Synthetic Molecular Machines for Active Self-Assembly: Prototype Algorithms, Designs, and Experimental Study

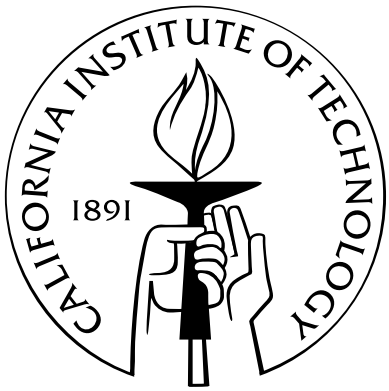
Thesis by

Nadine L. Dabby

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2013

(Defended February 4, 2013)

This manuscript is dedicated to

Yvette and Joseph Dabby (mom and dad): Of all the brilliant minds and beautiful souls that I have stumbled upon in this great wide world, there are no two people whom I admire more, or aspire more to be like. It is hard for me to imagine ever being as strong as you, or having as much integrity, or so much to give.

*

To Peter C. Beller: you are my eyes, my soul and my home.

*

And to my thesis advisor Richard Murray: Thank you for believing in me.

Acknowledgements

Sio credesse che mia risposta fosse
A persona che mai tornasse al mondo,
Questa fiamma staria senza piu scosse.
Ma perciocche giammai di questo fondo
Non torno vivo alcun, siodo il vero,
Senza tema dinfamia ti rispondo.

"The Inferno," Dante Alighieri [Alighieri, 2013]

It is fitting to invoke my first intellectual crush here. [1]

A short while after starting at Caltech in my first year of graduate school, I went to the Millikan library to photocopy an article that had been published in 1977 and had not yet been taken up by the digitization movement [Gillespie, 1977]. I remember looking out of the windows on the 8th floor toward the mountains in the north and thinking about how lucky I was to be getting paid to learn and ask questions and think. There was a peace that came with the promise of objectivity.

I came to this place via quite the circuitous path, asking some questions along the way [Dabby, 2003]. And there were many people who encouraged me onward (specifically **Kevis Goodman** and **Lynn Hejinian**, both faculty members of the amazing UC Berkeley English Department). Back then (not unlike now), it was not clear how I should proceed. I had many questions and I did not even know how to think about them in a productive way. I somehow had not managed to

[1] The epigraph to The Love Song of J. Alfred Prufrock [Eliot, 1995]. The passage translates to: "If I thought my answer were given to anyone who would ever return to the world, this flame would stand still without moving any further. But since never from this abyss has anyone ever returned alive, if what I hear is true, without fear of infamy I answer you."

jam in everything I needed to learn in my undergraduate career [2]. So I followed my instincts and I attempted to catch up on math and learn the thought process behind computer science.

Shuki Bruck is the reason I applied to the Computation and Neural Systems Ph.D. program at Caltech. I will never be sure of what compelled the admissions committee to accept me [3]. At Caltech I was pushed to think, to work hard and to push the limits of my creativity. It was tough: truth be told, brutal at times, and at every pitfall I found some new strength within myself that I did not know I had in me. That toughness has (somehow magically over time) given me a profound sense of confidence that I consider priceless.

This dissertation would absolutely not have been possible without the support and help of many people who believed in me.

The most important of these is **Ho-Lin Chen**, my mentor, colleague and friend. You have one of the most brilliant minds I have ever encountered and yet you are so humble about it.

My thesis committee has been a source of great strength and insight.

To **Richard Murray**: Thank you for giving me the most important gift of all – a room of my own [Woolf, 1957]. A key can be a magical thing: it opens doors and eyes and thoughts and minds, and in all those things: ideas, dreams, imaginations and opportunities. A refuge, a sacred place for quiet thoughts and a home for conversations. It can awaken a spirit and a soul simply by opening up the possibility of thinking freely. In your lab I have had the most important epiphany of all. You probably think that this was a small thing for you to do, but it has made all the difference in the world to me.

To **Niles Pierce**: Thank you for doing beautiful work. Thank you for opening up your lab to me and for guiding me through my experiments.

To **Al Barr**: I am very lucky that I first met you at the CNS interview weekend. I have really appreciated all of your thoughtful advice and feedback and the I have enjoyed working with you over these last few months. Thank you for your time and your patience.

To **Milan Stojanovic**, in a parallel universe you would have been by PhD. advisor and I would have been ecstatic about it. Thank you for being my would-be advisor. I have valued every con-

[2] While my formal education is coming to an end, I imagine that there is yet so much more to learn.

[3] Although I do have it on good authority that I was "an experiment"...

versation and every piece of wisdom you have tried to impart (even if I ignored some of it).

To **Deborah Fygenson**: Thank you for mentoring me. Your encouragement and advice have been invaluable to me.

To **Shin Shimojo**: Thank you for listening and for reminding me that I am not the only alien here.

I don't think graduation would even have been a possibility without a few guardian angels that have been watching out for me–**Paul W. K. Rothemund** and **Mathieu Desbrun**.

Thank you to the Murray Lab Biocircuits Subgroup (**Emzo de los Santos, Marcella Gomez, Clare Hayes, Victoria Hsiao, Vanessa Jonsson, Jongmin Kim, Joseph Meyerowitz, Dan Siegal-Gaskins, Vipul Singhal, Zachary Sun, Anandh Swaminathan, Anu Thubagere, Zoltan Tuza, Ophelia Venturelli**, and **Enoch Yeung**), an amazing group of people that have made my last few months at Caltech fun. Special thanks to Anu and Jongmin for patience with my Spex usage, to Zach and Clare for their friendship and to Jongmin, Zoltan and Dan for last minute o.d.e. aid.

