Programming chemical kinetics: engineering dynamic reaction networks with DNA strand displacement

Thesis by

Niranjan Srinivas

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 20, 2015)

ii

To the fond and loving memory of my grandmother, Rajalakshmi, affectionately known as Rajam *patti*.

We shall not cease from exploration,
and the end of all our exploring
will be to arrive where we started
and know the place for the first time.

- T. S. Eliot

Acknowledgements

No one has had a larger impact on my scientific thought process than my thesis advisor, Erik Winfree. He has not only been a close collaborator on the particular research projects we worked on together, but also a constant source of encouragement for my overall scientific development. Indeed, Erik has enthusiastically supported essentially anything that I have wanted to do for my larger scientific growth. He has also been a sounding board for my ideas and plans and thoughts and fears, both scientific and non-scientific, and always provided me with constructive feedback and thoughtful counsel. In addition to all of the above, Erik's unwavering encouragement was crucial in helping me persevere through the darkest periods of my graduate study, both professional and personal. I came to Caltech as a graduate student who was curious about everything from machine learning and artificial intelligence to brains and biology; it is to his great credit as a mentor that I leave Caltech as an even more curious independent scientist. For all his patient nurturing and guidance, I will always be grateful to Erik.

Paul W. K. Rothemund has always been generous with his time, energy, and expertise. I have lost count of the number of times I have walked in to Paul's office with some experimental data that seemed confusing and some plans that seemed hazy, and walked out with a clearer understanding of my own data and a better set of plans. In addition to my research projects, the diversity of topics about which I have had enlightening conversations with Paul continues to amaze me (although lately I have realized that I am far from alone in this). Indeed, Paul has shown me what being "interdisciplinary" really means. Lastly, Paul's guidance on managing the emotional side of doing science has been absolutely vital for me. I convey my heartfelt thanks to Paul for his generosity.

I thank the other members of my thesis committee, Georg Seelig, Richard Murray, Niles Pierce, and Michael Elowitz, for having been a constant source of constructive feedback. In particular, Georg has been a close collaborator; his perspective on our projects has always been insightful and

instructive. In thesis committee meetings, Richard raised important "medium" and "big picture" questions that enabled me to view my projects in different ways. Niles, along with his student Brian Wolfe, gave me excellent feedback on my work on strand displacement biophysics and on sequence design for my CRN-to-DNA efforts. Lastly, Michael asked a series of questions in my candidacy exam that helped me think more clearly about the scientific implications of my research projects.

David Soloveichik has been a close mentor, collaborator, and friend. His work has been the intellectual foundation for my efforts towards engineering chemical reaction networks with DNA strand displacement. In particular, his guidance was very important for me in my initial years of graduate study, when I was new to the field. I thank David for his help in shaping my development over the years.

Bernie Yurke was a collaborator on our work on the biophysics and kinetics of DNA strand displacement. However, his role in my work and my development is far bigger than that. I invariably met Bernie once a year at conferences, in addition to his visits to Caltech. Conversations with him were very valuable since he always had insightful suggestions about my work. I thank him for always being interested both in my work and in my development as a scientist, and for being generous with his time and ideas.

Joseph Schaeffer, another collaborator on our biophysics study, taught me a lot about the thermodynamics and kinetics of DNA. In addition to the content, his clarity of presentation and his feedback on my slides and writing significantly improved my ability to communicate my ideas and results. I thank him and our collaborators from Oxford, Tom Ouldridge, Petr Šulc, Ard Louis, and Jon Doye, for helping me reconcile the biophysics, molecular mechanisms, and kinetics of DNA strand displacement, tie up all the loose ends, and tell a unified story.

Yuan-Jyue Chen, Luca Cardelli, Andrew Phillips, and Neil Dalchau were fun collaborators on our effort to implement chemical reaction networks with nicked double-stranded DNA molecules. All that we learned working together was immensely helpful for my parallel effort focused on implementing dynamical behaviors. I thank them for their constant support and encouragement throughout my graduate study.

I am grateful to Andreas Krause for an excellent rotation experience in my first year. I learned a lot working with him and am very proud of my first research paper, which was a result of our collaboration.

my intellectual curiosity. Lastly, I thank Suchitra Mathur for helping me appreciate intellectual pursuits outside the sciences.

A lot of people at Caltech have supported me in ways that are not directly related to scientific pursuits but were absolutely crucial in enabling those pursuits. Lucinda Acosta made all administrative business related to research and travel at least ten times more efficient. On top of that she has been a warm and caring presence in the lab and gone way beyond the call of duty. For that she has my warm gratitude. Tanya Owen has been incredibly helpful with all matters related to the CNS program and CNS events. Laura and Daniel at ISP have made me feel welcome at Caltech and in America since the time I first arrived. Lastly, I thank Felicia Hunt and Helen Duong from the graduate office for all their help over the years.

During my time at Caltech I have been fortunate to find friends who are more like family. Teja, Maha, Batuk, Gopi, and Sri have always been there for me. Jay and Sushree's home has been another home for me in Pasadena. Britton has been a close friend ever since we went to check out a new Mexican restaurant called Chipotle. I thank all of them for their warmth and companionship.

In conclusion, I thank my family for everything. I thank my aunt and uncle, Nirmala and Thyagarajan, for having been more like parents to me. I thank my brother and sister-in-law, Aravind and Parul, and my parents, Usha and Srinivas, for being a warm, close-knit, and supportive family. Finally, I thank my wife, Bhargavi, for all her love, support, and companionship on our journey together.

Abstract

Over the last century, the silicon revolution has enabled us to build faster, smaller and more so-phisticated computers. Today, these computers control phones, cars, satellites, assembly lines, and other electromechanical devices. Just as electrical wiring controls electromechanical devices, living organisms employ "chemical wiring" to make decisions about their environment and con-trol physical processes. Currently, the big difference between these two substrates is that while we have the abstractions, design principles, verification and fabrication techniques in place for programming with silicon, we have no comparable understanding or expertise for programming chemistry.

In this thesis we take a small step towards the goal of learning how to systematically engineer prescribed non-equilibrium dynamical behaviors in chemical systems. We use the formalism of chemical reaction networks (CRNs), combined with mass-action kinetics, as our programming language for specifying dynamical behaviors. Leveraging the tools of nucleic acid nanotechnology (introduced in Chapter 1), we employ synthetic DNA molecules as our molecular architecture and toehold-mediated DNA strand displacement as our reaction primitive.

Abstraction, modular design and systematic fabrication can work only with well-understood and quantitatively characterized tools. Therefore, we embark on a detailed study of the "device physics" of DNA strand displacement (Chapter 2). We present a unified view of strand displace-ment biophysics and kinetics by studying the process at multiple levels of detail, using an intu-itive model of a random walk on a 1-dimensional energy landscape, a secondary structure kinetics model with single base-pair steps, and a coarse-grained molecular model that incorporates three-dimensional geometric and steric effects. Further, we experimentally investigate the thermody-namics of three-way branch migration. Our findings are consistent with previously measured or inferred rates for hybridization, fraying, and branch migration, and provide a biophysical expla-

nation of strand displacement kinetics. Our work paves the way for accurate modeling of strand displacement cascades, which would facilitate the simulation and construction of more complex molecular systems.

In Chapters 3 and 4, we identify and overcome the crucial experimental challenges involved in using our general DNA-based technology for engineering dynamical behaviors in the test tube. In this process, we identify important design rules that inform our choice of molecular motifs and our algorithms for designing and verifying DNA sequences for our molecular implementation. We also develop flexible molecular strategies for "tuning" our reaction rates and stoichiometries in order to compensate for unavoidable non-idealities in the molecular implementation, such as imperfectly synthesized molecules and spurious "leak" pathways that compete with desired pathways.

We successfully implement three distinct autocatalytic reactions, which we then combine into a *de novo* chemical oscillator. Unlike biological networks, which use sophisticated evolved molecules (like proteins) to realize such behavior, our test tube realization is the first to demonstrate that Watson-Crick base pairing interactions alone suffice for oscillatory dynamics. Since our design pipeline is general and applicable to any CRN, our experimental demonstration of a *de novo* chemical oscillator could enable the systematic construction of CRNs with other dynamic behaviors.

Contents

1 Introduction

> *...But in the real world, it's very hard to know which problems are important,*
> *and you never know whether at a given moment in history a problem is solvable...*
>
> - Steven Weinberg [1]

1.1 Motivation and context: the molecular programming perspective

Living organisms are fascinating. They start as a single cell, metabolize, grow, and reproduce. Some of them grow into complex multicellular creatures with precisely reproducible structures that perform important functions — leaves that carry out photosynthesis, eyes that enable vision, wings that power flight, and so on.

All this complexity arises from a single cell. Indeed, even the largest tree starts as a single seed. The seed contains all the information, and given favorable conditions and raw material, will execute its "developmental program" and grow into a tree. As foreseen by John von Neumann [2], growth and reproduction are essentially information processing tasks. Further, Alan Turing's work on chemical morphogenesis [3] was an early insight into the computations involved in biological growth.

If we[1] are to think of development as a program and the genome as software, we must identify features common to both[2]. First, software execution is reproducible: two copies of the same program in the same environment execute identically. The same is true for developmental programs: monozygotic twins share an (almost) identical genome and environment and develop nearly identically. Second, software is sensitive to small changes in crucial places: changing *one* loop control

---

[1]A clarification on the use of "we": "we", rather than "I", is used almost everywhere in this thesis both for stylistic purposes and to indicate the collaborative nature of the work presented. In some select places, "I" is used to emphasize the author's personal opinions or experiences.

[2]This part of the discussion was inspired by a popular talk given by Paul W. K. Rothemund.
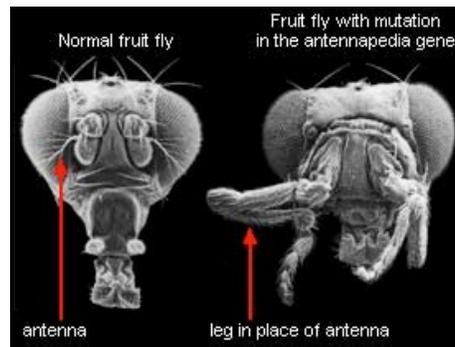
Figure 1.1: The developmental program, like software, is sensitive to small changes. Mutations in the *Antennapedia (Antp)* gene in the fruit fly *Drosophila melanogaster* can result in a leg growing out of the head rather than an antenna [4]. Figure adapted from Klug et al. [4], courtesy Paul W. K. Rothemund.

variable, for example, can make all the difference between a correct program and an incorrect one. Similarly, relatively small changes in the genetic code can result in dramatic changes in the growth process (Figure 1.1). Lastly, growth, like any well-written computer program, is intended to halt; indefinite growth is a serious disorder.

Inspired by the living world, where information directs matter, can we seek to program molecular systems? Maybe we could dream of programming biochemical machines one day, the way we program electromechanical machines today. In essence, that is the "big picture" vision underpinning this thesis. Although programming biochemical machines as complex as living organisms is a grand challenge with multiple intellectual and technical barriers to overcome, there are three reasons for cautious optimism.

First, the history of computing teaches us that multiple physical substrates can embody computation, and that learning to compute with a new substrate can revolutionize technology in ways we can scarcely imagine. Indeed, the earliest devices capable of performing calculations, such as the Antikythera mechanism [5–7], were fabricated over two thousand years ago and relied on mechanical components such as gears. Over the last century, we have learned to compute with silicon and replaced gears with diodes and transistors. This silicon revolution has resulted in faster, smaller, and more sophisticated computers which now control phones, cars, and satellites. It is quite conceivable that we can learn to compute with yet another substrate and thereby enable a molecular information technology.
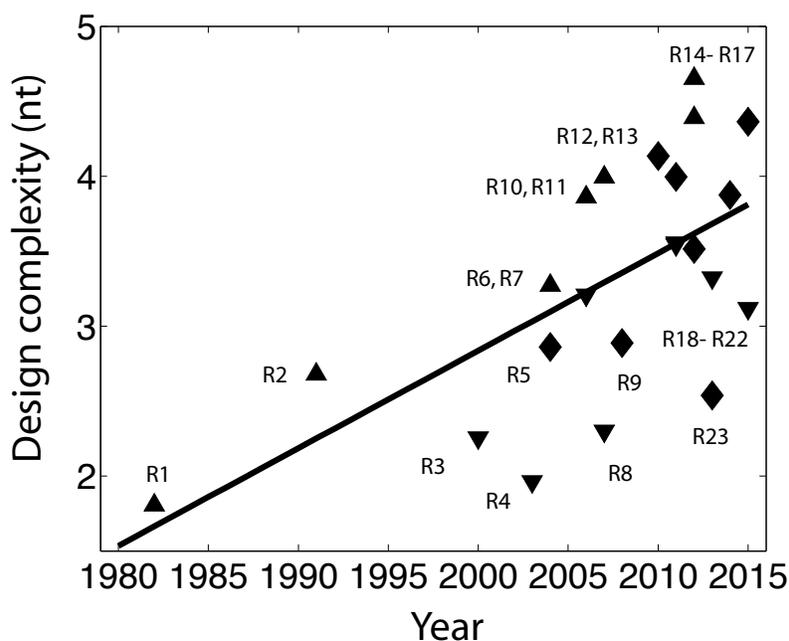
Figure 1.2: Moore's law in DNA nanotechnology: the exponential growth in design complexity of molecular systems that have been experimentally demonstrated over the years. Design complexity (plotted in log scale to the base 10) has been evaluated as the number of nucleotides of synthetic DNA incorporated in the experimentally demonstrated system. Studies in *structural* DNA nanotechnology are indicated by upward-pointing triangles, whereas studies in *dynamic* DNA nanotechnology are indicated by downward-pointing triangles. Studies that incorporate both structural and dynamic aspects are indicated by diamonds. The best fit line, considering all the points, is plotted in black. This figure is an updated and revised version of an earlier figure by Erik Winfree. References R1 - R23 are listed in Table 5.1.

Second, over the last thirty years, a diverse group of scientists and engineers have learned to use synthetic nucleic acids as programmable building blocks for engineering nanoscale structures and circuits [8–13]. Although this "nucleic acid nanotechnology" is still in its infancy, the complexity of engineered molecular systems has been increasing exponentially, following a growth trajectory that is reminiscent of Moore's law [14] (Figure 1.2). For a whirlwind tour of nucleic acid nanotechnology, see Section 1.4.1.

Third, our ability to synthesize and manipulate biomolecules — both single-stranded DNA oligonucleotides and genes (and therefore RNAs and proteins) — has been getting better, faster, and cheaper every year. Indeed, over the last 25 years, productivity of DNA synthesis has increased by nearly 5 orders of magnitude, while the cost of gene synthesis per base has dropped
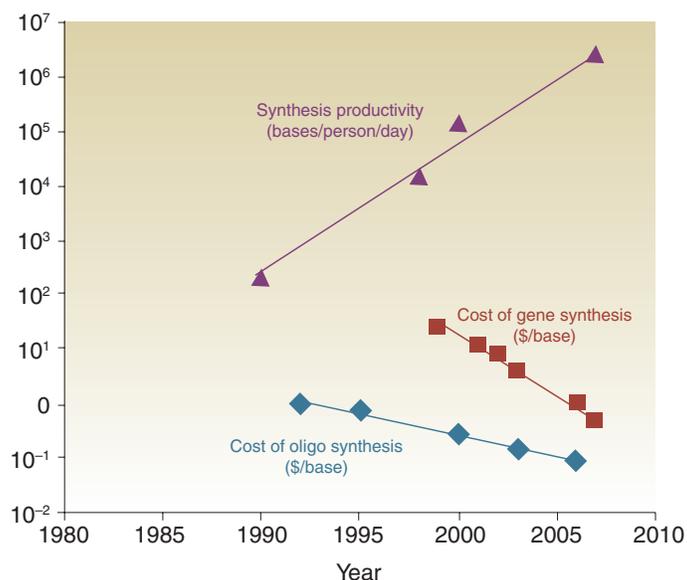
Figure 1.3: Economics of DNA synthesis. Since 1990, productivity has increased by nearly 5 orders of magnitude, while the cost of gene synthesis per base has dropped by nearly 3 orders of magnitude. This figure has been reproduced from Carlson [15].

by nearly 3 orders of magnitude [15] (Figure 1.3). In concert with synthesis, techniques to manipulate and visualize matter on the nanoscale have also improved dramatically in this time period. For example, high-speed atomic force microscopy has enabled striking visualization of dynamic molecular events [16].

How can we start programming with molecules? First, we will need a suitable programming language, one that is rich enough to allow interesting programs, yet simple enough and "natural" enough to permit a smooth molecular implementation. Second, we will need a molecular architecture to try to implement the programs we write. The living world is a natural place to look for inspiration, and one may be tempted to look for a language that captures all the rich phenomena in biology. Information processing happens in very different flavors in the living world; Figure 1.4 presents three distinct examples. Given this diversity, it seems unlikely, at least at the present time, that a single useful model of computation would meaningfully capture all of these different flavors of information processing.

In this thesis, we focus on programming chemistry. By this we mean that we attempt to systematically engineer prescribed dynamical behaviors in well-mixed chemical systems. Although

the first non-trivial dynamical behavior in a well-mixed chemical system, an oscillator, was discovered as early as 1921 [17], well-mixed solutions were widely considered incapable of exhibiting interesting dynamics until well into the second half of the twentieth century [18]. Such dynamic behaviors were considered possible only in inhomogeneous solutions with spatial differences in concentrations. In Section 1.2, we briefly discuss the historical development of experimental and theoretical tools to understand non-equilibrium dynamical behaviors in chemical systems.

These pioneering experimental studies led to theoretical studies that explored the class of dynamical behaviors that are, in theory, possible within the mathematical framework of "formal chemical reaction networks". A formal chemical reaction network (CRN) is a finite set of chemical reaction equations (with rate constants) between formal species. A brief description of the "lay of the land" with respect to these theoretical studies is provided in Section 1.3. In this thesis, we will employ formal CRNs, along with mass-action kinetics, as a programming language for specifying dynamical behaviors in well-mixed solutions. This approach is described in detail in Chapter 3.

A programming language is not of much use without an implementation architecture. In Section 1.4, we will introduce DNA nanotechnology and DNA strand displacement, our molecular architecture of choice for implementing the CRN programs we write.

Lastly, in Section 1.5, I list my scientific contributions to the work described in this thesis.

