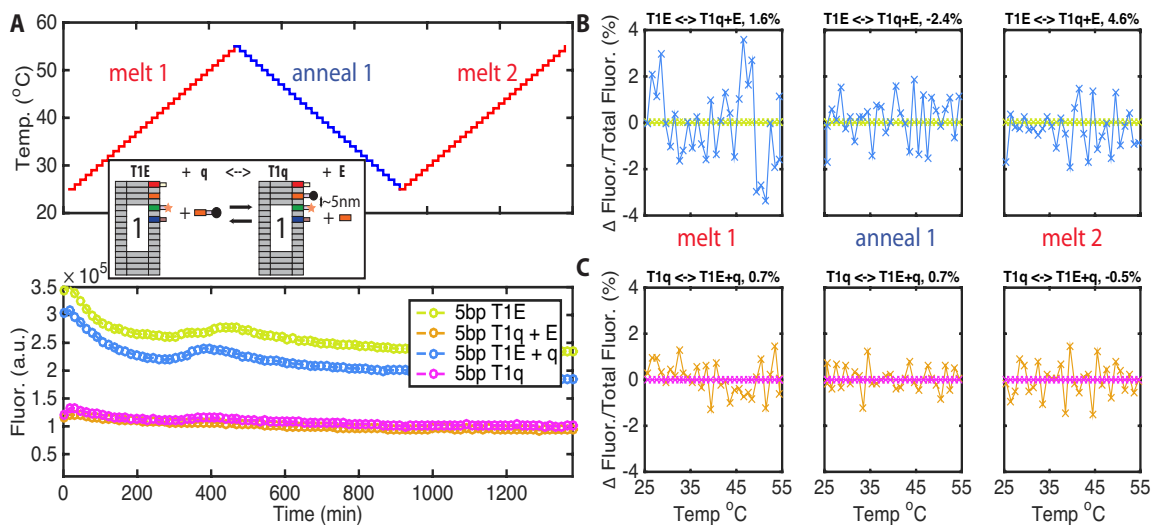


Figure S 3.30 Less than $\sim 5\%$ of the edge linkers dissociate from origami, even at high temperatures. We wanted to know whether linkers were being swapped in between origami components (*i.e.*, a given T2 linker has the same scaffold-complimentary region as the T1 linker) and thus contributing to low yields. To test this, we mixed T1 with a 5bp 3SE interface (labeled “T1E”) with the corresponding T2 quencher (“q”) linker in solution and tracked the fluorescence over a (A) typical annealing protocol (top). Fluorescence would decrease if T1 with the quencher (“T1q”) formed from T1E + q. Tile schematic shown in inset. (Bottom) Raw fluorescence data over time shown in the lower part of (A). (B) Fluorescence of T1E + q was normalized by T1E to show how the quencher influences fluorescence for the first melt (left), anneal (middle) and second melt (right). To obtain an estimate for how much linker exchange occurs in solution on a given section of a thermal protocol, we calculated the integral of the normalized difference in fluorescence of the sample divided by the fluorescence difference between “T1E” and “T1q” (*i.e.*, the fluorescence difference between fully fluorescent and fully quenched states). Estimates for the percentages of linkers swapped are listed above the plots. Negative values indicate the reverse reaction occurs. (C) The same system in (B), but now comparing the swapping of “q” with “E” and normalized by T1q.

3.5.5 Supporting Note 3.5: Two-dimensional Lattice with 6bp Floppy Linkers

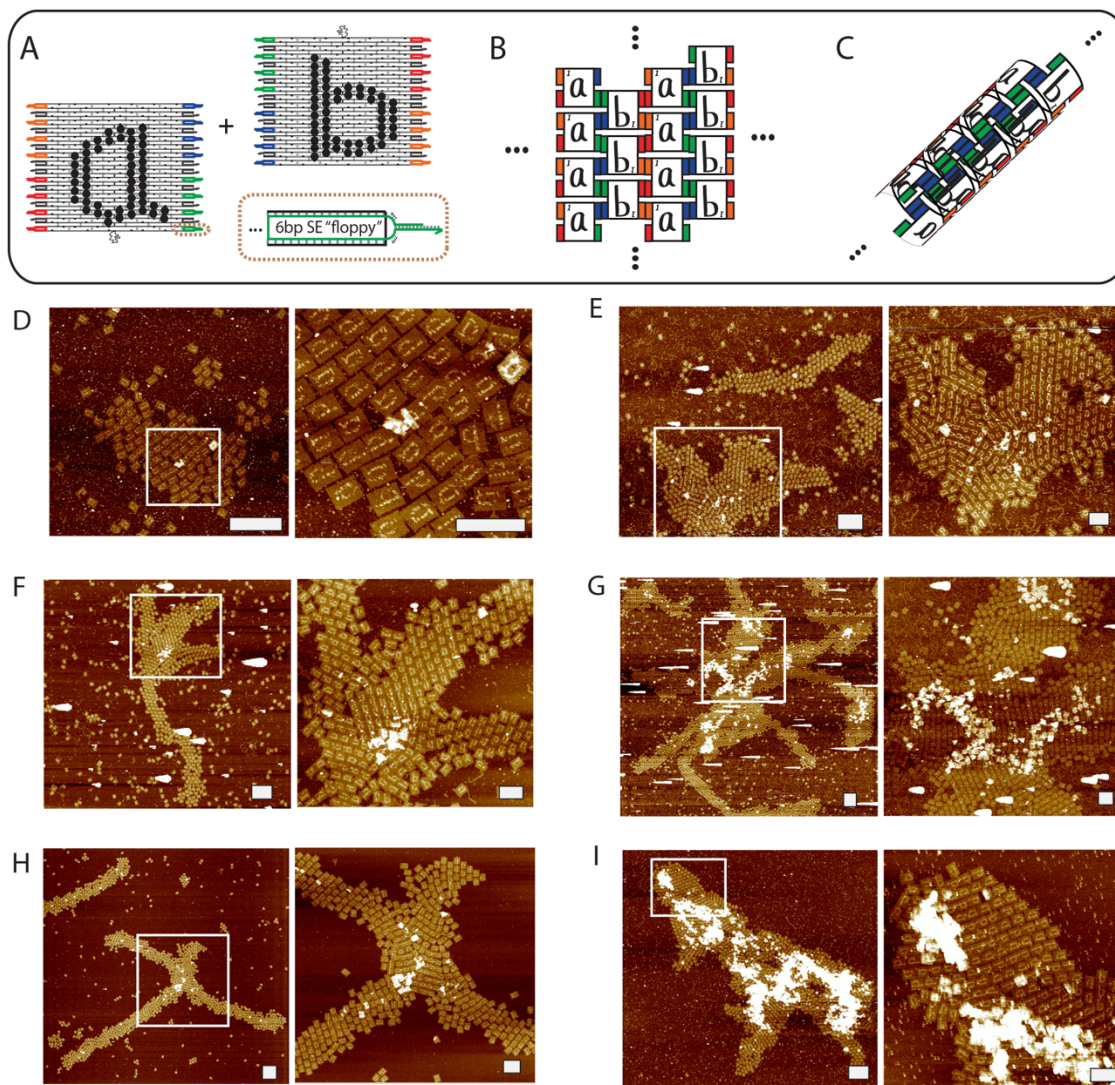


Figure S 3.31 Two-dimensional origami tile lattice with two tiles labeled “a” and “b” designed with four 6bp floppy SEs per interface. (A) Schematic of tiles “a” and “b” that form a lattice. These tiles are structurally identical to T1 and T2, except for the hairpin staples and linking schematic. For simplicity, linker edges are shown in a uniform color, although each sticky end sequence and scaffold-linker complementary region has unique sequences. All linkers were 6bp floppy linkers, with 4SE on each interaction interface. (B) Schematic of 2D lattice (intended) and (C) tube (unintended) as possible confirmations of the “a”-“b” lattice. (D-I) AFM images of origami lattices annealed with various protocols (see Table S5). White rectangles on the top images indicate zoomed-in region. Scale bars on the zoomed-out (top) images are 500nm and on the zoomed-in (bottom) are 200nm.

Table S 3.4 Concentration and Annealing Protocol for Infinite Lattice*

Image	$[A_0] = [B_0]$	Annealing Protocol
(D)	2nM	$55 \rightarrow 38^\circ C @ -\frac{1^\circ C}{r}$
(E)	2nM	$55 \rightarrow 36^\circ C @ -\frac{0.33^\circ C}{hr}$ then hold at 38 for 48hrs
(F)	2nM	$55 \rightarrow 36^\circ C @ -\frac{0.33^\circ C}{hr}$, then hold at 42 for 48hrs
(G)	2nM	$55 \rightarrow 36^\circ C @ -\frac{0.33^\circ C}{hr}$, then hold at 38 for 3weeks
(H)	5nM	$55 \rightarrow 40^\circ C @ -\frac{0.5^\circ C}{hr}$, then hold at 40 for 6hrs
(I)	5nM	$55 \rightarrow 25^\circ C @ -\frac{0.5^\circ C}{hr}$, plated @25

*All solutions were diluted to 1 nM before imaging and plated on the mica puck at 40 C, unless otherwise noted

3.5.6 Supporting Note 3.6: Design of Origami Edge Staples and Linker Sequences

Linker sticky ends and intra-linker complementary regions were designed using custom MATLAB software written to minimize spurious interactions. The script minimizes crosstalk between non-complementary SEs and intra-linker regions while keeping all desired SE interactions at approximately equal binding strength +/-20% (ΔG^0), according to the nearest neighbor model^{99, 132} (Figure S3.32). Intra-linker complementary regions were also unique, and their crosstalk minimized with one another and the SEs. The core MATLAB code we modified was used previously in Refs. ^{25, 118} and modified as needed for this work. The software is available upon request.

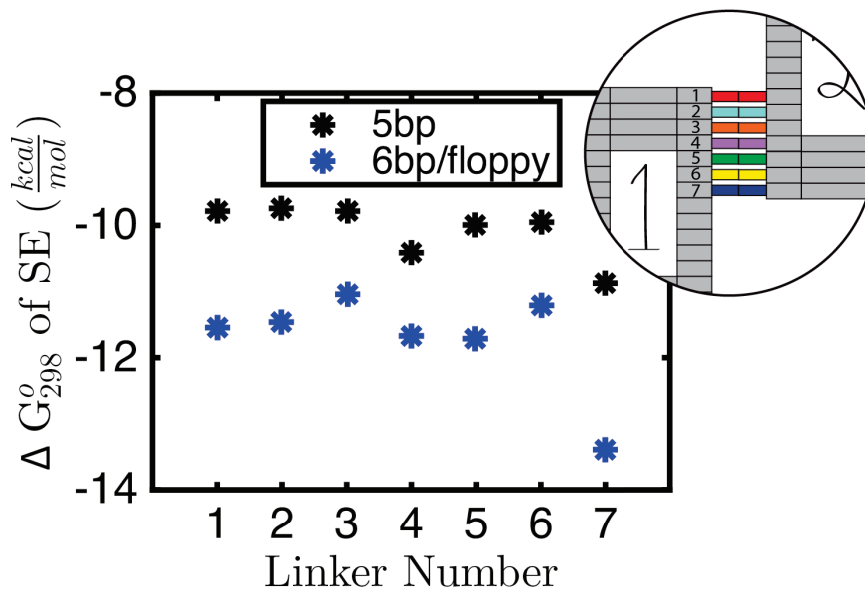


Figure S 3.32 Designed strength of SE, ΔG_{298}° , as calculated by the nearest neighbor model⁹⁹. The crosstalk between all SE strands was minimized and their relative strength was constrained to be within +/-15%. Linker number 1 corresponds to the top most linker on T1 while linker number 7 is the bottom most linker on T1, as depicted in inset. Note the 6bp SEs and 6bp floppy SEs have same sequence and thus same predicted free energy in our simple model.

3.6 ACKNOWLEDGMENTS

The authors would like to thank Abdul Mohammed, Joshua Fern, Dominic Scalise, Deepak Agarwal and Angelo Cangalosi for helpful discussions regarding this work and Jennifer Lu for designing the “A” and “B” hairpins for the infinite lattice. The full, published manuscript can be found at <http://dx.doi.org/10.1021/jacs.5b10502>.

CHAPTER 4 OPTIMIZING COMPONENT-COMPONENT INTERACTION ENERGIES IN THE SELF-ASSEMBLY OF FINITE, MULTICOMPONENT STRUCTURES

SUMMARY

Components involved in multicomponent biological self-assembly processes generally have interfaces that are optimized to enable the process to proceed quickly and reliably. In contrast, the principles for designing such interfaces for engineered multicomponent self-assembly processes, such as those involving nucleic acid components, are still being developed. Inspired by biological systems, here we use stochastic kinetic computer simulations to understand how to tune interface strengths to robustly self-assemble biomolecular target structures that are finite-sized and have multiple different components. We found that high yields of a desired product can be achieved across a broad range of isothermal assembly conditions by changing the interaction energies between components and that heterogeneous (*i.e.*, non-uniform) component-component interaction energies can improve self-assembly outcomes where the best assembling structures often include a strong-strong-strong-weak interface motif in closed ring substructures. We show that annealing is a robust strategy for self-assembling structures with a wide variety of interfacial energies, but structures with uniform interfacial energies are some of the least efficient assemblers. Finally, we apply this optimization process to an experimentally demonstrated self-assembly system, using measured kinetic and thermodynamic parameters and show that the initial design of the assembling complex could be improved (*i.e.*, made to assemble with higher yield) by 20-60% using different interface designs. This work suggests that this type of iterative, computational optimization can improve the design cycle for an engineered complex by suggesting next-generation complex designs and preventing the need to experimentally test many different designs.

4.1 INTRODUCTION

Nature elegantly assembles RNA and protein complexes such as the nuclear pore complex^{56, 136}, kinetochore⁵⁸ and ribosome¹³⁷, from individual components with high yields and with control over where and when components assemble^{93, 137, 138, 139}. Millions of years of selection pressures have largely determined biophysical properties of complexes, from characteristics such as thermostability¹⁴⁰, to interaction energies between components^{137, 138, 139}, to assembly pathways⁸⁹ and quaternary structures^{134, 141}. Complex formation highlights nature's strong control over both the folding and assembly landscape.

Without the advantage of millions of years of evolution, nanoscale engineers attempt to design self-assembly processes that rapidly produce desired structures with high yields from a set of components. To achieve this for a given system, the most easily controlled variable is the assembly protocol (*i.e.*, the reaction conditions)^{67, 68, 69, 115, 142}. Common protocols include isothermal assembly, such as rapid folding of DNA origami³⁴, altering the salinity of the assembly mixture²¹ and annealing (*i.e.*, cooling the solution temperature from above to below the structure's melting temperature). Annealing is the most widely-used protocol because the a temperature window that enables rapid assembly is likely to be found with little optimization^{34, 143}. In oligonucleotide-based self-assembly, for example, annealing is used to assemble 2D and 3D DNA "brick" structures consisting of hundreds of short ssDNA components^{68, 69} and to fold DNA origami nanostructures¹⁰. However, annealing is not guaranteed to result in rapid, high yield assembly and requires guess-and-check for improving outcomes. The resulting highly optimized thermal protocols are typically non-trivial (such as a thermal zig-zag) and produce low yields and/or non ideal assembly outcomes such as aggregates and intermediates^{67, 133, 142, 144}.

Another important design factor in nanoscale assembly processes that can be tuned and affects the assembly outcome is the interaction strengths between components^{115, 144}. Unlike designs of biological complexes, where there is strong evidence for a wide range of interaction

energies in a given complex^{81, 134}, many nanoscale engineers attempt to homogenize such energies, *i.e.* make all interaction strengths equal. For example, in assembling DNA-based nanostructures each component interface has the same number of DNA base pairs available to bind to a complementary interface^{10, 67, 68, 69, 115, 133, 144}. This naïve approach has typically resulted in low yields even with optimized assembly protocols^{67, 115}. More recently, studies of the self-assembly of DNA bricks are demonstrating the benefits of assembling components that have a wide distribution of interaction energies^{68, 69, 143}, because uniform interaction energies result in detrimental assembly phenomenon such as kinetic trapping¹⁴³. Similarly, a recent study of DNA-origami folding has shown that by using longer, stronger “staple” strands to influence early formation of long range interactions, the assembly pathways of the origami structure can be rationally designed, resulting in more robust assembly^{145, 146}. These increasing improvements in assembly outcomes through the use of heterogeneous interaction strengths between components in a finite-sized structure highlight the potential import of such heterogeneous interactions.

Instead of experimentally optimizing thermal protocols and designs of interfaces in a finite-sized self-assembling complex, which are laborious to design and test, we introduce a computational approach for optimizing the self-assembly of finite sized-structures and show that these techniques can be used to inform the next-generation of designs for a reported self-assembly system with measured kinetic and thermodynamic parameters. Here, “optimized” is considered the rapid production of target structures with high-yield across a broad range isothermal conditions. We further develop a coarse-grained model¹³² and use computational simulations to investigate: (1) how a self-assembly system can be optimized through iterative, small scale improvements such as those accomplished in a directed evolution algorithm and (2) will an optimized system result in better assembly under other circumstances as well?

A priori, it might seem that uniform interaction strengths in a target structure is the optimal design for an efficient self-assembly process^{64, 147}, however, we found that this is not a good strategy unless designing a 1D process. Independent of the assembly protocol, multicomponent structures with multidimensional topology are most efficiently formed when interfaces in a target structure have a wide distribution of energies, suggesting hierarchical assembly is a good strategy for complex design. Optimized structures were equally as affected by stoichiometric imbalances of components as structures with uniform interaction energies. Finally, we tested the optimization protocol using an experimental, DNA origami-based complex and corresponding interface designs with measured kinetic and thermodynamic parameters^{115, 144}, and found that we could suggest specific interface designs which would enhance the assembly of the complex. This work not only suggests that a distribution of interface strengths and their arrangement within the target structure is a critical, but historically overlooked, factor in improving assembly outcomes but also that such a computational, iterative optimization approach using real-world data might be useful in designing more complex structures.

4.2 MODEL AND METHODS

To computationally study biomolecular self-assembly processes, we adapted a coarse-grained kinetic model from *Chapter 2*. This model describes the assembly kinetics of biomolecular complexes, using finite-sized rectangular lattices in 2D and a cubic lattice in 3D as prototypical structures. It assumes short-ranged, specific, pairwise interactions between assembling structures, additive interface strengths and no interface crosstalk, non-specific interactions or component rotation (see Fig. 4.1A). Assembly begins with an initial, fixed number of each component which are depleted as the reaction progresses. All possible intermediate assemblies (*i.e.*, structures that have a connected subset of components in a complex) can form. We use regular rectangular lattices

in 2D and a regular cubic lattice in 3D as prototypical structures, which we expect to be good representative structures to a general class of such self-assembly problems.

The model implements a Gillespie algorithm¹⁴⁸ to describe reaction kinetics and assumes a constant, temperature-independent macroscopic forward reaction rate constant, k_{on} , similar to experimentally measured values for biomolecules^{70,71,72,73} and reverse reaction rate constant which depends on the standard Gibbs free energy, defined as $\Delta G^o = \Delta H^o - T\Delta S^o$, of the reaction. We set the standard enthalpy, ΔH^o , and standard entropy, ΔS^o , from experimentally measured values of two 5 base-pair DNA-DNA hybridization reactions through “sticky ends” in a DNA nanostructure assembly process²⁵, which have values comparable to other biomolecular interactions^{99,100,101}. The reverse reaction rate constant is mathematically defined as:

$$k_{off,b} = k_{on} \exp\left(\frac{b\Delta G^o}{RT}\right), \quad (4.1)$$

where R is the universal gas constant, ΔG^o is the standard Gibbs free energy (i.e., binding strength) through a single interface, T is the reaction temperature and the interface strength factor, b , is a tunable parameter that helps determine the strength of the interaction. We limited $0 \leq b \leq 2$ to act as a conservative estimate to the interaction strengths that are achievable with straightforward manipulation of the interface, for example by altering the number or type of contacts, length or sequence of the interacting nucleotides in an interface¹⁴⁴.

To describe the general state of the system, we use dimensionless variables. We define dimensionless time, τ :

$$\tau = k_{on}[X]_0 t, \quad (4.2)$$

which depends upon the initial component concentration, $[X]_0$, macroscopic forward reaction rate constant, and dimensional time t in seconds. Further, we define a dimensionless reaction temperature, η (high values of η correspond to low temperatures and *vice versa*):

$$\eta \equiv \log_{10}\left(\frac{k_{on}}{k_{off,1}}[X]_0\right). \quad (4.3)$$

In our simulations, we assemble both isothermally and *via* annealing. Isothermal assembly takes place at integral values of η in the range of -6 to 6. In annealing, the solution cools linearly from $\eta = -6$ to $\eta = 6$, in a stepwise fashion such that the solution is subject to 100 different values of η , each for durations of $\tau_{anneal}/100$, where τ_{anneal} is the total dimensionless time spent annealing. The dimensionless timescales and temperatures we use in these simulations are realistic, corresponding to a range of temperatures from about $T = 8^\circ\text{C}$ ($\eta = -6$) to $T = 58^\circ\text{C}$ ($\eta = 6$) for and timescales from 30 minutes ($\tau = 10$) to 2 days ($\tau = 1000$) for 10 nM of components, for example.

We define *yield* as the fraction of total starting material incorporated into the target structure, which for example can easily be determined by gel assay^{68, 69}.

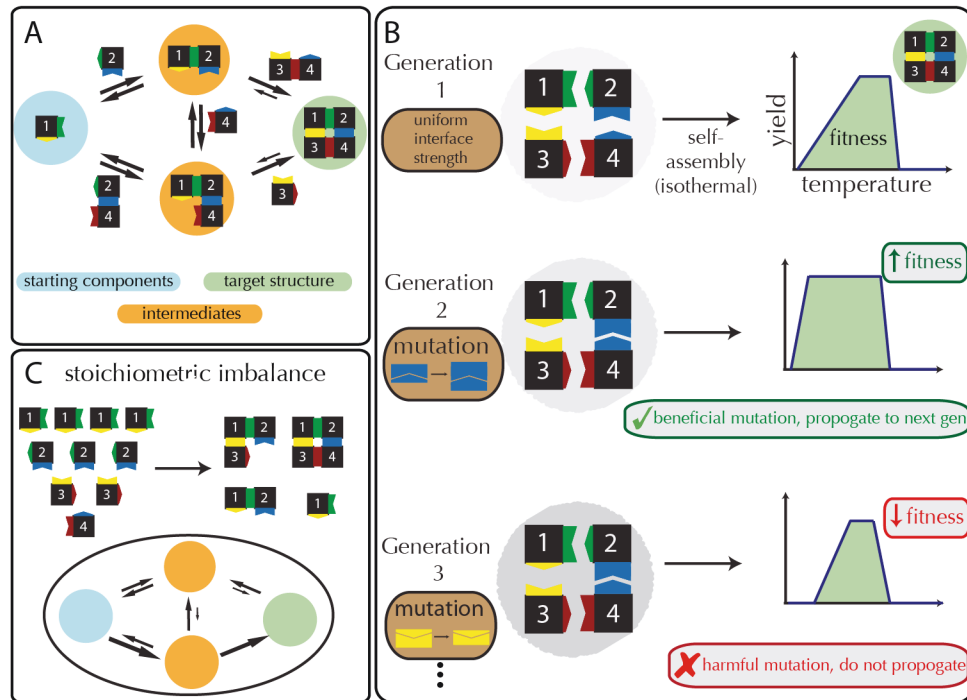


Figure 4.1. Simulated multicomponent self-assembly and component optimization shown with a 2x2 grid system. (A) Examples of possible reactions in a 2x2 grid system, where starting components

interact with starting components or intermediates, which then form target structures. Color choice and shape of interfaces on the square components indicate bond type and complementarity, respectively, while interface size indicates bond strength. Components and assemblies cannot be rotated or bind to themselves and are numbered to indicate uniqueness. Arrow size qualitatively indicates reaction propensity. (B) A directed evolution algorithm is used to find interfaces that allow robust assembly across a range of conditions. The algorithm begins with generation 1, where all interfaces have equal strength. We simulate isothermal self-assembly at many temperatures for a given period of time and plot assembly yield as a function of temperature. The fitness of a particular set of components is defined as the area under this curve (see Model and Methods in Section 4.2). Then we mutate the strength of one interface in the complex and measure the fitness of the new structure. If the fitness increased from the previous generation, as shown from generation 1 to 2, the mutation is retained, otherwise the interface returns to its previous strength. This process continues for a set number of generations, or different arrangements of interface strengths (typically 500). (C) Component stoichiometry shapes the assembly landscape (depicted in oval below the arrow), which influences the reaction outcome. Without stoichiometric imbalances (i.e., equimolar ratios of starting components), all starting material can be incorporated into the target structure, which is not the case for systems with imbalances. Depicted reaction outcomes are hypothetical.

4.2.1 Directed Evolution of Interfaces

One of the most important design considerations in building addressable self-assembly processes for regular structures that have high yields is the design of the interfaces between components. In addressable self-assembly processes, notably those involving nucleotides, there are a large number of possible sequences that can be chosen for each interface. This large library of potential interactions means not only that a large library of pairwise interactions with low crosstalk can be developed, but also that the energy of interaction between each pair of components can be tuned. In spite of this interfacial tunability, in practice building structures with *exactly* the designed interfacial energies is not possible, but building structures very close to designed energies is. This suggests that the best way to understand how to tune these structures is by studying the many different designs near one another in phase space which lead to a desired outcome.

In practice, one feasible method for optimizing such addressable structures is to iteratively improve the assembly efficiency by fabricating sets of components (*e.g.*, with different number and/or type of contacts), testing their assembly efficiency (*e.g.*, by using a gel or fluorescence assay

to rapidly characterize yields) then selecting for and tweaking the design of the best sets for the next generation design. Such a process guided by rational design could make an assembly processes faster and have higher yield.

An isothermal protocol is perhaps the most ideal assembly protocol, as there are many situations where annealing is not ideal or even possible. For instance, in assembling dynamic and reorganizable structures in physiological or other highly temperature sensitive conditions or protein complexes, which are typically subject to tight temperature ranges and well-defined environmental conditions. Furthermore, it is rarely the case that that a system which assembles efficiently *via* annealing conditions would be more desirable than one which assembles well isothermally. If a structure can assemble well isothermally, it most likely also can assemble well *via* annealing (even if cooling has to be very slow *e.g.* in the case of a tight optimal temperature window) but the opposite is not necessarily the case. Such examples highlight the importance enabling self-assembly under isothermal conditions thus informs the selection criterion for these simulations.

With such practical details in mind, we implemented a iterative optimization algorithm which we call the directed evolution of interfaces. The goal of this algorithm is to take a simple starting structure and iteratively mutate the interface strengths such that the optimized structure is one that assembles rapidly and produces high yield across a broad range of isothermal conditions. The algorithm starts by determining the “fitness” of a simple structure where all interface strength factors are equal ($b = 1$). Fitness is defined here as the integral of the *yield* versus η curve at time $\tau = 1000$ where assembly is isothermal. The algorithm then “mutates” the strength of one interface in the structure (*i.e.*, changes the value of b), where all interfaces have the same probability of being chosen to mutate, and then determines the fitness of this next generation structure. The mutation is propagated to the subsequent generation if it resulted in an increase in fitness of the structure as compared to the previous generation, otherwise the interaction strength returned to its previous

value (see Fig. 4.1B). The algorithm repeated this mutation/selection/propagation process for 500 generations. All components had initial concentration $[X]_0$ (*i.e.*, balanced stoichiometry).