Thank you to the Pierce Lab, another amazing group of people. In particular, I received a lot of help and support from **Victor Beck, Colby Calvert, Harry Choi, Lisa Hochrein, Jennifer Padilla, Maayan Schwarzkopf, Jonathan Sternberg**, and **Jeff Vieregg**.

Also thank you for support and help to **Tanya Owen, Karolyn Yong, Lucinda Acosta** and **Anissa Scott**.

Thank you to my friends, my mentors and colleagues who I have had the great pleasure of learning from during my time at Caltech: **Catherine Beni, Elisa Franco, Ashwin Gopinath, Greg Griffin, Rizal Hariadi, Ruxandra Paun, Joseph Schaeffer, Rebecca Schulman, David Soloveichik, Merrielle Spain, Sungwook Woo**, and **David Zhang**.

A special thank you to friends **Takeshi Irie** and **Nilesh Shah** for saying just the right things at the right time.

During my time at Caltech I had the great privilege of teaching at the Art Center College of Design. I have **Ann Marie Polsenberg Thomas** to thank for the opportunity. And I thank my colleagues and students there for teaching me, with particular thanks to **Julia Hur** and **Bruce Hubbard**.

To my parents, **Yvette** and **Joseph Dabby**: Life is about luck and then about choice. I feel so

lucky to be born into your world, to have learned even just a fraction of what you have to teach. And it wasn't so much anything you ever said, but more just watching the example that you set. Kind and strong, and ballsy enough to stand behind your name, your reputation and your beliefs even in difficult circumstances. You showed me the importance of family, love, community, and that it is only through this love that one can change the world, even if the part we are changing is only our small corner of the world.

To my sisters **Sharon Naomi Dabby** and **Lisa Dabby Joyce**: You were my first mentors, and you are now my closest friends.

To my husband **Peter C. Beller**: You are my soulmate and my partner in every possible way. Thank you for sitting with me in lab on Saturdays and for feeding my soul.

Thank you to my collaborators: **Ho-Lin Chen, Scott Goodfriend, Alexander Johnson-Buck, Kyle Lund, Anthony J. Manzo, Nicole Michelotti, Jeanette Nangreave, Renjun Pei, Joseph Schaeffer, Milan N. Stojanovic, Steven Taylor, Nils G. Walter, Erik Winfree, Damien Woods, Hao Yan, Peng Yin**, and **Bernard Yurke**.

This thesis was supported by generous funding from the NSF (Graduate Research Fellowship, Molecular Programming Project and other grants), and NIH (training grants).

Abstract

Computer science and electrical engineering have been the great success story of the twentieth century. The neat modularity and mapping of a language onto circuits has led to robots on Mars, desktop computers and smartphones. But these devices are not yet able to do some of the things that life takes for granted: repair a scratch, reproduce, regenerate, or grow exponentially fast–all while remaining functional.

This thesis explores and develops algorithms, molecular implementations, and theoretical proofs in the context of "active self-assembly" of molecular systems. The long-term vision of active self-assembly is the theoretical and physical implementation of materials that are composed of reconfigurable units with the programmability and adaptability of biology's numerous molecular machines. En route to this goal, we must first find a way to overcome the memory limitations of molecular systems, and to discover the limits of complexity that can be achieved with individual molecules.

One of the main thrusts in molecular programming is to use computer science as a tool for figuring out what can be achieved. While molecular systems that are Turing-complete have been demonstrated [Winfree, 1996], these systems still cannot achieve some of the feats biology has achieved.

One might think that because a system is Turing-complete, capable of computing "anything," that it can do any arbitrary task. But while it can simulate any digital computational problem, there are many behaviors that are not "computations" in a classical sense, and cannot be directly implemented. Examples include exponential growth and molecular motion relative to a surface.

Passive self-assembly systems cannot implement these behaviors because (a) molecular motion relative to a surface requires a source of fuel that is external to the system, and (b) passive systems

are too slow to assemble exponentially-fast-growing structures. We call these behaviors "energetically incomplete" programmable behaviors. This class of behaviors includes any behavior where a passive physical system simply does not have enough physical energy to perform the specified tasks in the requisite amount of time.

As we will demonstrate and prove, a sufficiently expressive implementation of an "active" molecular self-assembly approach can achieve these behaviors. Using an external source of fuel solves part of the the problem, so the system is not "energetically incomplete." But the programmable system also needs to have sufficient expressive power to achieve the specified behaviors. Perhaps surprisingly, some of these systems do not even require Turing completeness to be sufficiently expressive.

Building on a large variety of work by other scientists in the fields of DNA nanotechnology, chemistry and reconfigurable robotics, this thesis introduces several research contributions in the context of active self-assembly.

We show that simple primitives such as insertion and deletion are able to generate complex and interesting results such as the growth of a linear polymer in logarithmic time and the ability of a linear polymer to treadmill. To this end we developed a formal model for active-self assembly that is directly implementable with DNA molecules. We show that this model is computationally equivalent to a machine capable of producing strings that are stronger than regular languages and, at most, as strong as context-free grammars. This is a great advance in the theory of active self-assembly as prior models were either entirely theoretical or only implementable in the context of macro-scale robotics.

We developed a chain reaction method for the autonomous exponential growth of a linear DNA polymer. Our method is based on the insertion of molecules into the assembly, which generates two new insertion sites for every initial one employed. The building of a line in logarithmic time is a first step toward building a shape in logarithmic time. We demonstrate the first construction of a synthetic linear polymer that grows exponentially fast via insertion. We show that monomer molecules are converted into the polymer in logarithmic time via spectrofluorimetry and gel electrophoresis experiments. We also demonstrate the division of these polymers via the addition of a single DNA complex that competes with the insertion mechanism. This shows the growth of a

population of polymers in logarithmic time. We characterize the DNA insertion mechanism that we utilize in Chapter 4. We experimentally demonstrate that we can control the kinetics of this reaction over at least seven orders of magnitude, by programming the sequences of DNA that initiate the reaction.