1.2 Well-mixed chemical systems with complex dynamical behavior

For much of the twentieth century, most chemists believed that complex temporal dynamics, such as oscillations, could not possibly occur in well-mixed ("homogenous") solutions [18, 22–25]. For multiple accounts of different facets of the fascinating history of modern nonlinear chemical dynamics, see Irving Epstein's and John Pojman's excellent textbook [24], a couple of interesting perspectives by Epstein and Showalter [23] and Sagues and Epstein [25], or an article by Anatol M. Zhabotinsky [22] (of Belousov-Zhabotinsky (B-Z) oscillator fame [26, 27]) where he both recounts this history and tries to understand why belief in this impossibility was so pervasive in the face of significant evidence to the contrary. In addition to these accounts, A. T. Winfree [18] traces the history of the B-Z oscillator [26, 27], and in the process summarizes for the first time in English "much that seems common knowledge in Russia about the origins of the cerium/bromate oscillator".
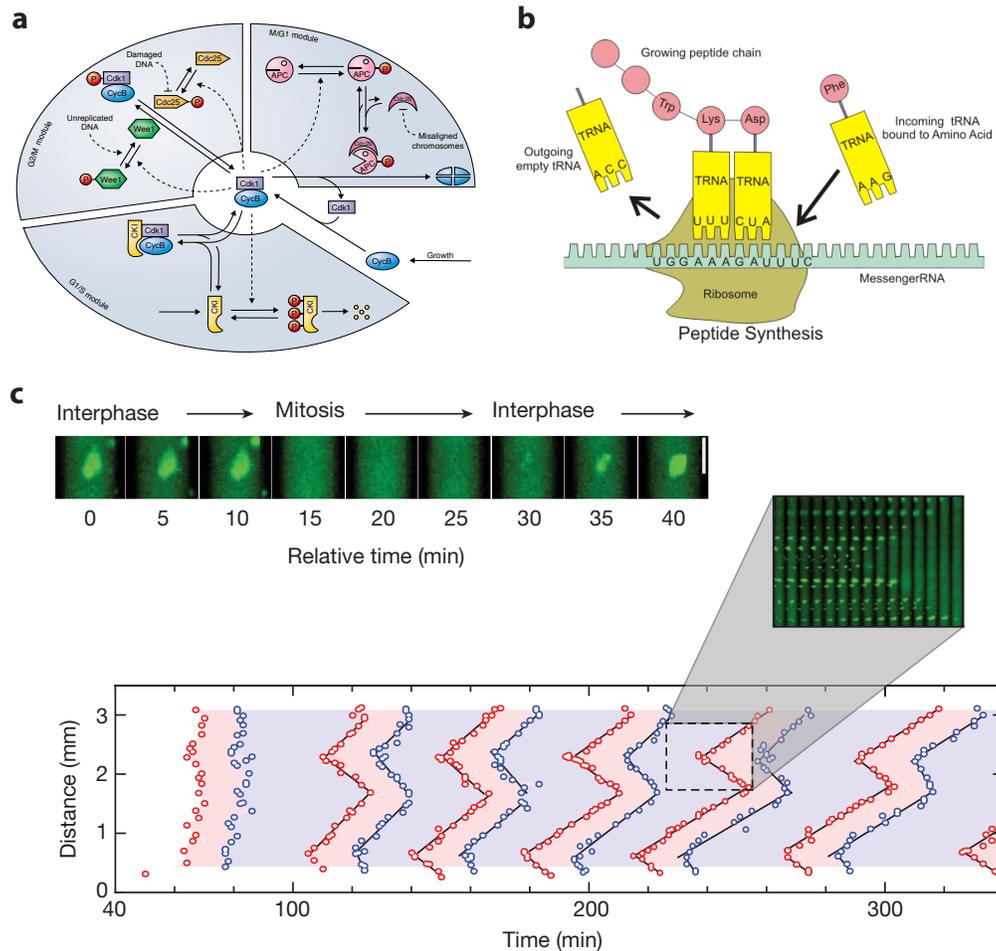
Figure 1.4: Information processing in living systems occurs in many different flavors. a. "Wiring diagram" of the chemical reactions in the Cyclin-dependent-kinase (Cdk) network that regulates the eukaryotic cell cycle, as presented by Tyson et al. [19]. They use the language of chemical kinetics (chemical reaction equations and rate constants) to model the system. b. Cartoon of protein synthesis by the ribosome, which "reads" three nucleotide subsequences, called codons, specified in the messenger RNA and adds the appropriate amino acid to the peptide being synthesized. Image from Wikimedia Commons [20]. c. As foreseen by Alan Turing [3], reaction-diffusion is an important mechanism for chemical patterning in biological systems, particularly in development. Chang and Ferrell Jr. [21] have recently argued that trigger waves of Cdk1 activation are essential for the rapid spread of mitotis through the cytoplasm in embryos of the frog *Xenopus laevis*. This enables rapid, coordinated mitotic events in distant parts of the large (1.2 mm in diameter) *Xenopus* eggs. Top: An example of nuclear breakdown and reformation in an *in vitro* extract with added sperm chromatin and GFP-NLS (nuclear localization signal fused to GFP). Bottom: Timing and position of nuclear breakdown (red points) and reformation (blue points) for individual nucleii at different locations in a 3 mm Teflon tube. Pink and blue regions of the plot denote mitosis and interphase, respectively. Inset panel shows frames from a video in montage form.

Here, I briefly paraphrase the history of nonlinear chemical dynamics. This re-telling is neither meant to be comprehensive nor detailed, and only intends to give the reader the flavor of the story and our current understanding of why the story unfolded the way it did.

According to E. N. Harvey's "A History of Luminescence: From the earliest times until 1900" [28], the first reported observations of an oscillatory chemical system were made as early as the end of the seventeenth century by Robert Boyle, who recorded periodic flashes of luminescence while studying the oxidation of phosphorus. A. M. Zhabotinsky [22] suggests that in those times, oscillating chemical systems did not attract much attention because chemical kinetics as a branch of study did not exist, and therefore there was no commonly accepted expectation for the "normal" course of a chemical reaction.

In 1828, Fechner [29] reported an electrochemical cell that generated an oscillating current. Ostwald [30] observed that the rate of dissolution of chromium in acid increased and decreased periodically. By the beginning of the twentieth century, multiple oscillating systems had been discovered: the periodic dissolution of an iron wire in nitric acid [31] and the periodic decomposition of hydrogen peroxide on the surface of metallic mercury [32–34].

It is worth noting, at this stage, that all the systems described above are "heterogeneous" or "not well-mixed" — we shall soon see that this will have important consequences. Theoretical attempts were soon made to understand these oscillating chemical phenomena. However, heterogeneous reactions have notoriously complicated mechanisms, which typically involve both phase transitions (e.g. solid to liquid) and transport processes. Therefore, understandably, theoretical investigations started with analysis of reactions in well-mixed solutions. Eventually (fast forwarding several decades) this line of enquiry led to the development of chemical kinetics.

The belief that oscillatory behavior is impossible in well-mixed closed chemical systmes arose from two (understandable) mistakes. First, since the first few oscillating chemical systems were all heterogeneous, it was mistakenly assumed that this was a necessary feature to generate such complex temporal dynamics. Second, and more serious, chemical oscillations were viewed as exactly analogous to physical oscillations that were, by then, well understood. A chemical oscillation was understood as the test-tube equivalent of a pendulum, which oscillates around its minimum free energy state. This had been shown to be impossible in the context of chemical reactions in a closed well-mixed solution, as it would violate the second law of thermodynamics.

However, a chemical oscillation is a far-from-equilibrium phenomenon, and the oscillation

occurs *on the way* to equilibrium rather than around it. Thus, a chemical oscillation is like a slide rather than a pendulum[3] (Figure 1.5).
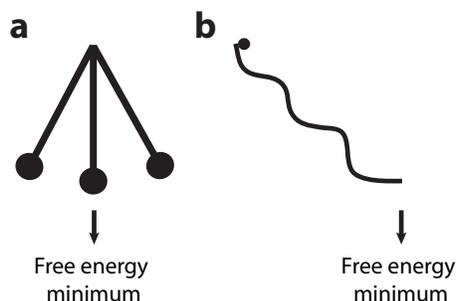
Figure 1.5: a. A pendulum oscillates around its free energy minimum. b. A chemical oscillation in a closed well-mixed solution is not analogous to the oscillation of a pendulum, but rather like the motion of the ball on the slide pictured here. The entire dynamic behavior happens far from equilibrium, but on the way towards it. Free energy decreases monotonically until it attains its minimum.

Meanwhile, in 1921, William C. Bray described the first homogeneous isothermal chemical oscillator [17], which involved redox reactions between iodate, iodine, and hydrogen peroxide. Belief in the impossibility of homogeneous chemical oscillators was so strong and pervasive in the field that for nearly five decades chemists would attribute Bray's oscillations to dust or bubbles (and hence heterogeneity of the reaction mixture) [35]. These objections were raised in spite of the fact that Bray expected and addressed these possible objections by using meticulously filtered and slowly stirred solutions at room temperature (with the reaction rate so slow that oxygen diffused out of the reaction mixture before bubbles could form).

Belousov's discovery of what is now known as the Belousov-Zhabotinsky (B-Z) oscillator [26, 27] was the next important experimental contribution. Belousov discovered this bromate-citric acid-cerium oscillator while trying to construct an inorganic version of the Kreb's cycle, where citric acid is a crucial intermediate. Belousov tried to publish his results several times, but was rejected each time; one editor even claimed that his "supposedly discovered discovery" was impossible [18]. Eventually he did publish a version of his findings in an obscure non-refereed booklet on radiation medicine [26]. Later, A. M. Zhabotinsky investigated Belousov's recipe for the oscillator, which was floating around in science departments in Moscow with the source having been

---

[3]In Section 1.2, for simplicity, we do not discuss *driven* systems in any detail. Driven systems are kept out of equilibrium indefinitely by the constant supply of free energy and/or matter, and are therefore neither like the pendulum nor like the slide shown in Figure 1.5.

forgotten, and succeeded in making various improvements and publishing the work in 1964 [27].

A major theoretical breakthrough was achieved by Field, Körös and Noyes when they developed an elementary-step level chemical reaction mechanism [36] for the B-Z oscillator. Remarkably (for that time), they showed that numerically simulating the 20 or so rate equations in their model with exactly the same laws of chemical kinetics that applied to "regular" reactions led to chemical oscillation [37].

In contrast to the B-Z oscillator, the mechanism of the Bray oscillator remains only partially understood even today [38, 39]. This is because the Bray oscillator involves very complicated hydroxyl radical chemistry. In hindsight, it seems reasonable to speculate that the path towards understanding non-equilibrium behaviors in chemical systems would likely have been smoother and faster if the first homogeneous chemical oscillator to be discovered had been the B-Z oscillator.

Once the principles behind non-equilibrium behaviors in chemical systems and the mechanistic details of particular chemical oscillators were both understood, several synthetic oscillators based on redox chemistries were engineered. Examples include the chlorite-iodate-arsenite oscillator [40], which was the first oscillator engineered intentionally (as opposed to serendipitously discovered), a hydrogen peroxide-sulfide reaction [41], a cobalt-catalyzed oxidation of benzaldehyde [42], and so on.

With the advent of molecular biology, synthetic chemical reaction networks with enzyme-based reactions have been engineered to exhibit non-equilibrium behaviors both *in vitro* and *in vivo*. These efforts essentially harness the machinery of the central dogma for engineering synthetic chemical reactions, in order to both understand the principles underlying natural cellular networks and to learn how to modify those networks for applications. For example, following the genetic toggle switch engineered by Gardner et al. [43] and the synthetic ring oscillator ("repressilator") built by Elowitz and Leibler [44], multiple *in vivo* genetic oscillators have been demonstrated [45–47]. However, these genetic circuits are (i) hard to scale up, because of the restricted number of orthogonal and well-characterized components available (at least as of today), and (ii) hard to model, because they are constructed *in vivo* and are therefore necessarily coupled to the large and complex endogenous genetic network of the organism [48, 49].

Ackermann et al. [50] proposed an *in vitro* biochemical oscillator using reverse transcriptase and T7 RNA polymerase. Despite theoretical investigations showing beautiful and sustained oscillatory behavior, successful experimental demonstration of the oscillator has not been reported,

even though partial progress had been made and each desired reaction step had been demonstrated experimentally [51]. Possibly, this oscillator was overcome by parasitic sequences arising out of accumulated sequence mutations [52, 53].

Inspired by both the *in vivo* and partially successful *in vitro* efforts, Kim et al. [54] proposed a simpler "genelet" architecture, based only on transcriptional machinery, for creating *in vitro* circuits [55]. This architecture, comprising RNA transcripts that regulate their own transcription from DNA promoters, has been used to experimentally demonstrate a bistable switch [56] and an oscillator [53]. Restricting the architecture offers two benefits. First, theoretical analysis of its capabilities is more tractable — Kim et al. [54] show that their genelet architecture is in principle capable of implementing any Boolean logic circuit or Hopfield neural network [57]. Second, the resulting *in vitro* genelet circuits are more accessible to mechanistic quantitative modeling [55, 58].

Montagne et al. [59] proposed another architecture that uses single-stranded DNA templates rather than genelets. This Polymerase-Exonuclease-Nickase (PEN) DNA tool box [60] has been used to demonstrate bistability and switchable memory [61], and multiple oscillators [59, 62]. Software is also available to automate the design process for engineering user-specified networks with this toolbox [63, 64]; however, it is currently not known what class of dynamical behaviors can, in theory, be achieved with this technology. For a more detailed introduction to enzyme-based *in vitro* DNA circuits, see reviews by Padirac et al. [65], Genot et al. [66], Franco et al. [67], Weitz and Simmel [68], and Kim [69].

The history of chemical systems demonstrating complex temporal dynamics raises two natural questions. First, over the last century we have learned that the mathematical formalism of chemical reaction equations between reacting species, along with mass-action kinetics, can explain a wide range of dynamic behaviors in chemical systems. What kinds of dynamical behaviors is this formalism of chemical reaction equations theoretically capable of capturing (possibly beyond the few behaviors, such as oscillations, that have been experimentally realized)? In other words, what is the class of dynamical behaviors that "formal" species, reacting with each other with prescribed rate constants, achieve? This question seeks to investigate the expressiveness of the language we use to model chemical systems, and is addressed in Section 1.3.

Second, the experimental attempts we have discussed above have either been serendipitous, or relied on tinkering with known chemistries to achieve a particular desired implementation of a particular desired behavior. They do not have an automated pipeline across the necessary

layers of abstraction with theoretical guarantees for each level of approximation, which is the kind of rigorous approach one observes and expects in other areas of engineering. Can we use the language of chemical reactions between formal species as a prescription for desired dynamical behaviors, and, given that prescription, implement it in a general way with a particular molecular architecture? Different facets of this question are addressed in Sections 1.3, 1.4 and in Chapter 3.

1.3 The language of formal chemical reaction networks

In this section, we present brief glimpses of the vast and diverse literature studying formal chemical reaction networks from different theoretical perspectives. A formal chemical reaction network (CRN) is a finite set of chemical reaction equations between formal species with specified rate constants. An example is provided below.

$$X + Y \xrightarrow{k_1} 2X \tag{1.1}$$

$$X \xrightarrow{k_2} \phi \tag{1.2}$$

$$\phi \xrightarrow{k_3} Y \tag{1.3}$$

Note that "open" systems are allowed - conservation of the number of molecules ("mass") is not necessary. A reaction with the appearance or disappearance of molecules should be physically interpreted as representing flows of matter and free energy that are not modeled by the particular formal CRN.

Depending on whether the state of the system is represented as a vector of molecular counts (integers) or as a vector of concentrations (real numbers), we have two distinct classes of formal CRNs - "discrete" state and "continuous" state, respectively. In either case, we could define a variety of "rate laws" that determine the rate of a given reaction. For the purposes of this section, we will restrict ourselves to the multiplicative rate law known as "mass-action".

In both the discrete and continuous state cases, we could have either "stochastic" or "deterministic" behavior. Informally, future states of a stochastic system are not completely determined by the past and present states - the system is a random, evolving process. In the deterministic case, the present completely determines the future. In this section, we will focus only on "discrete stochastic" and "continuous deterministic" systems, both with the mass-action rate law. It is worth noting

that there is also a vast literature on stochastic continuous state systems (e.g. Fokker-Planck and Langevin chemical equations [70, 71], based on stochastic differential equations) which is outside the scope of the present discussion.

1.3.1 Discrete stochastic CRNs

1.3.1.1 The model

A discrete stochastic CRN consists of a set of d species and m reactions. A vector $x \in \mathbb{N}^d$ specifies molecular counts of each species, and hence the state of the system. Each reaction α_i is a tuple $(r_i, p_i, k_i) \in \mathbb{N}^d \times \mathbb{N}^d \times \mathbb{R}^+$, where r_i is the vector of reactant stoichiometries, p_i is the vector of product stoichiometries, and k_i is the rate constant for reaction i. For our purposes we restrict reactions to being either unimolecular or bimolecular only.

A reaction $\alpha_i = (r_i, p_i, k_i)$ can occur in state x only if $x - r_i \geq 0$ (i.e. all coordinates of the vector are non-negative). If reaction α_i does occur, then state x is updated to to $x - r_i + p_i$. In some fixed volume V and starting from initial state x_0, the system evolves through a continuous time Poisson process [72–74], as detailed below.

Let $\#_x S$ denote the number of molecules of species S in state x. Given the current state x and a reaction α, we define a propensity function $\rho(x, \alpha)$ as follows.

If $\alpha : S_i \xrightarrow{k} \ldots$ is a unimolecular reaction, $\rho(x, \alpha) = k \cdot \#_x S_i$. Note that the unimolecular propensity is independent of the volume V, as one would expect.

If $\alpha : S_i + S_j \xrightarrow{k} \ldots$ is a bimolecular reaction with $S_i \neq S_j$, $\rho(x, \alpha) = k \cdot \frac{\#_x S_i \#_x S_j}{V}$.

Lastly, the propensity of the bimolecular reaction $\alpha : 2 S_i \xrightarrow{k} \ldots$ is $\rho(x, \alpha) = k \cdot \frac{\#_x S_i (\#_x S_i - 1)}{2V}$.

How does the propensity function ρ determine the kinetics of the CRN? From the current state x, no reactions are possible if $\rho(x, \alpha) = 0$ for all reactions α in the CRN. Else, the waiting time until the next reaction occurs is exponentially distributed with the rate parameter $\sum_\alpha \rho(x, \alpha)$. The probability that a particular α_i would be the next reaction is $\rho(x, \alpha_i) / \sum_\alpha \rho(x, \alpha)$.

If the volume V and the molecular counts of all species S_i are increased such that the concentration $\#S_i / V$ stays constant (the "thermodynamic limit"), then the stochastic trajectories converge to that predicted by the deterministic continuous-state mass action formulation [75, 76].

1.3.1.2 Computation with discrete stochastic CRNs

A number of formal models of computation that have been studied in theoretical computer science are closely related to discrete stochastic CRNs. These include Petri nets [77], vector addition systems [78] and population protocols [79]. The connections between discrete stochastic CRNs and some of these models of computation are discussed by Cook et al. [80].

What kind of computation are discrete stochastic CRNs capable of? Accommodating a probability of error $\epsilon > 0$ (which may be made arbitrarily small), discrete stochastic CRNs are capable of Turing-universal computation [80, 81]. The proof relies on simulating register machines (which are known to be Turing-universal [82]) with discrete stochastic CRNs. Essentially, species act as registers do - by storing non-negative integers in their molecular counts. Reactions, like instructions in a register machine, are used to increment and decrement the molecular counts of the appropriate species.

In contrast, for $\epsilon = 0$, universal computation is provably impossible [80, 81]. Indeed, Angluin et al. [83] show that the natural formulation of deterministic predicate computation by discrete stochastic CRNs corresponds exactly to the semilinear predicates, a relatively restricted class.

1.3.2 Continuous deterministic CRNs

In this model, we define the state of the system as a vector of real-valued concentrations rather than integer-valued molecular counts. Instead of formally specifying the continuous deterministic CRN model with mass-action kinetics, we simply state that it is the same as "standard" chemical kinetics [84], along with the mass action rate law. According to the mass-action rate law in this setting, the instantaneous rate of the general reaction $\mathrm{aS_i + bS_j} \xrightarrow{k} \cdots$ is defined to be $k \cdot [S_i]^a \cdot [S_j]^b$.

Given that the dynamics of a continuous deterministic CRN is determined by differential equations and reaction rates, it might seem that numerically solving the differential equations for the particular rate constants involved is necessary to answer any questions about the CRN's dynamical properties. However, it turns out that it is possible to make some deductions about the dynamical properties of a CRN directly from its network topology. ("Network topology" refers to the structure of a graph whose edges represent reactions and nodes represent complexes of reactants and products.) These deductions are either (i) true for all non-zero values of the rate constants, or (ii) there exists a set of non-zero rate constants for which they are true. Some examples of such

deductions could be: given the topology of a CRN, does it admit a positive equilibrium, where the concentrations of all species are positive? Does it admit an unstable positive equilibrium? For a nice introduction to this graphical view of chemical reaction networks and the main results (including the deficiency zero and deficiency one theorems [85]), see Martin Feinberg's lecture notes [86] .

Since for every continuous deterministic CRN with mass-action kinetics there exists a corresponding set of coupled polynomial ODEs that govern its dynamics, it is natural to wonder about the relationship between the class of ODEs that arise from CRNs and the general class of all polynomial ODEs. Korzuhin [87] and Klonowski [88] show that mass-action CRNs can approximate arbitrary polynomial ODEs. Given a system of polynomial ODEs with nonnegative integer powers, they explicitly construct a formal CRN some of whose species will approximate the solution to the system of ODEs on the positive orthant, up to arbitrary accuracy over any time interval. In fact, the CRN constructed has particularly nice properties: (i) all reactions follow conservation of mass, (ii) have at most two reactants and two products, and (iii) no reactions are autocatalytic.

Kerner [89]'s technique of introducing new variables "based on appropriate collectives of variables and parameters" to convert, without changing the dynamics, a very general ODE into a polynomial ODE with only quadratic nonlinearities offered a route to generalize these transformations. Consequently, Kowalski [90] showed how to construct mass action CRNs whose dynamics approximate the polynomial systems obtained from Kerner [89]'s technique. Wilhelm [91] extends this approach by showing that (i) a reaction with more than two reactants can be interpreted as the limit of a sequence of bimolecular steps and (ii) improves Kowalski [90]'s transformation to show that for any ODE system for which Kerner [89]'s technique applies, one may construct a mass action CRN with only bimolecular steps that approximates it.