4.2.2 Directed Evolution of Interfaces of a DNA Origami-Based Complex with Experimentally Measured Kinetic and Thermodynamic Parameters

This optimization algorithm has the same basic structure as the directed evolution of interfaces algorithm as described above, *i.e.*, it starts with an initial complex design and through many mutations of the interfaces arrives at a complex design that enhances the fitness of the structure, however, it uses experimental measurements of the self-assembly of DNA origami components based on two sources in the literature. Specifically, this algorithm optimizes the tetramer (2x2) DNA origami-based complex, which has a measured value for additivity of $\alpha = 0.58$, as reported in Ref. ¹¹⁵ (see SI Note 4.2 for additivity calculation). The algorithm can mutate an interface to any one of 24 different designs, with measured forward and reverse reaction rate constants, as reported in Ref. ¹⁴⁴ (see Table S4.1-4.2 for measured rate constants). To reflect specific experimental conditions, we use dimensional temperature, time and concentration values. The algorithm isothermally assembles the components for 72 hours at the four different temperatures reported in Ref. ¹⁴⁴ (30°C, 35°C, 40°C, 45°C) and selects for the structure's fitness, which is defined as the area under the yield versus temperature curve (see SI Note 4.2 for more details). The algorithm iterated through 500 designs and all components had initial concentration of $[X]_0 = 5$ nM.

4.2.3 Stoichiometric Imbalance

Variations in synthesis, concentration measurement or mixing often mean that the concentrations of components can vary significantly from one assembly process to another; such

variation can influence assembly outcome^{10, 68, 69}. We simulate this variation, which we assume to be unbiased by setting the initial concentration of each component j to be $\gamma_j[X]_0$, where $[X]_0$ is the desired concentration and γ_j is a Gaussian random variable with mean $\mu = 1$ and standard deviation σ constrained to $\gamma_j > 0$. We consider values of σ between 0 and 0.25. Stochastic stoichiometric simulations are performed in replicates of 10 (see Fig. 4.1C).

4.3 RESULTS AND DISCUSSION

To investigate the iterative optimization for the self-assembly of finite structures that produce efficient, high-yield assembly, we begin our study by simulating the directed evolution of interfaces (see Model and Methods in Section 4.2) for regular grid structures with various dimensionality and numbers of components including: 1x5 line structure in 1D, 2x2, 3x3 and 4x4 square grid structures in 2D and a 2x2x2 cubic structure in 3D.

Uniform interaction strength between components leads to high yields of 1D structures but not 2D or 3D structures. The interfaces that evolved between the components self-assembling 1D structures were all strong, consistent with previous observations that such interfaces maximize the yield of these structures under a variety of assembly conditions. In the 1x5 line structure, the optimized structure had uniform interface strength factors of ~ 2 (the highest possible), which increased the fitness of the structure by more than 130% (see Fig. S4.1-4.2). However, such uniformly strong interfaces do not lead to optimal yields of higher dimensional structures. The optimized 2x2 had three strong interfaces and one weak interface ($b_1 = b_2 = 2$, $b_3 = 1.99$, $b_4 = 0.78$). Structures with these interfaces were 80% fitter than the original structures, assembling with virtually 100% yield at almost all temperatures considered. Similarly, mixtures of strong, medium and weak interface strengths were observed in the optimized 3x3, 4x4, and 2x2x2 assemblies; these structures increased fitness by 120%, 220%, and 180%, respectively, in comparison to the initial

structures (see Fig. 4.2B-D, S4.2). This suggests that both higher dimensional target structures and those with more components show an increased benefit from optimizing interaction strengths; in such structures the efficient assembly range, which we define as the range of assembly conditions that produce >80% yield after $\tau = 1000$ (shown as the purple bar underneath plots in Fig. 4.2), shrinks with these two variables when structures have uniform interfacial strengths.

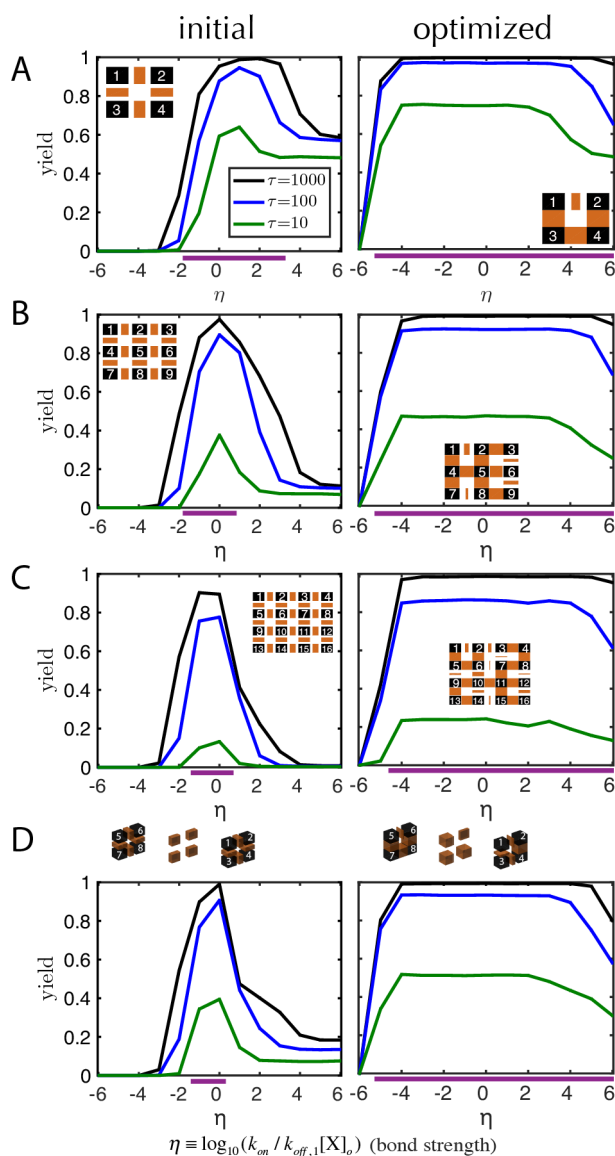


Figure 4.2. A mixture of strong and weak interfaces between components can enhance the fitness of 2D and 3D structures. Yield as a function of assembly temperature for the initial (left) and optimized

right) structure for (A) 2x2 grid, (B) 3x3 grid, (C) 4x4 grid, and (D) 2x2x2 cube target structures. Inset diagrams show complex structure: black squares are components and the thickness of the brown bands indicate the strength of interaction between them. Different component types each have different numbers. All interface strengths are equal initially. The 2x2x2 diagram is split into three separate cartoon diagrams to more easily visualize the interface strengths, where the left and right diagrams depict different sets of 4 components in the target structure. The size of the brown block is proportional to the interface strength. Purple lines underneath plots indicate efficient assembly regime, or where target structure $yield > 80\%$ after $\tau = 1000$.

Optimized structures assemble faster or just as fast as initial structures. Our next goal was to determine whether optimizing a complex using the selection criteria in the directed evolution of interfaces simulations *i.e.*, yield across a broad range of isothermal conditions at a particular long assembly time, also increased the speed of structure formation. To do this, we compared the rates of formation, defined mathematically as $yield/\tau$, for the optimized assemblies to those of the initial assemblies at the three reaction times tested and found that optimized assemblies assemble faster (up to $\sim 10^4$ faster) or just as fast as initial assemblies, at all isothermal conditions for all structures tested (see Fig. S4.3). This suggests that by optimizing structures using this selection criteria, structures gain the added bonus of increasing the speed of assembly for free.

Strong-strong-strong-weak interface strengths in a four component ring is common to efficient assemblers. Turning a 2D assembly problem into a pseudo-1D problem has been shown to be a good design strategy for multicomponent structures. In protein rings, which have evolved to contain one significantly weaker interface than the rest, assembly occurs through the pathway where components with strong interfaces first form the chain and finally the component with the weak interface completes the ring⁸¹. Our findings here not only corroborate these results in a simple 2x2 structure (see Fig. 4.2B), but also in more complex 2D and 3D structures. We found that optimal interface design commonly includes three strong interfaces and one significantly weaker interface in closed-ring substructures in a complex. For example, the optimized 3x3 square grid structure has four of four possible (4/4) closed four-component rings which share this motif (see Fig. 4.2B), the optimized 4x4 square grid has 5/9 (see Fig. 4.2C), and the optimized 2x2x2

cube has $5/7$ (see Fig. 4.2D). In the optimized 4x4 structure, for example, the four, four-component rings (quadrants) at the corners share this motif but there are minimal interfaces between the quadrants, suggesting a hierarchical assembly process where the four quadrants form and then those are assembled into a larger structure.

Multiple directed evolution of interfaces simulations converged to similar optimized structures for the 3x3 assembly. A common result in using genetic algorithms is that structures discovered through the optimization represent “local optima” such that different, fitter variants may be discovered by rerunning the process. We simulated the directed evolution of the 3x3 square grid three times from the same initial conditions (*i.e.*, uniform interaction strength) to see whether some iterations would produce fitter variants than others. We found that while the fitness values do converge to approximately 120% for all systems (see Fig. S4.4), the specific interaction strengths differ but are widely heterogeneous for all structures and suggest that all structures form *via* hierarchical assembly pathways. There are multiple “right” answers in terms of interaction strengths that produce structures which assemble efficiently. Further, the additional simulations of the 3x3 square grid converged to the most fit structures having interface strengths with 3/4 and 4/4 having the strong-strong-strong-weak interface strength motif. This suggests that a simple design strategy of strong-strong-strong-weak for interface strengths in higher dimensional structures, or more generally designing structures that assemble *via* hierarchical pathways, might be useful to implement in practice.

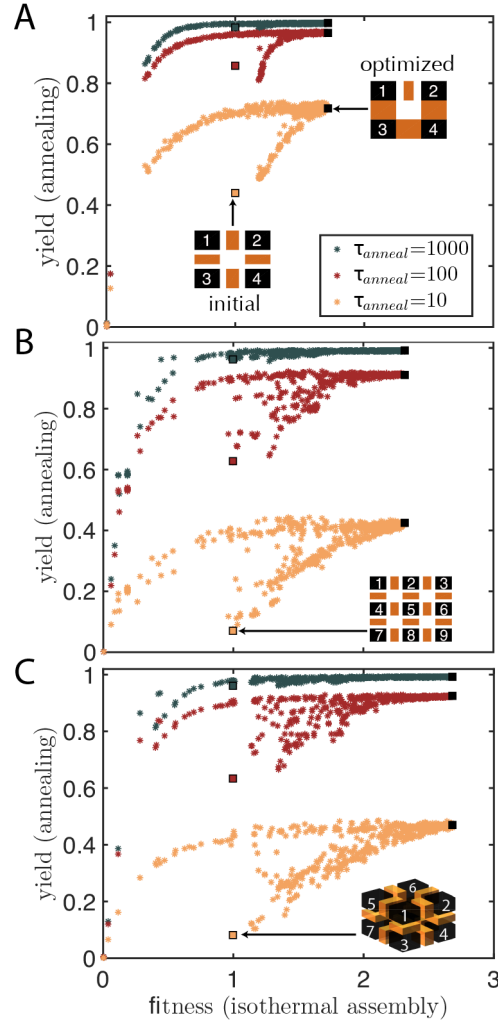


Figure 4.3. Structures with high fitness assemble via annealing and produce high yield. Yields of target complex after annealing as a function of fitness (isothermal assembly conditions) after various annealing assembly times, τ_{anneal} . The (A) 2x2, (B) 3x3 and (C) 2x2x2 systems are depicted. Interface strengths are equal initially, depicted in the inset diagrams and indicated by square markers outlined with black. All other generations shown as dots except the most fit generation is shown as filled-in black squares. The cartoon diagram for the optimized 2x2 depicted structure depicted in (A). Annealing protocol starts at $\eta = -6$ and steps toward $\eta = 6$ over a series of 100 steps (100 values of linearly increasing η), spending $\tau_{anneal}/100$ at each value of η . Yields are recorded at the end of the anneal ($\eta = 6$). Components are numbered to indicate uniqueness.

Structures with high fitness also assemble well when annealed. Many self-assembly processes rely on non-isothermal assembly conditions to achieve robust assembly^{10, 68, 69, 143} and is a growing body of computational studies, largely guided by principles observed in practice, that

annealing is the best method for promoting error-free, robust assembly of finite structures, because conditions that promote robust assembly are essentially guaranteed to be found^{85, 132, 143}, with longer annealing protocols typically resulting in higher yields⁶⁸. With this knowledge, our goal was to determine if optimized structures, or those with high fitness, also led to efficient assembly when annealed. To test this, we simulated the annealing of all 500 structures (generations) in the directed evolution of interfaces simulation for the 2x2, 3x3 and 2x2x2 systems (see Model and Methods in Section 4.2 for annealing protocol) and found that assemblies with high fitness values also have high yields after slow annealing (see Fig. 4.3), while the optimized structure produced the highest yields at intermediate times ($\tau_{anneal} \leq 100$) relative to all other structures at intermediate times. After long assembly times ($\tau_{anneal} = 1000$) yields approached 1 for all structures with fitness values approximately >0.5 , >0.75 and >1 for the 2x2, 3x3 and 2x2x2 assemblies, respectively. This suggests that annealing is an excellent method to assemble structures, as even less fit assemblies can produce high yields. However, assemblies with very low fitness values (≤ 0.5) do not assemble with high yield even at long assembly times likely due to the fact that one or more of their bonds are too weak to form, even at low temperatures. For assemblies with intermediate fitness values subject to shorter anneals, yields vary widely and do not necessarily produce higher yield above a fitness value of ≥ 0.5 . Interestingly, the least efficient assembler at intermediate times, excluding those structures with very low fitness values ≤ 0.25 , is the structure with uniform interaction strengths.

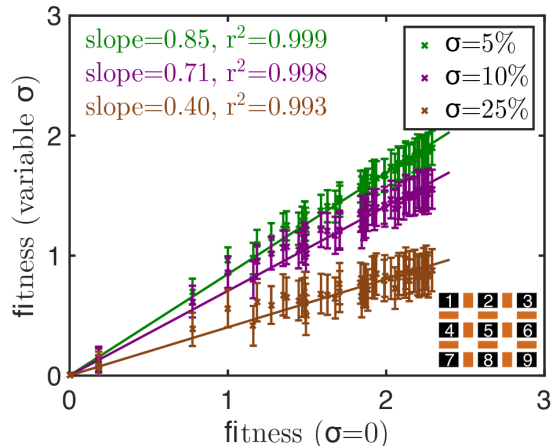


Figure 4.4. Fitness with variable stoichiometry (σ) as a function of fitness with no stoichiometric imbalance for 3x3 square grid system. Inset diagram depicts the target structure for generation 1, where all interface interaction energies are equal. Components are numbered to indicate uniqueness. Lines indicate linear regression, with statistics inset in plot. Ten simulations performed for each assembly condition. Error bars indicate one standard deviation of the reported quantity.

Structure formation rates under annealing and isothermal can vary widely. We next wanted to determine which protocol, annealing or isothermal assembly, produced structures most rapidly. To do this, we investigated the rate of structure formation for the two assembly methods for the population of all 500 structures in the 2x2, 3x3 and 2x2x2 assemblies and found that, on average, when isothermally assembled under moderate conditions ($-2 \lesssim \eta \lesssim 2$), isothermal assembly was faster than annealing, especially at shorter the reaction times, but under isothermal extremes ($\eta \lesssim -4$ and $\eta \gtrsim 4$), annealing was faster (see Fig. S4.5). Further investigation of formation rates for the optimized structure and the initial structure showed similar trends (see Fig. S4.6). Notably, for the optimized structure rates were about equal between the two methods at intermediate and long times if isothermal assembly took place under ideal conditions ($-4 \lesssim \eta \lesssim 4$) but isothermal assembly was slightly faster at short times.

Increasing stoichiometric imbalance decreases the fitness of all structures in a uniform fashion. To investigate the interplay between fitness and stoichiometric imbalance, a real-world experimental factor that could affect assembly outcomes, we selected a subset of 50

assemblies at random from the 500 generations of the 3x3 square grid assemblies, each with different interfacial interaction energies and thus presumably different values of fitness and tested their fitness when the assembly began with a stoichiometric imbalance of components. Specifically, we varied the number of starting components (at random, see Model and Methods in Section 4.2), such that each component did not necessarily begin with a stoichiometric quantity of material (see Fig. 4.1C). We hypothesized that a stoichiometric imbalance of starting components would decrease the fitness of the assemblies, however, given that structures with heterogeneous bond strengths could possibly have an increased reliance on one more components in the pathway to form target structures, it was unclear whether random stoichiometric perturbations would equally affect all structures. In fact, we found a strong linear correlation ($r^2 > 0.99$) between the fitness of isothermally assembled structures with stoichiometric balance and the fitness of isothermally assembled structures with stoichiometric imbalance ($\sigma > 0\%$). Increasing the values of σ (extent of imbalance) decreased the slope (all < 1) of the linear trend line (see Fig. 4.4). The decreasing slope indicates that the fitness of structures with stoichiometric balance are more affected by changes in σ than structures with lower fitness values. Still, structures with high fitness values (>1) maintain their superior fitness when subject to stoichiometric imbalance, even when the imbalance is as high as $\sigma = 25\%$. This suggests that by selecting for structures with greater fitness, structures also gain some robustness to stoichiometric imbalance.

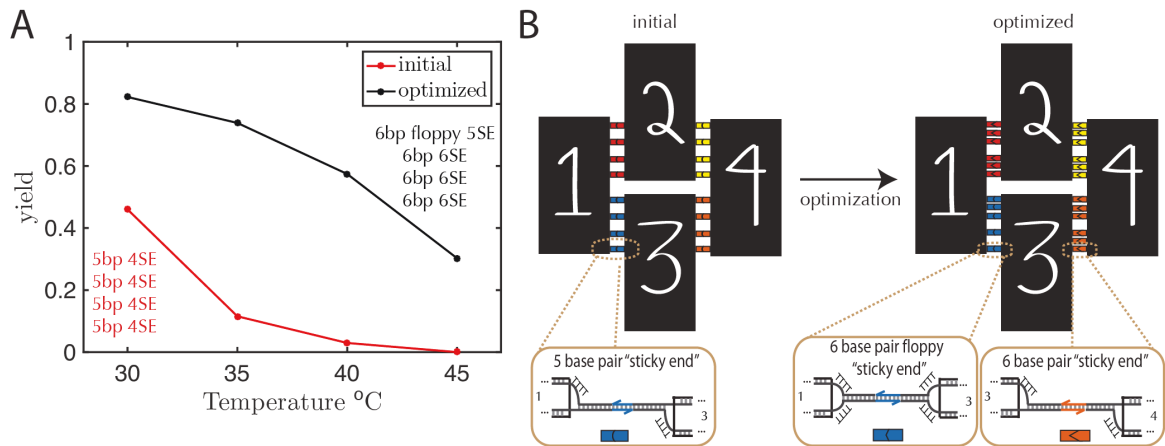


Figure 4.5. Optimizing the self-assembly system where four different DNA-origami components assemble into a tetramer complex. By using experimentally measured values of the additivity between interfaces, as reported in Ref. ¹¹⁵, and kinetic parameters such as the forward (k_{on}) and reverse (k_{off}) reaction rate constants at several different temperatures for a wide variety of interface designs, as reported in Ref. ¹⁴⁴, one can computationally determine the optimal set of interfaces for this system (see SI Note 4.1). The types of interfaces available in this design space vary in the number of linkers (*i.e.*, dsDNA connections that form) in an interface, the length of the “sticky end” (ssDNA region through which two components can specifically bind to one another) of the linker, and whether the linker has a flexible region (or is “floppy”). The naïve design approach assumes all interfaces have the 5bp 4SE design, which means that there are 4 linkers, each with 5 base pair sticky ends (SEs). (A) Yield of tetramer complex at a range of temperatures after 72 hours of isothermal assembly for the naïve design and the optimized designed after 500 generations. Inset text indicates the designs of the four interfaces. (B) Cartoon schematic of tetramer complex. Black components with white letters represent the four DNA origami components and hairpins on surface for component differentiation. Left diagram depicts naïve structure, with all interfaces having the same, 5bp 4SE design and right diagram shows optimized structure, with three interfaces having the 6bp 6SE design and one interface having the 6bp floppy 5SE design. Linkers are colored according to the interface. Zoomed-in view shows the linking architecture of the linkers in an interface and cartoon schematic, with the SE colored. 3’ end of DNA is indicated with arrows. Initial component concentration $[X]_0 = 5$ nM, the same as in the described experiments. These simulations assumed no component stoichiometric variability.

Further investigation of stoichiometric imbalance of in the formation of different initial structures suggested that systems with high stoichiometric imbalances should be assembled under conditions that slightly favor reverse reactions (*i.e.* nucleation limited). For example, increasing σ from 0% to 25% decreased the mean yield by about 10-60% at neutral and forward driven assembly conditions $\eta \geq 0$, but remained approximately constant in nucleation limited conditions for all

initial structures tested (see SI Figs. S4.7-S4.16). To understand whether the yields were being affected by a limiting single component or a general bias in the assembly pathways, we plotted the ratio of the number of target structures formed to the number of components for the component type which started the simulation with the fewest quantity and found that the ratio as a function of η changed little with increasing stoichiometric imbalance (see Figs. S4.7, S4.9, S4.11, S4.13, S4.15) and were almost identical to such curves for systems with stoichiometric balance. This suggests that the component with the lowest starting concentration indeed limits the assembly process. It is worth noting that here we randomly altered the number of starting components for each component type, however, Ref ¹⁴⁹ suggests a way to increase yields by intentionally introducing stoichiometric variability between different types of components in a target structure for a process self-assembling finite-sized structures.

Optimizing a DNA origami-based biomolecular complex using measured kinetic and thermodynamic parameters. Our final goal in this work was to take a known, reported structure with measured additivity and many potential different interface strengths and optimize it using our directed evolution of interfaces algorithm. Specifically, we take the structure reported in Ref. ¹¹⁵, which is a tetrameric complex made from four different DNA origami components designed to have four approximately equal interface strengths (*i.e.*, same number of dsDNA linkers per interface, each with the same length of “sticky end”, or single stranded region used to specifically bind a linker with the complementary sequence). Based on the reported free energies of reactions ¹¹⁵, this structure has an additivity between interfaces of $\alpha = 0.58$ (see SI Section 4.1). We select from possible interface strengths using the interface designs for such origami components reported in Ref. ¹⁴⁴, with measured forward (k_{on}) and reverse (k_{off}) reaction rate constants at different temperatures (see SI Section 4.2). The simulation parameters here are informed by the experimental

details described in Refs. ^{115 144} and reported in dimensional quantities (see Materials and Methods, SI Section 4.3 for computational details).

We found that the initial structure could be optimized from a structure with all interfaces having the same linking architecture, specifically four linkers with 5 base pair (bp) sticky ends (SEs) (or in shorthand “5bp 4SE”), to a structure where three interfaces have 6bp 6SE and one interface has a 6bp 5SE with a “floppy” portion in the linking architecture (see Fig. 4.5, and Fig. 4.5b insets for linker designs). With this optimization, simulations indicate that yield could be improved from 30-60% at the range of temperatures tested. Furthermore, for an analogous 3x3 structure, similar to the type of structure reported in Ref. ⁶⁷, yields could be improved by up to 50% (see Fig. S4.17). Interestingly, both the tetramer and 3x3 design include multiple types of interface architectures in the optimized design, suggesting that even for measured values of additivity and thermodynamic and kinetic parameters, heterogeneity in interfaces is a good design strategy.

4.4 CONCLUSION

The simulations described here, namely, directing the evolution of interfaces, suggest a simple computational method for guiding the design cycle of engineered complexes which could potentially eliminate the need of to design and experimentally test many different structural combinations. An iterative optimization approach like the one described here for finite-sized structures is feasible for almost any structure in practice. This method can be used for structures with a wide variety of starting conditions (*e.g.*, numbers and types of components, interface designs, dimensionality) and can be subject to different numbers of iterations (generations) to improve a structure in a more practical situation. For example, one could optimize isothermal assembly at a single temperature or ask whether this optimization would lead to efficient assembly *via* annealing.

We showed that iterative optimization of local contacts in a structure can enable structures to isothermally assemble faster, achieve higher yield, and do so across a wider range of

temperatures, effectively eliminating the need to guess-and-check or finely tune assembly protocols. With such isothermal optimization, structures also assemble well under an annealing protocol, and can do so just as rapidly at longer assembly times, where even many less fit structures can achieve rapid, high-yield assembly.

Designing target structures with uniform interaction strengths is a poor design strategy, unless designing a 1D structure, because the resulting landscape is vast and relatively flat: many potential intermediate structures can readily be formed at even slightly forward driven conditions. In comparison, well-designed (heterogeneous) interaction strengths enable rapid, high-yield assembly, suggesting hierarchal assembly pathways.

Our coarse-grained model makes simplifying assumptions for the self-assembly process which might result in an incomplete theory in practice. Here we do not consider other non-specific or off-target interactions which could affect the assembly outcomes in ways this model does not predict. Furthermore, modeling interface additivity using a single value might not be the best method, where real world systems often exhibit cooperative^{76, 150} or non-cooperative binding¹¹⁵ and depend on the number and type of ligands present.

4.5 SUPPORTING INFORMATION

4.5.1 Supporting Information Note 1: Directed Evolution of Interfaces Simulations

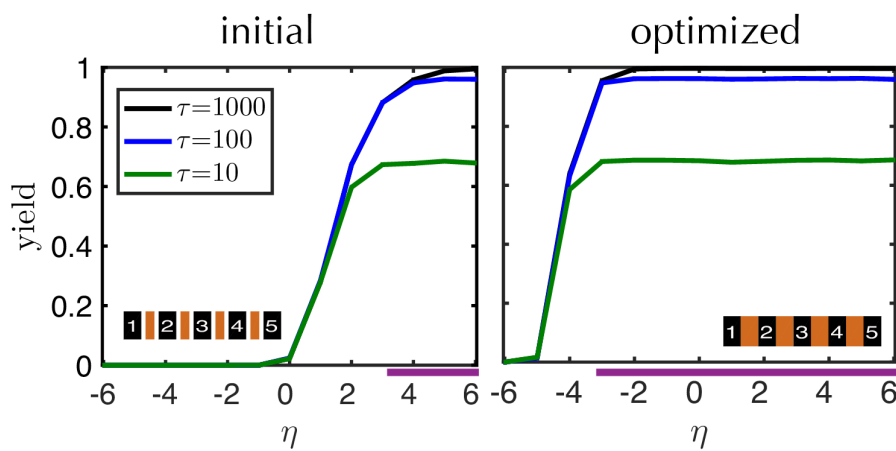


Figure S 4.1 Efficient assembly is achieved through strong interaction strengths in 1D assembly. Yield as a function of isothermal assembly condition for the initial (left) and optimized (right) 1x5 line structure. Inset diagrams represent complexes, where black squares represent individual components and width of brown blocks represent interface strengths. Components are numbered to indicate uniqueness. All interface strengths are equal initially. Purple lines underneath plots indicate assembly funnel regime, or where target structure $yield > 80\%$ after $\tau = 1000$. Here, there is no stoichiometric variability.