In addition, we review co-authored work on programming molecular robots using prescriptive landscapes of DNA origami; this was the first microscopic demonstration of programming a molecular robot to walk on a 2-dimensional surface. We developed a snapshot method for imaging these random walking molecular robots and a CAPTCHA-like analysis method for difficult-to-interpret imaging data.

Contents

Acknowledgements **iv**

Abstract **viii**

1 Introduction **1**

1.1 From Information to Activation 1

1.2 Review of Fundamental Components 4

1.2.1 Three-way versus Four-way Branch Migration 4

1.2.2 Hybridization Chain Reaction 5

1.2.3 Other Four-way Branch Machines 7

1.2.4 Programming Biomolecular Pathways 8

1.3 Related Work 8

1.3.1 Theory of Active Self-Assembly 8

1.4 Summary of Thesis 10

1.4.1 Programming Molecular Robots Using Prescriptive Landscapes 10

1.4.2 A Model for Active Self-Assembly in DNA Systems 10

1.4.3 A Synthetic Polymer that Grows Exponentially Fast 10

1.4.4 The Kinetics of Toehold-Mediated Four-way Branch Migration 11

2 Molecular Robots Guided by Prescriptive Landscapes **12**

3 A Computational Model for Active Self-Assembly in DNA Systems **22**

3.1 Abstract 22

3.2 Introduction . . . 23
3.2.1 Motivation . . . 26
3.2.2 Contextual Background . . . 28
3.2.3 Review of Self-Assembly Models . . . 29
3.3 Model . . . 31
3.3.1 Formal Model Description . . . 31
3.3.2 A Molecular Implementation . . . 33
3.4 Proofs of the Model's Expressive Power . . . 36
3.4.1 Analyzing the Theoretical Growth Speed of Polymers . . . 37
3.5 Conclusions . . . 41

4 Toward a Synthetic Polymer that Grows Exponentially Fast 42
4.1 Abstract . . . 42
4.2 Introduction . . . 43
4.3 Molecular Implementation . . . 46
4.4 Exponential Growth Results . . . 48
4.4.1 Exponential Growth Mechanism Controls . . . 48
4.4.2 The Kinetics of Parallel Insertion. . . . 50
4.4.3 Imaging with Atomic Force Microscopy . . . 56
4.4.4 Time Lapse Experiments . . . 57
4.5 Analysis . . . 60
4.5.1 Parameter Fitting . . . 64
4.6 Methods to Generate Other Behaviors . . . 65
4.6.1 Division . . . 65
4.6.2 Treadmilling . . . 66
4.7 Discussion . . . 66
4.8 Materials and Methods . . . 71

5 The Kinetics of Toehold-Mediated Four-way Branch Migration 75

5.1 Abstract . . . 75

5.2 Introduction . . . 76

5.3 System Description . . . 78

5.4 Experimental Data . . . 79

5.5 Energy Landscapes and Elementary Step Simulations . . . 86

5.6 Mechanistic Model . . . 90

5.6.1 Closed Model . . . 93

5.6.2 Open Model . . . 95

5.6.3 Results . . . 96

5.7 Discussion and Conclusions . . . 99

5.8 Materials and Methods . . . 101

6 Conclusion and Future Directions 106

A Partial Supplementary Material for Molecular Robots Guided by Prescriptive Landscapes 109

A.1 Supporting Discussion . . . 109

A.2 Materials and Methods . . . 112

A.2.1 Abbreviations . . . 112

A.2.2 Preparation of Spiders . . . 112

A.2.3 Surface Plasmon Resonance (SPR) . . . 113

A.2.4 Preparation of Spider-Origami Arrays . . . 115

A.2.5 Atomic Force Microscopy . . . 117

A.2.6 DNA Sequences . . . 120

B Supplementary Material for A Synthetic Polymer that Grows Exponentially Fast 132

B.1 Exponential Growth System Experiments . . . 133

B.1.1 DNA Sequences Final Version 6-3v1 . . . 133

B.1.2 Leakage in Various Designs . . . 134

B.1.3 Joining . . . 137
B.1.4 Kinetics . . . 138
B.2 Division Experiments . . . 144
B.2.1 Division Kinetics . . . 145

C Supplemental: The Kinetics of Toehold-Mediated Four-way Branch Migration 147
C.1 Experimental Details . . . 147
C.2 Modeling and Sample Code . . . 151
C.3 Experimental Data . . . 155
C.3.1 Completion Levels . . . 158
C.3.2 Displacement Strand Controls . . . 162
C.3.3 Reverse Experiments . . . 164
C.4 Reaction Coordinate . . . 167
C.5 Trajectory Simulations . . . 167

Bibliography 169

List of Figures

1.1 A schematic of three-way branch migration. 4
1.2 A schematic of four-way branch migration. 5
1.3 A schematic of the Hybridization Chain Reaction as described by [Dirks and Pierce, 2004]. 6
1.4 A schematic of a Rickettsia-like autonomous polymerization motor as described by [Venkataraman et al., 2007]. 7
1.5 An overview of a how a pattern is computed in the nubot model. 9

2.1 Schematics of molecular walker and origami prescriptive landscape 14
2.2 Spider movement along three tracks with schematics and AFM images 16
2.3 Schematics and AFM images of the spider moving along a track over time 17
2.4 Spiders imaged on origami tracks in real-time using super-resolution TIRF microscopy 19

3.1 The language hierarchy. 25
3.2 Growth of mass versus time in mammalian embryos. 27
3.3 Model abstraction 32
3.4 A system that implements insertional polymer growth in logarithmic time 34
3.5 A system that implements division in a polymer 35

4.1 Schematic of our insertional polymer implementation using DNA. 45
4.2 Legend of DNA hairpins and complexes. 47
4.3 Non-denaturing gel control experiments show if there are undesired interactions between reactants. 50
4.4 Spectrofluorimetry experimental design 51