So far we have discussed transformations of ODE systems into chemical systems with mass-action kinetics. A natural related question is whether dynamical systems that correspond to physical or electrical systems may be approximated by chemical systems with mass-action kinetics. In Section 1.2, we briefly discussed a key difference between mechanical and chemical oscillators - some mechanical oscillators (such as a pendulum) oscillate around their minimum free energy state, whereas chemical oscillation is a non-equilibrium phenomenon that occurs on the way to the minimum free energy state. Apart from this, another difference is that while position and velocity in mechanical systems and voltage and current in electrical systems can both be either

positive or negative, concentrations can only be positive. Despite these differences, Samardzija et al. [92] provide a link between physical and chemical dynamical systems through a transformation that converts dynamical systems arising from mechanics or electrical circuits into chemical systems that exhibit, qualitatively, similar dynamic behavior. They illustrate their scheme with the harmonic and van der Pol [93, 94] oscillators, the Lorenz [95] and Rössler [96] attractors, and an RLC-circuit in series.

In Section 1.3.1.2, we noted that discrete stochastic CRNs can implement universal computation if one allows for an (arbitrary small) margin of error. Can continuous deterministic CRNs implement computation? Hjelmfelt and colleagues present a theoretical proposal that uses reversible catalytic chemical reactions to implement McCulloch-Pitts neurons [97, 98] (linear threshold units with binary output). With multiple neurons communicating with each other through excitatory and inhibitory catalytic reactions, they provide chemical implementations of Boolean logic gates [99], binary adders and finite space stack machines [100], associative memories [101] that are similar to Hopfield networks [57], and argue that, in principle, their techniques can be used to construct a chemical implementation of an (arbitrarily large) finite-space version of a Turing machine.

A different way to understand the class of behaviors possible in chemistry is to ask what behaviors are learnable by chemical kinetics. By this we mean: given data that arises from experiment or an (unknown) simulation, can we find out if it was generated by a continuous deterministic CRN with mass action kinetics? For one precise formulation of this problem, see Hárs and Tóth [102]; for a broad overview of this approach in systems and synthetic biology, see Engl et al. [103].

Lastly, while synthetic biology and nucleic acid nanotechnology have so far largely relied on implementing computations using the digital abstraction [104–109], analog approaches have recently been demonstrated [53, 59, 65, 110–113]. For a review of the benefits of using analog rather than digital computation for biomolecular circuits, see Sarpeshkar [114].

1.4 DNA strand displacement as a candidate architecture

1.4.1 Nucleic acid nanotechnology

In one of the early perspectives on "nanotechnology", Eric Drexler [115] articulated a vision to enable molecular engineering at the nanoscale. He suggested that one could, in principle, self-

assemble protein machines that fabricate other molecular machines; his inspiration was the ribosome, which is a self-assembled protein machine that in turn manufactures other molecular machines. However, this vision remains inaccessible today, partly because the protein folding problem remains intractable.

With the benefit of hindsight, it is tempting to argue that proteins were simply the wrong molecules to *start* with. In contrast to proteins, nucleic acids are simpler and better understood. (Here, we do not consider other biopolymers like lipids and carbohydrates, although they may also be viewed with a similar lens [116].) In particular, DNA has well-understood thermodynamic [117–120] and mechanical properties [121]. The stability of a DNA double-helix can be predicted from the (independent) contributions of each of its nearest-neighbor base pair stacks [117–119]. This fact has two immediate consequences. First, the secondary structure description is a very useful abstraction for the structure of a DNA molecule. Second, for a large class of structures, fast algorithms and efficient software can be developed for predicting folding from DNA sequence [120, 122, 123], and this in turn permits sequence design given a prescribed structure [123–129]. Moreover, as noted in Section 1.1, DNA synthesis is relatively better, faster, and cheaper.

The study of nucleic acid nanotechnology was initiated by two distinct streams of pioneering thought. In 1982, Ned Seeman proposed that synthetic DNA molecules could be designed to form immobile three-armed and four-armed junctions, which could in turn be used to create three-dimensional lattices [130]. He envisioned using these DNA lattices as scaffolds for facilitating the crystallographic study of protein structure. Although this particular application requires several technical challenges to be overcome [131], such as increasing the resolution of the lattice itself, and designing larger unit cells to incorporate the proteins to be studied, the original idea of self-assembling structures from synthetic DNA molecules has led to the development of *structural* DNA nanotechnology [10, 12, 132–135] and the emerging field of RNA nanotechnology [13, 136–138].

In addition to forming static structures, the idea that synthetic nucleic acids could be used to perform computation was first demonstrated by Len Adleman [139], who exploited the massive parallelism provided by many DNA molecules bumping into each other in a test-tube to solve a seven-node Hamiltonian path problem. As with Seeman's original purpose, multiple challenges to scaling up seriously limit the possibility of using DNA molecules to solve NP-hard problems (at least in the way that Adleman envisioned it) [140]. However, the insight that DNA molecules

could be programmed to perform information-rich functions has resulted in much interesting work: DNA has been taught to play a perfect tic-tac-toe strategy interactively against a human player [141], implement cellular automata [142], execute enzyme-free logic circuits [105], act as a catalyst [143], and as a programmable building block for dynamic self-assembly [144]. In particular, one reaction mechanism called toehold-mediated DNA strand displacement [145–148] has been a major workhorse of *dynamic* DNA nanotechnology [8].

1.4.2   DNA strand displacement

In much the same way that complex electrical circuits can be built with simple primitives, a molecular reaction mechanism called DNA strand displacement [145–148] has been leveraged to implement feedforward Boolean logic circuits [105, 107], catalytic reactions [143, 144, 149], and entropy driven networks [143]. Indeed, the rapid growth in complexity of these experimental systems has led to the widespread use of automation and software. For example, Visual DSD [150–152] is a commonly used custom programming language created for analyzing, verifying, simulating, and de-bugging strand displacement circuits by enumerating and simulating the corresponding DNA strand displacement reactions.

Although the experimental systems described above can be modeled and understood using the CRN formalism, their particular use of strand displacement may not generalize to a strategy for implementing arbitrary formal CRNs. However, since strand displacement has proven to be a technology that enables a wide variety of chemical systems to be built experimentally and yet naturally yields itself to abstract analysis in the form a programming language, it could potentially be a good general purpose molecular architecture for implementing formal CRNs.

However, for implementing prescribed dynamical behaviors in a test tube, our molecular architecture should meet two requirements. First, it should offer a simple and understandable way for tuning rate constants across a wide dynamical range. This is quite easy to do with DNA strand displacement: rate constants can be varied over a million-fold range simply by changing the length of a particular sub-sequence of the reacting DNA species called the "toehold".

Second, design pipelines work in an engineering process only when the parts and tools being employed are well-understood and quantitatively characterized. Therefore, in order to have any hope of succeeding in our endeavor, we need to quantitatively understand the biophysics and molecular mechanisms underlying DNA strand displacement, so that we can effectively design

and de-bug our experiments.

1.5 Summary of contributions

The scientific contributions presented in this thesis are summarized below.

Chapter 2 presents our work on the "device physics" of toehold-mediated DNA strand displacement. Even though DNA strand displacement has been a major workhorse of dynamic DNA nanotechnology, the biophysics and molecular mechanisms underlying strand displacement had not been studied in detail with a view to explaining strand displacement kinetics. In particular, state of the art models of strand displacement biophysics predicted a blunt-end strand displacement rate about 3 orders of magnitude faster than experimental measurements. We resolve this discrepancy and present a unified view of strand displacement biophysics and kinetics by studying the process at multiple levels of detail, using an intuitive model of a random walk on a 1-dimensional energy landscape, a secondary structure kinetics model with single base-pair steps, and a coarse-grained molecular model that incorporates three-dimensional geometric and steric effects. Further, we experimentally investigate the thermodynamics of 3-way branch migration. Two factors explain the dependence of strand displacement kinetics on toehold length: (i) the physical process by which a single step of branch migration occurs is significantly slower than the fraying of a single base pair, and (ii) initiating branch migration incurs a thermodynamic penalty, not captured by state-of-the-art nearest neighbor models of DNA, due to the additional overhang it engenders at the junction. Our findings are consistent with previously measured or inferred rates for hybridization, fraying, and branch migration, and provide a biophysical explanation of strand displacement kinetics. Our work paves the way for accurate modeling of strand displacement cascades, which would facilitate the simulation and construction of more complex molecular systems. Chapter 2 is based on our published research manuscript: Srinivas et al. [148]. My own personal contributions to this work are listed in Chapter 2.

Chapters 3 and 4 present our work on programming dynamical behaviors in chemical systems using DNA strand displacement. Inspired by living organisms, which employ chemical reaction networks (CRNs) with complex non-equilibrium dynamics to control physical processes, we seek to engineer CRNs with similar capabilities. Although relatively simple CRNs exhibiting complex dynamics, such as the B-Z oscillator, are well known, the systematic design and fabrication of

CRNs exhibiting prescribed dynamic behavior has proved challenging. Here, we demonstrate a general DNA-based technology to engineer CRNs with, in principle, any dynamic behavior that can be specified by a set of chemical reaction equations with prescribed rate constants (up to scaling). We successfully implement three distinct autocatalytic reactions, which we then combine into a *de novo* chemical oscillator. Unlike biological networks, which use sophisticated evolved molecules (like proteins) to realize such behavior, our test tube realization is the first to demonstrate that Watson-Crick base pairing interactions alone suffice for oscillatory dynamics. Since our design pipeline is general and applicable to any CRN, our experimental demonstration of a *de novo* chemical oscillator could enable the systematic construction of CRNs with other dynamic behaviors. Chapters 3 and 4 are based on a research article in preparation. My own personal contributions to this work are listed in Chapter 4.

2 Understanding the Biophysics and Kinetics of Toehold-Mediated DNA stand displacement

> *My dear sir, in this world it is not so easy to settle these plain things.*
> *I have ever found your plain things the knottiest of all.*

- Herman Melville, *Moby Dick*

My investigation into the biophysics, kinetics, and molecular mechanisms involved in toehold-mediated DNA strand displacement began as a quick (or so we thought) rotation project in Erik Winfree's lab. As we went deeper and deeper into the rabbit hole, we realized that understanding the dependence of strand displacement kinetics on toehold strength was much more complex than we had originally imagined. Indeed, this was one of those projects where the process of writing the paper clarified our thoughts greatly and helped us appreciate the nuances and form a coherent picture in our minds.

When I began my rotation, we had a working problem statement. Joseph Schaeffer had used the stochastic simulator he developed, Multistrand [153], to run simulations of strand displacement systems whose kinetics had been experimentally characterized by Zhang and Winfree [147]. To his surprise he found that the acceleration in strand displacement rate due to toehold strength predicted by his simulations was about three orders of magnitude lower than the experimental value. However, at this time these simulations were preliminary and it was not clear how sensitively this apparent discrepancy depended on various choices that had to be made (such as calibration data sets, dangle energy models, and the like). A back-of-the-envelope calculation on a simplified state space and energy model reproduced the discrepancy. A candidate explanation was also available: Bernard (Bernie) Yurke had hypothesized that state-of-the-art free energy mod-

els of strand displacement neglected a crucial "bump" that arises from volume exclusion between single-stranded DNA overhangs co-localized around a co-axial junction.

Although the following seems naive in hindsight, we expected that a single large thermodynamic effect of this sort would completely account for the discrepancy, and that we could test for such a large effect with very simple experiments. After careful investigation it turned out that the discrepancy is indeed not an artifact, and that it arises from a combination of thermodynamic and kinetic factors. Pinning down exactly which effects were important and how much of the discrepancy they accounted for proved to be the challenging part. One of our naive assumptions did turn out to be true: although the thermodynamic component was not as large as we had anticipated, it was significant and relatively easy to demonstrate experimentally. Indeed, Erik and I designed the DNA complexes for the temperature-dependent absorbance experiments in one afternoon and they demonstrated the effect quite dramatically (a melting temperature shift of $8°C$) in the very first experiment I did. This was very satisfying and Bernie's hypothesis proved to be right! Our biggest experimental challenge during this project turned out to be instrumentation: our spectrophotometer kept breaking down.

Our collaborators, Tom Ouldridge, Petr Šulc, Ard Louis and Jon Doye at Oxford University helped us identify the molecular mechanisms responsible for the thermodynamic and kinetic effects that we predicted and measured experimentally. These insights arose from simulations based on their coarse-grained molecular model of DNA, OxDNA [154, 155], which incorporates geometric and steric effects (unlike the models that myself, Joseph Schaeffer and Erik Winfree contributed, which were at a higher level of abstraction). Our collaboration with our colleagues at Oxford greatly contributed to this work and helped us tell a coherent, unified story. This collaboration arose entirely by chance, when Erik ran into Tom's poster at a conference and found that Tom's molecular model predicted a free energy "bump" consistent with our models and experimental measurements.

Lastly, I now list my contributions to this work. I came up with the Intuitive Energy Landscape (IEL) model, with feedback from Joseph and Erik. During the writing phase Tom also provided important feedback on certain choices in this model. Apart from the preliminary simulations with Multistrand, which were run by Joseph, I performed all other Multistrand simulations and analyzed the results, with feedback from Joseph and Erik. Erik and I designed the temperature-dependent absorbance (UV melt) experiments and I performed all the experiments and analyzed

the data. I wrote all the sections of the paper except for those based on the coarse-grained molecular modeling, which were written by Tom. All authors gave valuable feedback on all aspects of the paper. In particular, Erik and Tom contributed significantly to the introduction, discussion, and conclusions sections, which should be considered jointly written. This chapter was published as: Niranjan Srinivas*, Thomas E. Ouldridge*, Petr Šulc, Joseph M. Schaeffer, Bernard Yurke, Ard A. Louis, Jonathan P. K. Doye and Erik Winfree, "On the biophysics and kinetics of toehold-mediated DNA strand displacement", Nucleic Acids Research 41 (22): 10641–10658 (2013). * indicates that those authors contributed equally to this work.

Dynamic DNA nanotechnology often employs toehold-mediated strand displacement for controlling reaction kinetics. While the dependence of strand displacement kinetics on toehold length has been experimentally characterized and phenomenologically modeled, detailed biophysical understanding has remained elusive. Here, we study strand displacement at multiple levels of detail, using an intuitive model of a random walk on a 1-dimensional energy landscape, a secondary structure kinetics model with single base-pair steps, and a coarse-grained molecular model that incorporates three-dimensional geometric and steric effects. Further, we experimentally investigate the thermodynamics of 3-way branch migration. Two factors explain the dependence of strand displacement kinetics on toehold length: (i) the physical process by which a single step of branch migration occurs is significantly slower than the fraying of a single base pair, and (ii) initiating branch migration incurs a thermodynamic penalty, not captured by state-of-the-art nearest neighbor models of DNA, due to the additional overhang it engenders at the junction. Our findings are consistent with previously measured or inferred rates for hybridization, fraying, and branch migration, and provide a biophysical explanation of strand displacement kinetics. Our work paves the way for accurate modeling of strand displacement cascades, which would facilitate the simulation and construction of more complex molecular systems.

2.1 Introduction

Recent advances in DNA nanotechnology have enabled the construction of two and three-dimensional nanoscale structures [10, 132, 134, 135, 156–158]. Nucleic acids have predictable double-helical structure and generally well-understood thermodynamic [117–120] and mechanical [121] properties, which makes them excellent engineering materials. In addition to static structures, dynamic nanoscale devices such as circuits [104, 105, 107, 159], catalysts [143, 160], autonomous molecular motors [145, 161–163], and reconfigurable nanostructures [8, 9, 160, 164] have been engineered using DNA. Inspired by experimental advances, theoretical schemes have been proposed [110, 111] to engineer arbitrarily complex chemical dynamics using DNA. If successful, such efforts could enable dynamic DNA circuits to actively control nanoscale devices.

Unfortunately, the biophysical understanding of key kinetic phenomena remains underdeveloped relative to our knowledge of static properties, limiting the development of dynamic DNA nanotechnology. Here, we study the biophysical basis of a molecular mechanism called toehold-mediated strand displacement, which is central to many dynamic DNA devices built to date. Toehold-mediated strand displacement enables control over the kinetics of molecular rearrangement, allowing the engineer to program when and where specific steps take place in a molecular machine.

Figure 2.1 illustrates strand displacement using domain notation. A domain is a set of contiguous nucleotides designed to be either fully bound or fully unbound in stable configurations. Our system initially comprises a two-stranded complex (S) and a single-stranded *invader* (X). S consists of an *incumbent* strand (Y) bound to a *substrate* strand which has a single-stranded overhang called a *toehold*. The invader is fully Watson-Crick complementary to the substrate and may bind reversibly to it using the toehold domain (h). This binding is reversible because the toehold may *fray* and eventually dissociate. (We use the term 'fraying' to describe the disruption of base pairs at the end of a duplex; if all base pairs fray, the duplex melts or dissociates. Conversely, 'zippering' refers to when a new base pair forms at the end of an existing duplex.) Once the toehold is bound, the *overhanging* branch migration domain (b) of the invader may compete with the incumbent for binding with the substrate. As the incumbent and invader exchange base-pairs with the substrate, the branch point of the three-stranded complex moves back and forth. This *three-way branch migration* (henceforth, *branch migration*) process has previously been modeled as an unbi-
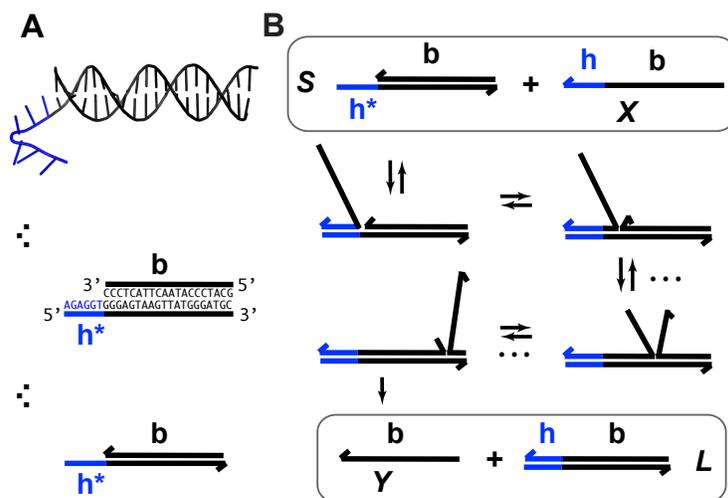
Figure 2.1: (A) Domain notation. Arrows indicate 3′ ends; * indicates Watson-Crick complementarity. (B) Toehold h mediates the displacement of the incumbent (Y) by the invader (X). Dots indicate branch migration intermediates which are not shown.

ased random walk [147, 165, 166] since each step causes no net change in base pairing. Eventually, the incumbent may dissociate, completing strand displacement. Overall, displacement is thermodynamically driven forward by the net gain in base pairs due to the toehold.

While it is known that bimolecular rate constants for strand displacement can vary a million-fold depending on the length of the toehold, little is known about the underlying biophysics or how to generalize the results to other reaction conditions or molecular modifications. For example, what are the mechanisms and intermediate states involved in a single step of branch migration? How do the kinetics of strand displacement depend on the length of the branch migration domain, or on the temperature and buffer conditions? How different is strand displacement in RNA and DNA? What is the effect of sequence mismatches at different positions along the branch migration domain? To begin answering these questions, a fundamental biophysical understanding is necessary.