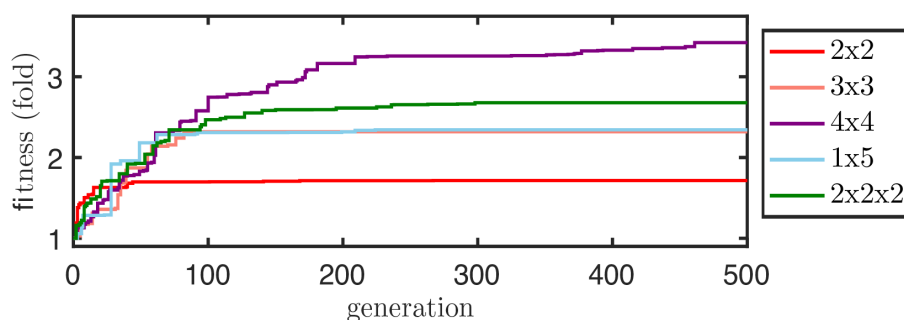


Figure S 4.2 Directed evolution of interfaces in simulation increases the fitness of all structures. The highest fitness observed at each generation, where fitness is defined as the area under the yield as a function of η curve after $\tau = 1000$ divided by (uniform interaction strengths). Only the mutations that lead to an increase in fitness are shown.

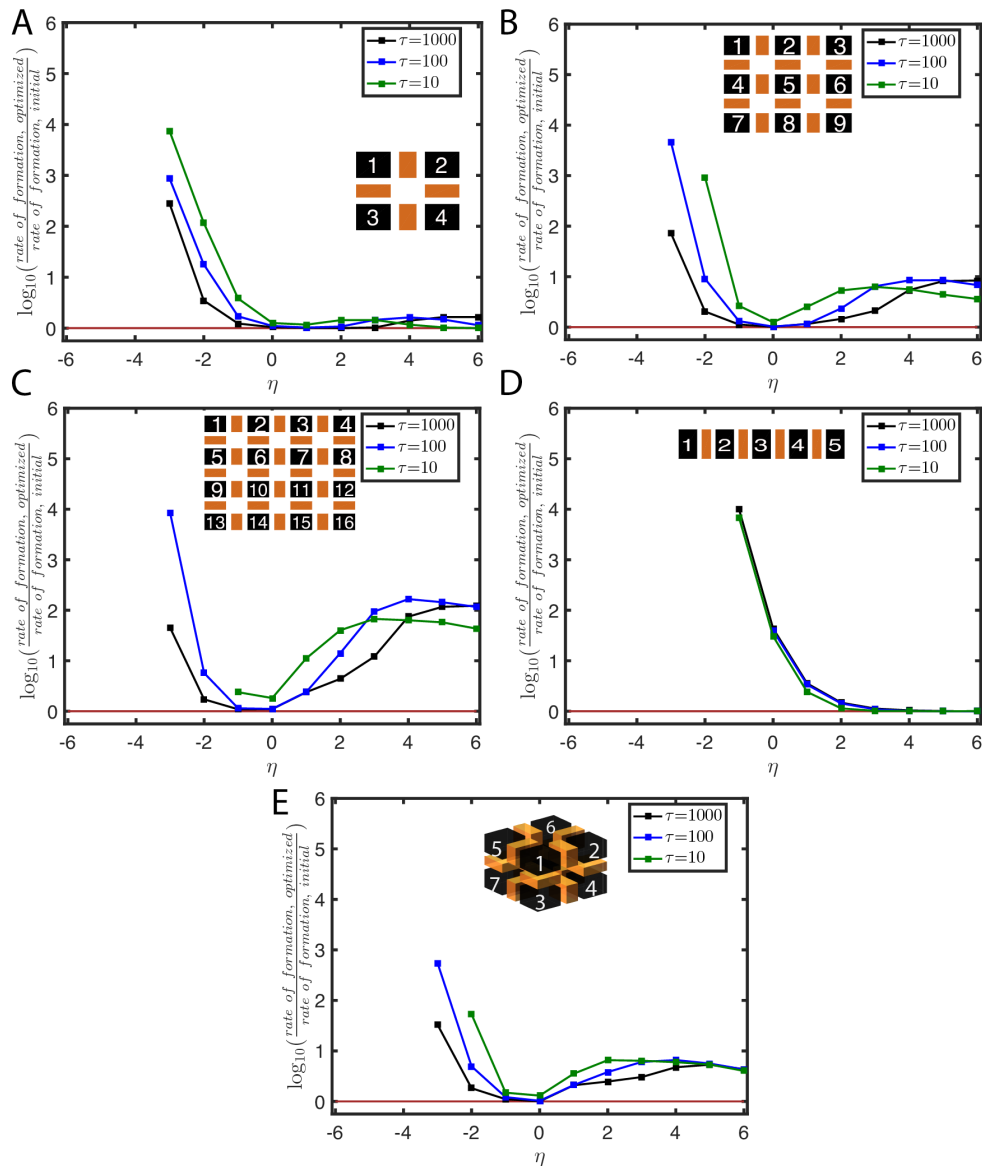


Figure S 4.3 The rate of structure formation of optimized structures is faster than or just as fast as the rate of formation of initial structures. To compare the formation rates of structures, defined as the yield of structures normalized by time, $yield/\tau$, at various isothermal conditions, we compared the formation rates of the optimized and initial structure. Specifically, we plot the \log_{10} of the ratio of the rate of formation for optimized structures (see Fig. 4.2 and S4.1) to the rate of formation for initial structures at various dimensionless temperatures for the (A) 2x2, (B) 3x3, (C) 4x4, (D) 1x5 and (E) 2x2x2 assembly. Inset diagram depicts initial structure. The initial structure formation rate is never faster than optimized structure formation rate (values always above or equal to 0). Brown line has value 0 and is shown as a guide for the eye, representing equal formation rates. Values not shown when either rate of formation for the initial structure is zero, or when both rates are zero (e.g., at $\eta \leq -4$ for all structures).

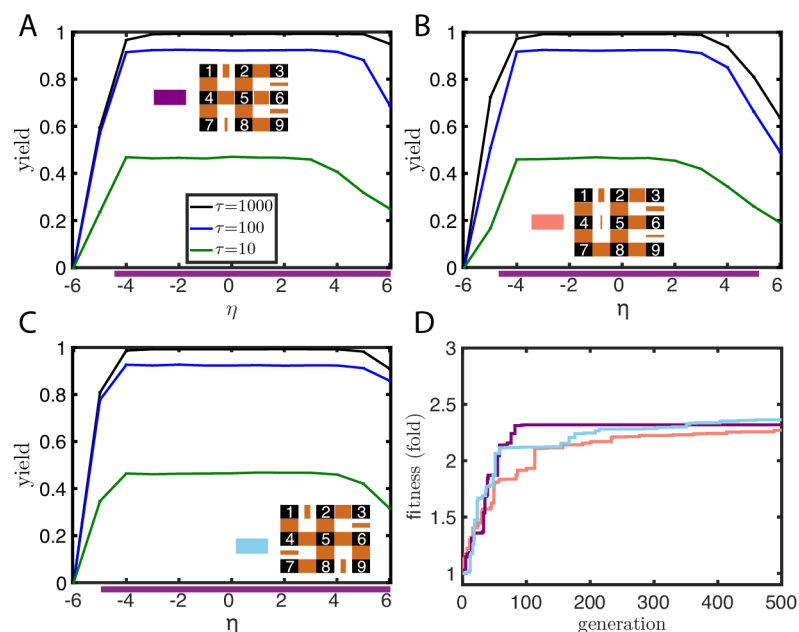


Figure S 4.4 Multiple different optimized structures can be found. Multiple directed evolution simulations from identical initial structures (*i.e.*, equal bond strengths at each interface, see Figure 4.2B, left) converge to structures with different bond strengths in the 3x3 grid assembly. Three optimized structures shown with yield as a function of assembly condition, η , in (A-C) which all produce similar increases in fitness (D). Color next to tile in (A-C) corresponds to line color in (D). Inset diagrams represent target structure, where black squares represent individual components and brown blocks represent interface strengths. The size of the brown block is proportional to the interface strength. Purple lines underneath plots in (A-C) indicate an efficient assembly regime, or where target structure yields are >80% after $\tau = 1000$. In these simulations, there is no stoichiometric variability.

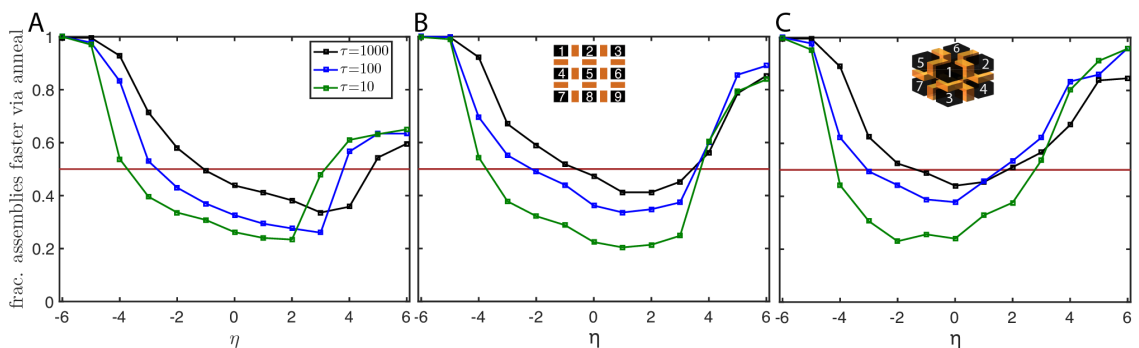


Figure S 4.5 On a population scale, structure formation rates higher, especially at short assembly times, or just as high via isothermal assembly versus via annealing at ideal isothermal conditions. To understand how, on a population scale, the rates of assembly *via* annealing compare to the rates *via* isothermal assembly, we plotted the fraction of the 500 generations tested that assembled faster when annealed versus isothermally assembled various isothermal conditions and assembly times for the (A) 2x2, (B) 3x3, and (C) 2x2x2 structures. We found that, on average, isothermal assembly was faster when subject to favorable isothermal conditions, which are typically modestly reverse to forward driven conditions ($-2 \lesssim \eta \lesssim 2$), but at thermal extremes annealing is faster. Brown line has value 0.5 and is shown as a guide for the eye, representing a fraction of assemblies that are equally fast forming *via* annealing as by isothermal assembly.

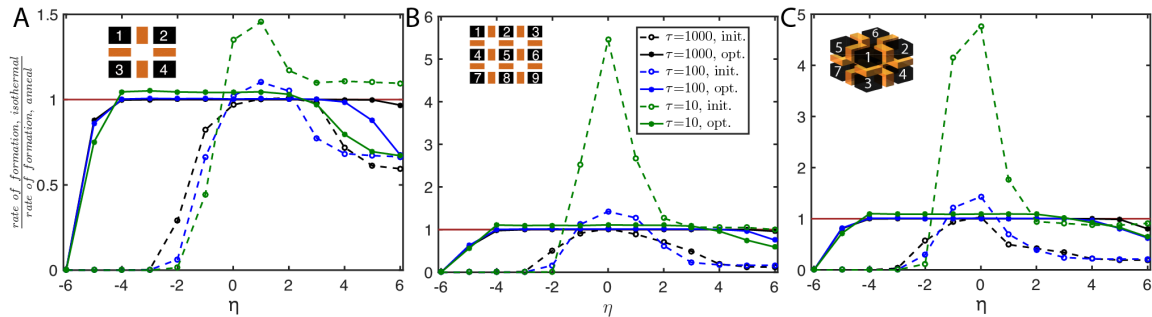


Figure S 4.6 Structure formation rates higher, especially at short assembly times, or just as high via isothermal assembly versus via annealing at ideal isothermal conditions. We compared the formation rates for the initial and optimized structures *via* isothermal assembly at various conditions versus *via* annealing by plotting the ratio the rates at various assembly times for the (A) 2x2, (B) 3x3, and (C) 2x2x2 structures. Dashed lines indicate the initial structure and (as depicted in the inset) and solid lines indicate the optimized structure. Colors correspond to assembly times. For regions where isothermal assembly is efficient (e.g., $\eta = 1$ for the 2x2 complex), target structure production is more rapid *via* isothermal assembly at short assembly times ($\tau = 10$), especially for the non-optimized structure, but at long assembly times ($\tau = 1000$) rates tend toward one another. However, in assembly conditions where isothermal assembly is not efficient, for example for $\eta \leq -4$ for the 3x3 assembly, annealing is faster. Brown line has value 1 and is shown as a guide for the eye, representing equal formation rates. Note that the scale in the y-axis is different for different plots.

4.5.2 Supporting Information Note 2: Stoichiometric Imbalances Simulations

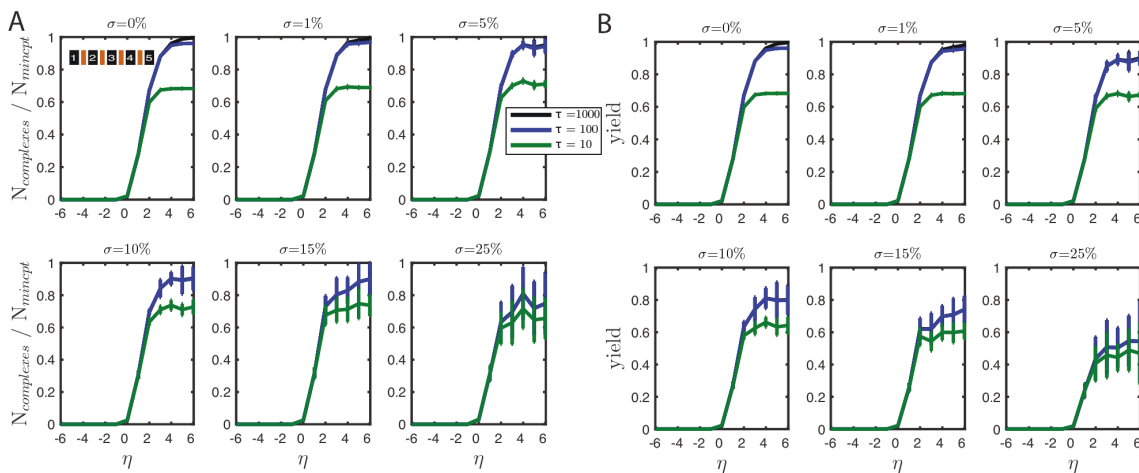


Figure S 4.7 Stoichiometric imbalances limit yield in the 1x5 line system. (A) The ratio of the number full target complexes to the number starting components for the species with fewest material (*i.e.*, the species that limits yield), $N_{\text{complexes}}/N_{\text{mincpt}}$ versus isothermal assembly conditions, η , at various values of σ . (B) Yield versus assembly conditions, η , at various values of σ . Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

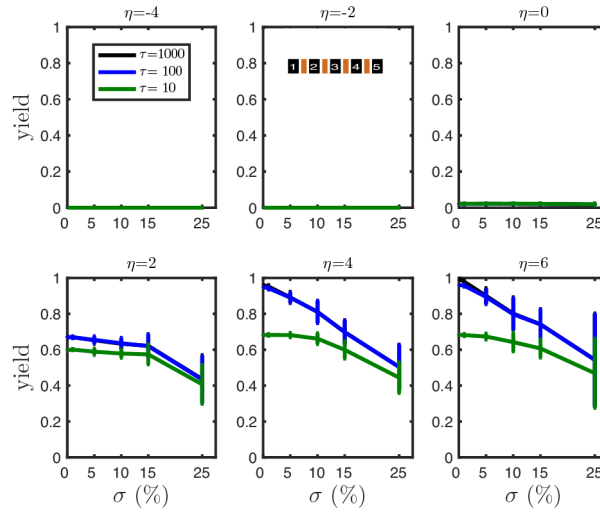


Figure S 4.8 Stoichiometric imbalances limit yield for a fixed assembly time in the 1x5 line system. Yield versus stoichiometric variability, σ , at various isothermal assembly conditions, η . Error bars indicate 10 simulations at a given η , and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

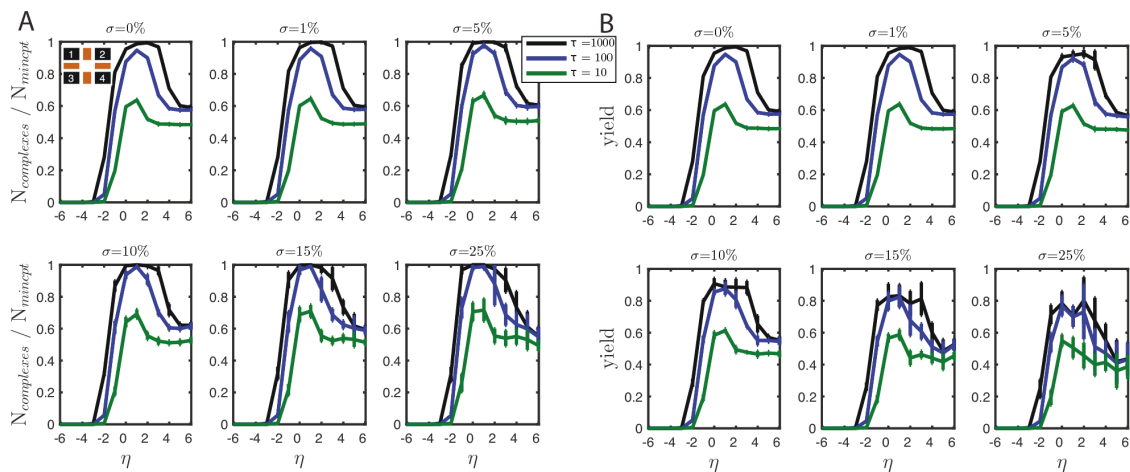


Figure S 4.9 Stoichiometric imbalances limit yield in the 2x2 grid system. (A) The ratio of the number full target complexes to the number starting components for the species with fewest material (*i.e.*, the species that limits yield), $N_{complexes}/N_{mincpt}$ versus isothermal assembly conditions, η , at various values of σ . (B) Yield versus assembly conditions, η , at various values of σ . Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

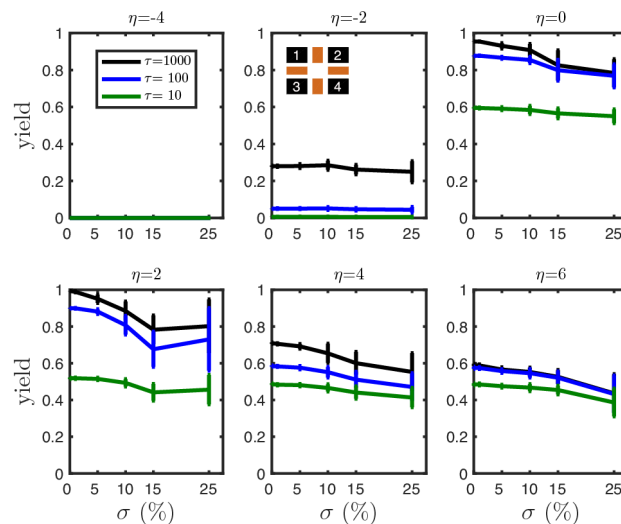


Figure S 4.10 Stoichiometric imbalances limit yield for a fixed assembly time in the 2x2 grid system. Yield versus stoichiometric variability, σ , at various isothermal assembly conditions, η . Error bars indicate 10 simulations at a given η , and τ value. Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

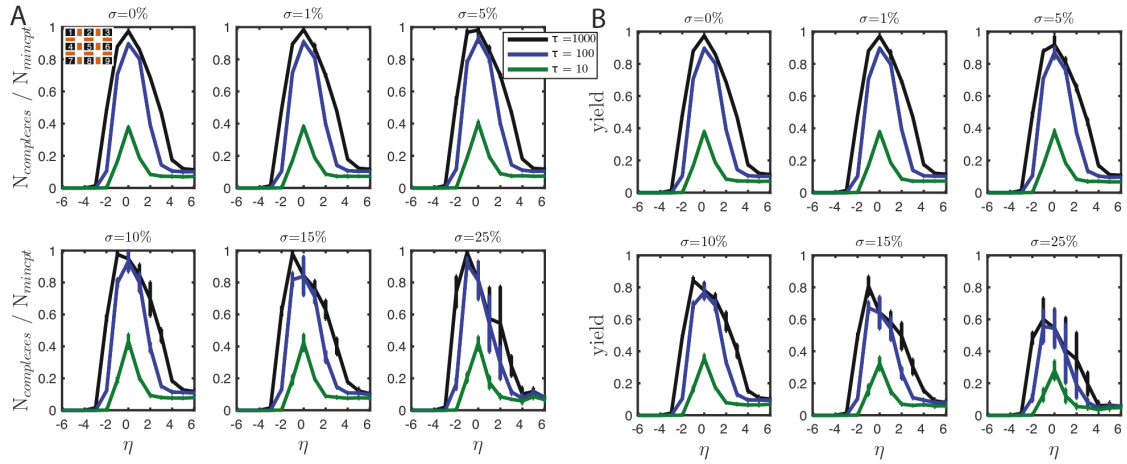


Figure S 4.11 Stoichiometric imbalances limit yield in the 3x3 grid system. (A) The ratio of the number full target complexes to the number starting components for the species with fewest material (*i.e.*, the species that limits yield), $N_{\text{complexes}}/N_{\text{mincpt}}$ versus isothermal assembly conditions, η , at various values of σ . (B) Yield versus assembly conditions, η , at various values of σ . Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent intercomplexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

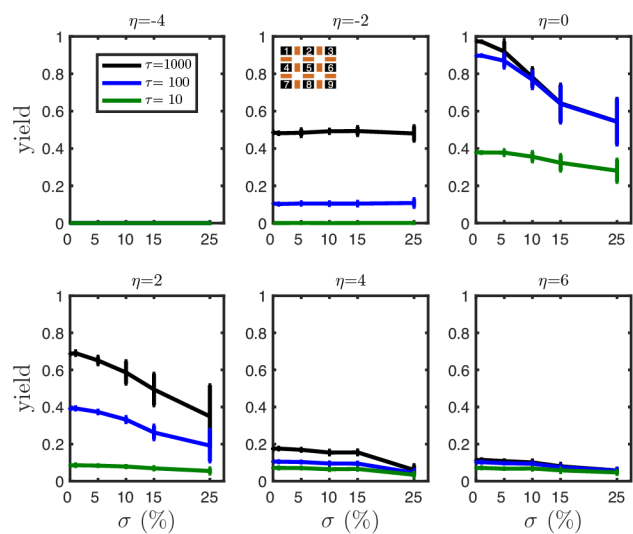


Figure S 4.12 Stoichiometric imbalances limit yield for a fixed assembly time in the 3x3 grid system. Yield versus stoichiometric variability, σ , at various isothermal assembly conditions, η . Error bars indicate 10 simulations at a given η , and τ value. Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

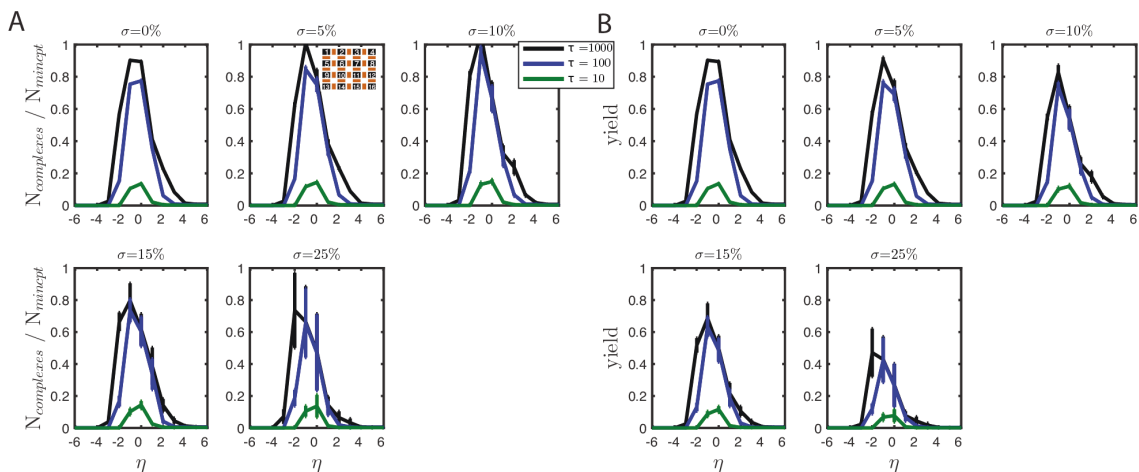


Figure S 4.13 Stoichiometric imbalances limit yield in the 4x4 grid system. (A) The ratio of the number full target complexes to the number starting components for the species with fewest material (*i.e.*, the species that limits yield), $N_{\text{complexes}}/N_{\text{mincpt}}$ versus isothermal assembly conditions, η , at various values of σ . (B) Yield versus assembly conditions, η , at various values of σ . Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$). $\sigma = 1\%$ omitted.

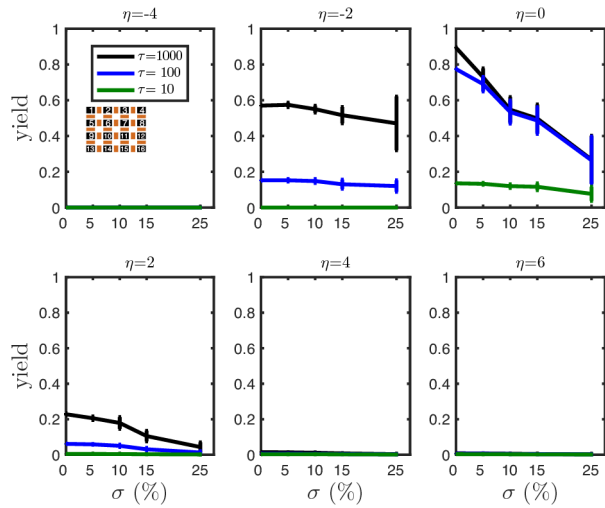


Figure S 4.14 Stoichiometric imbalances limit yield for a fixed assembly time in the 4x4 grid system. Yield versus stoichiometric variability, σ , at various isothermal assembly conditions, η . Error bars indicate 10 simulations at a given η , and τ value. Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

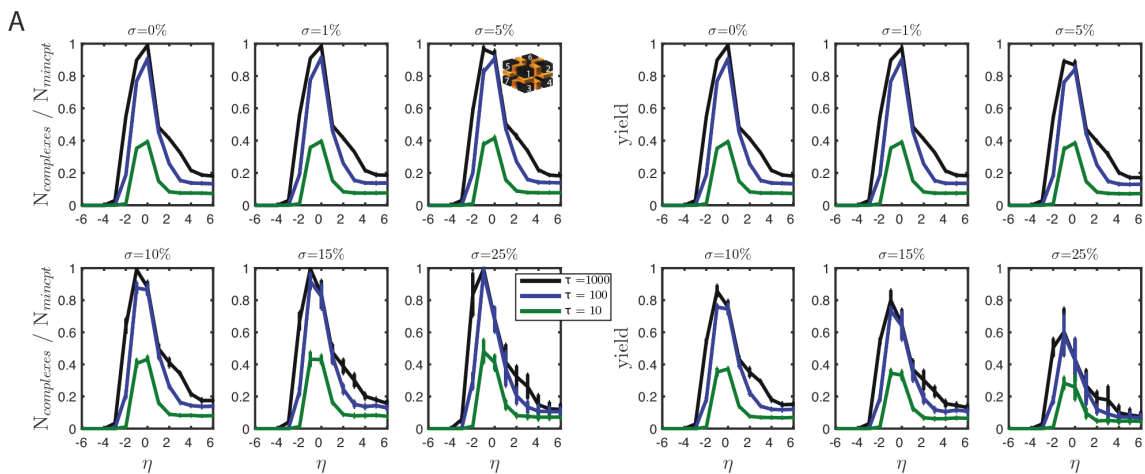


Figure S 4.15 Stoichiometric imbalances limit yield in the 2x2x2 cube system. (A) The ratio of the number full target complexes to the number starting components for the species with fewest material (*i.e.*, the species that limits yield), $N_{complexes}/N_{mincpt}$ versus isothermal assembly conditions, η , at various values of σ . (B) Yield versus assembly conditions, η , at various values of σ . Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

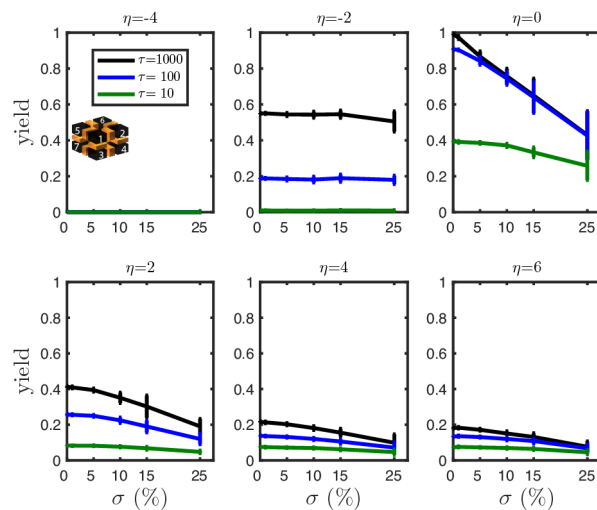


Figure S 4.16 Stoichiometric imbalances limit yield for a fixed assembly time in the 2x2x2 cube system. Yield versus stoichiometric variability, σ , at various isothermal assembly conditions, η . Error bars indicate 10 simulations at a given η , and τ value. Error bars indicate 10 simulations at a given η , σ and τ value. Inset diagrams represent complexes, where black squares represent individual components and brown blocks represent interface strengths. Components are numbered to indicate uniqueness. Here, all bond strengths are equal ($b = 1$).