4.5 Linear polymer growth kinetics examined via fluorescence. 53
4.6 Linear polymer growth kinetics are observed when inactivated Hairpin 3L is substituted for Hairpin 3 53
4.7 Exponential polymer growth kinetics examined via fluorescence. 54
4.8 Linear fit of the 10% completion time as a function of relative Initiator concentration. 54
4.9 The average length of the polymer shrinks with increase in Initiator concentration. . 55
4.10 Normalized binned distributions of lengths of polymers with different Initiator concentrations. 56
4.11 Atomic Force Microscopy images of the polymer and the leak product. 57
4.12 Gel time-lapse studies of linear polymer growth. 59
4.13 Gel time-lapse studies of exponential polymer growth. 60
4.14 The ordinary differential equation model of the exponential growth system. 63
4.15 A system that implements division. 66
4.16 Nondenaturing gel illustrates polymer division. 67
4.17 Division electrophoresis gel lane traces 68
4.18 Division electrophoresis gel binned data 69
4.19 Combining the insertion and division behaviors can result in treadmilling. 70

5.1 Fluorescence kinetics experimental design 80
5.2 Kinetic traces of the reactions between Reporter and Complexes 82
5.3 Plot of toehold length $(m + n)$ versus $\log(k_1^{fit})$ 85
5.4 Energy landscapes for toehold-initiated four-way branch migration 88
5.5 Summary of phenomenological and mechanistic models. 91
5.6 Plot of adjusted $\Delta G^{\circ}_{k_1}$ versus $\log_{10}(k_1)$ 98
5.7 A toehold-mediated four-way strand exchange mechanism 100

B.1 Combinatorial gel for 4-4v1 design. 134
B.2 Combinatorial gel for 5-4v2 design. 135
B.3 Combinatorial gel for 6-3v2 design. 136

B.4 Time lapse reaction of the experimental system with equivalent concentrations of Initiator and all hairpins tests joining . . . 137
B.5 SYBR Gold-stained linear and exponential final values gels. . . . 138
B.6 Gel time-lapse studies of exponential polymer growth. . . . 139
B.7 SYBR Gold-stained time lapse gels. . . . 140
B.8 Gel time-lapse studies of exponential polymer growth. . . . 141
B.9 SYBR Gold-stained time lapse gels. . . . 142
B.10 An additional gel time-lapse studies of polymer growth . . . 143
B.11 SYBR Gold-stained division gels. . . . 144
B.12 Exponential polymer division kinetics examined via fluorescence. . . . 145
B.13 Linear fit of the 10% completion time as a function of relative Initiator concentration. 145
B.14 Exponential polymer division kinetics examined via fluorescence. . . . 146
B.15 Linear fit of the 10% completion time as a function of relative Initiator concentration. 146

C.1 Long toehold traces for fitting k_2 . . . 155
C.2 Slow traces utilized a mean squared error linear fit to find k_1. . . . 156
C.3 Medium-speed traces utilized a mean squared error (fminunc) function to fit k_1. . . . 157
C.4 Mean squared error k_1 fit for the $(m = 4, n = 6)$ reaction . . . 158
C.5 Mean squared error k_1 fit for the $(m = 6, n = 2)$ reaction . . . 159
C.6 Mean squared error k_1 fit for the $(m = 6, n = 4)$ reaction . . . 160
C.7 Mean squared error k_1 fit for the $(m = 6, n = 6)$ reaction . . . 161
C.8 Displacement strand control experiments . . . 162
C.9 Experiment to measure reverse reaction rates . . . 164
C.10 Reversible four-way branch migration traces to fit the value of $k_r(0, 3)$. . . . 166
C.11 Reaction coordinate for the open loop case where $(m = 2, n = 2)$ . . . 167
C.12 Trajectory simulation rates compared to experimentally fit and calculated rates . . . 168

Chapter 1

Introduction

> science is the same as poetry
> only it uses the wrong words.
>
> Science, Robert Kelly (in May Day)

A sequence of DNA or a string of code is a line of poetry. These structures share a common trait: the ability of the individual units to transcend their symbolic nature. These seemingly unrelated fields share a profound emphasis upon sequences of information. This thesis is motivated by a deep philosophical question that underlies all of biology: how is it that a biological system derives all of its complexity from a simple sequence? While this is a very large question, we might approach it by trying to reason as to how information is transformed into an active entity in organic molecules.

1.1 From Information to Activation

A human's genome consists of approximately three billion base pairs [Venter et al., 2001]. This implies that the program our cells are running, which gives rise to our very complicated bodies and brains, utilizes less than two gigabytes of information[1]. There is a long history of great minds

[1]To calculate this number, first observe that each DNA base requires two bits to encode ($00, 01, 10, 11$ map to A, C, G, T). Thus, one byte is equivalent to a sequence of four bases. The genome consists of three billion base pairs or 6×10^9 bases of DNA. When we divide this number by four bases per byte, the result is 1.5×10^9 bytes or 1.5 gigabytes. Contrast this number to the amount of space one would need to store a low resolution movie – by today's standards that requires at least a few gigabytes of storage space.

having drawn comparisons between machines and man[2], but there are some notable differences.

The computer on which this report was written has about three hundred gigabytes of storage space, but it is not capable of doing many things that we people can do. For example, it cannot grow to five times its initial mass while operating at full capacity[3] and it cannot function if I impale the motherboard with a railroad spike[4]. A human's ability to function despite injury itself pales in comparison to the robustness of other members of the animal kingdom[5]. On the other hand, computers are capable of crunching out computations that are very difficult for a human to do.