However, the commonly accepted view of strand displacement biophysics appears at odds with kinetic measurements [146, 147]. Consider strand displacement with a 1-base toehold and a 20-base branch migration domain. Let us assume that the rate constant for the formation of the toehold base-pair is on the order of 10^6 /M/s. (This value for 1-nt association is within a factor of two of experimental values for 6-mer, 10-mer, 15-mer, and 20-mer hybridization rates [147,

167], after accounting for the linear length-dependence for short oligonucleotides [168].) Once the toehold has bound, there are two possibilities: (i) the toehold base pair could dissociate, leading to the dissociation of the invader, or (ii) the nearest base pair of the substrate-incumbent complex could fray, allowing the invader to compete to replace that base pair and complete the first step of branch migration. Assuming, as may seem reasonable, that the rate at which either base pair frays is similar, process (ii) should be approximately half as fast as process (i). This is because, once the substrate-incumbent base pair frays, there is a 50% chance of the invader replacing the frayed base pair, and a 50% chance of returning to the initial step. Once the first step of branch migration is complete, subsequent forward and backward steps are assumed to occur at the same rate. Therefore, the probability of successfully completing the remaining steps of branch migration before going back to the toehold-only-bound state is 1/20, from the gambler's ruin analysis (see Section 14.2 of Feller [169]). The lifetime of the three-stranded intermediate should be independent of concentration. Therefore, at low enough concentrations, the overall reaction can be modeled as an instantaneous second order process, as reported by Zhang and Winfree [147]. Under these conditions, the overall effective rate constant (k_{eff}) is given by a hybridization rate constant for the toehold, multiplied by a success probability of displacement once the toehold is bound. Note that, even though the time spent in the three-stranded intermediate is small, k_{eff} depends strongly on the probability of displacement once bound: $k_{\text{eff}} \approx 10^6 \times (1/3) \times (1/20) = 1.6 \times 10^4$ /M/s for a 1-base toehold. This is over three orders of magnitude larger than the 8 /M/s value measured by Zhang and Winfree [147]. This is a large unexplained discrepancy, despite the approximate nature of our calculation.

We now summarize the experimental evidence for the exponential acceleration in k_{eff} with toehold length, which was first reported by Yurke and Mills [146]. Zhang and Winfree [147] further characterized this exponential acceleration and confirmed that it saturates in the long-toehold limit. Relevant data from both studies (Figure 2.2A) suggests that the exponential acceleration in itself is not an artifact of particular sequences, although details may well depend on the experimental system (sequences, modifications for read-out, etc.) or conditions (salt, temperature, etc.). The kinetics of zero-toehold ("blunt end") strand displacement was investigated by Reynaldo et al. [170], whose measurement of 3.6 /M/s at 30 °C is similar to the 1.4 /M/s at 25 °C reported by Zhang and Winfree [147].

Zhang and Winfree [147] built a phenomenological model for predicting k_{eff} from toehold se-

quence (Figure 2.2B). They modeled branch migration using two macro-states I and J, containing the first and second half of the isoenergetic branch migration intermediates, respectively, which are connected by a simple transition with rate constant k_b (see Supplementary Figure S1 and Supplementary Section S1 for more details). Their model fits the data, but it is hard to physically interpret the transition between I and J. Branch migration is generally thought of as a random walk through many isoenergetic states, rather than a single reversible transition characterized by a first order rate constant. Without a more nuanced understanding of the process, it is unclear whether the fitted value of $k_b = 1.0$ /s can be justified on more fundamental biophysical grounds.

In this work, we model branch migration at a more detailed level that explicitly includes intermediates, thereby highlighting important thermodynamic and kinetic features of the process that are not evident from the phenomenological approach.

First, we analyze a 1-dimensional (single-pathway) model of toehold-mediated strand displacement called the intuitive energy landscape (IEL) model. Systematically exploring the parameter space of this simple model suggests that some combination of two factors could explain the dependence of strand displacement rate on toehold length: (i) the branch migration process is not isoenergetic and contains a free energy penalty for intermediate states, and (ii) branch migration is slow relative to fraying of the toehold.

To verify that these factors are not already implicit features of more detailed models of DNA secondary structure thermodynamics, we simulated the experiments of Zhang and Winfree [147]
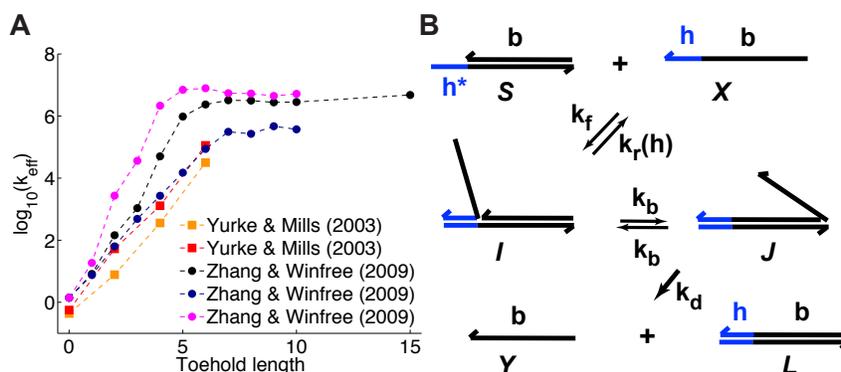
Figure 2.2: (A) Dependence of k_{eff} on toehold length, measured by Yurke and Mills [146] (at 20 °C, 1 M Na$^+$) and Zhang and Winfree [147] (at 25 °C, 12.5 mM Mg^{++}). Each curve is from a different toehold sequence. (B) The phenomenological model of Zhang and Winfree [147]. k_d is assumed to be very large relative to other rate constants.

using a secondary structure kinetics (SSK) simulator called Multistrand [153]. Multistrand incorporates extensive thermodynamic information — from state-of-the-art nearest-neighbor (NN) thermodynamic models of DNA secondary structure [117–120] — combined with a minimally parameterized rate model. Multistrand predicts that k_{eff} would increase by a factor of $10^{3.8}$ as toehold length increases from 0 to 15, in contrast to the experimentally observed factor of $10^{6.5}$. This quantitative discrepancy confirms that factors (i) and (ii) suggested by the IEL analysis are not already implicitly incorporated in more detailed models of DNA secondary structure thermodynamics.

We then experimentally investigate possibility (i) suggested by the IEL, by measuring the relative stability of complexes which mimic the geometric structure of branch migration intermediates. Our experiments provide evidence in support of a free energy penalty for branch migration intermediates that is not predicted by the NN models of DNA.

Finally, we use a recently proposed coarse-grained molecular model of DNA (oxDNA) [154, 155, 171], which incorporates more physical detail, including geometric and steric effects. In addition to correctly predicting the length-dependence of toehold-mediated strand displacement rates, oxDNA independently predicts the thermodynamic penalty for branch migration intermediates, suggesting it arises from local steric effects between single-stranded overhangs around the branch point. Further, oxDNA predicts that branch migration is indeed slower than fraying of the toehold, as it is geometrically more complex and necessarily involves more thermodynamically unfavorable steps.

2.2 Materials, Methods, and Results

2.2.1 Intuitive Energy Landscape model

Through a simple approximate calculation, we argued that the current view of strand displacement biophysics is at odds with kinetic measurements [146, 147]. We now perform a more rigorous intuitive analysis by building a simple single-pathway model called the Intuitive Energy Landscape (IEL) model. The IEL is simple enough that its kinetic predictions may be analytically or numerically calculated, for a given parameterization. We systematically explore the IEL's parameter space to obtain biophysical intuition and identify key thermodynamic and kinetic features essential for predictions to match experimental data.

State space. The IEL considers an invading strand and a substrate-incumbent complex con-
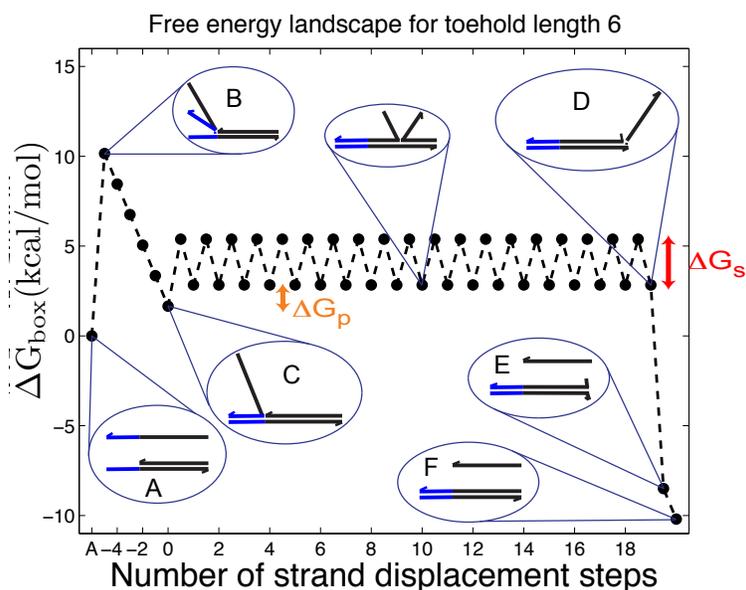
Figure 2.3: Free energy landscape of the IEL at 25 °C for a 6-base toehold. States A-F and the sawtooth amplitude (ΔG_s) and plateau height (ΔG_p) parameters are described in the text. $\Delta G_s = 2.6$ kcal/mol and $\Delta G_p = 1.2$ kcal/mol are used for illustration.

tained in a virtual box of volume V in solution, which defines a concentration u of 1 molecule per volume V. The state space of the IEL is illustrated in Figure 2.3. State A corresponds to the invader being unattached to the substrate-incumbent duplex. State B represents the formation of a first base pair within the toehold. For each additional toehold base pair that zips up, we define a new state, with state C indicating the fully-formed toehold. From here, each step of branch migration involves replacing an incumbent-substrate base pair with an invader-substrate base pair. We describe the stepping between these intermediates using a simple model in which the system must pass though a single effective transition state of raised free energy, so each complete branch migration step is shown as a single tooth of the "sawtooth" pattern between states C and D. This unknown effective transition state could be potentially as simple as a frayed substrate-incumbent base pair, but could also be more complex. The final stage of successful displacement involves the dissociation of the incumbent (state E) followed by the formation of the final base pair between invader and substrate (state F). Subtleties relating to the zero-toehold case are discussed in Supplementary Section S2 (see Supplementary Figure S2).

Energy model. The IEL models the free energy of the virtual box (ΔG_{box}) relative to state A.

ΔG_{box} is defined as in Figure 2.3. Initial binding (state A to B) incurs a free energy penalty of $\Delta G_{\text{init}} = \Delta G_{\text{volume}} + \Delta G_{\text{assoc}} \sim 9.95 + 1.90$ kcal/mol [120] (at concentration $u = 50$ nM) due to the reduction in the entropy of the box caused by lost translational and orientational degrees of freedom. ΔG_{assoc} is the free energy cost of association at a standard concentration of $u_0 = 1$ M, and $\Delta G_{\text{volume}} = RT \ln(u_0/u)$ is a correction for the actual concentration, where R is the universal gas constant and T is the temperature in Kelvin. Figure 2.3 shows the first base pair of the toehold adjacent to the helix, where it interacts favorably with the adjacent duplex end. For simplicity, we neglect sequence-dependent interaction strengths and use the average value $\Delta G_{\text{bp}} \sim -1.7$ kcal/mol for this and all other base pairs, as taken from the nearest-neighbor (NN) model [117–119] of DNA secondary structure thermodynamics. Formation of each successive base pair in the toehold therefore contributes ΔG_{bp}. Our effective transition states, which are local free energy maxima along the sawtooth, are raised ΔG_s above the branch migration intermediates with fully-paired substrate. As the physical details of branch migration steps are not well-understood, the effective "sawtooth amplitude" (ΔG_s) is an adjustable parameter. Changing ΔG_s allows branch migration and toehold melting rates to be independently modulated; for $\Delta G_s > |\Delta G_{\text{bp}}|$, branch migration is slow relative to fraying.

We also introduce a final parameter, a plateau height ΔG_p, which captures how the free energy of branch migration intermediates could vary with the structure of the branch migration junction. In particular, there is an asymmetry between state C and all other intermediates of branch migration: only one single-stranded overhang is present at the junction in state C, whereas an overhang protrudes from both sides for all other intermediates. We thus introduce a parameter ΔG_p that accounts for a possible free energy penalty due to the additional overhang, which could conceivably arise from entropic or electrostatic effects. Henceforth, IEL($\Delta G_s, \Delta G_p$) denotes a sawtooth amplitude of ΔG_s and a plateau height of ΔG_p (both in kcal/mol).

In an intuitive model like the IEL, one could indeed consider a range of alternative perturbations to the current biophysical understanding of strand displacement. ΔG_s and ΔG_p, however, parameterize in a simple way aspects of the branch migration process that are known to be poorly characterized. Furthermore, as we show later, experiments and more detailed modeling provide physical justification for the effects that these parameters represent.

Rate model. The system may undergo a transition to either of its neighboring states on the one-dimensional landscape. To ensure that stochastic simulations will eventually converge to the

thermodynamic (Boltzmann) equilibrium over the states, transition rates must satisfy detailed balance, i.e.,

$$\frac{k_{ij}}{k_{ji}} = e^{-\frac{\Delta G_{\text{box}}(j) - \Delta G_{\text{box}}(i)}{RT}}. \tag{2.1}$$

Here k_{ij} is the transition rate from i to j. Equation 2.1 only fixes the relative rates; in principle, for each pair of adjacent states i and j, an independent scaling factor could be chosen. Consequently, the number of candidate rate models is enormous. In the spirit of minimal parameterization in order to avoid over-fitting the data, we use only two independent scaling factors: k_{bi} for all bimolecular and k_{uni} for all unimolecular transitions.

For unimolecular transitions (all except A $\rightleftarrows$ B and D $\rightleftarrows$ E in Figure 2.3), we use a model in which all energetically downhill steps have a constant rate k_{uni}. This is a sensible first approximation given that these steps involve forming a base pair at the end of a duplex or stepping down from the transition states of branch migration, which may resemble forming a base pair at the junction.

In other words, for $\Delta G_{\text{box}}(i) > \Delta G_{\text{box}}(j)$,

$$k_{ij} = k_{\text{uni}} \tag{2.2}$$

and therefore,

$$k_{ji} = k_{\text{uni}} \, e^{-\frac{\Delta G_{\text{box}}(i) - \Delta G_{\text{box}}(j)}{RT}}. \tag{2.3}$$

This model is a continuous-time variant of the Metropolis [172] scheme for calculating thermodynamic averages.

For bimolecular transitions (A $\rightleftarrows$ B or D $\rightleftarrows$ E), the IEL assumes that complexes join at the constant rate $k_{\text{bi}} \times u$, and calculates the reverse rate by detailed balance. If i to j is a join step,

$$k_{ij} = k_{\text{bi}} \, u = k_{\text{bi}} \, e^{-\frac{\Delta G_{\text{volume}}}{RT}} u_0 \tag{2.4}$$

and

$$k_{ji} = k_{\text{bi}} \, e^{-\frac{\Delta G_{\text{box}}(i) - \Delta G_{\text{box}}(j) + \Delta G_{\text{volume}}}{RT}} u_0 \,. \tag{2.5}$$

We choose k_{bi} to be 3×10^6 /M/s, based on the hybridization rate constant fitted by Zhang and Winfree [147]. For simplicity, we choose k_{uni} such that the dissociation rate of the last base pair
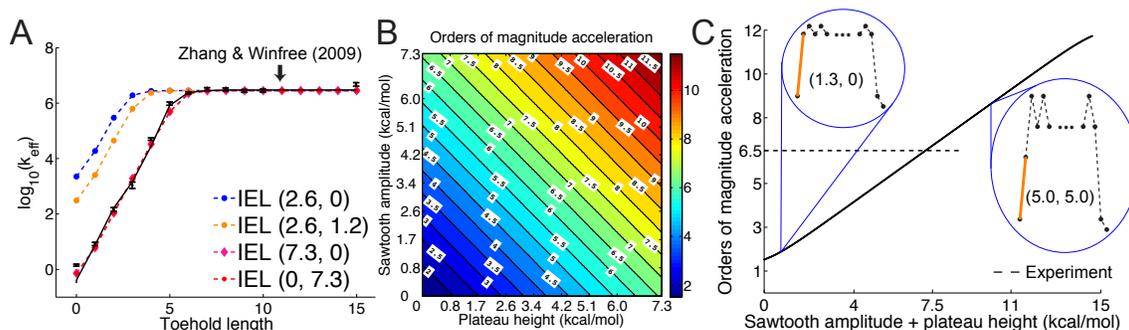
Figure 2.4: (A) Predictions of IEL(ΔG_s, ΔG_p) for different values of the sawtooth amplitude ΔG_s and plateau height ΔG_p. ΔG_{s+p} needs to be as high as 7.3 kcal/mol (> 4 base-pair stacks) to match experiment [147]. (B) Contour plot of orders of magnitude acceleration ($\mathcal{A}_{15,0}$) as a function of (ΔG_s, ΔG_p) for $\Delta G_s \in [0, 7.3]$ kcal/mol and $\Delta G_p \in [0, 7.3]$ kcal/mol. (C) Scatter plot of $\mathcal{A}_{15,0}$ vs ΔG_{s+p} using all the data in (B). The points plotted show almost no vertical spread, meaning that all points lie on a line. This indicates that ΔG_{s+p} is the predictive quantity. Inset: IEL(1.3, 0) and IEL(5.0, 5.0) for a 1-base toehold. The bimolecular initial binding step, which is identical in both landscapes, is marked in orange to highlight the contrast between the landscapes.

of the toehold (state B to A) is approximately equal to the fraying rate for every other base pair of the toehold (steps from state C towards B). This yields $k_{\mathrm{uni}} = 7.5 \times 10^7$ /s (see Supplementary Section S2 for details). This choice of k_{uni}, at the very least, ensures that two somewhat similar processes have similar rates.

Analytic and numerical calculations. We use an analytic formula for calculating absorption probabilities for a 1-dimensional random walk with absorbing boundaries [173] to calculate k_{eff} as a function of toehold length h for various values ΔG_s and ΔG_p (Figure 2.4(A)). (For details, see Supplementary Section S2.) To quantify the extent of control provided by toeholds, we define

$$\mathcal{A}_{h_2, h_1} = \log_{10}(k_{\mathrm{eff}}(h_1)) - \log_{10}(k_{\mathrm{eff}}(h_2)) \tag{2.6}$$

to be the orders of magnitude acceleration in k_{eff} as toehold length increases from h_1 to h_2. We will be most interested in $\mathcal{A}_{15,0}$, which corresponds to the difference between the leak rate and the maximal rate measured in Zhang and Winfree [147]. (We choose 15 rather than ∞ because the length-dependence of hybridization rates [168] suggests that experimental values for $\mathcal{A}_{h,0}$ may not be bounded, but for our purposes there is insignificant change past length 15.) The experimental results of Zhang and Winfree [147] can be matched by the IEL, but only with surprisingly large values of ΔG_s and ΔG_p. IEL(2.6, 0), with a plausible value of $\Delta G_s = 2.6$ kcal/mol (~ 1.5 base-pair

stacks) and an a priori choice of $\Delta G_p = 0$ kcal/mol, predicts $\mathcal{A}_{15,0} = 3.1$. This is 3.4 orders of magnitude smaller than the experimentally observed value of 6.5. Increasing either ΔG_s or ΔG_p serves to increase the predicted $\mathcal{A}_{15,0}$, as shown by the contour plot in Figure 2.4(B). The slope of contour lines suggests that $\Delta G_{s+p} \stackrel{def}{=} \Delta G_s + \Delta G_p$ is the key quantity. Indeed, a scatter plot of the dependence on ΔG_{s+p} (Figure 2.4(C)) using all the data in Figure 2.4(B) produces almost no vertical spread. $\Delta G_{s+p} = 7.3$ kcal/mol matches the experimentally observed value of $\mathcal{A}_{15,0} = 6.5$.

Other perturbations of the IEL model were unable to match the experiments. For example, we considered an alternative to the Metropolis method for setting unimolecular rates, the Kawasaki [174] method, which scales both uphill and downhill transition rates based on the corresponding change in ΔG_{box}. After rescaling k_{uni} so that dissociation of the last toehold base pair still occurs at the same rate as fraying, the Kawasaki method predicts a value for $\mathcal{A}_{15,0}$ within 10% of the Metropolis method (see Supplementary Figure S3). Going further, and considering that k_{bi} and k_{uni} may not be chosen ideally, we first note that logically, a uniform change to both rates will cancel and thus have no effect on $\mathcal{A}_{15,0}$. In contrast, IEL predictions do depend on the ratio $k_{\text{uni}}/k_{\text{bi}}$, which substantially affects the probabilities that initial contacts lead to successful zippering of the toehold and to successful branch migration before dissociation (see Supplementary Figure S4 and Supplementary Section S2). However, even implausibly low values of $k_{\text{uni}}/k_{\text{bi}}$ cannot account for the data without a large value of ΔG_{s+p}.