4.5.3 Supporting Information Note 3: Simulation Parameters from Experimental Measurements of the DNA origami-based complex

Calculating the additivity between components

Thermodynamic measurements of reactions in the self-assembly of finite-sized structures form the basis of these calculations have recently been reported¹¹⁵. In this work, the authors self-assemble a tetramer complex from four different components, made using the DNA origami technique¹⁰, and measured the free energy of the reactions. In this system, the interface between the components provides the orientation and binding specificity of the self-assembling components. The interface between two components consisted of four linkers, each with 5 base pair (bp) “sticky ends” (SEs), where the SE sequence determines the specificity of the reaction, where all SE sequences were unique. Minimal crosstalk (*i.e.*, unintended reactions) between interfaces was observed. We use these measurements to establish an estimate for the additivity of interfaces (see Table S4.1 for measurements).

From these measurements, we can obtain approximate values of additivity, α , where α is the multiplier of free energy for a reaction occurring through two interfaces as opposed to one. For example, in the case where a reaction that occurs through two interfaces is equal in free energy to the sum of the energies of the two reactions that occur through a single interface, the additivity is one. Additivity is restricted to $\alpha \geq 0$ and is defined mathematically:

$$\alpha = \frac{\Delta G_{A-B}}{(\Delta G_A + \Delta G_B)}, \quad (4.4)$$

where ΔG_A and ΔG_B are the free energy of reactions through interface A and interface B alone, while ΔG_{A-B} is the reaction through both interfaces simultaneously.

Table S 4.1 Thermodynamic parameters of Interfaces from Ref. ¹¹⁵ taken at 25°C

Reaction Type	Reaction	G^0 (kcal/mol)
Dimerization	$T_1 + T_2 \rightarrow C_{12}^*$	-11.5
Dimerization	$T_1 + T_3 \rightarrow C_{13}$	-10.95
Dimerization	$T_2 + T_4 \rightarrow C_{24}$	-10.69
Dimerization	$T_3 + T_4 \rightarrow C_{34}$	-11.7
Tetramerization	$C_{12} + C_{34} \rightarrow C_{1234}$	-12.5
Tetramerization	$T_1 + C_{234} \rightarrow C_{1234}$	-13.1

* T_1 denotes Tile 1 and C_{12} denotes the complex consisting of Tile 1 and Tile 2

Here we can calculate α by taking the energy of the reaction through two interfaces in the reaction where T_1 binds to C_{23} and the energies of the respective dimer reactions through a single interface, $T_1 + T_2 \rightarrow C_{12}$ and $T_1 + T_3 \rightarrow C_{13}$. This becomes $\alpha = \frac{\Delta G_{1-3}}{(\Delta G_1 + \Delta G_3)} = \frac{-13.1}{(-11.5 - 10.95)}$ and thus

$\alpha = 0.58$. We can perform this analysis again using different sets of reactions to obtain a second measurement of α for this system. We can do this for the energy of adding two dimers to make a tetramer versus the dimer reactions alone and we find that $\alpha = \frac{-12.5}{(-10.95 - 10.69)}$ which is $\alpha = 0.58$.

Both calculations produce a consistent additivity value of $\alpha = 0.58$, which is the value we use in the simulations.

Here we assume this additivity value would be constant and independent of the linker architecture, number of linkers per interface and sticky end length. We expect this is a fair assumption given that we expect the major factor in determining additivity is the rigidity of components and alignment of binding sites between components, which is largely independent of interface design.

Forward and reverse reaction rate constants for various interface designs

Measurements of dimerization kinetics of two different DNA origami tiles (a subset of two components in Ref ¹¹⁵) have recently been studied¹⁴⁴. Here, the authors use a variety of interface designs and determine the kinetic and thermodynamic properties of the dimerization reaction. Different interface designs included changing the number of linkers per interface, the number of base pairs in a SE and the architecture of the linker. In our simulations, we use the forward and reverse reaction rate constants at 30°C, 35°C, 40°C, and 45°C for the various interface designs reported in Ref. ¹⁴⁴ (see values in Tables S4.2-4.3).

Table S 4.2 Forward Reaction Rate Constants, k_{on} ($M^{-1}s^{-1}$), taken from *Chapter 3* (Ref. ¹⁴⁴)

	30°C	35°C	40°C	45°C
<i>5bp 3SE A</i>	9.44E+05	8.60E+05	9.89E+05	1.78E+05
<i>5bp 3SE B</i>	7.49E+05	7.49E+05	2.35E+05	36969
<i>5bp 3SE C</i>	6.83E+05	7.49E+05	5.42E+05	1.12E+05
<i>5bp 4SE</i>	1.30E+06	1.19E+06	7.84E+05	77526
<i>5bp 5SE A</i>	2.17E+06	1.98E+06	1.14E+06	1.63E+05
<i>5bp 5SE B</i>	2.87E+06	2.87E+06	1.98E+06	7.15E+05
<i>5bp 6SE</i>	3.00E+06	2.87E+06	1.89E+06	6.22E+05
<i>5bp 7SE</i>	2.61E+06	2.61E+06	2.17E+06	9.01E+05
<i>6bp 3SE A</i>	2.49E+06	2.27E+06	1.80E+06	1.14E+06
<i>6bp 3SE B</i>	2.27E+06	1.98E+06	1.43E+06	7.49E+05
<i>6bp 3SE C</i>	1.89E+06	1.50E+06	1.04E+06	4.94E+05
<i>6bp 4SE</i>	3.00E+06	2.61E+06	2.17E+06	1.80E+06
<i>6bp 5SE A</i>	4.15E+06	3.61E+06	3.29E+06	2.49E+06
<i>6bp 5SE B</i>	4.55E+06	3.96E+06	3.61E+06	2.74E+06
<i>6bp 6SE</i>	4.99E+06	4.35E+06	3.96E+06	3.00E+06
<i>6bp 7SE</i>	4.55E+06	3.96E+06	3.61E+06	2.87E+06
<i>6bp floppy 3SE A</i>	1.25E+06	1.04E+06	5.94E+05	1.70E+05
<i>6bp floppy 3SE B</i>	1.19E+06	9.88E+05	4.30E+05	1.76E+04
<i>6bp floppy 3SE C</i>	9.88E+05	7.84E+05	4.50E+05	1.23E+05
<i>6bp floppy 4SE</i>	1.43E+06	1.25E+06	1.08E+06	3.92E+05
<i>6bp floppy 5SE A</i>	1.72E+06	1.64E+06	1.19E+06	6.52E+05
<i>6bp floppy 5SE B</i>	1.19E+06	1.19E+06	1.08E+06	7.84E+05
<i>6bp floppy 6SE</i>	2.49E+06	2.27E+06	1.98E+06	1.50E+06
<i>6bp floppy 7SE</i>	2.27E+06	2.17E+06	1.98E+06	1.50E+06

*Nomenclature of interfaces can be found in Ref. ¹⁴⁴

Table S 4.3 Reverse Reaction Rate Constants, k_{off} (s^{-1}), taken from *Chapter 3* (Ref. ¹⁴⁴)

	30°C	35°C	40°C	45°C
<i>5bp 3SE A</i>	4.30E-03	2.17E-02	3.87E-02	7.32E-02
<i>5bp 3SE B</i>	8.12E-03	2.44E-02	3.26E-02	6.91E-02
<i>5bp 3SE C</i>	5.11E-03	1.93E-02	2.30E-02	3.65E-02
<i>5bp 4SE</i>	3.22E-03	1.45E-02	2.17E-02	4.10E-02
<i>5bp 5SE A</i>	1.61E-03	1.02E-02	1.72E-02	3.45E-02
<i>5bp 5SE B</i>	4.77E-04	6.83E-03	1.22E-02	2.44E-02
<i>5bp 6SE</i>	1.14E-03	7.23E-03	1.15E-02	2.44E-02
<i>5bp 7SE</i>	3.57E-04	6.83E-03	9.66E-03	1.93E-02
<i>6bp 3SE A</i>	1.80E-03	4.55E-03	6.83E-03	1.63E-02
<i>6bp 3SE B</i>	1.70E-03	5.11E-03	1.08E-02	2.05E-02
<i>6bp 3SE C</i>	2.41E-03	5.74E-03	1.22E-02	2.17E-02
<i>6bp 4SE</i>	1.20E-03	2.55E-03	7.66E-03	1.37E-02
<i>6bp 5SE A</i>	1.28E-03	2.27E-03	7.23E-03	1.37E-02
<i>6bp 5SE B</i>	1.14E-03	2.55E-03	7.23E-03	1.53E-02
<i>6bp 6SE</i>	8.50E-04	2.15E-03	6.44E-03	1.63E-02
<i>6bp 7SE</i>	9.55E-04	2.41E-03	6.83E-03	1.83E-02
<i>6bp floppy 3SE A</i>	1.02E-02	8.12E-03	1.45E-02	8.60E-03
<i>6bp floppy 3SE B</i>	1.70E-03	3.22E-03	5.42E-03	5.74E-03
<i>6bp floppy 3SE C</i>	2.70E-03	7.66E-03	1.22E-02	2.74E-02
<i>6bp floppy 4SE</i>	3.41E-03	2.70E-03	4.06E-03	7.23E-03
<i>6bp floppy 5SE A</i>	2.27E-03	2.15E-03	3.04E-03	5.42E-03
<i>6bp floppy 5SE B</i>	2.87E-03	6.83E-03	7.66E-03	1.08E-02
<i>6bp floppy 6SE</i>	1.61E-03	3.61E-03	5.42E-03	1.02E-02
<i>6bp floppy 7SE</i>	1.70E-03	6.44E-03	6.83E-03	1.29E-02

*Nomenclature of interfaces can be found in Ref. ¹⁴⁴

Directed Evolution Simulations with Measured Parameters

To optimize the self-assembly of a complex using measured parameters of k_{on} , k_{off} and α , we perform the same general procedure in the directed evolution of interfaces simulations, with two notable differences. First, we mutate an interface by choosing, at random, one of the 24 possible interfaces listed in Tables S4.2-4.3, which defines the k_{on} and k_{off} of the reaction. Second, we use dimensional values of temperature, time and concentration which correspond to values used in experiments in Refs. ^{115, 144}. Temperature varies from 30 to 45 °C in increments of 5 °C, with component concentrations of 5 nM and isothermal assembly times of 72 hours.

When an interface is mutated, the forward and reverse reaction rate constants that are affected by said mutation must be updated. All reactions that occur through the single interface that is mutated will be updated with k_{on} and k_{off} values corresponding to their new interface design (at a given temperature) according to Tables S4.2-4.3. For reactions that occur through multiple interfaces, the updates to the forward and reverse reaction rate constants is less straightforward.

When a reaction that occurs through two or more interfaces of a single type takes place, we assume that the k_{on} of the reaction is equal to the k_{on} for the reaction through a single interface of that type because it is unclear how adding multiple interfaces would change the forward reaction kinetics. Likely, as seen in Ref. ³, the forward reaction rate constant would increase as the number of contacts increases, however, at some number of contacts the k_{on} saturates, thus we expect that the reaction rate would either increase or stay the same. As a conservative estimate, we choose the reaction rate to stay the same. When a reaction that occurs through two or more interfaces of a different types takes place, we assume that the forward reaction occurs with a k_{on} equal to the reaction through the constituent single interfaces with a higher k_{on} value, because we expect that the reaction rate will not decrease down with an increasing number of contacts ¹⁴⁴, but how fast the speed up will be is not known.

The reverse reaction rate constant through two or more interfaces, depends on the number and type of interfaces through which a reaction occurs. Using the formula that relates the free energy to the equilibrium constant, we can calculate the reverse reaction rate constant as a function of the number of interfaces in a reaction. The free energy as a function of the temperature and equilibrium constant for a reaction through a single interface is $\Delta G = -RT \ln\left(\frac{k_{on}}{k_{off,1}}\right)$ where a reaction through multiple interfaces is $b\alpha\Delta G = -RT \ln\left(\frac{k_{on}}{k_{off,b}}\right)$, where α is additivity from Eqn. 1, where k_{on} is the forward reaction rate constant for the given reaction and $k_{off,1}$ and $k_{off,b}$ are the reverse reaction rate constants for the reaction occurring through a single interface and b identical interfaces. Using these equations to solve for the reverse rate constant we find:

$$k_{off,b} = k_{on} \left(\frac{k_{on}}{k_{off,1}} \right)^{-\alpha b} \quad (4.5)$$

In the case that a reaction occurs through more than one type of interface, $k_{off,1}$ is calculated as an average of the reverse rate constants for the reactions occurring through one interface for each type of interface.

All code is available upon request.

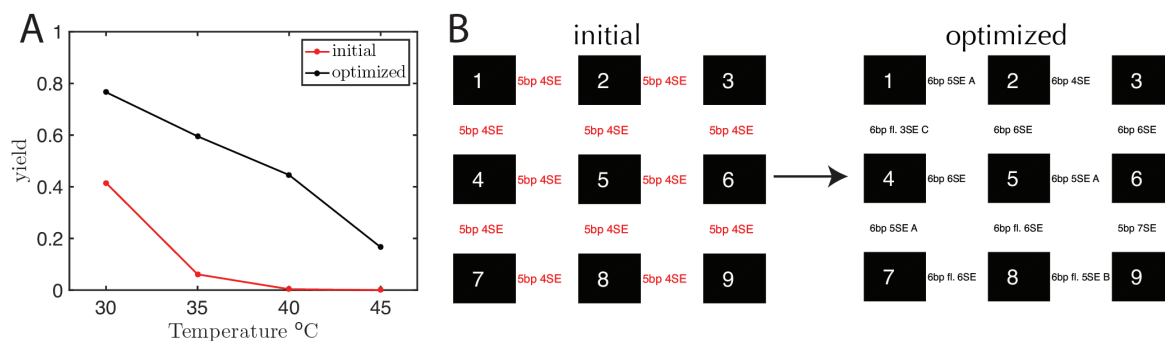


Figure S 4.17 Optimizing the self-assembly system where nine different DNA-origami components assemble into a complex. By using experimentally measured values of the additivity between interfaces, as reported in Ref. ¹¹⁵, and kinetic parameters such as the forward (k_{on}) and reverse (k_{off}) reaction rate constants at several different temperatures for a wide variety of interface designs, as reported in Ref. ¹⁴⁴, one can computationally determine the optimal set of interfaces for this system (see SI Note 4.1). The types of interfaces available in this design space vary in the number of linkers (*i.e.*, dsDNA connections that form) in an interface, the length of the “sticky end” (ssDNA region through which two components can specifically bind to one another) of the linker, and whether the linker has a flexible region (or is “floppy”). The initial design approach assumes all interfaces have the 5bp 4SE design, which means that there are 4 linkers, each with 5 base pair sticky ends (SEs). (A) Yield of complex at a range of temperatures after 72 hours of isothermal assembly for the initial design and the optimized designed after 500 generations. Inset text indicates the designs of the four interfaces. (B) Schematic of nine-mer complex. Left diagram depicts initial structure, with all interfaces having the same, 5bp 4SE design, (interface design listed in text between components whose color corresponds to color in (A)) and right diagram shows optimized structure. In the diagram, “fl.” is an abbreviation for “floppy.” Initial component concentration $[X]_0 = 5$ nM, the same as in the described experiments. These simulations assumed no component stoichiometric variability.

4.6 ACKNOWLEDGEMENTS

The authors would like to thank Dr. Deepak Agrawal, Angelo Cangialosi, Joshua Fern, Sam Schaffter and Dominic Scalise, for their helpful discussions in this work as well as Dr. Jeff Gray and Dr. Marc Ostermeier for their valuable input and feedback.

CHAPTER 5 STABLE DNA-BASED REACTION-DIFFUSION PATTERNS

SUMMARY

We demonstrate reaction-diffusion (RD) systems that generate stable patterns of DNA oligonucleotide concentrations within agarose gels, including linear and “hill” (i.e. increasing then decreasing) shapes in one and two dimensions. Our reaction networks are driven by enzyme-free DNA strand-displacement reactions, in which reactant DNA complexes continuously release and recapture target strands of DNA in gel; a balance of these reactions produces stable patterns. The reactant complexes are maintained at high concentrations by liquid reservoirs along the gel boundary. We monitor our patterns using time-lapse fluorescence microscopy, and show that the shape of our patterns can be easily tuned by manipulating the boundary reservoirs. Finally, we show that two overlapping, stable gradients can be generated by designing two sets of non-interacting release and recapture reactions with DNA strand-displacement systems. This paper represents a step toward the generation of scalable, complex RD patterns for programming the spatiotemporal behavior of synthetic materials.

5.1 INTRODUCTION

Gradients are ubiquitous as drivers of spatially differentiated behavior and communication in biological systems. For example, chemical stripes of different mRNA molecules generated by reaction-diffusion (RD) processes in the embryo of the fruit fly *Drosophila* act as chemical blueprints to direct the growth the embryo^{151, 152}, and Turing patterns are believed to be important in the patterning of animal digits. A variety of spatial concentration patterns also arise during ubiquitous intercellular signaling processes^{153, 154}. *In vitro*, spatial patterns of the concentrations of chemical and biological species have been generated to study and control biological systems

including chemotaxis¹⁵⁵, angiogenesis¹⁵⁶, stem cell proliferation and differentiation¹⁵⁷, axon growth¹⁵⁸, cell culture¹⁵⁹, the behavior of cells in hydrogels¹⁶⁰ and protein expression¹⁶¹ to name a few. Chemical gradients have also been used to control reactivity, direct mechanical actuation and pattern synthetic materials and are increasingly of interest for signal processing within materials, orchestrating spatially differentiated chemical or mechanical behaviors or in self-repair^{162, 163, 164}.

Chemical gradients are often produced using patterning methods that encode variations of density of a molecule along a surface or within a 3-dimensional material^{165, 166, 167}. Variations of chemical concentrations can also be generated using flows and/or diffusion across membranes in microfluidic devices^{168, 169, 170, 171}.

While lithographic or light-driven processes can generate patterns of molecules of high complexity^{166, 167}, the resulting patterns cannot evolve or regenerate over time as materials are consumed or molecules mix due to diffusion. Dynamic, or “bottom-up” methods for building spatial patterns are often limited in the complexity of what can arise.

Reaction-diffusion (RD) systems can produce a wealth of spatiotemporal chemical concentration dynamics such as waves and oscillations¹⁷² using chemical systems such as the well-studied inorganic Belousov-Zhabotinsky (BZ) reaction^{172, 173}. Several enzymatic RD networks have also been developed recently that generate patterns from the bottom-up^{174, 175, 176, 177, 178, 179}, both with transcriptional circuits¹⁸⁰ and with the Polymerase, Exonuclease, Nickase (PEN) toolbox^{181, 182}. Enzyme-based systems are typically limited to tight temperature ranges and buffer conditions and are more challenging to scale up the number of components that can be combined in the same solution¹⁸³.

An alternative bottom-up approach to implement chemical reaction networks in a RD system is to use enzyme-free DNA strand-displacement reactions¹⁸⁴, which have been shown to be capable of programming large reaction networks⁴². Further advantages for implementing DNA-

based strand displacement include the fact that reaction rate constants can be tuned^{42, 44, 45, 184, 185, 186, 187, 188} and reactions have potential to operate at a wide range of temperatures (*e.g.* at room temperature and in cell media¹⁸⁹). While several DNA reaction-diffusion systems have been built^{190,191}, none to date have shown spatiotemporally stable gradients that can persist over time and potentially reform after being perturbed. Such chemical patterning systems could make it possible to design materials with heterogeneous structure that could regenerate, and such systems could enable the modular design of more complex programmable patterns such as a stick figure through the design of chemical reaction networks.

Here we demonstrate that stable spatial patterns of soluble DNA molecules can be generated using the coupled process of diffusion and enzyme-free DNA strand displacement reactions. These systems overcome the homogenizing effect of diffusion through the use of a sustained energy source: a steady supply of reactant species. These systems produce millimeter-scale heterogeneous patterns in one and two dimensions in a hydrogel substrate that can either continuously grow or remain stable over time. DNA components with high concentrations in liquid reservoirs diffuse into the gel and participate in specific strand-displacement reactions designed to continuously release and recapture target strands of DNA. In principle, such patterns remain stable as long as the high fuel reservoirs are maintained. To underscore this point, we show patterns remain stable for a period of over 30 hours. We also show that this process of gradient formation can be easily extended to create more complex patterns involving multiple species with different sequences.

5.2 RESULTS AND DISCUSSION

System design and mechanism. Our goal was to develop a reaction-diffusion (RD) system that would lead to the formation of a spatiotemporally stable (*i.e.* unchanging) gradient using DNA-

based strand displacement systems. RD processes can be described by a set of partial differential equations of the form:

$$\frac{\partial C_j}{\partial t} = R_j + \nabla^2 C_j, \quad (5.1)$$

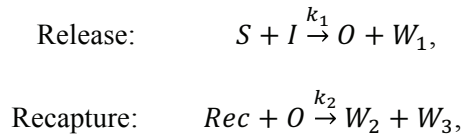
where C_j , D_j and R_j are the concentration, diffusion coefficient and the total rate of the reactions involving species j , respectively, thus there is one such equation for each of the species in the system. We used this model to develop a simple set of reactions that would produce a gradient with a well-defined stable state.

Our model for the generation of stable gradients was a coupled set of reactions that would (1) release the species forming the gradient and (2) recapture that species. Given such a set of reactions of the form:

$$d[O]/dt = k_{rel} - k_{cap}[O], \quad (5.2)$$

where k_{rel} is the rate of release of the output species and k_{cap} is the rate of recapture, $[O]$ is stable at the concentration level k_{rel} / k_{cap} . Such a mechanism could be used to create spatial patterns if the rate of release or recapture varied as a function of space.

To build this system we devised a set of reactions for release and recapture:



where S , I and Rec are input species and W_1 , W_2 and W_3 are waste products. We call S , I , Rec , and O the Source, Initiator, Recapturer and Output, respectively. When $[S]$, $[I]$ and $[Rec]$ are unchanging, the change in $[O]$ over time follows Equation 2 where $k_{rel} = k_1[S][I]$ and $k_{rec} = k_2[Rec]$ such that $[O]$ reaches a steady state concentration that is the ratio of these two terms.

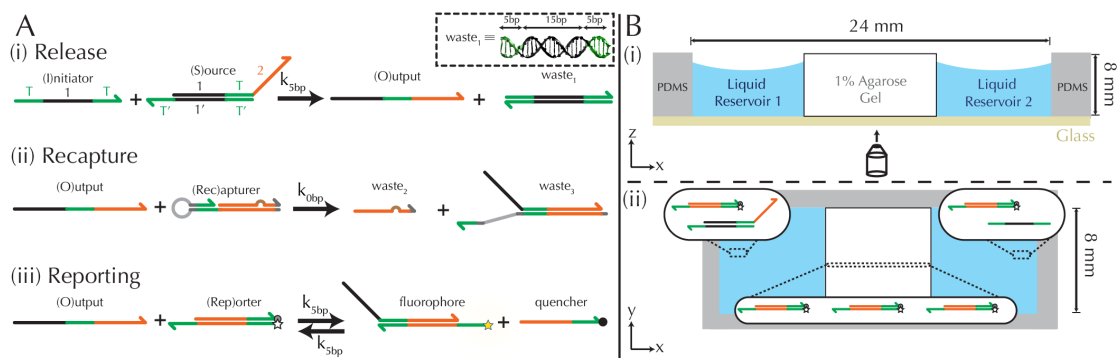


Figure 5.1. Schematic of DNA strand displacement reactions and the setup of reaction diffusion systems. (A) Three strand displacement reactions are used in different combinations and regimes to create and monitor different types of gradient patterns. (i) Release: The Output strand is initially sequestered within a Source complex but can be released by reacting with an Initiator strand *via* a relatively fast 5 base-pair (bp) toehold mediated reaction. (ii) Recapture: The Output strand is recaptured by a Recapturer complex in a relatively slow reaction. (iii) Reporting: The concentration of Output is “read” by reversibly reacting with a reporter complex whose strands have an attached fluorophore-and quencher. Inset diagram is a diagram of the structure of a waste₁ complex. Green regions indicate toeholds (5bp) and the black region indicates the “1” domain (15bp). Complementary sequences are denoted by an apostrophe (*e.g.*, sequence 1 is fully complementary to 1’) and share the same color. 3’ ends are labeled with an arrow. Brown bump in Recapturer complex indicates a single base pair mismatch. (B) (i) Side-view schematic of the RD cell (*x-z* plane), whose exterior is formed from PDMS cast around a negative mold. The RD system inside consists of a 1% agarose hydrogel between two liquid reservoirs, each containing a solution of DNA species. An optical microscope is used to image the cell through a glass coverslip bound to the PDMS (Methods in Section 5.4). (ii) Top-view schematic of an RD cell (*x-y* plane). Inset diagrams depict initial conditions in a typical “hill” gradient formation experiment: Liquid Reservoir 1 contains of Source and Reporter species, Liquid Reservoir 2 contains Initiator and Reporter species, and the hydrogel contains Reporter species.

Opposing release and recapture reactions could also be used to build stable concentration gradients in an environment where the concentrations of the reactions varies in space. By controlling this variation as well as the relative concentrations of the reactants and the rates of the reactions, a variety of gradients that change in time or which remain in a far-from-equilibrium stable configuration over time may be created. For example, if *S* and *I* are allowed to react only at a single location to release *O*, and *Rec* is present throughout the substrate, a stable gradient of *O* around the location will form that will remain stable as long as *S*, *I* and *Rec* are supplied to the system as fuel. And because the release and recapture process can emulate the dynamics of a proportional controller, the gradients that form can be restored if perturbed, so long as the input

species continue to be supplied. Further, release-recapture systems of this kind also serve as building blocks for a system in which complex networks of chemical reactions and controlled diffusion can form stable or dynamic gradient patterns of virtually arbitrary complexity.