Computers are traditionally suited to particular kinds of computation that differ from those encountered in biology. One difference between computer architectures and nature's architectures is the mode of information transfer. Whereas in computers, the information encoded in the program is fundamentally separate from the hardware used to execute it[6], in biology, software and hardware are the same thing. Of interest here are the implications of the unification of "form and function" of information in biology. Every day we encounter examples of phenomena that we are as of yet unable to reproduce in electronics, and we don't yet know the full capabilities of simple organic materials.

In 1953, Watson and Crick solved the structure of DNA and proposed that the double helix suggests a replication mechanism for genetics [Watson and Crick, 1953]. The nucleic acids DNA and RNA epitomize the blurring of form and function at the molecular scale: these molecules store information in their sequences, and they can perform enzymatic reactions such as ligation (attaching two strands to form a longer strand) and cleavage (cutting of one strand into two)[7].

DNA has been used as a material to build computing devices [Adleman, 1994, Winfree, 2000],

[2]In fact, Shannon and McCarthy allude to this in their introduction to *Automata Studies*: "Among the most challenging scientific questions of our time are the corresponding analytic and synthetic problems: How does the brain function? Can we design a machine which will simulate a brain? Speculation on these problems, which can be tracked back many centuries, usually reflects in any period the characteristic machines used" [Shannon et al., 1956].

[3]An average male grows from 13 kg at two years of age to 70 kg at twenty years of age [Kuczmarski et al., 2000].

[4]I refer to the case of Phineas Gage [Harlow, 1999], the railway worker who survived an accident in which a large iron rod was driven all the way through his frontal lobe. Humans are capable of carrying on (almost normally) after sustaining a rather large blow to their motherboards.

[5]To list a few examples: a newt's ability to regenerate its tail, a flatworm's ability to regenerate its head, a starfish's ability to regenerate its entire body from a severed leg [Alvarado and Tsonis, 2006].

[6]In fact, computers can be built from water pipes, wires, vacuum tubes, transistors, and even legos.

[7]All of these abilities have led to theories that nucleic acids may have begun life on Earth [Joyce, 1989].

circuits [Lederman et al., 2006, Seelig et al., 2006, Yin et al., 2008] and self-assembled two and three-dimensional structures [Chen and Seeman, 1991, Douglas et al., 2009, Rothemund, 2006, Yin et al., 2008]. An alternate line of research has utilized nucleic acids to engineer biological circuits [Elowitz and Leibler, 2000, Hasty et al., 2002]. For the most part, these endeavors in molecular programming either translate traditional computational problems into molecules, or hijack a cell's materials to control a behavior. While DNA computers have been proven to be Turing universal [Beaver, 1996, Rothemund, 1996, Smith and Schweitzer, 1996], they cannot compete with their digital counterparts in solving classic computational problems[8]. Meanwhile, the synthetic biology approach does not leave room to discover what molecules are capable of in the absence of four billion years of evolution.

What is the limit of behavioral complexity that can be achieved with a molecule as simple as DNA? We know that in biology, there is some nondeterministic encoding [Metzger et al., 2008] and overlaying of information [Breitbart et al., 1987]. We have encountered operons and regulatory networks [Jacob et al., 1960]. We have some idea that cells are running programs, that these programs are stochastic in nature, and that these systems have evolved out of some random mixture of molecules over a billion years ago. But we do not yet understand the programs that cells are running.

The DNA nanotechnology community has addressed this question by engineering programmable assemblies and motors. DNA has been made to programmably self-assemble [Douglas et al., 2009, Liu et al., 1994, Mao et al., 2000, Rothemund, 2006, Rothemund et al., 2004, Winfree et al., 1998] and disassemble [Yin et al., 2008]. An alternate line of research inspired by cellular machinery has resulted in the construction of several nucleic acid based motors [Bath and Turberfield, 2007]. The simplest actuators made of DNA are molecular switches that toggle between two or more conformations [Mao et al., 1999, Simmel and Yurke, 2002, Yurke et al., 2000]. DNA walkers can move relative to a track [Bath et al., 2005, Omabegho et al., 2009, Pei et al., 2006, Sherman and Seeman, 2004, Shin and Pierce, 2004, Tian et al., 2005, Yin et al., 2008], and can generate

[8] As Soloveichik, et al. eloquently phrase the issue, "... shoehorning the design of synthetic chemical circuits into familiar but possibly inappropriate computing models may not capture the natural potential and limitations of the chemical substrate" [Soloveichik et al., 2010].

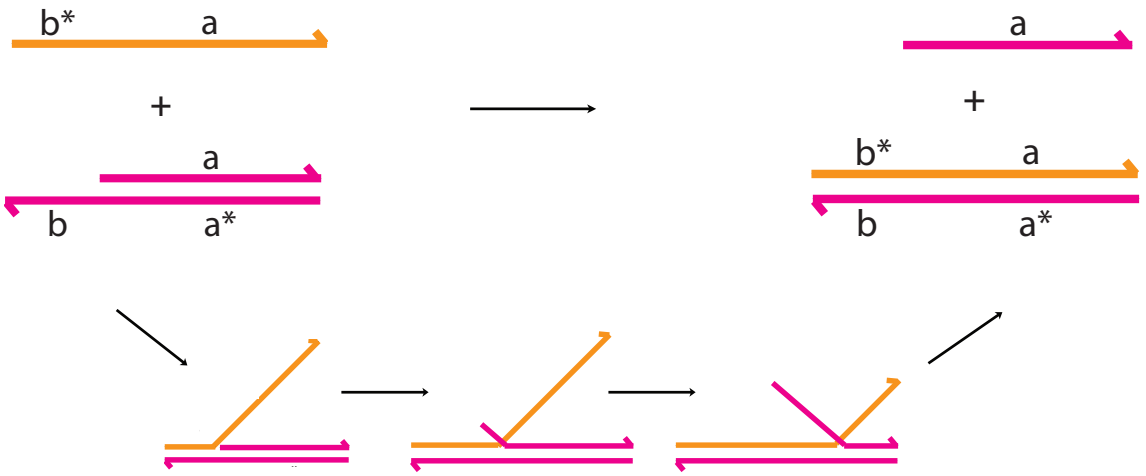


Figure 1.1: A schematic of three-way branch migration.

directional motion without tracks [Venkataraman et al., 2007]. In addition, DNA devices can coordinate two moving parts [Green et al., 2008] and change shape [Andersen et al., 2009, Lubrich et al., 2008, Yurke et al., 2000].