These features of the IEL can be understood through simple analytical approximations. Conceptually, we can split the strand displacement process into an attachment step (A→B), followed by success or failure of zipping up the toehold, followed by either dissociation of the toehold or successful displacement. We first treat the case of long toeholds, for which in the IEL model k_{eff} saturates at

$$k_{\text{eff}}(\infty) \approx k_{\text{bi}} \cdot p_{\text{zip}} \ , \tag{2.7}$$

with $p_{\text{zip}} = k_{\text{uni}}/(k_{\text{uni}} + k_{\text{bi}}\lambda)$ and $\lambda = e^{-(|\Delta G_{\text{bp}}| - \Delta G_{\text{assoc}})/RT} u_0$ giving the approximate probability that after making the first base pair, the invader does not dissociate and the remaining toehold bases zip up. (For long toeholds successful displacement is guaranteed once the toehold is formed.) For shorter toeholds that do not saturate k_{eff}, the probability of toehold dissociation

before branch migration dominates, and we can derive

$$k_{\text{eff}}(h) \approx \frac{k_{\text{uni}}}{2b} \cdot e^{\frac{h\,|\Delta G_{\text{bp}}| - \Delta G_{s+p} - \Delta G_{\text{assoc}}}{RT}}\, u_0 \,. \tag{2.8}$$

for $h \neq 0$ and, there being two ends where branch migration could start, twice the given value for $h = 0$.

We clearly see that the slope in Figure 2.4(A) is governed by ΔG_{bp}, while the role of b (Supplementary Figure S5) and ΔG_{s+p} in determining

$$\mathcal{A}_{15,0} \approx \log_{10} \frac{k_{\text{eff}}(\infty)}{k_{\text{eff}}(0)} \approx 0.9 + \log_{10}\left(\frac{2b\, e^{\Delta G_{s+p}/RT} \lambda}{k_{\text{uni}}/k_{\text{bi}} + \lambda} \right) \tag{2.9}$$

is immediate from their influence on $k_{\text{eff}}(0)$. The more subtle effect of changing $k_{\text{uni}}/k_{\text{bi}}$ results partly from effects on p_{zip}; a slightly more accurate approximation can explain the behavior seen in Supplementary Figure S4. Details on this and other derivations are given in Supplementary Section S2.

Matching both strand displacement and branch migration rates. The IEL analysis suggests that we can reconcile kinetic models of displacement at the base pair level with the experimental measurements of Zhang and Winfree [147] by incorporating a thermodynamic penalty for initiating branch migration (ΔG_p) and/or slowing down the rate of branch migration through a relatively large sawtooth amplitude (ΔG_s) associated with each step of branch migration. Although ΔG_{s+p} is constrained to be at least 7.3 kcal/mol in order to match the data, the individual contributions of ΔG_s and ΔG_p are not determined by the analysis.

To compare the inferences of the IEL analysis with experimental measurements, we need to account for the increase in initial binding rate (A→B) with toehold length, as observed for short oligonucleotide hybridization [168]. (For simplicity, the IEL assumes that the initial binding rate is independent of toehold length.) To this end, we also explored a variant of the IEL, called the Augmented Energy Landscape (AEL) model, that includes the linear increase in the formation rate of the first base pair (see Supplementary Section S3). For the AEL, we find that a smaller value of $\Delta G_{s+p} = 5.6$ kcal/mol is required due to the contribution of the binding rate to the overall acceleration. A good fit to the experimental data shown in Figure 2.4 is obtained for $k_{\text{bi}} = 3.3 \times 10^5$ /M/s and $k_{\text{uni}} = 8.2 \times 10^6$ /s (Supplementary Table S1).

We have not yet presented evidence to suggest that the plateau height ΔG_p is non-zero. However, if $\Delta G_p = 0$, a sawtooth amplitude of 7.3 kcal/mol (5.6 for the AEL) would be required to account for the data, which implies an average branch migration step time of ~ 3.0 ms (~ 1.6 ms for the AEL). This is much slower than experimentally inferred step times on the order of 12 - 20 µs [165, 166]. Therefore, simultaneously matching both measured branch migration and strand displacement rates requires a significant thermodynamic penalty to initiating branch migration.

The IEL analysis raises two important questions. First, could the necessary values of ΔG_s and ΔG_p represent features missing in the IEL's simplified thermodynamic landscape that are implicitly already present in more detailed models? If this is not the case, are ΔG_s and ΔG_p purely phenomenological parameters used to fit the data, or do they represent real physical effects that arise from the molecular properties of DNA?

2.2.2 Secondary structure kinetics model
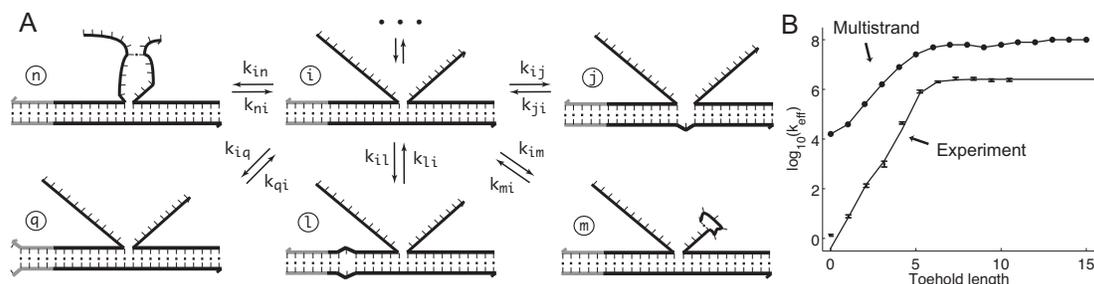
Figure 2.5: (A) Example states and elementary steps in Multistrand [153], a secondary structure kinetics (SSK) simulator. States illustrated are each adjacent to state i, as they differ from i by only one base pair. Transition rates are chosen to obey detailed balance. Dots and arrows at the top indicate other possible elementary steps from state i (not shown). (B) Multistrand predictions of experimentally measured [147] strand displacement rates as a function of toehold length. Experimental data points and error bars are from Zhang and Winfree [147]; the fitted line is their phenomenological model. Standard errors for Multistrand simulations are under 1% (not shown).

As a first step towards answering these questions, we use a secondary structure kinetics (SSK) simulator called Multistrand [153] to study strand displacement as a random walk on a more complex energy landscape, the NN secondary structure model [117–120], that incorporates a wealth of existing thermodynamic knowledge. Multistrand extends the Kinfold simulator [175] from single-strand landscapes to landscapes for multiple interacting nucleic acid molecules. Code implementing the Multistrand model is available for public download (see Supplementary Section S4

for details).

State space. Multistrand considers a set of strands in a virtual box of volume V in solution, which defines a concentration u of 1 molecule per volume V. The state space consists of all possible sets of Watson-Crick base pairs (such as states in Figure 2.5(A)), with two restrictions: (i) no base can have more than one pairing interaction and (ii) secondary structures containing pseudo-knots are not allowed. Structures without pseudo-knots are tree-like and have nested base pairing [120]. The size of Multistrand's state space grows exponentially in the number of bases [120].

Energy model. Let $\Delta G_{\text{box}}(i)$ be the free energy of the virtual box in state i, relative to a completely unstructured state with no base pairs. $\Delta G_{\text{box}}(i)$ is the sum of free energies of each isolated complex c, $\Delta G(c)$, in state i. The free energy of each complex is estimated using the NN model [117–119], which has been extended to multiple interacting nucleic acid strands [120]. NN parameters were measured [118, 119] in $1\,\text{M}\,\text{Na}^+$, which is roughly thermodynamically equivalent to the $12.5\,\text{mM}\,\text{Mg}^{++}$ used by Zhang and Winfree [147], according to salt corrections [176, 177] to the NN model. The NN model assumes that the free energy contribution of each base pair is dependent only on the identity and orientation (5′ or 3′) of its nearest neighbors on either side, and calculates $\Delta G(c)$ by summing contributions from each sub-structure ("loop") closed by a base-paired section where

$$\Delta G(c) = (L - 1)\Delta G_{\text{init}} + \sum_{\text{loop} \in c} \Delta G(\text{loop}). \tag{2.10}$$

L is the number of strands in complex c and $\Delta G_{\text{init}} = \Delta G_{\text{assoc}} + \Delta G_{\text{volume}}$ is, like in the IEL, the free energy cost of bringing two separate strands together. Stabilizing contributions to $\Delta G(\text{loop})$ mainly arise from base-pair stacks, and destabilizing contributions from the entropic cost of closing loops. At duplex ends in either interior or exterior loops, the first overhanging nucleotide contributes a *dangle* energy term [178]. When two duplex ends abut, as at a nick, the duplexes are known to *coaxially stack* onto each other [179–182], thereby stabilizing the structure; this interaction is not explicitly incorporated into Multistrand, although in some cases dangle energy terms partially account for it. Multistrand's energy model is identical to that used in NUPACK [123] and is thus very similar to those in Vienna RNA [183] and Mfold [184].

Rate model. Multistrand allows transitions between states i and j if they differ by a single base pair. The rate models we have explored for Multistrand are identical to the IEL, except for scaling factors. Like the IEL, Multistrand's predictions are not particularly sensitive to the choice between

standard unimolecular rate models: The predicted orders of magnitude acceleration in k_{eff} between toeholds 0 and 15 differ by less than 3% between Metropolis and Kawasaki (Supplementary Figure S6(B)). For Metropolis, $k_{bi} = 1.26 \times 10^6$/M/s and $k_{uni} = 4.4 \times 10^8$/s were calibrated [153] by fitting Multistrand simulations to experimentally measured DNA-DNA hybridization [167] and zippering [168] rates, respectively. Given the rate model, energy model, and current state, the choice and timing of the next transition is determined using a Gillespie algorithm [73]. Note that Multistrand allows any initial first pair of nucleotides to interact, each at the standard bimolecular rate, resulting in an increase in hybridization rates with increasing toehold lengths.

In principle, each i to j transition could have both forward and reverse rates scaled arbitrarily but equally, thus obtaining a distinct kinetic model that still satisfies detailed balance. However, doing so would amount to treating the microscopic barriers that impede transitions between the states of the NN model as different in every case. In the absence of additional information justifying such differences, the Metropolis approach constitutes a minimally parameterized rate model in which microscopic barriers associated with downhill processes (which typically involve base-pair formation) are assumed to be uniform, and their effects incorporated in the value of k_{uni}. As such, the Metropolis approach allows us to ask whether the free-energy landscape given by the NN model is sufficiently accurate and precise that it captures the factors that determine relative kinetics, and that finer details are only important in setting an overall rate constant k_{uni}.

Comparing Multistrand predictions to data. We simulated the "average strength toehold" experiments of Zhang and Winfree [147] and calculated k_{eff} as a function of toehold length. For technical details regarding the simulations, see Supplementary Section S4 and Supplementary Table S2. Multistrand captures the qualitative dependence of k_{eff} on toehold length, despite its minimally parameterized rate model. However, Multistrand predictions quantitatively diverge from experiment in two ways (Figure 2.5(B)): (i) the saturation value of k_{eff} for toeholds longer than 5 bases is roughly 20 times larger and (ii) increasing toehold length from 0 to 15 increases k_{eff} by a factor of $10^{3.8}$, in contrast to $10^{6.5}$ observed in experiment.

The first issue could be addressed by uniformly decreasing both k_{uni} and k_{bi}, which would shift the predicted curve down to match the data for toeholds longer than 5 bases. Indeed, this rescaling is also physically reasonable, as there is considerable variation among hybridization and fraying rates in the literature for nucleic acids [167, 168, 185–190].

Scaling k_{uni} and k_{bi} uniformly would not resolve the second challenge. However, like the IEL,

Multistrand is particularly sensitive to the ratio $k_{\text{uni}}/k_{\text{bi}}$. Decreasing k_{uni} slows down all unimolecular transitions - which brings the branch migration rate closer to experimentally inferred values, but makes fraying too slow. Two parameters, k_{uni} and k_{bi}, are simply not enough [153] to simultaneously match the four distinct time scales involved: rates of hybridization, fraying, branch migration, and branch migration initiation. Therefore, even unrealistically low choices of $k_{\text{uni}}/k_{\text{bi}}$ are unable to match observed acceleration in strand displacement rates due to toehold length (see Supplementary Figure S7).

The SSK analysis confirms that understanding what the IEL's ΔG_s and ΔG_p represent requires examining features not present in the NN model. Multistrand models branch migration as a fray–and–replace process and interprets the IEL's sawtooth transition state as one in which the substrate-incumbent base pair at the junction is frayed. This choice, when coupled with a k_{uni} calibrated to match fraying rates [168], results in a branch migration rate that is much faster than experimentally inferred step times [165, 166].

Indeed, the thermodynamics of the branch migration junction, e.g. states i and j in Figure 2.5(A), is not well characterized in the standard nearest-neighbor secondary structure model, as it involves overhangs, dangles, and coaxial stacking. Reflecting the lack of consensus, tools like NUPACK [123], Vienna RNA [183] and Mfold [184] offer several ways of treating dangle contributions; however, none of the three "dangle options" in the NUPACK energy model [120] improved Multistrand predictions (Supplementary Figure S6(A)).

2.2.3 Measuring relative stability of branch migration intermediates

For the IEL to match measured hybridization, fraying and branch migration rates, and hence strand displacement rates, a non-zero plateau height (ΔG_p) was necessary. We hypothesized that the initiation of branch migration incurs a thermodynamic cost due to the second overhang it engenders at the junction, even though the nearest neighbor model for DNA secondary structure thermodynamics predicts no such effect. We now present experimental evidence in support of this hypothesis by investigating the free energy landscape of branch migration. The biggest experimental challenge in measuring the relative stability of branch migration intermediates is that they cannot be easily isolated. Indeed, they are interchangeable as branch migration proceeds back and forth, with individual step times just tens of microseconds [165, 166].

System description. To overcome this issue, we designed immobile complexes Xi:Yj com-

prising hairpin `Xi` and strand `Yj` (Figure 2.6). `Xi` and `Yj` have poly-`T` overhangs of length i and j respectively. Varying (i, j) from $(20, 0)$ to $(0, 20)$ with $i + j = 20$ yields complexes which are "frozen snapshots" of branch migration, with no expected branch migration possible. `X20:Y00` mimics the binding of the invader by the toehold, while `X19:Y01` represents the displacement of 1 base, and so on. `X10:Y10` represents the "half way stage" of branch migration, while `X00:Y20` captures nearly successful displacement. Measuring the relative stability of these frozen snapshots is expected to be indicative of the relative free energies of branch migration intermediates. Experiments involving these complexes will henceforth be referred to as the *strand displacement snapshot* study. To investigate the consequences of short overhangs at the junction, we designed complexes `Xi:Yi` (for i = 0, 1, 2, 5 and 10). These experiments will be referred to as the *local overhang* study. All our complexes have the same base pairs at the junction and poly-`T` overhangs, while branch migration typically involves different bases at each step. We can therefore study the thermodynamic consequences of junction geometry, without the complication of sequence dependence.

Temperature dependent absorbance experiments. We measure the UV absorbance (at 260 nm) of each complex between 20 °C and 90 °C, at four different concentrations. Since the absorbance of single-stranded DNA (ssDNA) is higher than that of double-stranded DNA (dsDNA), and the fraction of ssDNA is dependent on the temperature, a temperature-dependent absorbance curve is obtained at each concentration (Supplementary Figure S8). Each complex `Xi:Yj` exhibits two transitions: the bimolecular, lower temperature transition and the unimolecular, higher temperature transition due to the hairpin in `Xi` closing or opening. The unimolecular transition was identified both by it being independent of concentration and by control melts involving the hairpins only (data not shown). At the concentrations chosen, the bimolecular and unimolecular transitions are distinct.

For each complex, we infer the enthalpy (ΔH°) and entropy (ΔS°) of formation by fitting the smoothed and normalized temperature-dependent absorbance curves (Figure 2.7) to a two-state

Domain	Sequence	Length
d	CCTCATCATACTACG	15
e	CTCCATGTCACTTC	14

Table 2.1: Sequences for domains from Figure 2.6, listed 5′ - 3′.
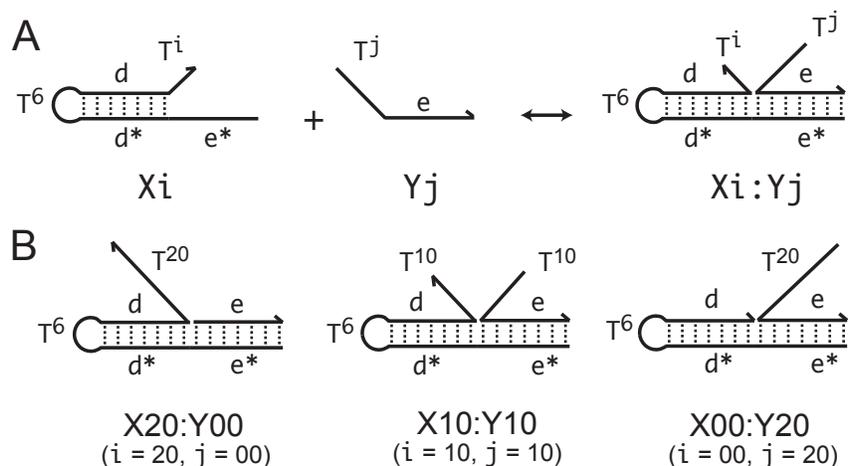
Figure 2.6: (A) Complex Xi:Yj comprises hairpin Xi and strand Yj, with poly-T overhangs of length i and j respectively. Domains d and e are designed to be orthogonal to each other and the overhangs (sequences in Table 2.1). (B) Varying (i, j) from $(20, 00)$ to $(00, 20)$ with $i + j = 20$ mimics the geometry branch migration intermediates (X20:Y00 (start), X10:Y10 (middle) and X00:Y20 (end) respectively). No branch migration is intended in these complexes.

model [191]:

$$\frac{[X_i : Y_j]}{[X_i][Y_j]} = \mathrm{e}^{-(\Delta H^\circ - T\Delta S^\circ)/RT} \tag{2.11}$$

where ΔH° and ΔS° are assumed to be temperature independent. We perform this fitting using a Bayesian analysis and confirm our findings using a simpler descriptive "leave-one-concentration-out" approach. Details are provided in Supplementary Section S5; see Supplementary Figure S9 and Supplementary Tables S3, S4 and S5.

From ΔH° and ΔS° we can calculate the free energy of formation, $\Delta G^\circ(T) = \Delta H^\circ - T\Delta S^\circ$. By comparing the free energies of different complexes, we infer the contribution of the poly-T overhangs. Our two-state assumption means that $\Delta G^\circ(T)$ is assumed to be linear in T – deviations from linearity limit the accuracy of extrapolation from values around the melting temperature of the complexes. Values of ΔG° at 55 °C, which is closer to the melting temperature, are plotted in Supplementary Figure S10.

Second overhang causes de-stabilization due to a local effect. Unlike the predictions of NN thermodynamic models [118, 120] of DNA, the immobile complexes we designed are not all of equal free energy; a second overhang at the junction causes a thermodynamic penalty (Figure 2.8A). At 25 °C, we infer a free energy penalty of ~ 2.0 kcal/mol (3.4 RT) as branch migration

proceeds from 0 to 10 steps (X20:Y00 vs. X10:Y10), with the majority (~ 1.5 kcal/mol) arising from the first step (X20:Y00 vs. X19:Y01). An approximately symmetric decrease is inferred for steps 11 to 20 (X10:Y10 vs. X00:Y20).

Since the de-stabilization due to an additional overhang plateaus so quickly, we suspect that the penalty is due to local effects at the junction, which is supported by the *local overhang* study (Figure 2.8B). Two 1-base overhangs on either side of the junction (X01:Y01) result in a penalty of ~ 1.4 kcal/mol relative to no-overhangs (X00:Y00). Lengthening the overhangs increases this penalty, but each additional base contributes progressively less, with an overall penalty of 3.0 kcal/mol and 3.2 kcal/mol respectively for 5-base (X05:Y05) and 10-base (X10:Y10) overhangs.