To develop this gradient forming system, we consider how to (1) generate patterns using strand displacement reactions between DNA oligonucleotide strands and strand complexes and control these patterns through design of reaction rates and reactant concentrations and (2) construct a system in which these patterns can be stable over long periods of time so long as external reservoirs that contain supplies of chemical fuel are occasionally replenished.

To do so, we consider a system in which S and I diffuse into a 1-dimensional reaction chamber from two reservoirs on either end that contain high concentrations of S and I respectively. In one case, S is maintained at a high concentration in one reservoir, and I is maintained at a high concentration in the opposing reservoir. This setup controls when S and I meet within the substrate and can react to release O . This release process first happens as the diffusion gradients of S and I expand and meet, which produces a gradient of O that emerges and spreads in a controlled fashion. If Rec were present throughout the substrate and within the reservoirs, its reaction with the substrate would recapture it, limiting the gradient's spread and causing it to stabilize over time.

Design of a reaction-diffusion process via DNA strand displacement. Given such a process for assembling a stable reaction-diffusion process could be a modular component of more complex patterning systems, our next goal was to design a system for RD and the molecules for the process that would react and diffusion with rates that would recapitulate the designed dynamics.

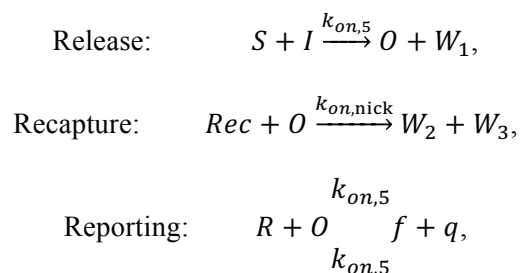
The gradients that we have designed require that the concentration of S and I remain constant at the boundaries in order to maintain the formation of a stable pattern. Waste produced from the reaction must also be able to exit the substrate. To develop a system in which these boundary conditions can be satisfied, we developed a system in which the pattern is formed within

a hydrogel and fluid reservoirs to each side of the hydrogel that may be periodically exchanged maintain the boundary concentration conditions and permit waste to be removed. We used a standard hydrogel, a 1% agarose gel for patterning. We expected DNA to diffuse mostly unobstructed by the agarose mesh because the mean pore size in 1% agarose hydrogel is estimated to be $\sim 580 \text{ nm}$ ¹⁹², whereas the length of a typical DNA complex involved in a simple strand displacement reaction is $\sim 10 \text{ nm}$, has a diameter of $\sim 2.5 \text{ nm}$ and has (for ssDNA) a hydrodynamic radius of $\sim 2 \text{ nm}$ ¹⁹³, so we expect the diffusion coefficient for such species to be similar to a solution-based diffusion constant of about $150 \mu\text{m}^2/\text{s}$ ^{194, 195}. In designing the reaction process, we assumed that all ssDNA and dsDNA would diffuse at approximately these rates, even though there are likely to be slight differences in diffusion coefficients for species of different lengths or secondary structures.

The RD system that we designed relies on the fact that diffusion rates are slower than reaction rates. Because the diffusion rates of biomolecules are relatively slow, this suggested that patterns would be easiest to visualize across size scales of millimeters or more.

We next chose rates for release and recapture reactions that would produce a sharp gradient using reasonable concentrations of the inputs (Fig. S5.1). To build a simple release reaction, we sequester an Output, O in an inert form within a Source complex, S , which rapidly releases O in the presence of an Initiator molecule, I (Fig. 5.1Ai). The reaction was mediated by a 5bp toehold, to ensure the reaction's speed. The release of O is balanced by a recapture reaction that constantly consumes O through a Recapturer complex, Rec , which binds to and sequesters O back into an inert state (Fig. 5.1Aii). We designed this reaction to occur much more slowly, *via* a strand displacement process that is initiated at a nick in the Rec dsDNA backbone, which has a rate constant of $500 (\text{M s})^{-1}$. Together, release and recapture processes cycle molecules of O between their released (free) and recaptured (inert) state, forming a gradient faster than diffusion mixes the components. To

experimentally monitor the concentration of released O , a fast, reversible reporting reaction was designed. In this reaction, a Reporter complex R , which consists of a fluorescent strand f , and a quencher strand q , rapidly and reversibly reacts with O through 5bp toeholds. In the unreacted state, the quencher strand prevents the fluorescing of the Reporter complex. However, when the Reporter complex successfully reacts with the Output, the Output displaces the quencher strand and the resulting complex fluoresces (Fig. 5.1Aiii). The resulting reactions are as follows:



where W are waste products and $k_{on,5}$ and $k_{on,nick}$ are the second order reaction rate constants. The strand displacement processes that underlie each reaction are shown in Figure 5.1A. The subscript in the reaction rate constant indicates the length of the toehold in the reactions. In our models we assume reaction rates are not significantly affected by the hydrogel and use approximate values for the forward reaction rate constants in solution: $k_{on,5} = 5 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ and $k_{on,nick} = 500 \text{ M}^{-1}\text{s}^{-1}$ ⁴⁵ (SI Fig. S5.1, SI Note S5.4).

We next designed the DNA sequences that would be the physical implementation of the reactions. To do so, we augmented sequences reported in Ref.⁴² with longer toeholds and additions of clamps to mitigate leak reactions¹⁹⁶, and used NUPACK⁴⁶ to aid in our sequence design (see SI Note 5.6 for sequences).

To mitigate the leak reactions that occur between species we PAGE purified the complexes (see Methods in Section 5.4). However, it has been observed that some small leak still occurs between species in the reaction, namely the Source and the Reporter¹⁹⁷, likely a result of fraying ends of complexes^{198, 199}, synthesis errors (*e.g.*, truncation, deletion, transversion)²⁰⁰ or

imperfections in complex purification. In our models, we account for this leak by assuming that a leaky source complex, at an initial concentration of 4.7% of the Source complex can react reversibly with a Reporter complex with a forward rate of $50 \text{ M}^{-1}\text{s}^{-1}$ ¹⁹⁷. Specifics can be found in Supplementary Note S4, and reactions seen in SI Fig. S5.10.

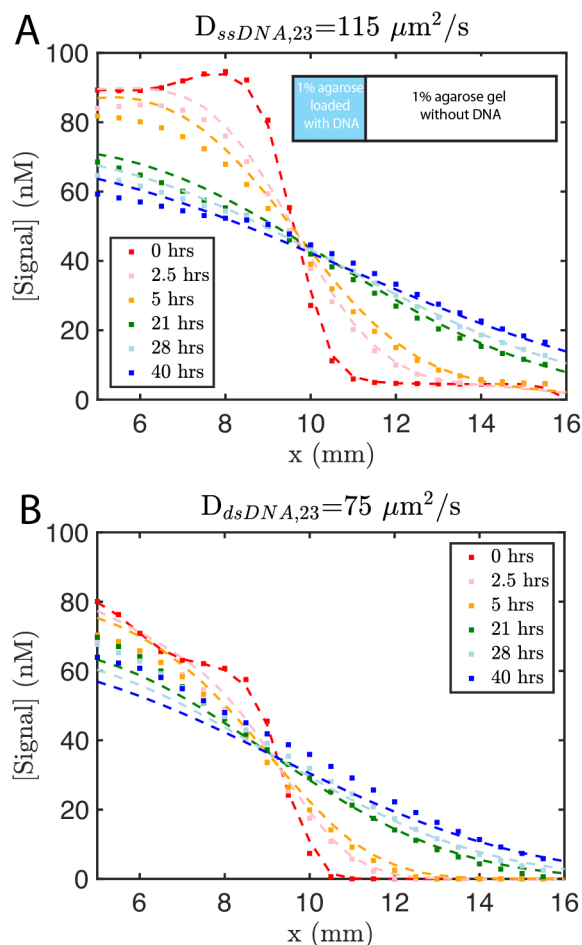


Figure 5.2. Diffusion coefficient measurement for ssDNA and dsDNA in 1% agarose. Experimental data (square markers) and simulation of the best fit to the diffusion coefficient (dashed line) for (A) ssDNA and (B) dsDNA. The initial conditions for both experiments are shown in the inset in (A). Diffusion coefficients were fit to experimental values using a least squares fitting method; simulations show the predictions of diffusion with the best fit. The system consisted of a 1% agarose gel that initially contained no DNA as shown and was cast and left at room temperature to gel for approximately 30 minutes prior to adding the gel with DNA. After this set time, the remaining $\sim 1/3$ of the reaction cell was loaded with reaction buffer that contained ~ 75 to 100 nM of fluorescently labeled DNA. 0 hours indicates the time at which imaging started, which was about 30 minutes after the fluorescently labelled DNA in agarose was added to the reaction channel. Starting concentration profile used in the simulation was taken from the initial experimental

concentration profile (*i.e.*, at 0 hours). Both DNA species are 23 base pairs long. Sequences are listed in SI Note 5.6. We found the diffusion coefficients to be $D_{ssDNA,23} = 115 \pm 1 \mu\text{m}^2/\text{s}$ and $D_{dsDNA,23} = 75 \pm 3 \mu\text{m}^2/\text{s}$, where the uncertainty represents one standard deviation of the measured quantity.

Construction of a reaction-diffusion reaction cell and reaction monitoring. To characterize RD patterns forming on the scale of millimeters such that external fluid reservoirs would resupply reactants and exchange waste, we designed a reaction cell composed of three portions of approximately equal volume. The outer portions were two liquid reservoirs with a 1% agarose hydrogel in the middle (Fig. 5.1B and SI Note S5.5). We assembled different gradients by changing the composition of the liquid reservoirs, which determined the boundary conditions for the RD process and characterized the gradients that formed within the hydrogel compartment. This set up allows energy to be supplied to the system in the form of reactants within introduced in the liquid reservoirs that diffuse into the hydrogel. We can exchange the reaction buffer in these reservoirs repeatedly without perturbing the pattern, enabling the gradients that form to be maintained as long as the solution in the reservoirs maintains the boundary conditions required for the pattern. We measure the intensity of fluorescence using time-lapse microscopy (see Methods in Section 5.4) and convert intensity to concentration of Output using a calibration curve (SI Fig. S5.8, Methods).

To understand and design the gradient patterns that would form as a result of DNA complex strand displacement and diffusion, we built simple computational models using the measured diffusion rates and assumed reaction rate constants and found that indeed, stable gradients can be formed and remain stable for many hours using these sets of diffusion coefficients and reaction rate constants, as long as the liquid reservoirs were periodically replenished (SI Fig. S5.2, Methods).

Gradient generation using DNA diffusion. Our first goal in performing experiments was to verify that the DNA complexes we designed could produce gradients generated through the

diffusion of different DNA complexes between reservoirs of different concentrations of a size and shape consistent with their diffusion coefficients.

Because the diffusion constant of the species in the particular environment of the 1% agarose gel substrate in such a system determines the dynamics of gradient formation as well as its final state, we first measured the bulk diffusion coefficients of species within our systems to improve on our initial estimates based on previously measured diffusion constants. Because the diffusion constants for DNA molecules depend sublinearly on its length^{194, 195}, and most complexes and strands in our system are of relatively similar size, we chose to use measurements of one 23 base-pair single-stranded DNA molecule (Fig. 5.2A) and one 23 base-pair double-stranded complex (Fig. 5.2B) as estimates for the diffusion rates of each of the single- and double-stranded species in our system. To measure diffusion rates for each species, we built a two-chamber agarose system where approximately 1/3 of the total length of the cell contained a 1% agarose gel “reservoir” which initially contained DNA and the remaining 2/3 was comprised of 1% agarose gel without DNA (Methods in Section 5.4). Initially, the reservoir contained approximately 100 nM of the respective species with a fluorescent label to allow it to be tracked. We followed the spread of the species using time-lapse fluorescence microscopy (Methods in Section 5.4) and fit a diffusion rate constant using standard diffusion equations to the changes in the concentration of the species within the agarose gel. We found $D_{ssDNA,23} = 115 \pm 1 \mu\text{m}^2/\text{s}$, or ~75% the reported value for DNA of similar size in solution^{194, 195}, and $D_{dsDNA,23} = 75 \pm 3 \mu\text{m}^2/\text{s}$. We used these values in combination with the reaction rates in order to design and model reaction-diffusion processes involving DNA strand displacement within an agarose gel.

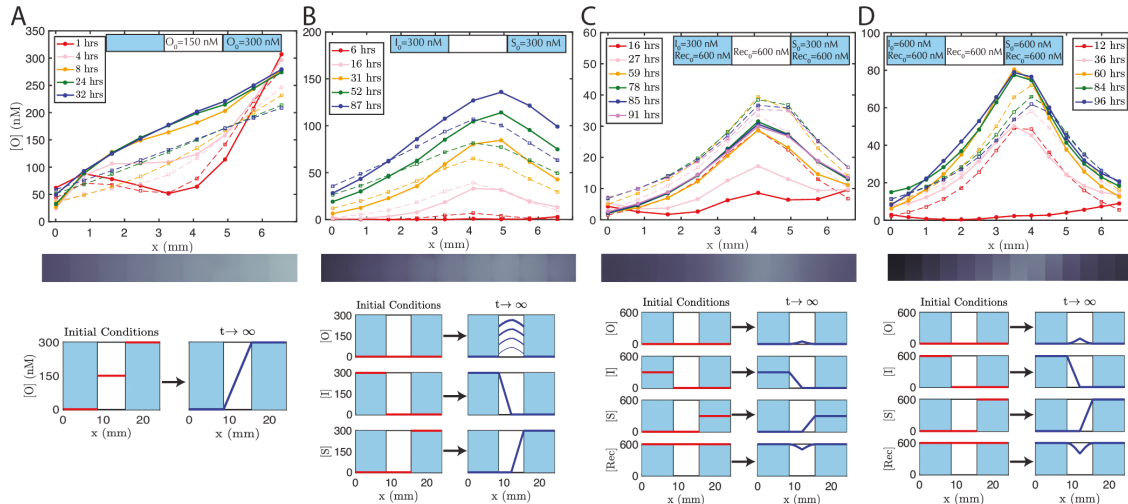


Figure 5.3. Designed processes of pattern generation in a RD system using DNA strand displacement reactions. Diffusion (and pattern formation) occurs within an agarose hydrogel. Boundary conditions are controlled by liquid reservoirs on opposing sides of the agarose hydrogel, which contain defined concentrations of reactants and which were periodically replenished without disturbing the pattern formation process. Initial conditions established through loading the liquid reservoirs and the agarose hydrogel with reactants. Solid lines show experimental values drawn from micrographs (see Methods in Section 5.4) and dotted lines are the predictions of zero parameter-fit simulations. Various types of 1D patterns can be formed including (A) linear gradient, (B) hill gradient that grows over time and (C-D) stable hill gradients. Note that the y-axis scale is different from plot to plot. The bottom panel shows initial conditions and expected concentrations after a long time of the different species in the reaction. The initial concentration of the reporter complex is $R_0 = 200$ nM in both the liquid reservoirs and the agarose hydrogels for all systems. Fluorescence micrograph depicted underneath each plot shows a visualization of the state at the longest time listed in the legend. Initial conditions (left, red line) and concentration profile after long reaction time (right, blue line) of the RD system are depicted beneath plots, where the two blue regions indicate liquid reservoirs and the white region indicates hydrogel. $[O]$, $[I]$, $[S]$, $[Rec]$, indicate the concentration of Output, Initiator, Source and Recapturer, respectively. These plots are meant to be qualitative guides to help the reader gain intuition about how the concentration profiles of various key reactants might look after long periods of time and assume that the liquid reservoirs were periodically exchanged (*i.e.*, concentrations in liquid reservoirs at $t \rightarrow \infty$ equal to I.C.). Increasing thickness of blue line in (B) indicates increasing time.

We next characterized how gradients of the Output species, within our system would form given two boundary reservoirs with the DNA species at different concentrations. We fabricated a three-chamber system that we designed that contained a 1% agarose gel in the center and two liquid reservoirs on either end. We placed a buffer solution containing 300 nM of Output in one liquid reservoir and a buffer solution with no Output in the opposing reservoir so that the stable state of the system would be a linear concentration gradient spanning from one hydrogel gel edge to the

other. To accelerate gradient development by reducing the quantity of Output which must diffuse into and across the distance of the gel, the hydrogel was initially loaded with 150 nM of Output. For the linear gradient and all gradients described hereafter, both the reservoirs and the hydrogel contained 200 nM of Reporter to enable the reaction to be followed by fluorescence imaging. The expected linear gradient mostly formed by about 8 hours and remained stable over at least 24 hours (Fig. 5.3A). The dynamics of formation were also consistent with simulations that used the reaction rates diffusion constants we measured. We also verified that the expected linear gradient formed in response to different initial concentrations of Output in the reservoirs and hydrogel, and where the total amount of DNA in the hydrogel would need to change to reach steady state (SI Fig. S5.3).

Coupled release and diffusion form growing gradients. To verify that designed gradient patterns could also be formed using coupled reaction and diffusion, we tested how a gradient of Output formed within the agarose gel substrate loading buffer containing 300 nM of Initiator species into one reservoir and buffer containing 300 nM of Source species into the opposite reservoir. These species can diffuse into the gel and react to release the Output species, and the maximum amount of Output species arises where the Initiator and Source species (on average) arrive *via* diffusion in approximately equal times from their respective reservoirs, forming a gradient. The height and width of the gradient both increase as Output continues to be released and diffuses outward. In experiments, the expected hill-shaped of Output formed (Fig. 5.3B), in a manner consistent with the predictions of simulations. As the reservoirs were replenished Output concentration continues to rise over at least 87 hours, producing a gradient where Output concentrations reached 150 nM at the peak.

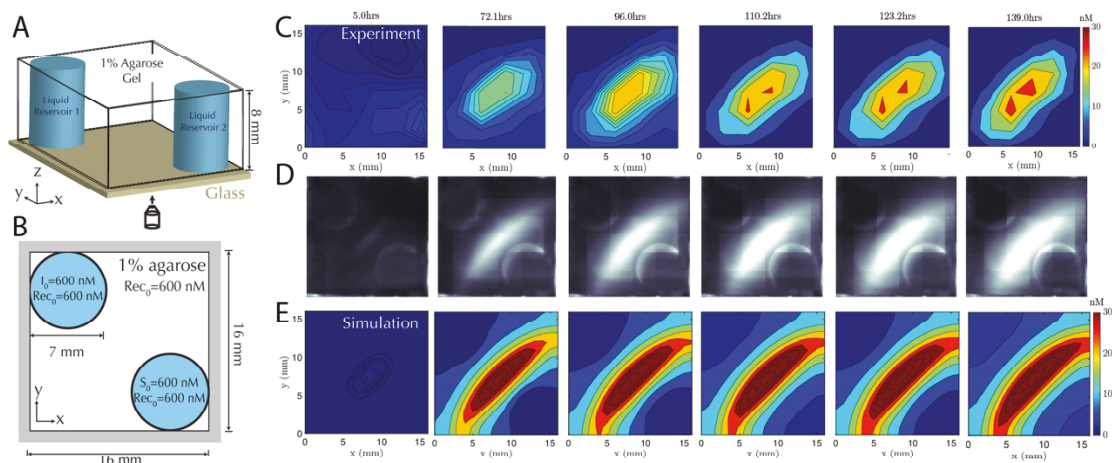


Figure 5.4. Stable two-dimensional gradient. (A) Schematic of the two-dimensional RD cell. Two cylindrical liquid reservoirs are at opposing corners of the RD cell. (B) Schematic of the top view of the 2D RD cell with initial and boundary conditions for the reactants. Reporter concentration is $R_0 = 200$ nM in both cylindrical liquid reservoirs and the agarose gel. (C) Filled contour plots depicting the concentrations of the Output at times ranging from 5 hours (left) to 139 hours (right). The stable gradient within the hydrogel takes approximately 110 hours to develop and is stable until the experiment was terminated at 139 hours. Exchange of the buffer in the reservoirs occurred in the experiment after 22, 48, 71, 93 and 116 hours. (D) Corresponding fluorescence micrographs (see Methods in Section 5.4 and SI Note S5.2). (E) Filled contour plots depicting the simulated values of Output concentration profile from 5 to 139 hours.

Stable gradients form through balanced release and recapture. To build gradient patterns that emerge and stabilize, we designed a system in which Output species would be both released and recaptured within the hydrogel. We added 300 nM of Initiator species into one reservoir, 300 nM of Source species into the second reservoir and 600 nM of Recapturer species into both reservoirs and the hydrogel portion of the RD cell (see Methods in Section 5.4). The high concentration of Recapturer ensured that it would not be depleted significantly through interaction with the Output, and that its concentration could remain stable across the substrate as the reservoirs are replenished periodically. A gradient of Output species emerged over 30-60 hours with higher concentrations of Output near the middle of the hydrogel and lower concentrations of the edges. After 60 hours, the shape of the gradient reached a shape that remained stable for an additional 30 hours. Zero parameter-fit simulations matched experiments closely: in both the simulations and

experiments, gradients formed the same final stable, shape and the approach to stability and the time scale at which a stable shape is achieved are very similar.

Stable gradient height is controlled by boundary conditions. Higher concentrations of Source and Initiator on the boundary would be expected to increase the rate of Output release within the hydrogel and thus increase gradient height. To test whether the resulting change in gradient height would correspond to the change in these concentrations, we assembled a stable gradient using the Source, Initiator and Recapturer species where Source and Initiator Concentrations in their two reservoirs were 600 nM, double the concentrations used in the first experiment (Fig. 5.3D). The resulting gradient shape stabilized at least as quickly as the gradient formed using lower concentrations of the Source and Initiator, consistent with predictions, and remained stable for 30 hours. The Output concentration at the gradient peak also increased in height over the corresponding peak concentration formed in the first stable system, as predicted by simulations performed using the higher concentrations of Source and Initiator

Stable reaction-diffusion in two dimensions. Programmed reaction-diffusion processes should enable the formation of patterns in 1, 2 and 3 dimensions. Having characterized the formation of 1-dimensional patterns of DNA species, our next goal was to characterize the formation of 2-dimensional patterns using a similar process. To enable control over the boundary conditions in such a way that they could be maintained over time we fabricated a square RD cell (Methods in Section 5.4) containing two cylindrical liquid reservoirs positioned in opposite corners of the hydrogel (Fig. 5.3A, B). We first tested the potential for building patterns within this setup by creating two RD systems by adding Source to the buffer within one reservoir and Initiator molecules the liquid reservoir in the other. We tested whether patterns formed according to the predictions of our reaction-diffusion network with the predicted rates and diffusion constants by studying how patterns formed within substrates of two different sizes. The dynamics of gradient

growth followed the predictions of simulations in both cases, and as expected, gradients arose more quickly in the smaller system, where the Source and Initiator needed to diffuse a smaller distance from their reservoir to react and release the output (Figs. S5.5, S5.6).

To test that stable, two dimensional patterns of Output concentration could form, we loaded one reservoir with 600 nM of Source complex and the other with 600 nM of Initiator species. 600 nM of Recapturer was added to both reservoirs and to the hydrogel. The gradient grew over time at rates in close agreement to simulations, producing a peak concentration of about 30 nM in 90 hours (Fig. 5.3C-E).

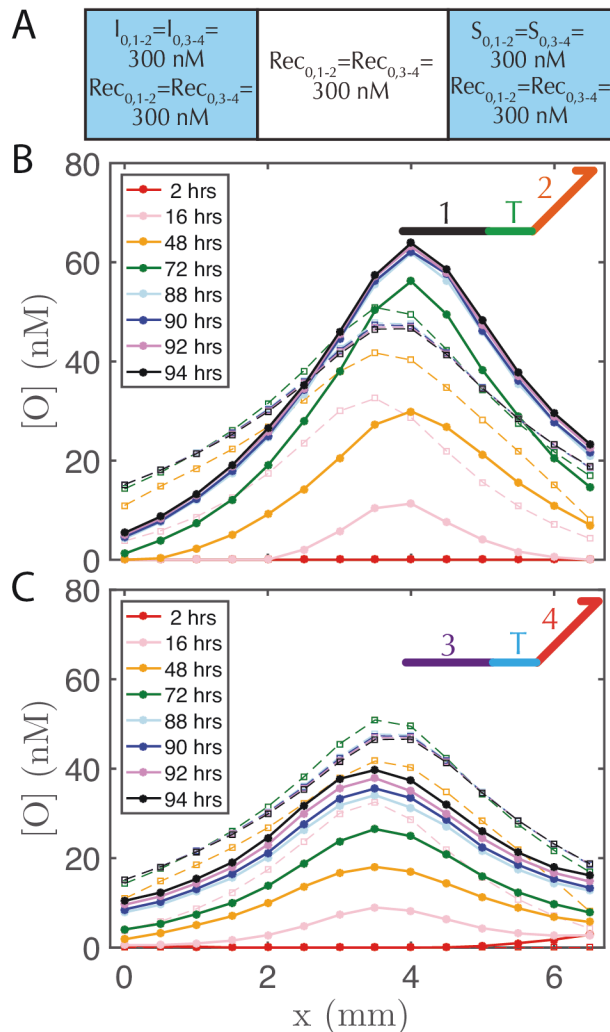


Figure 5.5. Two stable gradients of Output species in one hydrogel with two non-interacting (orthogonal) sets of reactants. (A) Initial conditions of reactants in the 1-2 and 3-4 system in the liquid reservoirs and hydrogel. The two sets of reactions are called systems 1-2 and 3-4, for their recognition domains (see inset cartoon diagrams of Output species in B-C). Initial reporter concentration is $R_{0,1-2} = R_{0,3-4} = 200 \text{ nM}$ in both liquid reservoirs and agarose hydrogel and is not depicted for clarity. Concentration profile for the Output strand with the (B) 1-2 system and (C) 3-4 system. The Reporter in the 1-2 system has a FAM fluorophore, whereas the Reporter complex in the 3-4 system has a Texas-Red fluorophore, so the respective fluorescence profiles (which are then converted to concentration profiles) were measured using non-overlapping filters for FAM and Texas Red channel (see Methods in Section 5.4). Buffer exchanged after 48 hours.

DNA systems with different sequences form similar gradient patterns. DNA strand-displacement systems are of interest as a substrate for programming RD processes not only because the reaction rates of the components can be controlled, but also because multiple reaction processes

involving different sequences can operate together with minimal crosstalk or as coupled reaction processes to produce more complex patterns. To verify that such scaling is feasible, we developed two non-interacting (*i.e.*, orthogonal) sets of reactions for the Release, Recapture and Reporting of Output Species and tested whether they could be executed in parallel. To do so, we first designed a set of complexes that release and recapture an Output species with a different sequence than the Output species shown in Fig. 5.1A. Within this new set, the 1 domain is replaced by a new 3 domain, the 2 domain is replaced by a 4 domain. The toehold sequence is also changed. We termed the initial set of complexes the 1-2 system and the new set of complexes the 3-4 system, after the labels on the sequence domains within each system. We designed the 3-4 system in the same fashion as the 1-2 system, and also used NUPACK⁴⁶ to ensure no unintended interactions between recognition domains. To separately measure the concentration profiles of the 1-2 and 3-4 system we ensured that the emissions from each Reporter could be observed in two separate, non-overlapping filter cubes (Methods). For the Reporter in the 1-2 system, a 3' Iowa Black FQ quenches 5' fluorescein (FAM) fluorophore (Figure 1A(ii)) and for the 3-4 system, a 3' Iowa Black RQ quenches 5' Texas Red® fluorophore (see sequence information in SI Note S5.6).