This thesis attempts to bridge the self-assembling and dynamic / motor trajectories of DNA nanotechnology by utilizing insertion and deletion primitives to actively grow and shrink assemblies and to further the goal of programming molecular components that can output complex behaviors like those we see in life.

1.2 Review of Fundamental Components

1.2.1 Three-way versus Four-way Branch Migration

In three-way branch migration only one strand in one duplex is traded for a single strand with the same sequence (Figure 1.1). Four-way branch migration is the process by which two double-stranded oligonucleotides that share the same stem sequence simultaneously exchange strands (Figure 1.2).

Many developments in DNA nanotechnology rely on three-way branch migration to implement

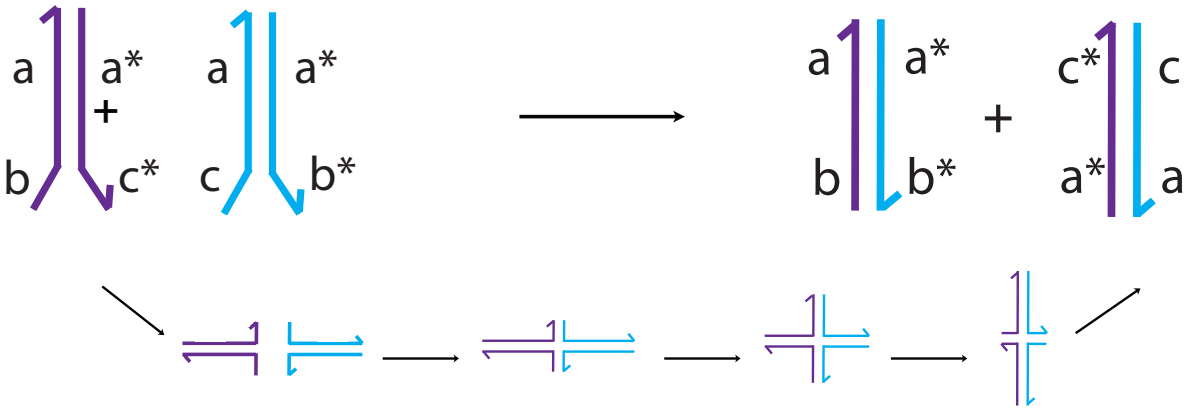


Figure 1.2: A schematic of four-way branch migration.

switches [Lubrich et al., 2008, Simmel and Yurke, 2002, Yurke et al., 2000], circuits [Seelig et al., 2006, Yin et al., 2008], motors [Omabegho et al., 2009, Shin and Pierce, 2004, Yin et al., 2008], assembly [Yin et al., 2008] and amplification [Dirks and Pierce, 2004]. Three-way branch migration is easier to initiate than four-way branch migration (owing to the entropic penalty of bringing two complexes together), and three-way branch reaction rates can be controlled over six orders of magnitude [Zhang and Winfree, 2009]. But it appears that four-way branch migration may give us finer control over this rate, and greater range (seven or more orders of magnitude) [Dabby et al., 2013]. Four-way branch migration has been used to perform directional motion via insertion [Venkataraman et al., 2007]. The capabilities of four-way branch migration have not been fully explored.

1.2.2 Hybridization Chain Reaction

Dirks and Pierce make use of three-way branch migration in their Hybridization Chain Reaction construction [Dirks and Pierce, 2004]. Their construction, which triggers the polymerization of DNA monomers, uses two single-stranded DNA hairpins that have the same 18 base-pair stem sequence and one toehold that is complementary to the other hairpin's loop sequence. These hairpins are caught in a kinetic trap that causes them to react with each other very slowly in the absence of an initiator strand. The initiator consists of a toehold, which is complementary to one hairpin's

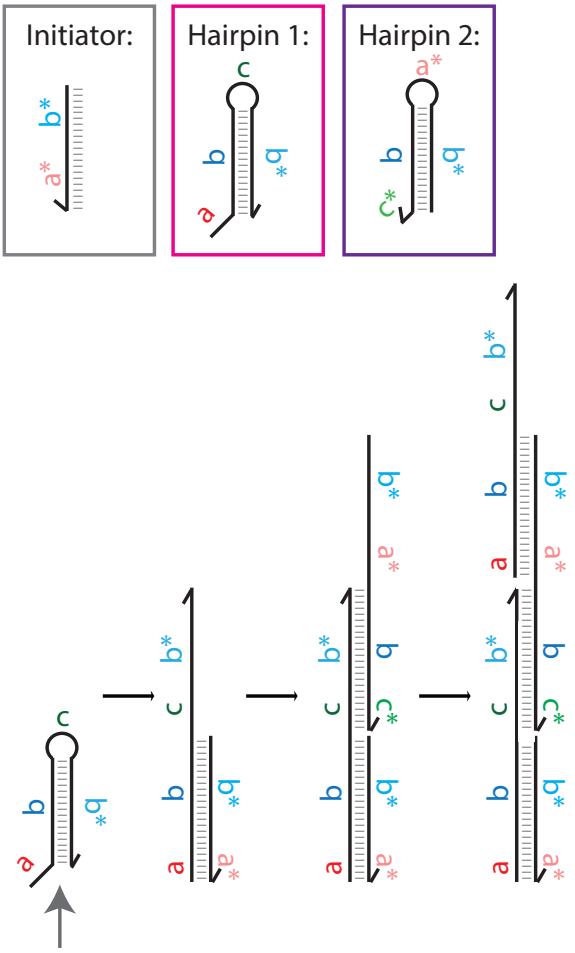


Figure 1.3: A schematic of the Hybridization Chain Reaction as described by [Dirks and Pierce, 2004].

toehold, and its adjacent stem sequence. When the initiator is added to the solution of monomers, it binds to the toehold of the first hairpin and undergoes a strand displacement reaction that opens the hairpin. The newly exposed sticky end of the hairpin can then undergo a similar reaction with the second hairpin.