Our experiments suggest that current NN models of DNA do not capture the free energy landscape of strand displacement accurately enough to capture the kinetics of branch migration. This explains in part the inability of SSK models like Multistrand to match experimentally observed toehold-mediated acceleration.
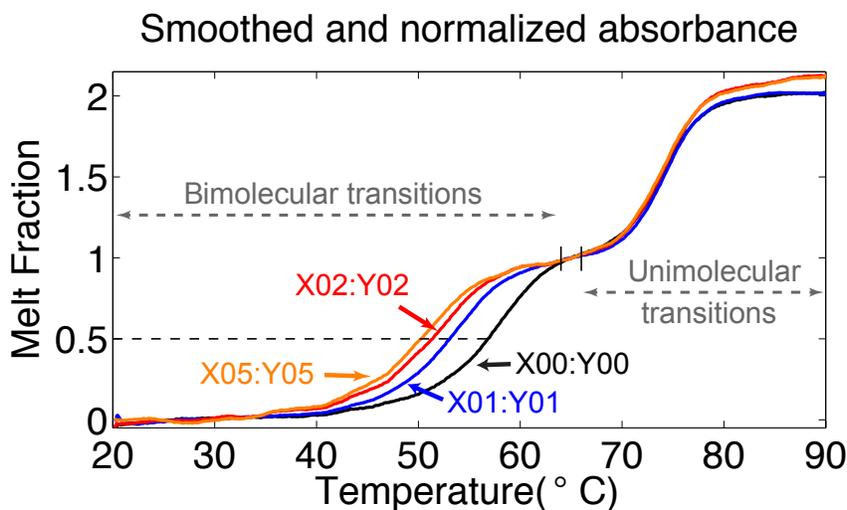
Figure 2.7: Smoothed and normalized UV absorbance data while annealing (at 200 nM). The lower temperature transition is the (bimolecular) formation of the complex, while the higher temperature transition is the (unimolecular) formation of the hairpin. The mean absorbance between 20 °C and 35 °C is normalized to 0 and that between 64 °C and 66 °C (indicated by vertical lines) to 1. The temperature range whose mean absorbance is normalized to 1 is concentration-dependent (Supplementary Table S3). Data acquired by annealing and melting are essentially superimposable. The dashed line indicates the halfway point of the bimolecular transition.
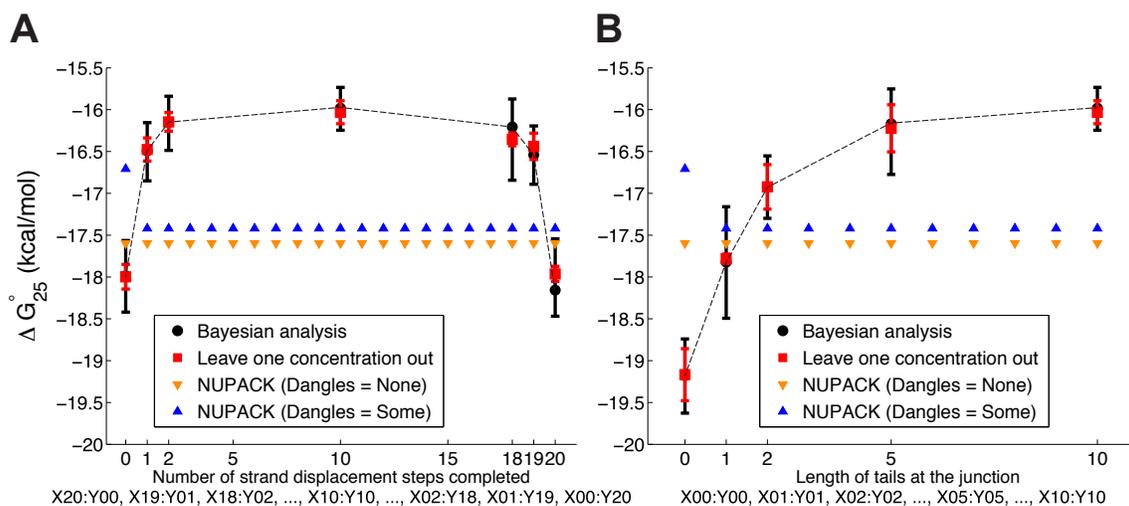
Figure 2.8: ΔG_{25}° of formation for complexes in the *strand displacement snapshot* study (A) and the *local overhang* study (B). Black error bars indicate Bayesian posterior means and 99% confidence intervals, while red error bars indicate means and standard deviations of leave-one-concentration-out least square fits. NUPACK predictions with dangles options "some" and "none" are provided for comparison.

2.2.4 Coarse-grained molecular modelling

A 3-dimensional model of DNA at the nucleotide level. Although it is possible to tune the IEL to agree with experimental data, and, moreover, the observed destabilisation of duplexes by ss-
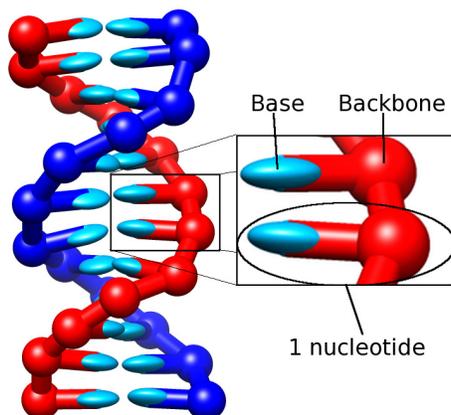
Figure 2.9: A 12-bp DNA duplex as represented by oxDNA. The enlarged section shows two rigid nucleotides, highlighting the backbone and base parts. The planarity of bases is captured through the orientational dependence of interactions. Image reproduced from Ouldridge et al. [192].

DNA overhangs appears to support a plateau during displacement, a physical explanation of the parameters required is important. Atomically detailed or coarse-grained molecular models have the potential to provide this kind of insight [193]. Here, we consider a recently proposed coarse-grained molecular model of DNA (oxDNA) [154, 155, 171], specifically the parameterization of Ouldridge [154]. Code implementing the model is available for public download (see Supplementary Section S6). In this model, illustrated in Figure 2.9, each nucleotide is a 3-dimensional rigid body, so that the state space of N model nucleotides has $6N$ spatial dimensions and $6N$ momenta. Pairs of nucleotides interact through a number of pairwise effective interactions (shown in Supplementary Figure S11), representing chain connectivity, excluded volume, hydrogen-bonding and stacking interactions between bases. The combination of nearest-neighbour stacking within a strand and hydrogen-bonding between complementary bases drives the formation of helical duplexes. We note here that oxDNA explicitly considers stacking interactions between pairs of bases. This is distinct from the base-pair stacks and coaxial stacks that appear in typical NN models, which involve four bases (two from each strand). Multiple interactions in oxDNA then contribute to the effective strength of base-pair and coaxial stacking, when the model is compared to NN descriptions.

OxDNA captures the thermodynamic and mechanical changes associated with the formation of duplexes from single strands, under high salt conditions. Specifically, it quantitatively reproduces the oligomer length-dependence of melting temperatures for the duplex transition, the widths of transitions, the elastic moduli of duplexes and the short persistence length of single strands. OxDNA includes the Watson-Crick rules of complementary base-pairing, but no further sequence dependence. Duplex formation was therefore fitted to the behaviour of SantaLucia's NN model [119] when averaged over sequence, and we can only compare directly to the average-strength toehold data of Zhang and Winfree [147]. OxDNA was fitted at $0.5\,\mathrm{M}$ [$\mathrm{Na^+}$], where electrostatic interactions are strongly screened – the repulsion of phosphates is therefore incorporated into the backbone excluded volume for simplicity. The experiments of Zhang and Winfree [147] were performed in a buffer of $12.5\,\mathrm{mM}$ [$\mathrm{Mg^{2+}}$], which is known to result in similar duplex formation thermodynamics to high monovalent salt buffers [119]. It is plausible that kinetics of strand displacement might be significantly different in these two buffers – the migration of Holliday junctions, for example, is known to vary between magnesium and sodium buffers [194, 195]. Given that oxDNA was fitted only to the duplex formation thermodynamics, however, there is no reason
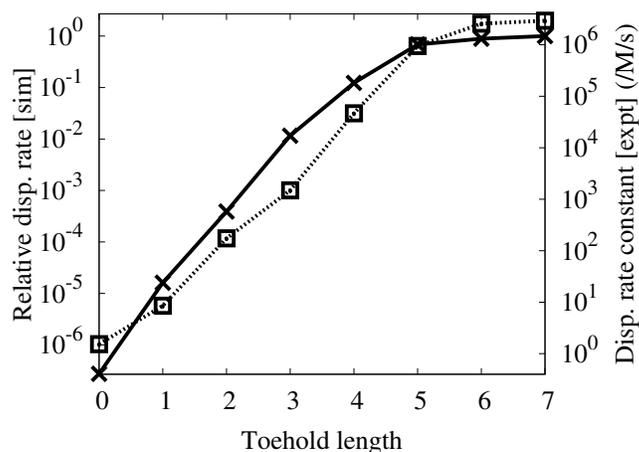
Figure 2.10: Rate of displacement, as a function of toehold length, observed in simulations (crosses, left axis). Details on the measurement errors of these values are provided in Supplementary Section S6. Also shown (squares, right axis) are the experimental data for the bimolecular rate constant of strand displacement for an average-strength toehold of varying length, taken from Figure 3(b) of Zhang and Winfree [147]. The logarithmic scales of both axes are identical up to a constant normalizing offset.

to assume that its kinetics and the representation of transition intermediates are more suited to one buffer than the other.

Although oxDNA contains many parameters, none were adjusted from those presented previously [154] for the purposes of this particular study. The parameter values employed have been seen to produce physically reasonable behaviour for a range of systems. Such systems include DNA tweezers [154], a DNA walker [196] and the interaction of kissing hairpins [197] and overstretching of DNA under tension [198]. Two of these systems, the tweezers and the walker, depend upon strand displacement and so oxDNA has previously been shown to reproduce this phenomenon. Furthermore, a barrier to initiating branch migration was predicted when studying the tweezers [154], although it was not studied in great detail.

Simulation techniques. We provide a concise summary of our simulation techniques here; for details, see Supplementary Section S6. We use two algorithms to simulate oxDNA: the 'Virtual Move Monte Carlo' algorithm (VMMC) of Whitelam *et al.* [199, 200] and the rigid-body Langevin Dynamics (LD) algorithm of Davidchack *et al* [201]. The first approach randomly attempts and accepts moves of clusters of nucleotides (illustrated in Supplementary Figure S12) in a manner that ensures the system samples from a Boltzmann distribution. The second incorporates noise

and damping terms into Newton's equations in a self-consistent manner, thereby generating states drawn from the canonical ensemble. The sampling of states is more efficient with VMMC and so it is used for all thermodynamic averages. Langevin algorithms, however, are explicitly dynamical, so they naturally give kinetic information that is harder to infer from VMMC. Consequently, all kinetic results quoted in this work use the LD algorithm.

Processes like strand displacement are computationally demanding to simulate, and thus require enhanced sampling techniques. We use umbrella sampling [202] to improve equilibration of thermodynamic averages in VMMC simulations. This method involves applying an artificial bias to lower free-energy barriers, thereby accelerating the transitions between (meta)stable states. For LD simulations of kinetics, we use forward flux sampling (FFS) [203, 204] to obtain accurate estimates of the relative rates of strand displacement for different lengths of toehold, as shown in Supplementary Figure S13.

OxDNA's energy, mass and length scales imply a time scale. Results in this Section are quoted in terms of this time scale for completeness. As we are using a coarse-grained model with an approximate model of dynamics, however, the absolute times and rates reported should not be over-interpreted. We focus on relative rates, which should be affected in a similar manner by the approximations in oxDNA and the algorithms used to simulate it.

Kinetic simulations of displacement. We consider a slightly truncated system (Supplementary Table S6) based on that used by Zhang and Winfree [147]. For computational simplicity, we removed the majority of the tail of the incumbent strand, which was used to interact with a reporter. Further, to simplify the order parameter for displacement, we use an approach in which only the expected (native) base pairs between the incumbent and the substrate or the invading strand and the substrate are given a non-zero binding strength. This simplification is reasonable because the sequences were designed to exhibit minimal secondary structure and cross-interactions when single-stranded. Additional simulations of toehold association were also performed to explore the consequences of this simplification: for further details, see Supplementary Section S6. Simulations of the three strands were performed in a periodic cell of volume 1.67×10^{-20} L for toehold lengths between 0 and 7 bases. We initialized the system at 25 °C with the incumbent fully bound to the substrate and the invading strand separate, and measured the overall rate of displacement using FFS. Order parameters and detailed results are given in Tables S7, S8, S9 and S10.
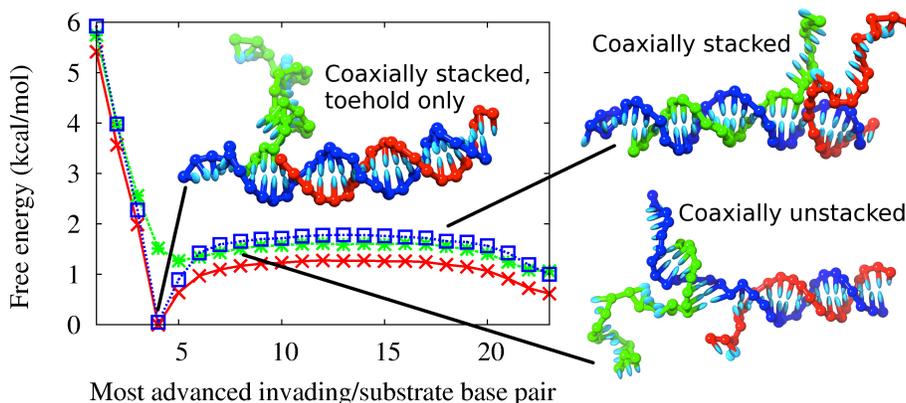
Figure 2.11: Free energy profile of displacement for a 4-base toehold. The red crosses show the free energy as a function of the index of the most advanced base pair between the invading strand and the substrate (base pair 1 is the base pair in the toehold furthest from the incumbent). These macrostates can be subdivided into those that involve coaxial stacking at the branch point (blue squares) and those that do not (green stars). The definition of which states are defined as stacked and which as unstacked is given in Supplementary Section S6. The simulation snapshots illustrate typical configurations of the macrostates indicated. In each of these images, the blue strand is the substrate, the red the incumbent and the green the invading strand.

Figure 2.10 reports the measured relative rates of displacement for different toehold lengths in simulation [147]. OxDNA agrees remarkably well with the experimental data: in particular, it shows the same exponential dependence on toehold length for short toeholds followed by a plateau for longer ones. The overall acceleration from 0-base to 7-base toehold is $10^{6.56}$, close to the experimental value of $10^{6.28}$. It would be unwise to put excessive emphasis on this agreement, but the base-pairing energies of oxDNA are fitted to the same secondary-structure free energies underlying the thermodynamics of Multistrand and the IEL, and therefore the predictions of these discrete models should be equally applicable to oxDNA as to real DNA. By analyzing oxDNA's representation of displacement, we now aim to physically justify the parameters used in the IEL.

Free-energy profile of displacement. We measure the free-energy profile of displacement to see whether oxDNA reproduces the experimental tendency of two ssDNA overhangs to destabilize a branch point and provides a physical explanation for it. We show the free energy of the three-stranded displacement complex as a function of the progress of branch migration, as measured by the identity of the base pair between invading and substrate strands closest to the 3′ end of the substrate, in Figure 2.11. We observe an increase in free energy of around 1.3 kcal/mol as branch migration is initiated, similar to the plateau height introduced to improve the IEL. Furthermore, as
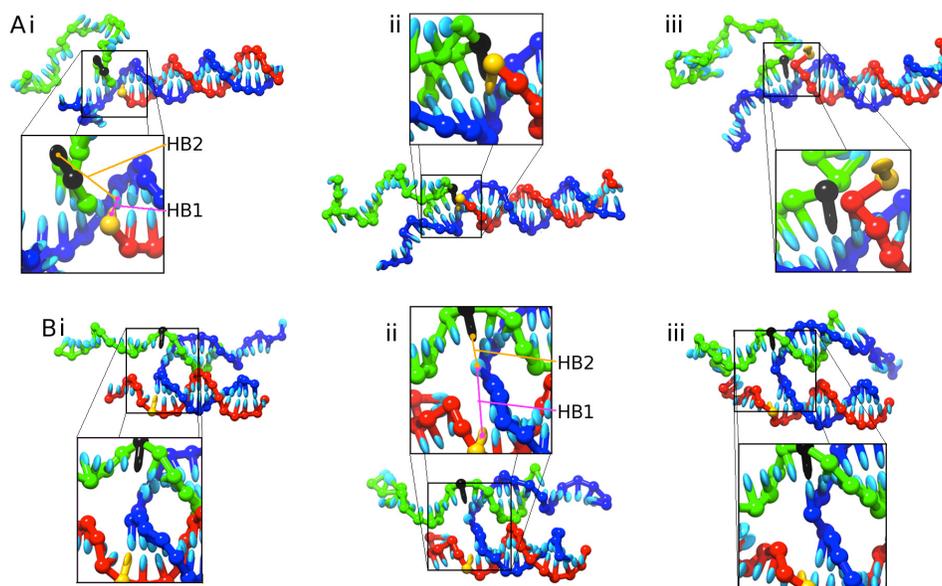
Figure 2.12: Typical branch migration steps as represented by oxDNA, for a toehold length of 3 bases. In each case, the nucleotide coloured black displaces the nucleotide coloured gold. A) Displacement via invasion: a base from the invading strand enters the incumbent/substrate duplex and competes with a base from the incumbent for base pairing. A.i) Initial state, with the gold-colored base on the incumbent bound to the substrate and the black-colored base on the invading strand unbound. A.ii) Metastable intermediate with both the gold and black bases competing for base pairing with the substrate. A.iii) Final state, with the black base bound to the substrate and the gold base unbound. B) Displacement via sequential disruption and formation of bonds: a base pair in the incumbent/substrate duplex breaks and the substrate base unstacks and moves across to the invading/substrate duplex. B.i) Coaxially unstacked initial state, with the gold-colored base on the incumbent bound to the substrate and the black-colored base on the invading strand unbound. B.ii) Metastable intermediate state, in which the base pair involving the gold base has broken, and the base from the substrate strand has unstacked. B.iii) Final state, with the substrate base now bound to the black base in the invading strand. Labelled distances between hydrogen-bonding sites are used as coordinates to monitor branch migration steps in detail later in Figure 2.13.

suggested by our strand displacement snapshot experiment, this barrier appears to saturate quite quickly: once a second ssDNA overhang of three or four bases has been created, there is negligible further increase in the destabilization. When the invading strand's single-stranded overhang is reduced to a few bases towards the end of branch migration, the free-energy penalty decreases (as in experiment).