To form stable, multiplexed gradients, we loaded one liquid reservoir with 300 nM Initiator, the other liquid reservoir with 300 nM Source and in each liquid reservoir and hydrogel loaded 300 nM Recapturer and 200 nM Reporter for each system (Fig. 5.5A). Both gradients achieved the expected shape and approached a final stable state, consistent with the predictions of simulations (Fig. 5.5B-C). Each gradient took approximately the same amount of time to form, but the 1-2 gradient had a maximum concentration of about 60 nM, whereas the 3-4 gradient was in better agreement with simulations and had a maximal concentration of about 40 nM. However, the 3-4 gradient did not appear completely stable after about 85-95 hours with increases in output concentration on the order of a few percent per hour over 88-94 hours, whereas the 1-2 gradient

appeared to have achieved a stable configuration by this time. These results thus demonstrate that multiple DNA-based reaction networks can operate in tandem such that the dynamics and final state quantitatively match the designed dynamics and final gradient state.

5.3 CONCLUSION

In this work we have shown how DNA strand displacement reactions, operating far from equilibrium can be used to engineer stable and dynamic gradients with quantitatively defined features in a reaction-diffusion system. While there are many methods for generating chemical gradients by coupling chemical reactions and diffusion, this work demonstrates how simple enzyme-free and yet programmable chemistries capable of complex logic can form patterns that can reach a stable, stationary state. Such patterns can also continue to evolve over time in a predictable fashion. Most aspects of programmability that are suggested by using strand-displacement systems remain to be explored, including altering the reaction rates (*e.g.*, changing the length of the toehold domain^{44,45}) and/or the diffusion rates (*e.g.*, by introducing hydrodynamic drag¹⁸² or by altering length of the DNA components¹⁹⁵), which could also serve to sharpen the spatial resolution. Our setup shows how gradient patterns can form within a hydrogel such that reservoirs external to the system can be refueled, enabling far-from-equilibrium gradient patterns to be maintained within the system over long periods of time. The limited spatial resolution (hundreds of μm) observed here and long timescales (hours) for gradients to develop, might also be addressed by miniaturizing such RD systems within a microfluidic device.

The balanced process of release and recapture demonstrated here is a primitive, which when combined with chemical logic, could be used to build reaction-diffusion systems that produce elaborate stable patterns such as a French flag, a human stick figure, and cellular automata using a structured, rational approach^{201,202}. This reaction mechanism should also allow stable patterns that

are perturbed by some external stimulus (*e.g.*, a high concentration of a reactant in a strand displacement system) to revert to the pattern before the perturbation.

To build more sophisticated systems, it will be essential to implement more sophisticated sequence design and purification methods so that components interact only with their intended targets with predictable rates^{46, 203, 204, 205}. Logic operations and the systematic design of more complex networks suggest the possibility of engineering patterns with virtually arbitrary shape²⁰¹ or even dynamics. Strands with more complex domains can be designed such that leak can be significantly reduced²⁰⁶.

One major advantage of a DNA-based RD system is the fact that through logic reactions, DNA can sense the current state of the environment and adapt accordingly. Such an interaction with the environment could lead to patterns that change over time as the environment does, creating new function. For example, one could envision using such a system at biological temperatures to interact and perform logic with biological systems²⁰⁷ or other downstream processes^{208, 209} at specific locations and times. From the starting point of this work, further spatial (*e.g.*, 3D) and temporal control (*e.g.*, by introducing a various types of circuits¹⁹⁷) of patterns could be achieved too. The use of a biocompatible agarose hydrogels for patterning and nontoxic biomolecules as substrates also suggests that in principle such patterns might be used within environments for the study and control of biological systems²¹⁰.

5.4 MATERIALS AND METHODS

5.4.1 DNA complex preparation

All DNA strands were ordered from Integrated DNA Technologies (IDT) with standard desalting except fluorescently modified strands, which were HPLC purified. Complexes were formed by mixing the component strands at equimolar ratio in TAE Mg²⁺ (40 mM Tris-Acetate, 1 mM EDTA buffer containing 12.5 mM Magnesium Acetate) and then placed in an Eppendorf Mastercycler

PCR, where the strands were annealed. Annealing consisted of holding the temperature at 95 °C for 5 minutes and then cooling the solution to 25 °C at a rate of -1 °C per minute. After annealing, each complex was purified by polyacrylamide gel electrophoresis (PAGE) using a 10% polyacrylamide run at 120 V for 90 minutes at 4 °C (see Ref. ¹⁹⁷ for more details). The bands corresponding to the complexes were identified using UV-shadowing at a wavelength of 254 nm. The band was then diced, combined with TAE Mg²⁺ buffer into a tube and shaken on a vortexer for about 12 hours, to promote complex migration into the aqueous solution. The solution was then centrifuged for 5 minutes at 3000x g and the supernatant removed, which was repeated twice to ensure separation of gel from solution. Concentration measurements were obtained using an Eppendorf Biophotometer. The extinction coefficient, ϵ , of a complex was approximated by the formula: $\epsilon = \epsilon_{top} + \epsilon_{bottom} - 3200N_{AT} - 2000N_{GC}$, where ϵ_{top} and ϵ_{bottom} are the extinction coefficients of the two strands that comprise the complex and N_{AT} and N_{GC} are the number of hybridized A-T and G-C base pairs in the complex ²¹¹.

5.4.2 Hydrogel preparation

Agarose gels with DNA complexes were prepared by mixing liquid agarose and complexes and then cooling the gels in devices to set. We prepared 1% agarose hydrogel (1 g / 100 mL) in TAE Mg²⁺ and left it to cool to 40 °C, after which we transferred the agarose solution to a glove box with PID fan temperature control (Coy Labs) set to 40 °C. Buffers, DNA complex solutions, pipettes and pipette tips were left in the glove box at least 30 minutes prior to sample preparation to achieve thermal equilibrium. The agarose was mixed with the DNA solution (typically with Reporter and any other complex that was required for the experiment) and the resulting mixed was transferred to the device in the desired well(s). After all wells were patterned, a piece of Scotch tape was adhered to the PDMS to seal the wells and the device was transferred to the refrigerator

at 4 °C for 15 minutes to set the gels. We found that 40 °C was hot enough for the agarose to remain a liquid and cool enough that DNA complexes did not melt. The reservoir solutions were added after the gel had been cooled to room temperature and all results were collected at room temperature. Glass coverslips were placed on top of the device to mitigate evaporation from the calibration and reaction wells during the reaction-diffusion process.

5.4.3 Reaction Monitoring

The reaction was monitored using time-lapse fluorescence microscopy on either an IX73 or IX71 (Olympus) optical microscope. Image sets were obtained every ~30 minutes with an exposure of 50-150 ms using a 4X, 10X or 20X objective (Olympus) and were captured by an Infinity 3 CCD camera (Lumenera Corporation) in non-overlapping FAM and Texas-Red channels (Chroma) on the IX73, or an iXon3 cooled EMCCD camera (Andor) using a FAM channel (Chroma) on the IX71. The center of the focal plane for each experiment was approximately 1 to 4 mm in the z direction above the glass slide, as determined by the minimum width of the light beam in the direction perpendicular to the optical axis. We post-processed the images via binning and dark frame correction to compress the data and eliminate some of the optical artifacts, respectively (Fig. S5.7).

5.4.4 Liquid Reservoir Exchange

Contents of the liquid reservoirs are removed with a transfer pipette and replaced with fresh reaction buffer (prepared <10 minutes prior to exchange). Reservoir solution exchanges were performed approximately every 24 hours unless otherwise stated.

5.4.5 Simulations

To perform the simulations, we used COMSOL Multiphysics® Version 4.4 and LiveLink™ for MATLAB. Models of reaction-diffusion channels were built using COMSOL with the “Transport of Diluted Species” physics. All hydrogel-PDMS and liquid-PDMS boundaries were simulated with no flux boundary conditions. Simulations were run using scripts written in MATLAB using COMSOL Java API commands, which is how we defined simulation parameters such as reactions and their rate constants, diffusion constants, initial conditions and boundary conditions, mesh size and buffer exchange times. Buffer exchange of the liquid reservoirs occurred in simulations at the same time points as those in our experiments. Buffer exchange consisted of replacing each liquid reservoir with their initial contents (unreacted species) while the concentration profiles of species in the hydrogel remained unchanged. Diffusion coefficients for the hydrogel domain were set to measured values. Diffusion coefficients in the liquid reservoirs were all equal and had a value of $150 \mu\text{m}^2/\text{s}$ ^{194, 195}. Reaction rate constants for intended reactions were set according to estimated values in solution (see Ref.⁴⁵) and reaction rate constants for leak reactions were estimated (see SI Note S5.4). COMSOL models and MATLAB scripts are available upon request.

5.5 SUPPORTING INFORMATION

5.5.1 Supplementary Note S1: Supplementary Data

1D Gradients: Simulations

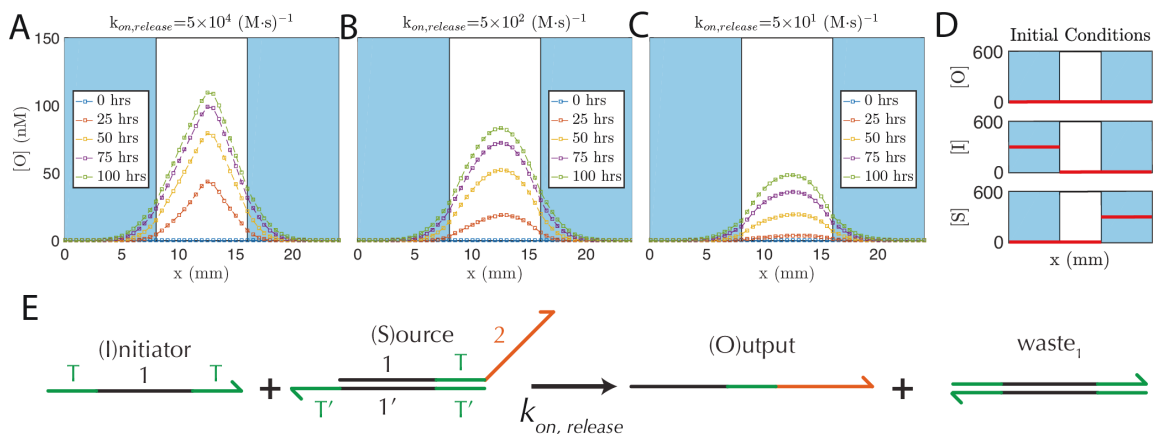


Figure S 5.1 Simulations of reaction and diffusion processes with different rate constants for the release reaction. To understand the influence of the release reaction rate constant on the gradients of that would form over time, we simulated the reaction and diffusion of Source and Initiator species in 1D with measured diffusion coefficients and different assumed reaction rate constants for the release reaction. We found that high reaction rate constants produce gradients with the highest peak concentrations and the largest changes per unit distance in concentrations. We simulated release rate constants of (A) $k_{on,release} = 5 \times 10^4 \text{ (M s)}^{-1}$, (B) $5 \times 10^2 \text{ (M s)}^{-1}$ and (C) 50 (M s)^{-1} , which correspond to reactions mediated by approximately 5bp, 3bp, 2bp toeholds respectively⁴⁵. (D) Initial conditions for the various species in the reservoirs and the agarose gel. The buffer in the liquid reservoirs was exchanged every 24 hours so that these conditions were maintained over time. Blue regions denote liquid reservoirs whereas the white region denotes hydrogel. (E) Schematic of the strand displacement process designed to occur at each of the rates considered in simulation.

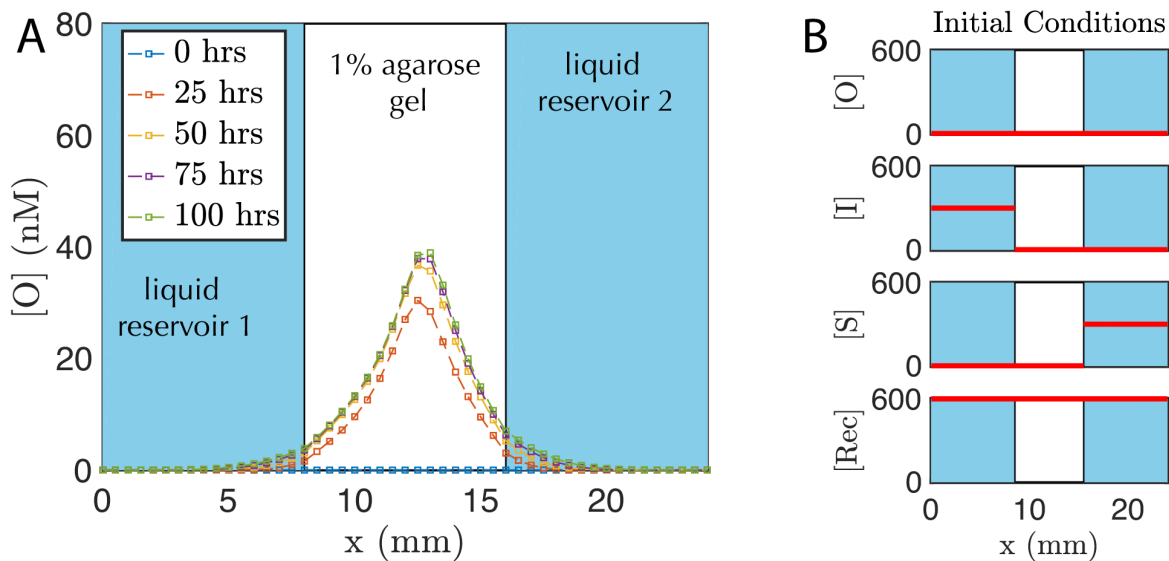


Figure S 5.2 Simulated reaction and diffusion of the release-recapture processes in 1D. To determine whether 1D gradients formed by release, recapture and reporting reactions in a 1% agarose hydrogel could be stable over time, we simulated the RD system with measured diffusion coefficients and assumed reaction rate constants based on measurements made in solution (see Simulation section in Methods of main text). (A) Simulations showed that gradients could remain stable once formed. Buffer exchange was simulated in the liquid reservoirs every 24 hours. (B) Initial conditions for the various species in the reservoirs and the agarose gel assumed in simulation. Blue regions denote liquid reservoirs whereas the white region denotes hydrogel. Reporter concentration is not shown but is initially 200 nM in both liquid reservoirs and in the hydrogel.

1D Gradients: Experiments and Simulations

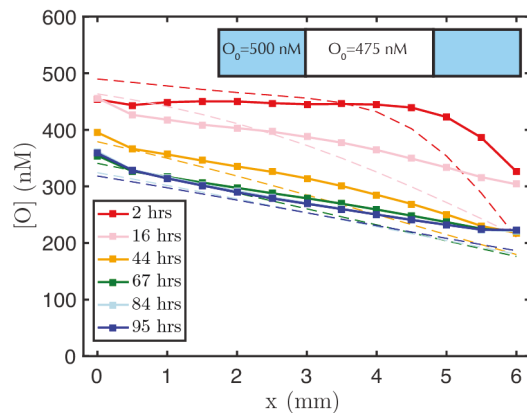


Figure S 5.3 A linear gradient forms when Output concentration is initially high within the hydrogel substrate. Inset shows reservoir concentrations of Output and the initial concentration of Output within the hydrogel. The Output concentration as a function of position within the hydrogel and time are shown. Dashed lines are results from a zero parameter-fit simulation whereas bold lines and squares indicate measured values from experiment. Initial conditions are depicted in the inset. Exact initial conditions in the gel were taken from the experimentally measured Output concentration at the initial time point and used as initial conditions for the simulation. Initial Reporter concentrations are not depicted for clarity and are $R_0 = 200$ nM in liquid reservoirs and agarose hydrogel. Buffer exchange of the reservoirs occurred after 24, 51 and 76 hours.

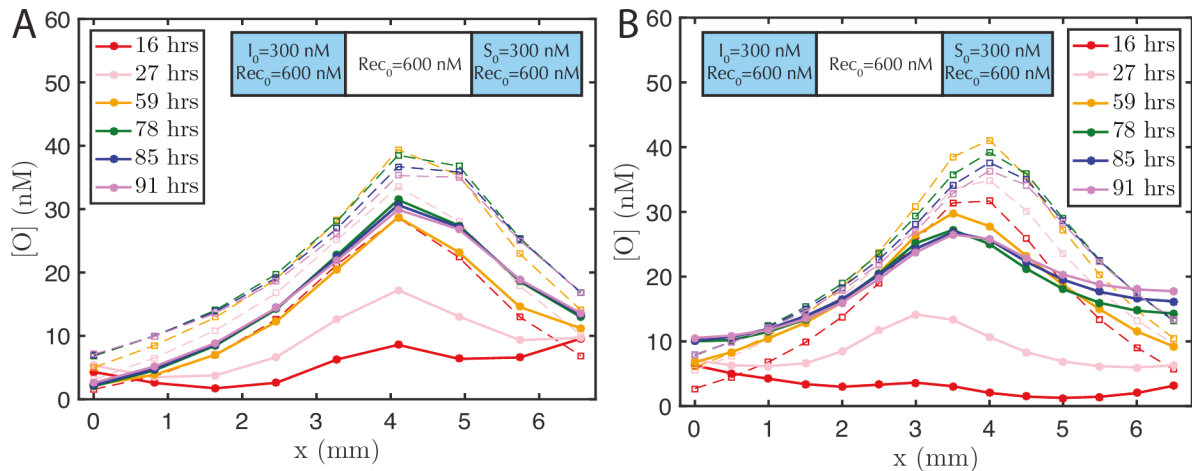


Figure S 5.4 Stable gradients form consistently. To characterize the inherent variations expected in the formation of gradient patterns within RD strand displacement systems, we performed two separate experiments in which we formed two stable hill gradients with identical initial and boundary conditions from separate Source, Initiator, Reporter and Recapturer stocks. The two gradients both stabilize in similar periods of time and have similar shapes but differ in peak height and shape. Such differences could be the result in differences in gels, reservoir heights and the purity and effective concentration of the component complexes. Initial conditions of the RD system (the same in both experiments) are depicted in the insets. Initial Reporter concentration is not depicted for clarity and is $R_0 = 200$ nM in the liquid reservoirs and agarose hydrogels for all systems. Inset is a cartoon schematic of the Output species (left) and is the same for both experiments. Buffer exchange occurred after (A) 24 and 52 hours and (B) 48 hours. Differences in simulations between the two figures reflect the different buffer exchange times. (A) 10X objective (IX71 microscope) and (B) 20X objective (IX73 microscope) were used to image the systems.

2D Gradients: Experiments and Simulations

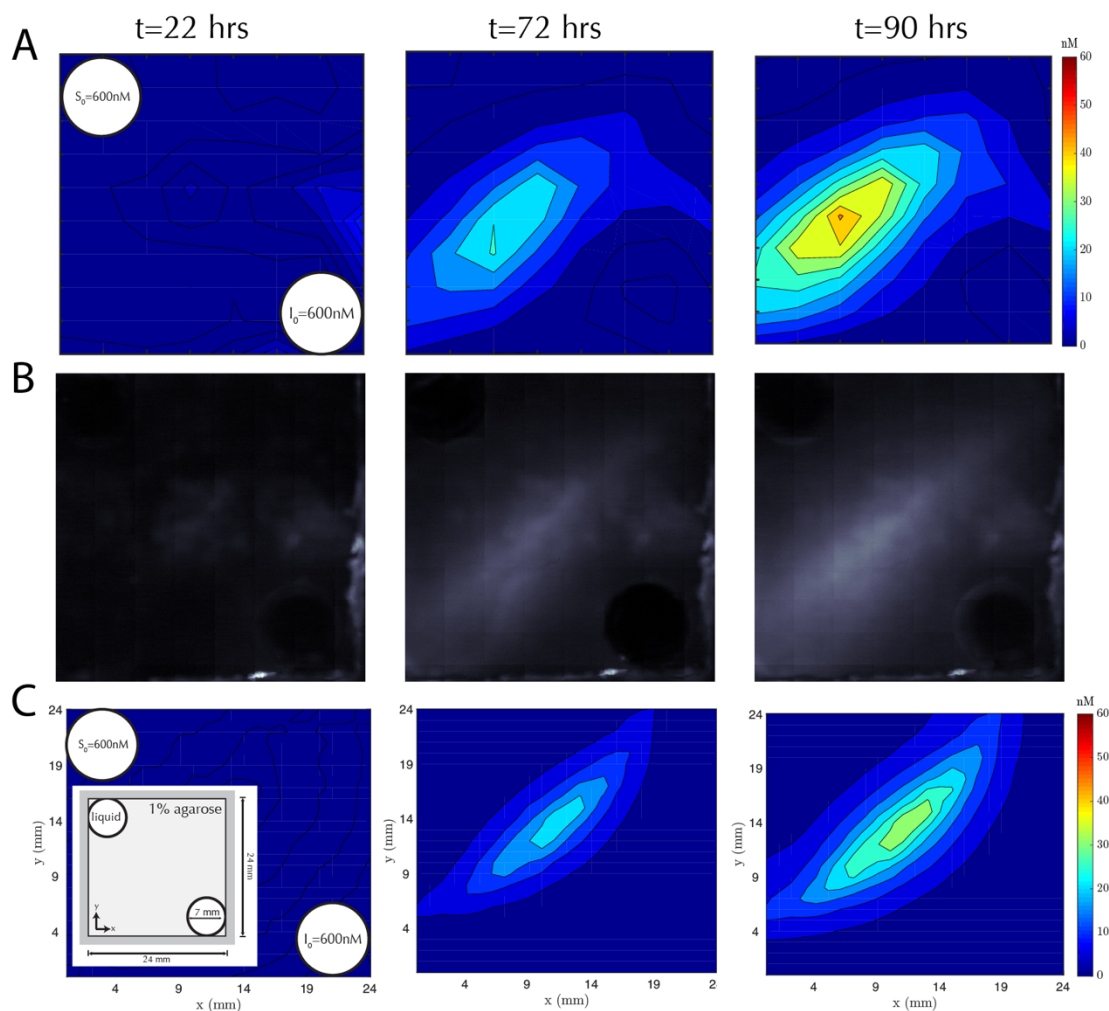


Figure S 5.5 Growing gradient in two-dimensions with large dimensions (24 x 24 x 8 mm). (A) Contour map of total concentration of Output species in experiment and corresponding (B) optical images and (C) simulations at three time points: 22, 72 and 90 hours. The first column is shown at ~22 hours, second column at ~72 hours and 3rd column at ~90 hours. Initial conditions (which are also refresh conditions) shown in the liquid reservoirs of the experimental and simulation contour map at 22 hours which also indicates well locations. Inset diagram in simulation contour map at 22 hours depicts the schematic for the reaction cell. Fluorescent micrographs are obtained in a raster fashion and stitched together as a mosaic, as the field of view of the 4X objective used is smaller than the RD cell. Dark frame correction was performed for the individual images (see SI Note S5.2). Buffer exchanged occurred after 23, 49 and 75 hours.

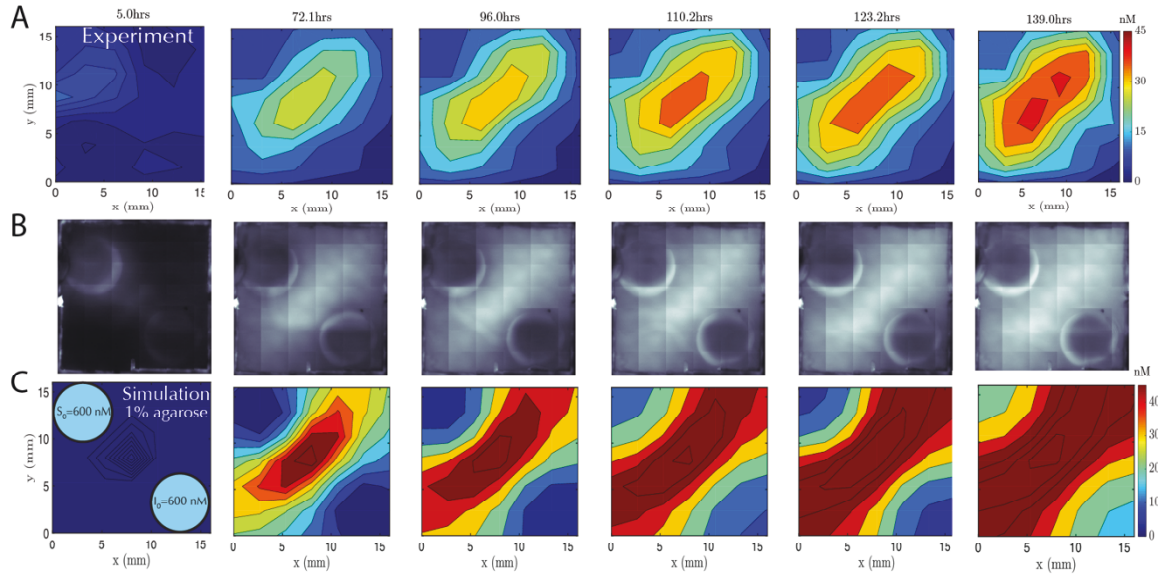


Figure S 5.6 Growing gradient in two-dimensions with small dimensions (16 x 16 x 8 mm). (A) Filled contour plots depicting the experimental values of Output concentration profile in from 5 hours (left) to 139 hours (right). Here the gradient continues to grow after 139 hours to develop, when the experiment was terminated. Buffer exchange occurred in the experiment after 22, 48, 71, 93 and 116 hours. (B) Corresponding fluorescent micrographs, which are obtained in a raster fashion and stitched together as a mosaic, as the field of view of the 4X objective used is smaller than the RD cell. Dark frame correction was performed for the individual images (see SI Note S5.2). (C) Filled contour plots depicting the simulated values of Output concentration profile from 5 to 139 hours. Leftmost plot in (C) depicts initial and boundary conditions of species. Reporter concentration is $R_0 = 200$ nM in both liquid reservoirs and in the agarose gel and is not depicted for clarity. See Figure 5.3 for RD cell cartoon diagram.

5.5.2 Supplementary Note S2: Image Processing

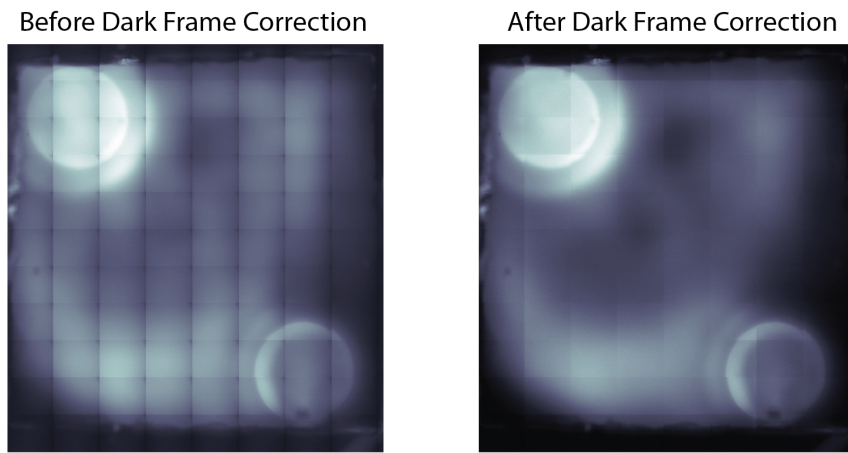


Figure S 5.7 Dark frame correction to reduce edge effects in image montages. For the sake of continuity in fluorescence images, we eliminated the majority of the edge effects (typically a darker ring around the outer pixels) by using standard dark frame correction algorithms. While still not perfect, the dark frame correction eliminates the majority of the imaging artifacts. A more sophisticated algorithm like flat fielding would likely eliminate more edge effects, however, it seemed infeasible to obtain the necessary uniformly-lit images for every image in the set. Immediately after image acquisition, we use a binning algorithm to compress the images, as a typical experiment captures $\sim 20,000$ images and an image captured on the 16-bit Infinity 3 CCD camera has 2752×2192 pixels (which would amount to ~ 240 Gb of data per experiment). The binning algorithm takes the mean intensity value of the nearest 4×4 pixels and stores the resulting value as a new pixel (resulting in a 16X compression).