The two hairpins will continue to polymerize until an equilibrium concentration of monomers is reached. While the HCR system has been applied to the amplification of nucleic acid probes [Choi, 2009], the system was modified to employ a four-way branch migration design to create an autonomous polymerization motor [Venkataraman et al., 2007]. The metastable fuel hairpins from

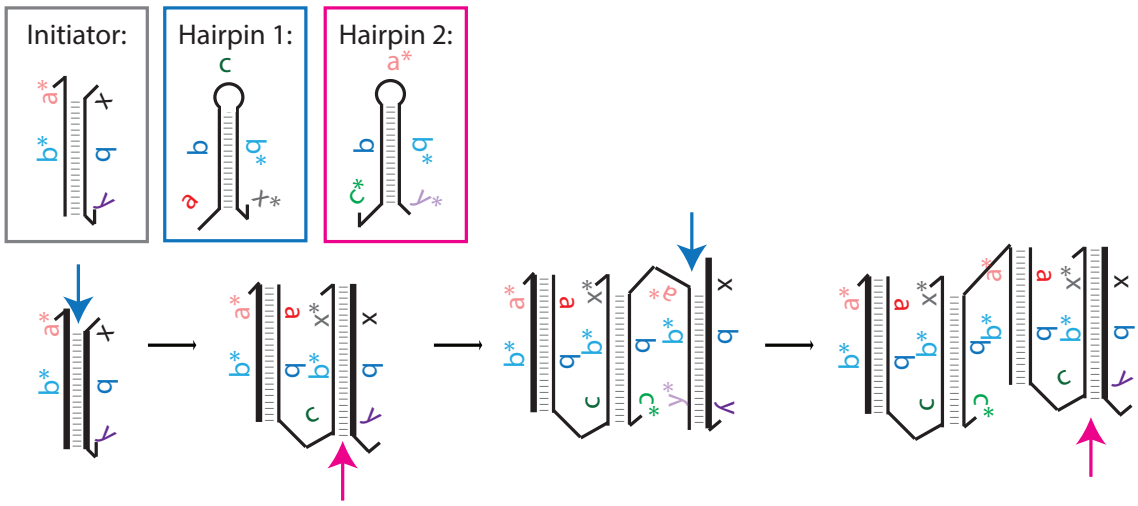


Figure 1.4: A schematic of a Rickettsia-like autonomous polymerization motor as described by [Venkataraman et al., 2007].

the HCR system were modified to include an extra toehold, and the initiator strand was replaced by an initiator complex that is composed of an "anchor" strand and a "rickettsia" strand. Upon mixing, the first hairpin binds to the sticky ends of the anchor-rickettsia complex, initiating a four-way branch migration in which the rickettsia strand is passed from the anchor to the hairpin. The second hairpin then binds to the newly exposed sticky ends and the rickettsia strand is passed to the second hairpin. The rickettsia strand continues to be passed back and forth between newly added hairpins as the polymer grows in its wake.

1.2.3 Other Four-way Branch Machines

Four-way branch migration machines make use of duplex DNA's ability to undergo genetic recombination via a Holliday Junction, a four-way branched DNA structure in which the opposite arms share common sequences. The junction can migrate along a duplex of DNA by breaking the basepairs in one pair of opposite arms and forming base pairs in the other pair [Holliday, 1964]. One of the first DNA nanomachines that made use of the Holliday Junction worked by converting torsional strain into the linear motion of a Holliday junction [Yang et al., 1998]. The device, which consisted of a Holliday junction connected on opposite arms by a closed loop of double

stranded DNA, was powered by the addition and removal of ethidium bromide (an intercalating dye that binds DNA between adjacent base pairs). Since then, the Holliday Junction has been employed in constructing a single molecule "nanometronome" [Buranachai et al., 2006], and in the design of a single molecule switch that can detect single nucleotide mismatches in RNA and DNA oligonucleotides [Buck et al., 2007].

1.2.4 Programming Biomolecular Pathways

Yin, et al. demonstrated a method for programming the pathway by which DNA self-assembles and disassembles that allows one to design a sequence of reactions that can implement dynamic functionality [Yin et al., 2008]. Molecules are initially trapped in a metastable state that allows other oligonucleotides to systematically catalyze their interactions with each other. The examples are implemented using three-way branch migration.

1.3 Related Work

1.3.1 Theory of Active Self-Assembly

One of the central questions that this work addresses is how to program global tasks through local interactions. Graph grammars [Klavins et al., 2004] allow for a systematic way to program molecular self-assembly using rule sets to synthesize a general graph. The approach taken by Klavins, et al. lacked a geometrical component, which we tried to address in our prior work on active self-assembly, and in the process we proved some very interesting theorems that highlight the power that can be gained from actively assembling units over passively assembling units.

The Nubot model [Woods et al., 2013] builds on the concept of graph grammars, by defining rule sets over two dimensional monomers (that we represent as disks of unit diameter centered on a point in a triangular grid). Two monomers can react with each other (according to a rule) to change state, make and break bonds, change relative position, appear and disappear.