The fact that the penalty saturates after around 4 bases suggests that the cause is local to the branch point. From looking at the branch migration intermediate in Figure 2.11 in which the helices are coaxially stacked at the junction, one can see that the branch point is densely packed
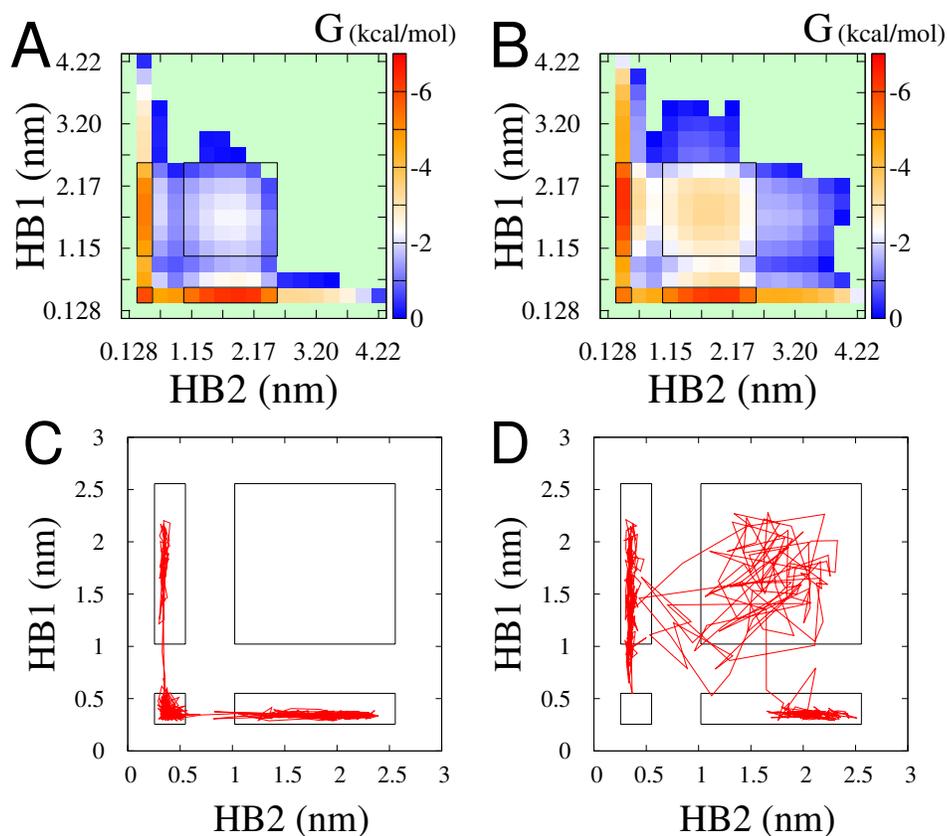
Figure 2.13: A) The free-energy landscape of the first branch migration step (for a 3-base toehold) as a function of the separation of hydrogen-bonding sites involved, obtained from umbrella sampling simulations. The distances HB1 and HB2 between hydrogen-bonding sites of nucleotides are illustrated in Figure 2.12 (A.i). In essence, they are distances of the competing bases in the invading (HB2) and incumbent (HB1) from their complement in the substrate. In these figures the black rectangles and squares highlight the same regions in each graph, roughly corresponding to the initial, final and intermediate states involved in branch migration. Free energy landscapes are measured in bins of $0.255\,\text{nm}$ – the labels indicate the values of HB1 and HB2 at the center of the bins. The arbitrary offset of G is chosen so that the most probable bin has a free energy $G_{\min} = -6.37\,\text{kcal/mol}$. All bins with a free energy greater than zero on this scale are shown in light green. B) An equivalent landscape to (A), but obtained at a later stage (step 11) of migration when branch migration via sequential disruption and formation of bonds is more common. C) An example trajectory showing the first step of branch migration occurring via invasion. D) An example trajectory from the 11th step showing branch migration via sequential disruption and formation of bonds.

Duplex length	T / °C	m	n	$\delta G_{m,n}$ (kcal/mol)
		10	0	1.46 [1.2]
6	25	0	10	1.38 [1.0]
		10	10	2.72 [3.2]
		10	0	0.87 [0.60]
8	55	0	10	0.94 [0.78]
		10	10	1.65 [1.79]

Table 2.2: Destabilizing $\delta G_{m,n}$ due to dangling ssDNA sections, as found in simulations. m is the length of the excess ssDNA attached to the hairpin, and n the length attached to the shorter strand. Values for $\Delta G_{m,n}$, the free energy of formation of the duplex, are extracted from simulations. The reported value, $\delta G_{m,n}$, is given by $\delta G_{m,n} = \Delta G_{m,n} - \Delta G_{0,0}$. $\delta G_{m,n}$ is then the destabilization of the duplex due to the ssDNA overhangs. The values in square brackets correspond to values inferred from our experimental studies for the closest equivalent lengths of dangling ssDNA. Note that $\delta G_{10,0}$ and $\delta G_{0,10}$ found here are compared to $\delta G_{0,20}$ and $\delta G_{20,0}$ from experiment – the local nature of the destabilisation makes this reasonable.

with nucleotides. In order to maintain coaxial stacking of helices at the junction, the two single-stranded overhangs must both take evasive action, unstacking and bending away from each other. The system can also reduce the overcrowding by breaking the coaxial stacking at the branch point, as in the coaxially unstacked image in Figure 2.11, but this carries a penalty itself. This overcrowding is worse with two ssDNA overhangs than one, resulting in a free-energy penalty for initiating branch migration and creating a second overhang.

Figure 2.11 also shows separate free-energy profiles for systems restricted to coaxially stacked or unstacked states (definitions of these states are provided in Supplementary Section S6). The existence of the overcrowding penalty makes the coaxially unstacked state (which is initially unfavourable, but involves minimal overcrowding) relatively more probable at later stages of branch migration. In the absence of coaxial stacking the plateau height is minimal, consistent with our diagnosis of overcrowding.

To complete the comparison between simulation and experiment, we have estimated the destabilizing effect of single-stranded overhangs on duplex formation within oxDNA. The systems are analogous to those used in our experimental studies and illustrated in Figure 2.6, but the hairpins have shorter stems of length 12 base pairs. Furthermore, to make comparisons at 25 °C and 55 °C for the chosen simulated concentration, we use complementary lengths for interstrand binding of 6 and 8 bases respectively ($\Delta G°$ can be inferred most accurately from experiments at ~ 55 °C, where the complexes melt, but we are most interested in the value at ~ 25 °C). The contribution of the single-stranded overhangs to the free energy of association $\Delta G°$ is expected to be indepen-

dent of the length of the duplexes, however, allowing a direct comparison of this property to be made. The sequences used are given in Supplementary Table S11 and further details are provided in Supplementary Section S6.

The results of the hybridization simulations are reported in table 2.2. The presence of two long ssDNA overhangs is destabilizing by $\sim 1.3\,\mathrm{kcal/mol}$ and $\sim 0.75\,\mathrm{kcal/mol}$ at $25\,°\mathrm{C}$ and $55\,°\mathrm{C}$ respectively, relative to the case with one single-stranded overhang. We conclude that oxDNA has a plateau height comparable to (but slightly smaller than) that found in experiment, with the caveat that we have not measured sequence-dependent effects at the branch point.
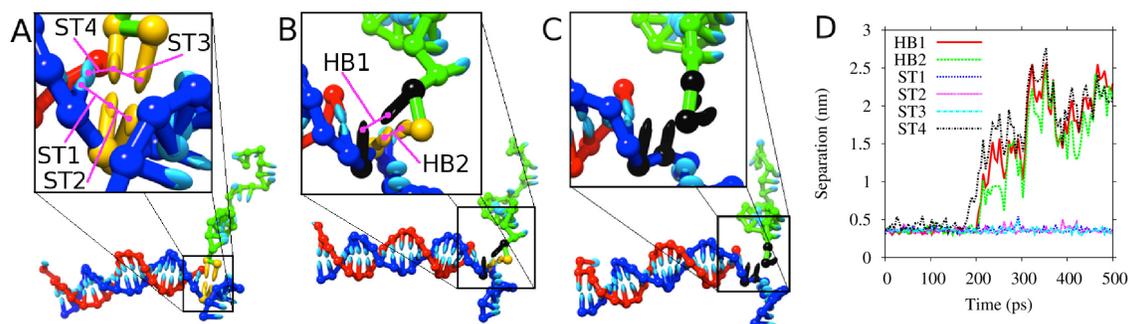
Figure 2.14: An example of melting for a 2-base toehold. As before, the blue strand is the substrate, the red the incumbent and the green the invading strand. A) Both base pairs of the toehold (shown in gold) are formed. B) one base pair (shown in black) is broken, one (gold) remains. C) Both base pairs (black) are broken. D) Plots of hydrogen-bonding and stacking site separation during this detachment, demonstrating the relatively minor disruption of interstrand stacking (only the stacking between incumbent and invading strand is disrupted) and rearrangement of structure necessary, and the proximity of the transition state to the fully bound state. The relevant distances for the stacking interactions are shown in A, and for the hydrogen-bonding in B. A, B and C correspond to times of 175 ps, 204 ps and 210 ps on the trajectory shown in D. Note that in this case, the system first breaks the intrastrand stack ST4, followed by base pairs HB1 and HB2. Trajectories with different orders are also observed.

Kinetics of branch migration and fraying. Here, we examine oxDNA's representation of the mechanism of branch migration in detail. Consistent with the literature [165, 166, 168], oxDNA predicts that branch migration is slow compared to the rate at which base pairs in the toehold fray, and provides a mechanistic explanation for this difference.

Typical branch migration steps observed in oxDNA are shown in Figure 2.12. Two types of process occur most commonly:

1. Invading and incumbent duplexes remain coaxially stacked at the branch point, and a single-stranded base from the invading strand enters the duplex region, competes for base-pairing

to the substrate with a base from the incumbent, and eventually displaces it. This process is illustrated in Figure 2.12 (A). We refer to this process as branch migration *via invasion*.

2. Invading and incumbent duplexes coaxially unstack at the branch point, and a base in the substrate is transferred from one duplex to the other. In this process, one base pair must fully break before the other forms some time later. An example of such a step is shown in Figure 2.12 (B). We refer to this process as branch migration *via sequential disruption and formation of bonds*.

Invasion is dominant at the start of branch migration, when the majority of systems are coaxially stacked. Branch migration via sequential disruption and formation of bonds becomes more relevant at intermediate stages, when coaxial stacking is less prevalent. The two mechanisms differ in whether or not the duplexes are coaxially stacked during branch migration, and whether the displacing base invades the duplex prior to disruption of the initial base-pairing. Invasion is geometrically infeasible from an unstacked state. However, the substrate strand could remain stacked at the branch point whilst 'flipping out' a base from the incumbent and only then replacing it with a base from the invader. This third type of process is rarer than the alternatives in oxDNA, but cannot be ruled out for real DNA. Importantly, all three processes require the disruption of favourable stacking interactions between neighbouring bases and considerable structural rearrangement, for *each* step of branch migration.

We also note that the branch migration pathways have relatively long-lived metastable intermediates (states represented by Figure 2.12 (A).ii and (B).ii). These intermediates can be clearly identified on free-energy landscapes for individual steps of the branch migration process that are plotted in Figure 2.13 (A) and (B). Typical trajectories of branch migration are shown in Figure 2.13 (C) and (D). Figure 2.13 (C), taken from the start of branch migration, clearly shows the system moving from one base-pairing configuration to another via a metastable intermediate in which both invading and incumbent bases are close to the substrate base (branch migration via invasion). Figure 2.13 (D), taken from a later stage of branch migration, shows a change in base-pairing occurring via the diffuse metastable intermediate in which both invading and incumbent bases are distant from the substrate base (branch migration via sequential disruption and formation of bonds). Note that trajectories of both types are possible at each stage, but that invasion is particularly dominant for the first step of branch migration. The fact that the intermediates are

metastable, despite their high free energy, indicates that they are not easily accessible from the typical configurations of the system.

By contrast, the fraying of a few base pairs in the toehold and the subsequent detachment of the invading strand can occur relatively easily. As illustrated in Figure 2.14 (A)-(C), it is not necessary to disrupt at least one stacking interaction for every base pair that is broken (as it is in branch migration). Further, transition states in fraying are much closer in configuration space to the typical states of the system than they are for branch migration, making them easier to access. This finding is corroborated by Figure 2.14 (D), which shows that the coordinates of the system do not have to deviate far from their typical values before detachment occurs. For comparison with Figure 2.13, Figure 2.15 (A) shows a free-energy landscape as a function of the separation of the base pairs in a two base-pair toehold. The minimum in the bottom left corresponds to the fully bound state, and the trajectory shown in Figure 2.15 (B) is the same as that in Figure 2.14, illustrating both base pairs breaking in quick succession and the system leaving the bound state.

Although the fine details of branch migration processes will be sensitive to detailed chemistry neglected by oxDNA, the conclusion that each step of branch migration necessarily involves the breaking of more stacking interactions and a greater structural rearrangement than fraying of base pairs in the toehold is likely to be a robust one. As a result, if toehold melting and branch migration are to be simultaneously characterised by the IEL, the sawtooth amplitude of branch migration should be larger than the cost of fraying a base pair. We have not directly attempted to infer rates or transition free energies for processes that would correspond to elementary steps in the IEL with oxDNA. As we have discussed, such elementary processes can be relatively complex in oxDNA, involving effects that cannot be captured at the secondary structure level, making the precise definition of rate constants difficult. Nonetheless, understanding the process at an effective secondary structure level is helpful: oxDNA then justifies tuning the IEL to use an effective sawtooth amplitude significantly larger than the free energy of a single base-pair stack to slow the rate of branch migration.

2.3 Discussion

We have argued that the kinetics of strand displacement can be explained by the destabilizing effect of single-stranded overhangs at the branch migration junction, and the relative slowness of branch migration compared to the fraying of a base pair in the toehold. How consistent are our
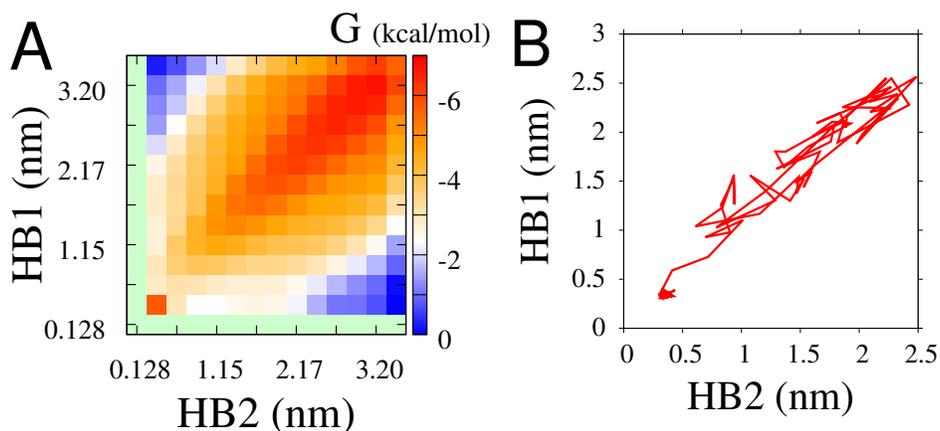
Figure 2.15: (A) Free-energy landscape of a system with a two base-pair toehold, with the system prevented from forming other base pairs between invader and substrate and also prevented from having either base pair separation exceed 3.7 nm. HB1 and HB2 are the base-pair separations defined in Figure 2.14 (B). Free energy landscapes are measured in bins of 0.255 nm; the labels indicate the values of HB1 and HB2 at the center of the bins. The arbitrary offset of G is chosen so that the most probable bin has a free energy $G_{\min} = -6.37$ kcal/mol. All bins with a free energy greater than zero on this scale are shown in light green. (B) The melting trajectory from Figure 2.14 projected onto the HB1, HB2 space.

findings and interpretations with the literature?

Our experiments infer two de-stabilizations: one due to the first ssDNA overhang at the junction (X20:Y00 vs. X00:Y00; ~ 1.2 kcal/mol at 25 °C) and another due to the addition of the second overhang (X10:Y10 vs. X20:Y00; ~ 2.0 kcal/mol). We have not investigated the sequence-dependence of these effects. Vasiliskov et al. [182] have measured the destabilization due to a single base overhang (analogous to X01:Y00 vs X00:Y00, which we did not measure) by immobilizing DNA oligonucleotides in 3-D polyacrylamide gel microchips. Most of their sequence-dependent values lie between 0.5 and 1.0 kcal/mol at 37 °C, which is similar to our temperature-extrapolated destabilization of 0.9 kcal/mol for X20:Y00 relative to X00:Y00. Note that we would expect the destabilization to be slightly less for X01:Y00 than X20:Y00 from the intuition derived from our experimental studies of two overhangs, which indicate that the destabilization is not fully saturated by a single-base overhang.

OxDNA suggests that the destabilization due to overhangs arises from steric interference, which can be relieved by breaking coaxial stacking at the junction. This suggests that ΔG_p (and, in fact, the free energy cost of introducing two overhangs to an overhang-free system) should not be significantly larger than the ΔG associated with the coaxial stacking transition in a nicked du-

plex. Several attempts to measure this quantity [180, 182, 205] have inferred values of $|\Delta G| \sim$ $0.2 - 2.3 \, \text{kcal/mol}$ at $37 \, ^\circ\text{C}$ and $[\text{Na}^+] = 1$ to $15 \, \text{mM}$. Increasing $[\text{Na}^+]$ to $100 \, \text{mM}$ was observed to increase stability of coaxial stacks by approximately $0.4 \, \text{kcal/mol}$. We therefore conclude that our temperature-extrapolated destabilization of $2.8 \, \text{kcal/mol}$ between X10:Y10 and X00:Y00 at $37 \, ^\circ\text{C}$ (and in high salt conditions) is not inconsistent with these data.

What do our findings imply for RNA? Walter et al. [206] have measured de-stabilization due to a single base overhang at a junction of duplexes as well as two single base overhangs on either side of a junction for RNA. They report a $2.8 \, \text{kcal/mol}$ de-stabilization for their analog of X01:Y01 vs X00:Y00 at $37 \, ^\circ\text{C}$, which is much higher than our extrapolation of $1.4 \, \text{kcal/mol}$ for DNA. Since the geometry (A-form helix) and thermodynamics of RNA are significantly different from that of DNA, we would indeed not expect analogous measurements for RNA to be quantitatively similar. For DNA, we find that X20:Y01 is significantly less stable than X01:Y01, and if this holds true for RNA we would expect ΔG_p for RNA to be significantly higher than the $2.0 \, \text{kcal/mol}$ at $25 \, ^\circ\text{C}$ we infer for DNA. However, *efn2*, a NN thermodynamics model for RNA that has been expanded to include experimentally-measured terms for coaxial stacking and 1-nt and 2-nt overhangs at junctions [207, 208], does not predict a free energy penalty similar to that which we infer for DNA strand displacement (see Supplementary Section S7 and Supplementary Figure S14), possibly due to incorporating these terms only in 'interior loops'.

Is the IEL's inferred value of ΔG_s consistent with measured branch migration rates? Using the experimentally measured plateau height $\Delta G_p = 2.0 \, \text{kcal/mol}$ and the corresponding inferred values of $\Delta G_s = 5.3 \, \text{kcal/mol}$ for the IEL and $\Delta G_s = 3.6 \, \text{kcal/mol}$ for the AEL, these models predict branch migration step times of $\sim 103 \, \mu\text{s}$ and $\sim 53 \, \mu\text{s}$ respectively. These values are now more in line with previous experimental measurements: Radding et al. [165] report an average branch migration step time of $12 \, \mu\text{s}$ at $37 \, ^\circ\text{C}$ in $10 \, \text{mM Na}^+$, while Green and Tibbetts [166] estimate an upper limit of $20 \, \mu\text{s}$ at $65 \, ^\circ\text{C}$ in $0.3 \, \text{M Na}^+$. Our estimates are therefore somewhat larger, although it would be expected that the low temperature, high salt conditions of Zhang and Winfree [147], which stabilize helices, would reduce branch migration rates. Indeed, such an effect is reported by Radding et al. [165]. Finally, we note that Zhang and Winfree [147] infer a branch migration step time of around $2.5 \, \text{ms}$ from their phenomenological model, about 12 to 24 times larger still than our estimates. We attribute this difference to the absence in their model of a free energy penalty for initiating branch migration, i.e. what we here call ΔG_p.

Are the IEL's values for k_{uni} and ΔG_{bp} compatible with measured fraying rates? Wetmur and Davidson [168] have inferred fraying rates from temperature-jump experiments. They predict 0.04 µs for fraying of a base pair at 25 °C, which is faster than our values (0.2 µs for the IEL, 2 µs for AEL) and consistent with fraying being at least an order of magnitude faster than branch migration steps. We note, however, that these reported values were not directly measured in experiments, and required modeling to extract. Futhermore, related measurements of the kinetics of RNA base pair fluctuations found significantly slower rates, with fraying times on the order of microseconds at 25 °C [185–187]. We therefore conclude that our parameterization is not inconsistent with well-established kinetic properties of nucleic acids.

We argue that the slowness of branch migration initiation relative to fraying is a key aspect in understanding strand displacement. In contrast, Reynaldo et al. [170] explained the low rate constant of zero-toehold displacement by positing that a certain number of base pairs, $n = 3$ or 4, are necessary to form a stable duplex, and thus shorter duplexes can be ignored or assumed to detach instantly. Our models differ by assigning favorable and increasingly strong thermodynamic energies for toeholds attaching by 1, 2, or more base pairs; there is no penalty for short duplexes. Instead, we can reinterpret their n as the value for which the rate of fraying n base pairs ($k_{\text{uni}}e^{-n|\Delta G_{\text{bp}}|/RT}$) equals that of initiating branch migration ($k_{\text{uni}}e^{-\Delta G_{s+p}/RT}$), which for the IEL gives $n = 4.3$. A similar calculation for the AEL gives $n = 3.8$.