5.5.3 Supplementary Note S3: Experimental Notes

Fluorescence Calibration to Output Concentration

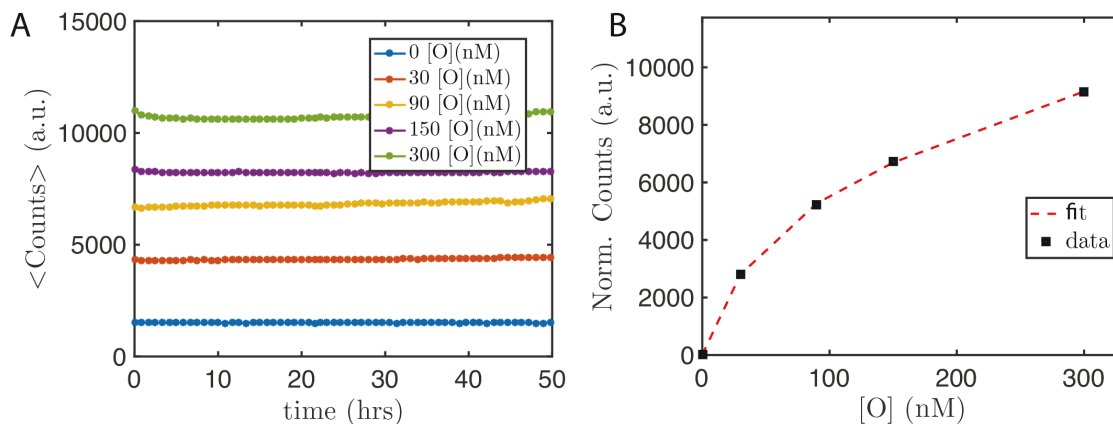


Figure S 5.8 Calibrating counts from calibration wells to concentration of Output species. (A) A typical calibration plot depicting counts from the calibration wells as a function of time. Counts are averaged across entire image. Wells have varying concentration of Output but have a constant amount of Reporter $[R]_0 = 200$ nM in each well. Images are typically taken with an exposure ranging from 50 to 150 ms. (B) The counts are averaged over a period of time, typically 20 hours, and then normalized to zero (such that the average counts at $[O] = 0$ nM are 0). The data is fit to the curve $C_N = \frac{a[O]^2 + b[O] + c}{[O] + d}$ where a, b, c and d are fitted parameters, NC is the normalized counts and $[O]$ is the Output concentration. This curve is then used to calculate the concentration of the total Output species (unbound Output species + Output species bound to the Reporter), where:

$$[O] = \frac{\sqrt{-4ac + 4adC_N + b^2 - 2bC_N + C_N^2} - b + C_N}{2a}.$$

Where the calibration wells were only fluorescent ssDNA or dsDNA, such as in the diffusion only experiments (see Figure 2 in the main text), the calibration curve was linear and thus fit to a linear equation, $C_N = m[O] + b$ (the fit largely depends on the gain setting used). All calibration wells are comprised of 1% agarose. Calibration wells measured 8x8x8 mm (512 μ L).

Preventing Evaporation of Material

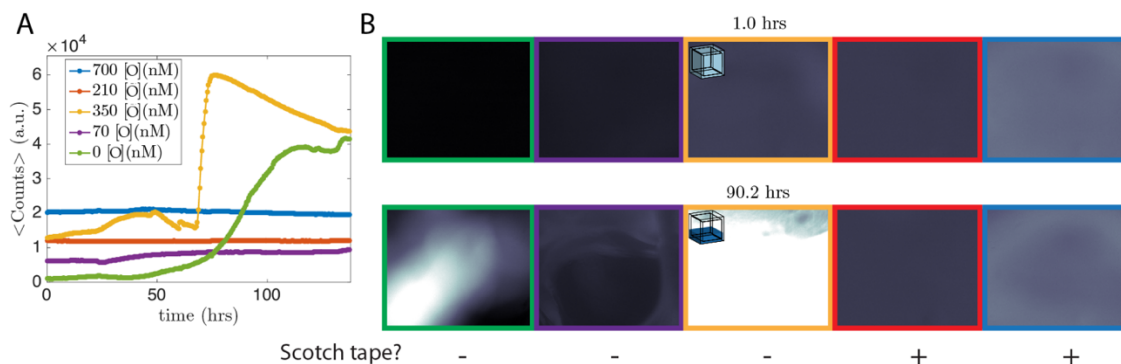


Figure S 5.9 Evaporation in calibration wells mitigated by using Scotch Tape. One significant factor in accurately converting fluorescence values to DNA concentrations was ensuring that material did not evaporate during time lapse imaging. Specifically, when covered only with a glass coverslip, almost all of the moisture in the calibration wells would evaporate in ~ 24 -100 hours, leaving the agarose hydrogel at 1-5% of its original volume. This would, in turn increase the fluorescence of the sample as the evaporation process concentrates the DNA species, seen in (A) mean fluorescence vs time of the calibration wells (yellow, green and purple lines) and in the fluorescent images in (B). However, the two calibration wells sealed with Scotch tape did not significantly change fluorescence over the course of 130 hours (red and blue lines). Thus, we sealed all calibration and reaction channels with Scotch tape to mitigate evaporation. Border color in optical images (right side) corresponds with line color in plot. Optical images shown after 1 hour (above) and 90.2 hours (below). Inset cartoon cube diagrams show (above) a calibration well full of agarose solution with DNA and (below) a well after some evaporation resulting in a solution that is more concentrated in DNA species.

5.5.4 Supplementary Note S4: Leak Reactions and Recapture Kinetics

Leak reactions, or unintended reactions between DNA species which result in Output release or recapture, are a ubiquitous and well documented phenomenon in DNA-based strand displacement systems. Strands and complexes that are less prone to leak can be designed but require more complexity²⁰⁶. Previous work¹⁹⁷ documented several types of leak reactions based on molecules similar to those used in this study and are the basis for reactions and leak parameters used in our models here. Such reactions are enumerated in Fig. S5.10.

Truncated complexes are a possible source of leak reactions. Such complexes can be created when one full length strand and one truncated strand hybridize to form a complex that has a smaller molecular weight than the full-length complex. Such leaky complexes are difficult to distinguish from full-length complexes in a PAGE purification process. Thus, leaky complexes are unintentionally introduced into all experiments. For example, leaky Source complexes occur when the bottom strand of a Source complex is missing bases on its 5' end (Fig. S5.10(i)), thus providing another reaction mechanism through which a Reporter complex can interact. Further examples of such leaky complexes and their reaction mechanisms can be seen in Fig. S5.10(ii-iii). Optimizing PAGE protocols to reduce the leak reactions is an ongoing effort, with some changes to protocols described here in Ref.¹⁹⁷.

The quantity of the leaky complexes has been measured in Ref¹⁹⁷, where the authors measured Source strands, similar to those used in this study, with leaky toehold domains to be anywhere between 0.5 and 4.7% of $[S]_0$, and react with rate constant $\sim 50 \text{ M}^{-1}\text{s}^{-1}$, which is the approximate rate constant for a strand displacement reaction mediated through a two base pair toehold, k_{2bp} . For the simulations here we assume that all leak reactions through a truncated strand occur at a reaction rate constant of $k_{2bp} = 50 \text{ M}^{-1}\text{s}^{-1}$ and that $[S]_{leak,0} = 0.047[S]_0$, as the bottom strand of the Source species in this paper is longer than in Ref.¹⁹⁷ and therefore we expect the

truncation error to be on larger end of the measured range. We also assume that a leaky Recapturer complexes (Fig. S5.9(iii)) also have an initial concentration of $0.047[Rec]_0$ and react with rate constants of k_{0bp} . Although leaky (*i.e.*, unintended), the leak pathway in a recapture reaction is slower and almost insignificant when compared to the primary mechanism of recapture.

While a strand displacement reaction mediated by a nick in a backbone of a DNA duplex, as seen in the Recapturer complex, would typically be expected to produce strand displacement kinetics with forward rates of 0.5 to $5 \text{ M}^{-1}\text{s}^{-1}$, (which approximately correspond to 0 and 1 bp toehold mediated reaction kinetics, respectively) we found that the Recapture reaction was best modeled with a reaction rate constant of $500 \text{ M}^{-1}\text{s}^{-1}$ more in line with a 3 bp toehold mediated reaction⁴⁵. Significant fraying of ends can explain the increase in kinetic rates. When analyzed with NUPCK at room temperature, the 1-2 and 3-4 Recapturer complexes have $\sim 50\%$ and $\sim 60\%$ probability, respectively, of having an unpaired base at the nick. To mitigate this fraying in a next generation design, a GC rich region that occludes fraying, for example, can be designed.

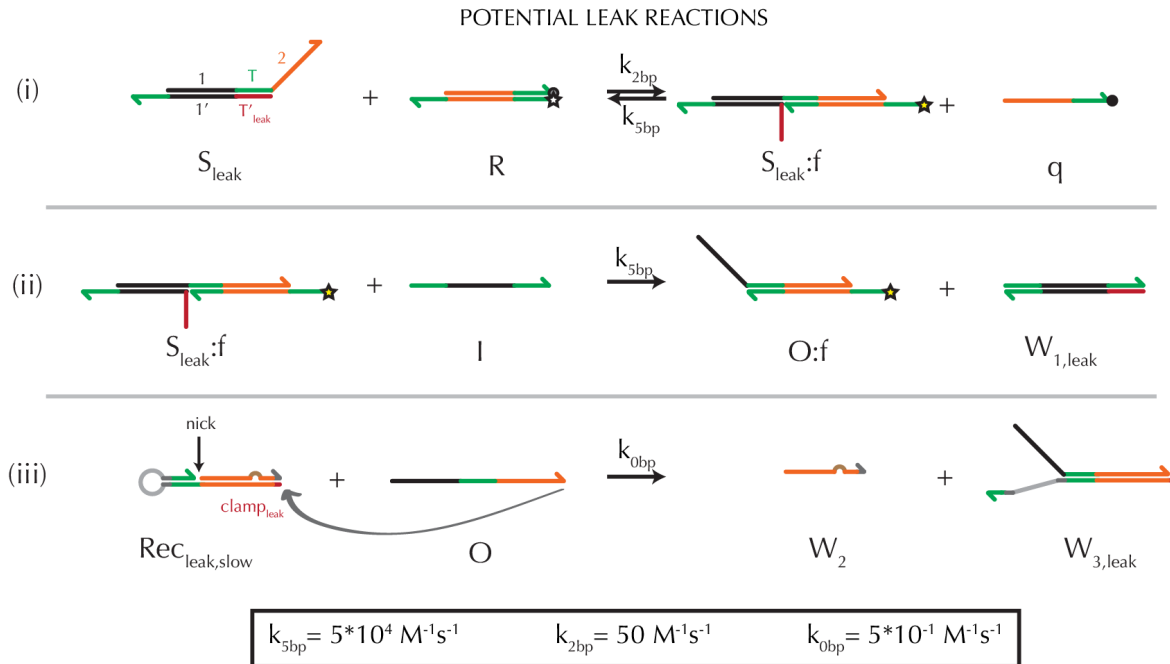


Figure S 5.10 Potential leak reactions considered in the model. (i) Reversible reaction between a leaky Source complex (S_{leak}) and the Reporter complex, where the leaky Source complex has a bottom strand that is missing some bases (shown in red and labeled T'_{leak}). Reaction rate constants shown above and below arrows and their values listed at the bottom of the figure and are estimates from Ref. ⁴⁵. We assume a leaky toehold has a -2bp truncation ¹⁹⁷ at the 5' end (IDT synthesizes DNA from 3' to 5', thus a truncation error will occur at the 5' end of a strand) and is shown in red. Although we PAGE purify the complexes, we found it challenging to distinguish between complexes with lengths that differ by only a few base pairs in our gels; the primary reason for PAGE purification is to ensure proper stoichiometric ratios of all strands in a complex. Such a source of error could be likely eliminated by purification methods which were able to differentiate between small changes in number of base pairs in a complex, or by ordering purified ssDNA strands. (ii) Irreversible reaction between a leaky Source complex with a fluorophore strand ($S_{leak}:f$) and an Initiator through a 5bp toehold-mediated strand displacement reaction to produce an Output:fluorophore and waste complex. (iii) Irreversible reaction between a leaky Recapturer with slow kinetics, where the 5' end of the bottom strand in the Recapturer has two missing bases, therefore the non-reactive clamp would be eliminated, exposing a potential 0 bp strand displacement reaction between an Output and the Recapturer. Gray arrow indicates the trajectory of the strand displacement reaction.

Table S 5.1 List of reactions modeled in RD system.

Release	$S + I \xrightarrow{5bp} O + W$	(1)
Recapture	$O + Rec \xrightarrow{k_{nick}} W_2 + W_3$	(2)
Reporting	$O + R \xrightarrow[k_{5bp}]{k_{5bp}} f + q$	(3)
LEAK (i)*	$S_{leak} + R \xrightarrow[k_{5bp}]{k_{2bp}} S_{leak}:f + q$	(4)
LEAK (ii)*	$S_{leak}:f + I \xrightarrow{k_{5bp}} O:f + W_{1,leak}$	(5)
Leak (iii)*	$Rec_{leak,slow} + O \xrightarrow{k_{0bp}} W_2 + W_{3,leak}$	(6)

*Leak reactions correspond to those listed in Fig. S5.10.

5.5.5 Supplementary Note S5: Reaction-Diffusion Device Fabrication

To fabricate the PDMS molds we used Sylgard 184 Silicone Elastomer Kit (Dow Corning) and mixed 10:1 of polymer:curing agent. We epoxied 8x8x8 mm wooden cubes (Amazon.com) to the bottom of a weigh boat to create the negative for the mold. Each calibration well was comprised of a single wooden cube and a reaction well was comprised of a linear chain of three cubes so that the 1D RD cell measured 24x8x8 mm (LxWxH) and the calibration wells (8 mm)³. Two dimensional molds were made from 4 or 9 cubes (dimensions of 16x16x8 or 24x24x8 mm, respectively) and the diameter of the cylindrical wells was 7 mm. The polymer and curing agent were well-mixed, poured into weigh boats and then placed in the desiccator for 90 minutes to eliminate entrapped air bubbles. The device was then cured for 2 hours at 65 °C. The PDMS mold was then extracted from the weigh boat. Dust particles were removed from the PDMS mold using Scotch Tape and the glass slide (48x65 mm, Ted Pella) was cleaned using 70% EtOH and dried with N₂. To attach the glass slide to the PDMS, the coverslip and PDMS were treated for ~45 seconds each using a high frequency generator (BD-20, Electro-Technic Products). The device was set at 1.5 hours at 85°C to help promote bonding of the glass to the PDMS.

5.5.6 Supplementary Note S6: DNA Sequences

Diffusion Measurement Experiment

ssDNA used in Diffusion Experiment (shown Figure 2A)

Rb28f: /5TEX615/GTATTGTTGAATTGTAGAGTATT

dsDNA used in Diffusion Experiment (shown Figure 2B)

Rb28f: /5TEX615/GTATTGTTGAATTGTAGAGTATT

Rb28f_full_comp: AATACTCTACAATTCAACAATAC

1-2 System

Reversible Reporter 5

Rv5q: CCACCAAACCTTCATCTCA/3IABkFQ/

Rb5f: /56-FAM/TGAGATGAAGTTTGGTGGTGAGA

Source 6_5

W6_5: CATAACACAATCACATCTCACCACCAAACCTTCA

Gb6(5bp): TGAGATGTGATTGTGTTATGAGATG

Initiator 6

W_6_: CATCTCATAACACAATCACATCTCA

Recapturer 5

Dv5: CACCACCAATCTTCACT

Db5: AGTGAAGTTTGGTGGTGGAGATGTTTTTACATCT

(base pair mismatch)

3-4 system

Reversible Reporter 28

Rv28q: TCTACAATTCAACAATAC/3IAbRQSp/

Rb28f: /5TEX615/GTATTGTTGAATTGTAGAGTATT

Source 27_28

W27_28: ACAACACTCTATTACAATACTCTACAATTCAAC

Gb6(5bp): TGAGATGTGATTGTGTTATGAGATG

Initiator 27

W_27_: ACAATACAACACTCTATTACAATAC

Recapturer 28

Dv28: ACTCTACAAATCAACAG

Db28: CTGTTGAATTGTAGAGTATTGTATTTTACAAT

(base pair mismatch)

5.6 ACKNOWLEDGEMENTS

The authors would like to thank Deepak Agrawal, Angelo Cangialosi, Qi Huang, Abdul M. Mohammed, and Samuel Schaffter for helpful discussions and useful input regarding this work, and Markela Ibo for her assistance with fabrication techniques for reaction-diffusion devices. This work was supported by NSF CCF grant 1161941 and DOE BES grant 90068952 (for support of J.F. and J.Z. and some materials and supplies) to R.S.

CHAPTER 6 CONCLUSIONS AND FUTURE DIRECTIONS

Overall, this work contributes to the growing body of literature which attempts to understand how to form, control and improve the self-assembly of DNA-based structures and patterns. Along with the future directions discussed in the conclusions sections of the chapters, there are some clear, actionable directions that one can envision as a result of this work, in addition to the conclusions in each of the chapters.

Per the results and discussions in *Chapters 2-4*, one could continue developing and testing the theory behind multicomponent self-assembly of finite sized structures. The efforts described in this thesis have helped slowly march toward the fabrication of what could be very useful technology: large, multi-origami structures with well-defined structure that rapidly assemble with high-yields. Although not yet achieved, there are many future studies directly stemming from this work that could help achieve this goal. For example, one could fairly easily test whether by simply changing the interface designs in a tetrameric structure one could enhance self-assembly efficiency at various assembly conditions. These results could be used to inform and improve the model outlined in *Chapter 2* and *4*. Requiring a more laborious effort, one could characterize the species in the reaction mixture after different reaction times and across a wide variety of conditions using the atomic force microscope or optical methods like those developed in our lab¹¹⁵. This type of study would allow us to improve models by directly comparing the results to those of experiments. Further, one could design and build components that have qualities more desirable for building larger structures than those used in our experimental study here: namely, increase the rigidity of components. Recently published, a good candidate for an experimental system could be the programmable DNA tiles used in Ref.²¹² where they demonstrate square grid structures, which form without curving and forming tubes in solution, with 9, 16, and 25 DNA origami tiles (similar to the model structures described in *Chapters 2* and *4*) that are able to achieve 15.6%, 15.0% and

32.4% yield, respectively. Following the discussion in *Chapter 5*, a long-term vision for such DNA-based reaction-diffusion systems is to couple downstream processes to such chemical patterns to control when and where more complicated processes occur, for example, cell differentiation, and incorporate “smart” reactions into the buffer such that when perturbed by an adverse chemical, for example, DNA sensors could quickly sense and act to remove such chemicals.

As with the trends in the cost of DNA sequencing and synthesis, the functional and structural complexity of nanoscale systems created from the broad field of DNA nanotechnology is nonlinear. For example, a nanoscale wireframe cube made from 6 synthesized ssDNA strands in 1991²¹³ to a solid cuboid with 710 strands of similar length in 2012⁶⁸. The complexity DNA-based devices, no matter how it is defined (e.g., the number of components or strand displacement reactions, the types of structural conformations or interactions in a device *etc.*), is increasingly nonlinearly with time. Like Moore, but with the benefit of many years of data, I predict this nonlinear trend will continue for at least the next decade.

REFERENCES

1. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738 (1953).
2. Seeman NC. Nucleic-Acid Junctions and Lattices. *J Theor Biol* **99**, 237-247 (1982).
3. Seeman NC, Kallenbach NR. Design of Immobile Nucleic-Acid Junctions. *Biophys J* **44**, 201-209 (1983).
4. Kallenbach NR, Ma RI, Seeman NC. An Immobile Nucleic-Acid Junction Constructed from Oligonucleotides. *Nature* **305**, 829-831 (1983).
5. Moore GE. Cramming more components onto integrated circuits (Reprinted from Electronics, pg 114-117, April 19, 1965). *P IEEE* **86**, 82-85 (1998).
6. KA W. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) (ed[^](eds).
7. Carlson R. The changing economics of DNA synthesis. *Nat Biotechnol* **27**, 1091-1094 (2009).
8. Carlson R. Time for New DNA Synthesis and Sequencing Cost Curves. <http://synbiobeta.com/> (2014).
9. Church GM, Gao Y, Kosuri S. Next-Generation Digital Information Storage in DNA. *Science* **337**, 1628-1628 (2012).
10. Rothemund PWK. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006).
11. Fu JL, Liu MH, Liu Y, Woodbury NW, Yan H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J Am Chem Soc* **134**, 5516-5519 (2012).
12. Linko V, Eerikainen M, Kostianen MA. A modular DNA origami-based enzyme cascade nanoreactor. *Chem Commun* **51**, 5351-5354 (2015).
13. Voigt NV, *et al.* Single-molecule chemical reactions on DNA origami. *Nat Nanotechnol* **5**, 200-203 (2010).

14. Andersen ES, *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73-U75 (2009).
15. Zadegan RM, Jepsen MDE, Hildebrandt LL, Birkedal V, Kjems J. Construction of a Fuzzy and Boolean Logic Gates Based on DNA. *Small* **11**, 1811-1817 (2015).
16. Zadegan RM, *et al.* Construction of a 4 Zeptoliters Switchable 3D DNA Box Origami. *Acs Nano* **6**, 10050-10053 (2012).
17. Kilchherr F, Wachauf C, Pelz B, Rief M, Zacharias M, Dietz H. Single-molecule dissection of stacking forces in DNA. *Science* **353**, 1116+ (2016).
18. Gu HZ, Chao J, Xiao SJ, Seeman NC. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202-U286 (2010).
19. Lund K, *et al.* Molecular robots guided by prescriptive landscapes. *Nature* **465**, 206-210 (2010).
20. Marras AE, Zhou LF, Su HJ, Castro CE. Programmable motion of DNA origami mechanisms. *P Natl Acad Sci USA* **112**, 713-718 (2015).
21. Gerling T, Wagenbauer KF, Neuner AM, Dietz H. Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components. *Science* **347**, 1446-1452 (2015).
22. Zhou LF, Marras AE, Su HJ, Castro CE. DNA Origami Compliant Nanostructures with Tunable Mechanical Properties. *Acs Nano* **8**, 27-34 (2014).
23. Schulman R, Winfree E. Programmable Control of Nucleation for Algorithmic Self-Assembly. *Siam J Comput* **39**, 1581-1616 (2009).
24. Mohammed AM, Schulman R. Directing self-assembly of DNA nanotubes using programmable seeds. *Nano Lett* **13**, 4006-4013 (2013).
25. Schulman R, Winfree E. Synthesis of crystals with a programmable kinetic barrier to nucleation. *P Natl Acad Sci USA* **104**, 15236-15241 (2007).
26. Bujold KE, Hsu JC, Sleiman HF. Optimized DNA "Nanosuitcases" for Encapsulation and Conditional Release of siRNA. *J Am Chem Soc*, (2016).
27. Douglas SM, Bachelet I, Church GM. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831-834 (2012).

28. Zhang Q, *et al.* DNA Origami as an In Vivo Drug Delivery Vehicle for Cancer Therapy. *Acs Nano* **8**, 6633-6643 (2014).
29. Krishnan S, *et al.* Molecular transport through large-diameter DNA nanopores. *Nat Commun* **7**, 12787 (2016).
30. Hariadi RF, Cale M, Sivaramakrishnan S. Myosin lever arm directs collective motion on cellular actin network. *P Natl Acad Sci USA* **111**, 4091-4096 (2014).
31. Langecker M, *et al.* Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures. *Science* **338**, 932-936 (2012).
32. Douglas SM, Marblestone AH, Teerapittayanon S, Vazquez A, Church GM, Shih WM. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res* **37**, 5001-5006 (2009).
33. Kim DN, Kilchherr F, Dietz H, Bathe M. Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures. *Nucleic Acids Res* **40**, 2862-2868 (2012).
34. Sobczak JPJ, Martin TG, Gerling T, Dietz H. Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature. *Science* **338**, 1458-1461 (2012).
35. Pinheiro AV, Han DR, Shih WM, Yan H. Challenges and opportunities for structural DNA nanotechnology. *Nat Nanotechnol* **6**, 763-772 (2011).
36. Fu J, *et al.* Assembly of multienzyme complexes on DNA nanostructures. *Nat Protocols* **11**, 2243-2273 (2016).
37. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with Rosetta. *Proteins* **55**, 656-677 (2004).
38. Gonen S, DiMaio F, Gonen T, Baker D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365-1368 (2015).
39. King NP, *et al.* Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **336**, 1171-1174 (2012).
40. Adleman LM. Molecular Computation of Solutions to Combinatorial Problems. *Science* **266**, 1021-1024 (1994).

41. Adleman LM. Computing with DNA. *Sci Am* **279**, 54-61 (1998).
42. Qian L, Winfree E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332**, 1196-1201 (2011).
43. Yurke B, Turberfield AJ, Mills AP, Simmel FC, Neumann JL. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605-608 (2000).
44. Zhang DY, Seelig G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nat Chem* **3**, 103-113 (2011).
45. Zhang DY, Winfree E. Control of DNA strand displacement kinetics using toehold exchange. *J Am Chem Soc* **131**, 17303-17314 (2009).
46. Zadeh JN, *et al.* NUPACK: Analysis and Design of Nucleic Acid Systems. *J Comput Chem* **32**, 170-173 (2011).
47. Doherty EA, Doudna JA. Ribozyme structures and mechanisms. *Annu Rev Biochem* **69**, 597-615 (2000).
48. Petsko GA. Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. *Nature* **401**, 115-116 (1999).
49. Noller HF. RNA structure: Reading the ribosome. *Science* **309**, 1508-1514 (2005).
50. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-2198 (1999).
51. Whitesides GM, Grzybowski B. Self-assembly at all scales. *Science* **295**, 2418-2421 (2002).
52. Whitesides GM, Boncheva M. Beyond molecules: Self-assembly of mesoscopic and macroscopic components. *P Natl Acad Sci USA* **99**, 4769-4774 (2002).
53. Zaher HS, Green R. Fidelity at the Molecular Level: Lessons from Protein Synthesis. *Cell* **136**, 746-762 (2009).
54. Herschlag D. RNA Chaperones and the RNA Folding Problem. *J Biol Chem* **270**, 20871-20874 (1995).

55. Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Contact order revisited: Influence of protein size on the folding rate. *Protein Sci* **12**, 2057-2062 (2003).
56. Alber F, *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695-701 (2007).
57. Burridge K, Fath K, Kelly T, Nuckolls G, Turner C. Focal Adhesions - Transmembrane Junctions between the Extracellular-Matrix and the Cytoskeleton. *Annu Rev Cell Biol* **4**, 487-525 (1988).
58. Cheeseman IM, Desai A. Molecular architecture of the kinetochore-microtubule interface. *Nat Rev Mol Cell Bio* **9**, 33-46 (2008).
59. Keseler IM, *et al.* EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Res* **39**, D583-D590 (2011).
60. Wilber AW, Doye JPK, Louis AA, Noya EG, Miller MA, Wong P. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. *J Chem Phys* **127**, (2007).
61. Li F, Josephson DP, Stein A. Colloidal Assembly: The Road from Particles to Colloidal Molecules and Crystals. *Angew Chem Int Edit* **50**, 360-388 (2011).
62. Das R, Baker D. Macromolecular modeling with Rosetta. *Annu Rev Biochem* **77**, 363-382 (2008).
63. Mandell DJ, Kortemme T. Computer-aided design of functional protein interactions. *Nat Chem Biol* **5**, 797-807 (2009).
64. Hormoz S, Brenner MP. Design principles for self-assembly with short-range interactions. *P Natl Acad Sci USA* **108**, 5193-5198 (2011).
65. Bray D, Lay S. Computer-based analysis of the binding steps in protein complex formation. *P Natl Acad Sci USA* **94**, 13493-13498 (1997).
66. Jankowski E, Glotzer SC. A comparison of new methods for generating energy-minimizing configurations of patchy particles. *J Chem Phys* **131**, (2009).
67. Rajendran A, Endo M, Katsuda Y, Hidaka K, Sugiyama H. Programmed Two-Dimensional Self-Assembly of Multiple DNA Origami Jigsaw Pieces. *Acs Nano* **5**, 665-671 (2011).

68. Ke YG, Ong LL, Shih WM, Yin P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **338**, 1177-1183 (2012).
69. Wei B, Dai MJ, Yin P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623-626 (2012).
70. Recht MI, Williamson JR. Central domain assembly: Thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J Mol Biol* **313**, 35-48 (2001).
71. Camacho CJ, Kimura SR, DeLisi C, Vajda S. Kinetics of desolvation-mediated protein-protein binding. *Biophys J* **78**, 1094-1105 (2000).
72. Evans CG, Hariadi RF, Winfree E. Direct Atomic Force Microscopy Observation of DNA Tile Crystal Growth at the Single-Molecule Level. *J Am Chem Soc* **134**, 10485-10492 (2012).
73. Wetmur JG. DNA Probes - Applications of the Principles of Nucleic-Acid Hybridization. *Crit Rev Biochem Mol* **26**, 227-259 (1991).
74. Woo S, Rothemund PWK. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nat Chem* **3**, 620-627 (2011).
75. Chakrabarty R, Mukherjee PS, Stang PJ. Supramolecular Coordination: Self-Assembly of Finite Two- and Three-Dimensional Ensembles. *Chem Rev* **111**, 6810-6918 (2011).
76. Perutz MF. Mechanisms of Cooperativity and Allosteric Regulation in Proteins. *Q Rev Biophys* **22**, 139-236 (1989).
77. Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA. Thermodynamic analysis of interacting nucleic acid strands. *Siam Rev* **49**, 65-88 (2007).
78. Grant J, Jack RL, Whitelam S. Analyzing mechanisms and microscopic reversibility of self-assembly. *J Chem Phys* **135**, (2011).
79. Sweeney B, Zhang T, Schwartz R. Exploring the parameter space of complex self-assembly through virus capsid models. *Biophys J* **94**, 772-783 (2008).
80. Villar G, *et al.* Self-Assembly and Evolution of Homomeric Protein Complexes. *Phys Rev Lett* **102**, (2009).

81. Deeds EJ, Bachman JA, Fontana W. Optimizing ring assembly reveals the strength of weak interactions. *P Natl Acad Sci USA* **109**, 2348-2353 (2012).
82. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the Folding Routes. *Science* **267**, 1619-1620 (1995).
83. Tsai CJ, Kumar S, Ma BY, Nussinov R. Folding funnels, binding funnels, and protein function. *Protein Sci* **8**, 1181-1190 (1999).
84. Hagan MF, Chandler D. Dynamic pathways for viral capsid assembly. *Biophys J* **91**, 42-54 (2006).
85. Reinhardt A, Frenkel D. Numerical evidence for nucleated self-assembly of DNA brick structures. *Phys Rev Lett* **112**, 238103 (2014).
86. Saiz L, Vilar JMG. Stochastic dynamics of macromolecular-assembly networks. *Mol Syst Biol* **2**, (2006).
87. Neumann MA, Leusen FJJ, Kendrick J. A major advance in crystal structure prediction. *Angew Chem Int Edit* **47**, 2427-2430 (2008).
88. Groschel AH, *et al.* Precise hierarchical self-assembly of multicompart ment micelles. *Nat Commun* **3**, (2012).
89. Marsh JA, *et al.* Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. *Cell* **153**, 461-470 (2013).
90. Kim AJ, Scarlett R, Biancaniello PL, Sinno T, Crocker JC. Probing interfacial equilibration in microsphere crystals formed by DNA-directed assembly. *Nat Mater* **8**, 52-55 (2009).
91. Whitelam S, Schulman R, Hedges L. Self-Assembly of Multicomponent Structures In and Out of Equilibrium. *Phys Rev Lett* **109**, (2012).
92. Gillespie DT. Exact Stochastic Simulation of Coupled Chemical-Reactions. *J Phys Chem-U S* **81**, 2340-2361 (1977).
93. Mulder AM, *et al.* Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30S Subunit. *Science* **330**, 673-677 (2010).

94. Korevaar PA, *et al.* Pathway complexity in supramolecular polymerization. *Nature* **481**, 492-496 (2012).
95. Powers ET, Powers DL. Mechanisms of protein fibril formation: Nucleated polymerization with competing off-pathway aggregation. *Biophys J* **94**, 379-391 (2008).
96. Pan J, Thirumalai D, Woodson SA. Folding of RNA involves parallel pathways. *J Mol Biol* **273**, 7-13 (1997).
97. Knowles TPJ, Oppenheim TW, Buell AK, Chirgadze DY, Welland ME. Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nat Nanotechnol* **5**, 204-207 (2010).
98. Gandra N, Abbas A, Tian LM, Singamaneni S. Plasmonic Planet-Satellite Analogues: Hierarchical Self-Assembly of Gold Nanostructures. *Nano Lett* **12**, 2645-2651 (2012).
99. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *P Natl Acad Sci USA* **95**, 1460-1465 (1998).
100. Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *Journal of medicinal chemistry* **45**, 2469-2483 (2002).
101. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. *Protein Sci* **1**, 169-181 (1992).
102. Pickett SD, Sternberg MJE. Empirical Scale of Side-Chain Conformational Entropy in Protein-Folding. *J Mol Biol* **231**, 825-839 (1993).
103. Marchi AN, Saaem I, Vogen BN, Brown S, LaBean TH. Toward larger DNA origami. *Nano Lett* **14**, 5740-5747 (2014).
104. Zhao Z, Liu Y, Yan H. Organizing DNA Origami Tiles into Larger Structures Using Preformed Scaffold Frames. *Nano Lett* **11**, 2997-3002 (2011).
105. Woo S, Rothmund PWK. Self-assembly of two-dimensional DNA origami lattices using cation-controlled surface diffusion. *Nat Commun* **5**, (2014).

106. Wei XX, Nangreave J, Liu Y. Uncovering the Self-Assembly of DNA Nanostructures by Thermodynamics and Kinetics. *Accounts Chem Res* **47**, 1861-1870 (2014).
107. Nangreave J, Yan H, Liu Y. Studies of Thermal Stability of Multivalent DNA Hybridization in a Nanostructured System. *Biophys J* **97**, 563-571 (2009).
108. Jiang SX, Yan H, Liu Y. Kinetics of DNA Tile Dimerization. *Acs Nano* **8**, 5826-5832 (2014).
109. Hariadi RF, Yurke B, Winfree E. Thermodynamics and kinetics of DNA nanotube polymerization from single-filament measurements. *Chem Sci* **6**, 2252-2267 (2015).
110. Han DR, *et al.* DNA Gridiron Nanostructures Based on Four-Arm Junctions. *Science* **339**, 1412-1415 (2013).
111. Bai XC, Martin TG, Scheres SHW, Dietz H. Cryo-EM structure of a 3D DNA-origami object. *P Natl Acad Sci USA* **109**, 20012-20017 (2012).
112. Seeman NC. DNA in a material world. *Nature* **421**, 427-431 (2003).
113. Castro CE, *et al.* A primer to scaffolded DNA origami. *Nat Methods* **8**, 221-229 (2011).
114. Li Z, Wang L, Yan H, Liu Y. Effect of DNA Hairpin Loops on the Twist of Planar DNA Origami Tiles. *Langmuir* **28**, 1959-1965 (2012).
115. Fern J, Lu J, Schulman R. The Energy Landscape for the Self-Assembly of a Two-Dimensional DNA Origami Complex. *Acs Nano* **10**, 1836-1844 (2016).
116. Wang RS, Kuzuya A, Liu WY, Seeman NC. Blunt-ended DNA stacking interactions in a 3-helix motif. *Chem Commun* **46**, 4905-4907 (2010).
117. Wang JC. Helical Repeat of DNA in Solution. *P Natl Acad Sci USA* **76**, 200-203 (1979).
118. Rothmund PWK, Papadakis N, Winfree E. Algorithmic self-assembly of DNA Sierpinski triangles. *Plos Biol* **2**, 2041-2053 (2004).
119. You Y, Tataurov AV, Owczarzy R. Measuring Thermodynamic Details of DNA Hybridization Using Fluorescence. *Biopolymers* **95**, 472-486 (2011).

120. Rice SA, Doty P. The Thermal Denaturation of Desoxyribose Nucleic Acid. *J Am Chem Soc* **79**, 3937-3947 (1957).
121. Dreyfus R, *et al.* Simple Quantitative Model for the Reversible Association of DNA Coated Colloids. *Phys Rev Lett* **102**, (2009).
122. Williams AP, Longfellow CE, Freier SM, Kierzek R, Turner DH. Laser Temperature-Jump, Spectroscopic, and Thermodynamic Study of Salt Effects on Duplex Formation by Dgatgc. *Biochemistry-U.S.* **28**, 4283-4291 (1989).
123. Porschke D, Uhlenbeck OC, Martin FH. Thermodynamics and Kinetics of Helix-Coil Transition of Oligomers Containing Gc Base Pairs. *Biopolymers* **12**, 1313-1335 (1973).
124. Nelson JW, Tinoco I. Comparison of the Kinetics of Ribooligonucleotide, Deoxyribooligonucleotide, and Hybrid Oligonucleotide Double-Strand Formation by Temperature-Jump Kinetics. *Biochemistry-U.S.* **21**, 5289-5295 (1982).
125. Gao Y, Wolf LK, Georgiadis RM. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res* **34**, 3370-3377 (2006).
126. Morrison LE, Stols LM. Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution. *Biochemistry-U.S.* **32**, 3095-3104 (1993).
127. Logan SR. The Origin and Status of the Arrhenius Equation. *J Chem Educ* **59**, 279-281 (1982).
128. Craig ME, Crothers DM, Doty P. Relaxation Kinetics of Dimer Formation by Self Complementary Oligonucleotides. *J Mol Biol* **62**, 383-& (1971).
129. Chen CL, Wang WJ, Wang Z, Wei F, Zhao XS. Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. *Nucleic Acids Res* **35**, 2875-2884 (2007).
130. Ouldrige TE, Sulc P, Romano F, Doye JPK, Louis AA. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res* **41**, 8886-8895 (2013).

131. Rothemund PWK, Ekani-Nkodo A, Papadakis N, Kumar A, Fygenon DK, Winfree E. Design and characterization of programmable DNA nanotubes. *J Am Chem Soc* **126**, 16344-16352 (2004).
132. Zenk J, Schulman R. An Assembly Funnel Makes Biomolecular Complex Assembly Efficient. *Plos One* **9**, (2014).
133. Liu WY, Zhong H, Wang RS, Seeman NC. Crystalline Two-Dimensional DNA-Origami Arrays. *Angew Chem Int Edit* **50**, 264-267 (2011).
134. Levy ED, Erba EB, Robinson CV, Teichmann SA. Assembly reflects evolution of protein complexes. *Nature* **453**, 1262-U1266 (2008).
135. Hansma HG, Laney DE. DNA binding to mica correlates with cationic radius: Assay by atomic force microscopy. *Biophys J* **70**, 1933-1939 (1996).
136. Suntharalingam M, Wentz SR. Peering through the pore: Nuclear pore complex structure, assembly, and function. *Dev Cell* **4**, 775-789 (2003).
137. Talkington MWT, Siuzdak G, Williamson JR. An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628-632 (2005).
138. Synowsky SA, van den Heuvel RHH, Mohammed S, Pijnappel WWMP, Heck AJR. Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol Cell Proteomics* **5**, 1581-1592 (2006).
139. Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV. Structural organization of the 19S proteasome lid: Insights from MS of intact complexes. *Plos Biol* **4**, 1314-1323 (2006).
140. DePristo MA, Weinreich DM, Hartl DL. Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat Rev Genet* **6**, 678-687 (2005).
141. Marsh JA, Teichmann SA. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annual Review of Biochemistry, Vol 84* **84**, 551-575 (2015).
142. Endo M, Sugita T, Katsuda Y, Hidaka K, Sugiyama H. Programmed-Assembly System Using DNA Jigsaw Pieces. *Chem-Eur J* **16**, 5362-5368 (2010).

143. Jacobs WM, Reinhardt A, Frenkel D. Rational design of self-assembly pathways for complex multicomponent structures. *P Natl Acad Sci USA* **112**, 6313-6318 (2015).
144. Zenk J, Tuntivate C, Schulman R. Kinetics and Thermodynamics of Watson–Crick Base Pairing Driven DNA Origami Dimerization. *J Am Chem Soc* **138**, 3346-3354 (2016).
145. Dunn KE, Dannenberg F, Ouldrige TE, Kwiatkowska M, Turberfield AJ, Bath J. Guiding the folding pathway of DNA origami. *Nature* **525**, 82-+ (2015).
146. Marras AE, Zhou L, Koliopoulos V, Su HJ, Castro CE. Directing folding pathways for multi-component DNA origami nanostructures with complex topology. *New J Phys* **18**, 055005 (2016).
147. Hedges LO, Mannige RV, Whitelam S. Growth of equilibrium structures built from a large number of distinct component types. *Soft Matter* **10**, 6404-6416 (2014).
148. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340-2361 (1977).
149. Murugan A, Zou J, Brenner MP. Undesired usage and the robust self-assembly of heterogeneous structures. *Nat Commun* **6**, (2015).
150. Williamson JR. Cooperativity in macromolecular assembly. *Nat Chem Biol* **4**, 458-465 (2008).
151. Akam M. The Molecular-Basis for Metameric Pattern in the Drosophila Embryo. *Development* **101**, 1-& (1987).
152. Green JBA, Sharpe J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* **142**, 1203-1211 (2015).
153. Bastiaens P, Caudron M, Niethammer P, Karsenti E. Gradients in the self-organization of the mitotic spindle. *Trends Cell Biol* **16**, 125-134 (2006).
154. Kholodenko BN. Spatially distributed cell signalling. *Febs Lett* **583**, 4006-4012 (2009).
155. Jeon NL, Baskaran H, Dertinger SKW, Whitesides GM, Van de Water L, Toner M. Neutrophil chemotaxis in linear and complex gradients of interleukin-8 formed in a microfabricated device. *Nat Biotechnol* **20**, 826-830 (2002).

156. Jeong GS, *et al.* Sprouting Angiogenesis under a Chemical Gradient Regulated by Interactions with an Endothelial Monolayer in a Microfluidic Platform. *Anal Chem* **83**, 8454-8459 (2011).
157. Chung BG, *et al.* Human neural stem cell growth and differentiation in a gradient-generating microfluidic device. *Lab Chip* **5**, 401-406 (2005).
158. Dertinger SKW, Jiang XY, Li ZY, Murthy VN, Whitesides GM. Gradients of substrate-bound laminin orient axonal specification of neurons. *P Natl Acad Sci USA* **99**, 12542-12547 (2002).
159. Shin Y, *et al.* Microfluidic assay for simultaneous culture of multiple cell types on surfaces or within hydrogels. *Nat Protoc* **7**, 1247-1259 (2012).
160. DeForest CA, Polizzotti BD, Anseth KS. Sequential click reactions for synthesizing and patterning three-dimensional cell microenvironments. *Nat Mater* **8**, 659-664 (2009).
161. Karzbrun E, Tayar AM, Noireaux V, Bar-Ziv RH. Programmable on-chip DNA compartments as artificial cells. *Science* **345**, 829-832 (2014).
162. He JK, Du YA, Villa-Urbe JL, Hwang CM, Li DC, Khademhosseini A. Rapid Generation of Biologically Relevant Hydrogels Containing Long-Range Chemical Gradients. *Adv Funct Mater* **20**, 131-137 (2010).
163. Sourjik V, Wingreen NS. Responding to chemical gradients: bacterial chemotaxis. *Curr Opin Cell Biol* **24**, 262-268 (2012).
164. Zarzar LD, Liu QH, He XM, Hu YH, Suo ZG, Aizenberg J. Multifunctional actuation systems responding to chemical gradients. *Soft Matter* **8**, 8289-8293 (2012).
165. Mayer M, Yang J, Gitlin I, Gracias DH, Whitesides GM. Micropatterned agarose gels for stamping arrays of proteins and gradients of proteins. *Proteomics* **4**, 2366-2376 (2004).
166. Lee SH, Moon JJ, West JL. Three-dimensional micropatterning of bioactive hydrogels via two-photon laser scanning photolithography for guided 3D cell migration. *Biomaterials* **29**, 2962-2968 (2008).
167. Hahn MS, Miller JS, West JL. Three-dimensional biochemical and biomechanical patterning of hydrogels for guiding cell behavior. *Adv Mater* **18**, 2679-+ (2006).

168. Dertinger SKW, Chiu DT, Jeon NL, Whitesides GM. Generation of gradients having complex shapes using microfluidic networks. *Anal Chem* **73**, 1240-1246 (2001).
169. Atencia J, Morrow J, Locascio LE. The microfluidic palette: A diffusive gradient generator with spatio-temporal control. *Lab Chip* **9**, 2707-2714 (2009).
170. Lin F, Saadi W, Rhee SW, Wang SJ, Mittal S, Jeon NL. Generation of dynamic temporal and spatial concentration gradients using microfluidic devices. *Lab Chip* **4**, 164-167 (2004).
171. Diao JP, *et al.* A three-channel microfluidic device for generating static linear gradients and its application to the quantitative analysis of bacterial chemotaxis. *Lab Chip* **6**, 381-388 (2006).
172. Vanag VK, Epstein IR. Design and control of patterns in reaction-diffusion systems. *Chaos* **18**, (2008).
173. Sagues F, Epstein IR. Nonlinear chemical dynamics. *Dalton T*, 1201-1217 (2003).
174. Semenov SN, Markvoort AJ, Gevers WBL, Piruska A, de Greef TFA, Huck WTS. Ultrasensitivity by Molecular Titration in Spatially Propagating Enzymatic Reactions. *Biophys J* **105**, 1057-1066 (2013).
175. Loose M, Fischer-Friedrich E, Ries J, Kruse K, Schwille P. Spatial regulators for bacterial cell division self-organize into surface waves in vitro. *Science* **320**, 789-792 (2008).
176. Miguez DG, Vanag VK, Epstein IR. Fronts and pulses in an enzymatic reaction catalyzed by glucose oxidase. *P Natl Acad Sci USA* **104**, 6992-6997 (2007).
177. Jee E, Bansagi T, Taylor AF, Pojman JA. Temporal Control of Gelation and Polymerization Fronts Driven by an Autocatalytic Enzyme Reaction. *Angew Chem Int Edit* **55**, 2127-2131 (2016).
178. Wrobel MM, *et al.* pH Wave-Front Propagation in the Urea-Urease Reaction (vol 103, pg 610, 2012). *Biophys J* **106**, 1548-1548 (2014).
179. Semenov SN, Markvoort AJ, de Greef TFA, Huck WTS. Threshold Sensing through a Synthetic Enzymatic Reaction-Diffusion Network. *Angew Chem Int Edit* **53**, 8066-8069 (2014).

180. Isalan M, Lemerle C, Serrano L. Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *Plos Biol* **3**, 488-496 (2005).
181. Padirac A, Fujii T, Estevez-Torres A, Rondelez Y. Spatial Waves in Synthetic Biochemical Networks. *J Am Chem Soc* **135**, 14586-14592 (2013).
182. Zadorin AS, Rondelez Y, Galas JC, Estevez-Torres A. Synthesis of Programmable Reaction-Diffusion Fronts Using DNA Catalyzers. *Phys Rev Lett* **114**, (2015).
183. van Roekel HWH, *et al.* Programmable chemical reaction networks: emulating regulatory functions in living cells using a bottom-up approach. *Chem Soc Rev* **44**, 7465-7483 (2015).
184. Seelig G, Soloveichik D, Zhang DY, Winfree E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585-1588 (2006).
185. Zhang DY, Turberfield AJ, Yurke B, Winfree E. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318**, 1121-1125 (2007).
186. Chen YJ, *et al.* Programmable chemical controllers made from DNA. *Nat Nanotechnol* **8**, 755-762 (2013).
187. Soloveichik D, Seelig G, Winfree E. DNA as a universal substrate for chemical kinetics. *Proc Natl Acad Sci U S A* **107**, 5393-5398 (2010).
188. Qian L, Winfree E. A simple DNA gate motif for synthesizing large-scale circuits. *J R Soc Interface* **8**, 1281-1297 (2011).
189. Duose DY, Schweller RM, Zimak J, Rogers AR, Hittelman WN, Diehl MR. Configuring robust DNA strand displacement reactions for in situ molecular analyses. *Nucleic Acids Res* **40**, 3289-3298 (2012).
190. Allen PB, Chen X, Ellington AD. Spatial Control of DNA Reaction Networks by DNA Sequence. *Molecules* **17**, 13390-13402 (2012).
191. Chirieleison SM, Allen PB, Simpson ZB, Ellington AD, Chen X. Pattern transformation with DNA circuits. *Nat Chem* **5**, 1000-1005 (2013).
192. Pernodet N, Maaloum M, Tinland B. Pore size of agarose gels by atomic force microscopy. *Electrophoresis* **18**, 55-58 (1997).

193. Doose S, Barsch H, Sauer M. Polymer properties of polythymine as revealed by translational diffusion. *Biophys J* **93**, 1224-1234 (2007).
194. Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS. Size-dependent DNA mobility in cytoplasm and nucleus. *J Biol Chem* **275**, 1625-1629 (2000).
195. Stellwagen E, Lu YJ, Stellwagen NC. Unified description of electrophoresis and diffusion for DNA and other polyions. *Biochemistry-Us* **42**, 11745-11750 (2003).
196. Yin P, Choi HMT, Calvert CR, Pierce NA. Programming biomolecular self-assembly pathways. *Nature* **451**, 318-U314 (2008).
197. Fern J, Scalise D, Cangialosi A, Howie D, Potters L, Schulman R. DNA Strand-Displacement Timer Circuits. *ACS Synthetic Biology*, (2016).
198. Leroy JL, Kochoyan M, Huynhdinh T, Gueron M. Characterization of Base-Pair Opening in Deoxynucleotide Duplexes Using Catalyzed Exchange of the Imino Proton. *J Mol Biol* **200**, 223-238 (1988).
199. Nonin S, Leroy JL, Gueron M. Terminal Base-Pairs of Oligodeoxynucleotides - Imino Proton-Exchange and Fraying. *Biochemistry-Us* **34**, 10652-10659 (1995).
200. Hecker KH, Rill RL. Error analysis of chemically synthesized polynucleotides. *Biotechniques* **24**, 256-260 (1998).
201. Scalise D, Schulman R. Designing modular reaction-diffusion programs for complex pattern formation. *TECHNOLOGY* **02**, 55-66 (2014).
202. Scalise D, Schulman R. Emulating cellular automata in chemical reaction-diffusion networks. *Nat Comput* **15**, 197-214 (2016).
203. Schaeffer JM, Thachuk C, Winfree E. Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands. In: *DNA Computing and Molecular Programming: 21st International Conference, DNA 21, Boston and Cambridge, MA, USA, August 17-21, 2015. Proceedings* (ed[^](eds Phillips A, Yin P). Springer International Publishing (2015).
204. Grun C, Werfel J, Zhang DY, Yin P. DyNAMiC Workbench: an integrated development environment for dynamic DNA nanotechnology. *J R Soc Interface* **12**, (2015).

205. Olson X, *et al.* Availability: A Metric for Nucleic Acid Strand Displacement Systems. *ACS Synth Biol*, (2016).
206. Thachuk C, Winfree E, Soloveichik D. Leakless DNA Strand Displacement Systems. In: *DNA Computing and Molecular Programming: 21st International Conference, DNA 21, Boston and Cambridge, MA, USA, August 17-21, 2015. Proceedings.* Springer International Publishing (2015).
207. You MX, Zhu GZ, Chen T, Donovan MJ, Tan WH. Programmable and Multiparameter DNA-Based Logic Platform For Cancer Recognition and Targeted Therapy. *J Am Chem Soc* **137**, 667-674 (2015).
208. Epstein IR, Xu B. Reaction-diffusion processes at the nano- and microscales. *Nat Nano* **11**, 312-319 (2016).
209. Franco E, *et al.* Timing molecular motion and production with a synthetic transcriptional clock. *P Natl Acad Sci USA* **108**, E784-E793 (2011).
210. Wu JD, Mao ZW, Tan HP, Han LL, Ren TC, Gao CY. Gradient biomaterials and their influences on cell migration. *Interface Focus* **2**, 337-355 (2012).
211. Puglisi JD, Tinoco I, Jr. Absorbance melting curves of RNA. *Methods Enzymol* **180**, 304-325 (1989).
212. Tikhomirov G, Petersen P, Qian L. Programmable disorder in random DNA tilings. *Nat Nanotechnol*, (2016).
213. Chen JH, Seeman NC. Synthesis from DNA of a Molecule with the Connectivity of a Cube. *Nature* **350**, 631-633 (1991).

CURRICULUM VITAE

John Zenk was born in Littleton, Colorado on October 18, 1989 to Elizabeth and Robert Zenk. He attended Littleton High School where he studied in the International Baccalaureate Program and graduated in May 2008. He then attended the University of Colorado, Boulder where he received a B.S. degree in Chemical and Biological Engineering in May 2012. While obtaining his B.S., John spent two summers as an intern doing part procurement and quality assurance work at SEAKR Engineering, an aerospace firm in the Denver area. For about 18 months while obtaining his undergraduate degree, John researched granular fluid flows using computational fluid dynamics in the lab of Dr. Christine M. Hrenya. In the summer after receiving a B.S., he worked as an intern with the Infectious Disease group at J. Craig Venter Institute, where he built a metabolic model of an infectious, and potentially lethal to humans, fungus. In September 2012, John entered graduate school at Johns Hopkins University in the Department of Chemical and Biomolecular Engineering (ChemBE) and started work in the laboratory of Dr. Rebecca Schulman. At Hopkins, John had the good fortune to serve as a teaching assistant for two semesters in the core “Modeling and Statistics” course and co-teach a one credit engineering elective called “Matlab Made Easy” to freshman and sophomore ChemBE undergraduates.