Using this model, we showed that a line of length n can be constructed with $O(\log n)$ states, in $O(\log^2 n)$ time. An arbitrary two-dimensional geometric shape with n pixels can be constructed

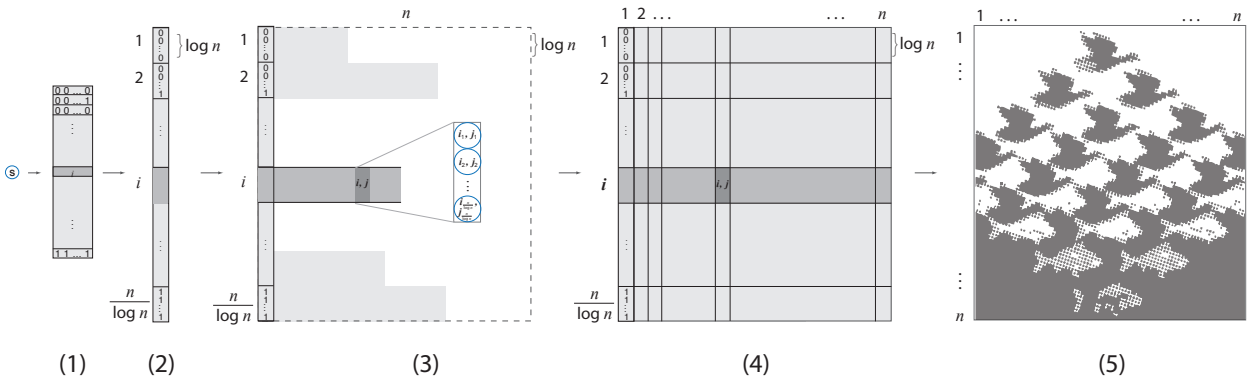


Figure 1.5: An overview of a how a pattern is computed in the nubot model: The shape begins with an initial seed that generates a vertical line of n nubots with unique states in logarithmic time. Each of these nubots then grows a horizontal line of length n, thus generating an $n \times n$ square. Each $\dfrac{1}{\log(n)} \times n$ strip of nubots acts as the input tape to a turing machine, which colors the pixels (nubots) either black or white depending on its unique address in the total assembly.

in $O(\log^2 n + \log n \times f(n))$ time using $O(\log n + g(n))$ states, where *f(n)* is the time required for a size *g(n)* Turing machine to determine if a given pixel is in the shape. It works as follows: the shape begins with an initial seed that generates a vertical line of n nubots with unique states in logarithmic time. Each of these nubots then grows a horizontal line of length n, thus generating an $n \times n$ square. Each $\dfrac{1}{\log(n)} \times n$ strip of nubots acts as the input tape to a turing machine, which colors the pixels (nubots) either black or white depending on its unique address in the total assembly. For many common shapes, *f(n)* and *g(n)* are polylogarithmic in *n*. This is exponentially faster than systems composed entirely of passive components (e.g. tiles).

The main limitation of the Nubot model is that it requires individual monomers to have Turing machine capability to perform the above complex tasks. Molecules are not individually capable of having large state spaces with large look-up tables, thus the Nubot model is not chemically implementable. In order to implement such a system today, one would require macro scale robots with relatively large onboard memories and complex actuation capability.

While the Nubot model is not chemically implementable, it motivates experimental efforts to construct systems with actively assembling components. In fact, it inspired Chapters 3 and 4 of this manuscript. In our ongoing development of a new model for active assembly, we seek to preserve the complex behaviors that our abstract system is capable of, but in a formulation that is simple

enough to implement experimentally.

1.4 Summary of Thesis

1.4.1 Programming Molecular Robots Using Prescriptive Landscapes

Taking some cues from Brooks's "Intelligence without Representation" [Brooks, 1991], we considered how we might imbue molecules with complex programs given their limited encoding space. We hypothesized that molecules might be capable of being programmed by their interaction with a surface. In this way one can program "local rules" such that the configuration and location of a molecule will determine the next step that the molecule can carry out. To this end we explored the ability of a pre-programmed surface to direct the behavior of a "molecular robot" [Lund et al., 2010, Lund, 2008] in Chapter 2.

1.4.2 A Model for Active Self-Assembly in DNA Systems

In Chapter 3 we describe a formal model for studying the complexity of self-assembled structures with active molecular components. In particular, we add an insertion primitive and we show a direct mapping of our model to a molecular implementation using DNA. We show that the expressive power of this language is stronger than regular languages, but at most as strong as context-free grammars. Here, we explore the trade-off between the complexity of the system (in terms of the number of unit types), and the behavior of the system and speed of its assembly. We find that we can grow a line of any given length n in expected time $O(log^3 n)$ using $O(log^2 n)$ monomers. If we grow a line with k insertion rules, either the expected final length is infinite or the expected length at time t is at most $(2t + 2)^{k^2}$, which is polynomial in t.

1.4.3 A Synthetic Polymer that Grows Exponentially Fast

In Chapter 4, we demonstrate the growth of a linear DNA polymer exponentially fast using a molecular insertion primitive to deterministically incorporate three hairpins into a linear structure.

We experimentally verify the exponential kinetics of the system and compare it to its linear counterpart. In addition, we implement a division primitive and show that we can initiate the exponential growth of populations of smaller sized polymers. Lastly we present a theoretical implementation of a treadmilling behavior using these two primitives from our model.

1.4.4 The Kinetics of Toehold-Mediated Four-way Branch Migration

Chapter 5 explores the kinetics of the insertion mechanism implemented in the prior chapter. We characterize the kinetics of toehold-mediated four-way branch migration. We found that by designing the toeholds that initiate the reaction, we can control the reaction rate over at least seven orders of magnitude. We propose a model for the design of the sequences of a four-way branch reaction that operates with desired kinetics. The ability to control the kinetics of these reactions should greatly facilitate the programming of dynamic behaviors mediated by four-way branch migration.