2.4 Conclusions

We explain the dependence of strand displacement kinetics on toehold length using two factors: (i) the physical process by which a single step of branch migration occurs is significantly slower than the fraying of a single base pair and (ii) initiating branch migration incurs a thermodynamic penalty not captured by state-of-the-art nearest neighbor models of DNA, due to the additional overhang it engenders at the junction. The slowness of branch migration relative to fraying is captured in the IEL by a sawtooth amplitude $\Delta G_s = 5.3$ kcal/mol, which is significantly larger than a single base-pair stack ($|\Delta G_{bp}| = 1.7$ kcal/mol). OxDNA provides physical justification for this by suggesting that branch migration steps are slower than fraying as each stage necessarily involves greater structural rearrangement and disruption of favorable stacking interactions. Initiating branch migration is slower than the average branch migration step because of the free

energy penalty (ΔG_p) incurred; our experiments infer $\Delta G_p \sim 2.0$ kcal/mol. OxDNA reproduces this penalty and suggests that it arises from steric interference of the single strands and duplexes at the branch migration junction. Specifically, the presence of an extra single-stranded overhang after the first step of branch migration causes additional disruption of (coaxial) stacking and reduction of conformational freedom as the strands are forced to bend away from each other.

Our work shows that toehold-mediated strand displacement involves four distinct time scales—rates of hybridization, fraying, branch migration, and branch migration initiation—coupled with the energetics of base pairing in the toehold. The IEL model captures these rates via the parameters k_{bi}, k_{uni}, ΔG_s, ΔG_p, and ΔG_{bp}. Presuming that these rates and energies are the essential determinants of strand displacement kinetics, the IEL analysis provides a framework for making kinetic predictions about different strand displacement-based systems. For example, the IEL predicts that for short toeholds, k_{eff} scales as approximately $1/b$, where b is the branch migration domain length (see Supplementary Section S2 and Supplementary Figure S15). However, the dependence of k_{bi}, k_{uni}, ΔG_s, ΔG_p, and ΔG_{bp} on reaction conditions (e.g. salt concentrations and temperature) are only partially understood, impeding quantitative predictions for differing reaction conditions. Nonetheless, our current understanding does suggest that determining just those five parameters should be sufficient to predict toehold-mediated strand displacement under new conditions.

We expect that the IEL framework is also suitable for describing other nucleic acids. Toehold-mediated strand displacement has been demonstrated with RNA [209, 210], but the dependence of kinetics on toehold length has not been characterized. Although several experimental studies have reported hybridization and fraying rates for RNA that are within the range observed for DNA [185, 187, 211–213], branch migration and branch migration initiation rates are not well understood. There is, however, some evidence that with 1-nt toeholds, strand displacement rates for RNA are considerably slower than those for DNA [214], while the average base pairing energy is stronger [207]; these observations are enough to predict that a plot of $\log_{10} k_{\mathrm{eff}}(h)$ for RNA would have a lower y-intercept but higher initial slope than that for DNA. Broadly, then, we would expect the qualitative features of RNA strand displacement kinetics to be similar to that of DNA; quantitative understanding will require further experiments.

Similarly, the effect of structural variations of toehold-mediated strand displacement, such as associative and remote toeholds [215, 216], or initial mismatches in the branch migration domain [217] could be understood in terms of their effects on the fundamental rates and energies.

These particular variants are likely to slow down branch migration initiation, and hence increase $\mathcal{A}_{15,0}$, which could in turn permit greater design flexibility while engineering nucleic acid devices and systems.

Although the IEL provides a simple framework for understanding strand displacement, it does not predict or explain the underlying biophysical mechanisms, nor can it easily be adapted to handle more complex structural variants or explain sequence-dependent behavior in larger strand displacement cascades and dynamic DNA nanotechnology [8], where the kinetics of both on- and off-pathway reactions (such as "leak" reactions analogous to zero-toehold strand displacement) are of great current interest. OxDNA is well suited to studying the underlying biophysical mechanisms and behaviors of structural variants, but it is computationally expensive for larger systems.

Efficient, quantitative simulation of sequence-dependent kinetics for both desired and spurious strand displacement pathways would facilitate design and optimization of large systems *in silico*. SSK models based on NN thermodynamics, such as Multistrand, would be well-suited for such purposes, but as found here, current implementations require adjustment to match critical thermodynamic and kinetic features. Thermodynamically, accurate coaxial stacking terms and penalties for overhanging single-stranded sections of DNA at junctions should be incorporated into the model. Incorporating the relative rates of fraying and branch migration within the Multistrand model faces several challenges. Because every secondary structure state has a well-defined free energy, branch migration intermediates at the 'top of the sawtooth' (frayed junctions with two tails) cannot simply be pushed to higher energies, as was done for the IEL using ΔG_s. Moreover, Multistrand currently incorporates only one of the two possible branch migration mechanisms suggested by oxDNA (sequential disruption and formation of bonds). Three independent approaches could conceivably be pursued to better capture branch migration kinetics. First, an additional unimolecular rate constant parameter could be used for transitions local to the junction. Second, new transitions, such as *shift* moves in Kinfold [175, 218], could be incorporated to model alternative branch migration pathways. Third, Multistrand's state space could be augmented, possibly by including features such as non-base-pair stacking interactions.

Supplementary Information. Supplementary Tables S1-S10, Supplementary Figures S1-S16 and Supplementary References [192, 219–221] are available in the appendix.

3 Adventures in programming and debugging molecular reaction networks

As soon as we started programming, we found to our surprise that it wasn't as easy to get programs right as we had thought. Debugging had to be discovered. I can remember the exact instant when I realized that a large part of my life from then on was going to be spent in finding mistakes in my own programs.

- Maurice Wilkes discovers debugging, 1949

I began working on engineering dynamic chemical reaction networks in July 2010, effectively right after I joined Erik Winfree's lab. This was a collaboration with David Soloveichik and Georg Seelig at the University of Washington. The foundation for this project was an earlier theoretical proposal by David, Georg and Erik to "translate" arbitrary chemical reaction networks (CRNs) to DNA [110]. Their proposal is essentially an algorithm which, given a set of chemical reaction equations and rate constants, provides a molecular DNA-based implementation with reactions being mediated through DNA strand displacement cascades.

The basic idea of the project was to investigate the experimental challenges involved in implementing chemical reaction networks with this theoretical proposal. If those experimental challenges could be identified, understood, and overcome, our work could open the door towards engineering essentially any dynamical behavior in a test tube. David, Georg, and Erik had already been working on the initial stages of this project when I started.

Personally, my own excitement about this project stems from two observations. First, since the scheme is completely general, it can be used to build not just one interesting dynamical behavior, but essentially any behavior. Therefore, any understanding we gain about overcoming experimental challenges specific to the scheme could be used to build a systematic "pipeline" for engineering arbitrary CRNs. Second, if we can learn how to engineer CRNs with prescribed dynamic behavior,

we could in principle use that knowledge to engineer chemical control modules for nucleic acid nanostructures, in much the same way we write computer programs to control electromechanical machines.

Since our goal was to learn how to program dynamical behaviors in chemical systems, we picked a challenging test case: a chemical oscillator. If we succeeded, this would be the first chemical oscillator consisting only of DNA, without the use of any complex, evolved molecules whose mechanisms of action we do not understand as well (such as proteins). Indeed, prior to our attempt, it was unclear whether the relatively simple DNA-DNA interactions based only on Watson-Crick complementarity would suffice to engineer oscillatory dynamics.

Although my investigation into the biophysics and kinetics of toehold-mediated DNA strand displacement (Chapter 2) began completely independently, there proved to be significant synergy between the two projects. What we learned about the biophysics and molecular mechanisms underlying strand displacement proved very helpful while de-bugging our experiments and re-designing our sequences.

Before we can hope to program molecular reaction networks, we need both a programming language and an implementation architecture. Section 3.1 introduces "formal" CRNs, our programming language of choice. Section 3.2 details the DNA strand displacement architecture we use to implement our CRN programs. Sections 3.4 to 3.8 discuss the multiple DNA sequence designs (which are particular instances of the DNA strand displacement architecture in section 3.2) whose performance we investigated experimentally. In essence, those sections summarize our adventures while de-bugging molecular reaction networks. A summary of my contributions to this effort is provided in the first part of Chapter 4.

3.1 The programming language: formal chemical reaction networks

For an introduction to formal chemical reaction networks and the diverse viewpoints with which they have been studied, see Section 1.3.

Chemical reaction equations, coupled with mass action kinetics, have been used to model and predict the dynamical behavior of well-mixed chemical systems for about a hundred and fifty years. Indeed, chemical systems that exhibit complex dynamical behavior, such as oscillations, feedback control and regulation, memory, logic, and chaos, can be accurately described using a
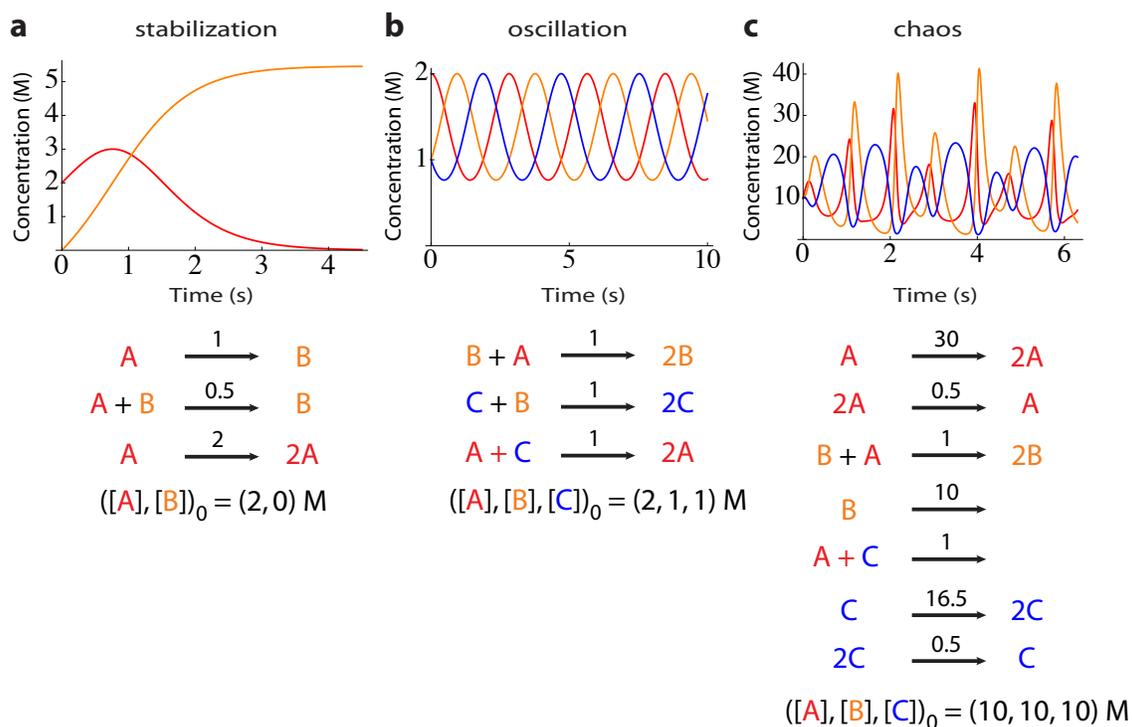
Figure 3.1: Examples of formal CRNs exhibiting different dynamical behaviors in the mass action setting (based on numerical solutions to mass action ODEs). This figure was inspired by one of David Soloveichik's slides introducing CRNs.

set of coupled chemical reactions with rate constants [19, 24, 222, 223]. This language of chemical reaction networks (CRNs) is thus a powerful descriptive tool.

In this section, instead of defining a CRN rigorously, we will loosely define a CRN as a finite set of chemical reaction equations (with rate constants) between (formal) species. For our purposes, we will assume that the dynamical behavior of our CRNs are specified by mass action kinetics, although, as described in Section 1.3, this is not the only choice. Figure 3.1 shows some example CRNs exhibiting different dynamical behaviors. Note that formal CRNs include open systems, and that there is no restriction for the number of molecules ("mass") to be conserved. Physically, a reaction with the appearance or disappearance of molecules should be interpreted as representing flows of matter and free energy that are not modeled by the particular formal CRN.

In this thesis, we think of CRNs as a prescriptive programming language rather than a descriptive modeling language. This approach was pioneered by Soloveichik et al. [110], who asked the

following "inverse" question: given a finite set of formal chemical reaction equations (with specified rate constants) between formal species $X_1, X_2, ..., X_n$, can we design a set of "real" molecules $M_1, M_2, ..., M_p$ that interact in a well-mixed solution to approximate the mass-action kinetics specified by the formal system?

Of course, this question regarding a general strategy for implementing arbitrary formal CRNs is interesting only if CRNs are capable of exhibiting a wide range of dynamical behaviors. This is indeed the case: given a system of polynomial ODEs with nonnegative integer powers, one may explicitly construct a formal CRN some of whose species will approximate the solution to the system of ODEs on the positive orthant, up to arbitrary accuracy over any time interval [87, 88]. In fact, the CRN constructed has particularly nice properties: (i) all reactions follow conservation of mass, (ii) have at most two reactants and two products, and (iii) no reactions are autocatalytic.

Soloveichik et al. [110] show that, given a formal CRN, it is indeed possible to engineer a molecular implementation that will, assuming certain "fuel" species are in large excess, approximate the mass action kinetics specified by the formal CRN (up to scaling rate constants). Indeed, they provide a construction which "compiles" any set of formal chemical reactions into a set of DNA strand displacement reactions, which approximate the prescribed dynamics up to arbitrary accuracy. The "DNA implementation" of the formal CRN has a larger set of interacting molecules, some of which represent the formal species of the formal CRN and approximate their dynamics, while the others are auxiliary species that mediate the desired reactions. Following this theoretical advance, other such CRN-to-DNA compilation schemes have been proposed [111, 224–226]. We describe our DNA strand displacement architecture, adapted from Soloveichik et al. [110], in Section 3.2.

3.2 DNA strand displacement architecture

As described in Section 1.4, DNA strand displacement is an excellent candidate architecture for implementing chemical reaction networks in a systematic way. Recently, Chen et al. [112] have experimentally demonstrated a CRN-to-DNA compilation scheme [111] to engineer several chemical reaction networks, from simple catalytic and auto-catalytic reactions to a distributed control algorithm ("approximate majority") for achieving consensus between multiple agents. However, despite this significant advance, the attractive goal of engineering the full spectrum of dynamic

a Desired dynamics Molecular program DNA architecture DNA dynamics

$$B + A \xrightarrow{k} 2B$$
$$C + B \xrightarrow{k} 2C$$
$$A + C \xrightarrow{k} 2A$$

signal species fuel species Concentrations of free A, B, C

b

c
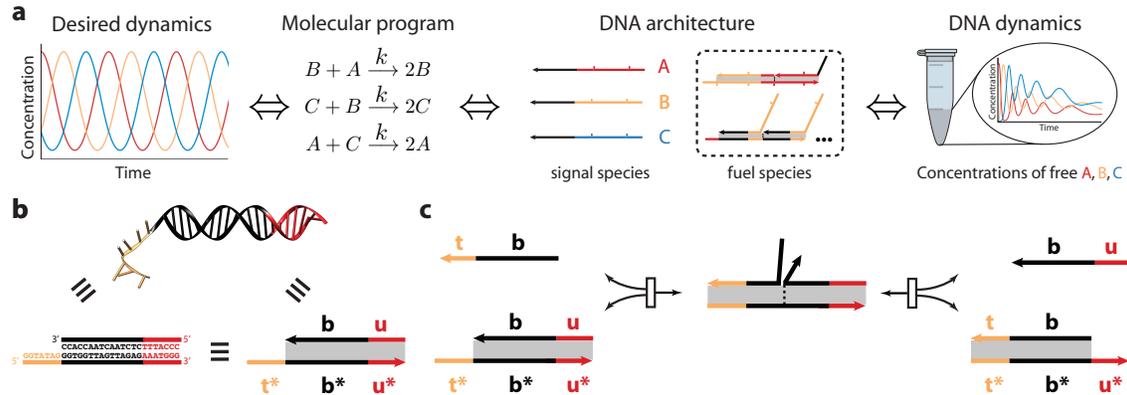
Figure 3.2: a. Overview of our CRN-to-DNA efforts. We start with a desired dynamical behavior (oscillation, in this case) and a CRN program that captures the desired dynamics. We then use the CRN-to-DNA scheme described in this chapter to translate the formal CRN into a DNA strand displacement implementation, where the formal species are represented by single strands of DNA called "signal" species. Desired reactions between signal species are mediated by "fuel" species which provide both logic and free-energy for the reaction. Some of the fuel species are mutli-stranded complexes which are pre-prepared and purified. In the regime where the fuel species are at high concentration, the signal species approximate the dynamics of the formal species in the original CRN. Our reactions are performed in "batch reactor" mode, which means that fuel species are not replenished. Therefore, the test tube dynamics is expected to deviate from idealized formal CRN dynamics. b. Domain notation. A "domain" comprises contiguously located bases whose binding and unbinding occurs as one logical unit. * indicates Watson-Crick complementarity. Arrows indicate 3' ends. c. Toehold exchange. "Short" (5-7 nucleotide) domains which bind fleetingly to their components at room temperature and reversibly co-localize distinct molecules are called "toeholds". Here toehold t reversibly co-localizes the molecules to form a three stranded intermediate, where the two b domains can exchange base pairs by a process called three-way branch migration. Eventually, either toehold u dissociates (leading to the products) or toehold t dissociates (leading to the reactants). Notice that the entire process is reversible and toehold u can also carry out toehold exchange.
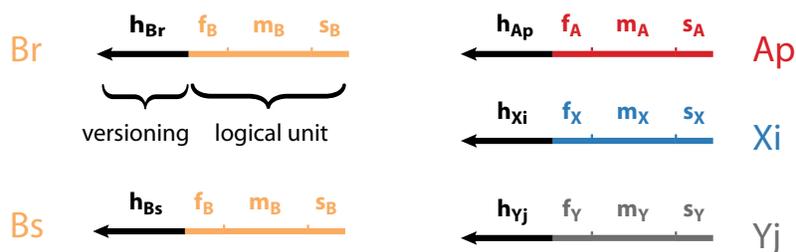
Figure 3.3: The formal species are represented by single-stranded DNA molecules (signal strands). Each signal strand comprises a history domain in black (versioning unit, e.g. h_{Br}) and a logical unit. The logical unit comprises three domains: the first toehold (e.g. f_B), a branch migration region (e.g. m_B), and the second toehold (e.g. s_B). Signal strands are designed to not interact with each other. Signal strands with the same logical unit (e.g. Br and Bs) represent the same formal species (B) and are designed to behave identically in solution.

behavior that CRNs are capable of (such as oscillations, chaos, etc.), rather than just the steady state end point, remains elusive.

We now describe our attempt to exploit a modified version of Soloveichik et al. [110]'s CRN-to-DNA scheme to engineer prescribed dynamical behaviors in chemical systems. Figure 3.2 provides a pictorial overview of our efforts.

Figure 3.3 illustrates the single-stranded representation of formal species employed by our scheme. Each formal species (e.g. B) is represented by single strands that contain a history domain (in black, e.g. h_{Br}) followed by 3 logical domains: a first toehold (e.g. f_B), a branch migration domain (e.g. m_B) and a second toehold (e.g. s_B). Strands that have identical logical domains (e.g. Br and Bs) are designed to behave identically in solution, as they both represent formal species B, regardless of their history domain. The reason for this will become clear once the mechanism for implementing reactions is illustrated.

Strands representing formal species ("signal strands") are designed to have orthogonal domains — they are not supposed to interact with each other directly. Desired reactions between signal strands are mediated by auxiliary species. Some of those auxiliary species are fuel species, which are present in large excess at the beginning of the reaction and perform the dual functions of both encoding the logical flow of the desired reactions and providing the required free energy to drive the intended reactions. This design principle ensures that (i) signal strands do not have any sequence inter-dependence and (ii) if a formal CRN, say CRN1, is extended to CRN2, then the DNA implementation of CRN1 may also be extended to a DNA implementation of CRN2 merely

by adding to the test tube fuel species necessary for the additional reactions in CRN2.

Figure 3.4 illustrates how a general bimolecular reaction of the form $B + A \rightarrow X + Y$ would be implemented. Logically, the DNA implementation is performed in two steps. First, the "react" step consumes reactants B and A and releases Flux_{ABi} — if and only if both reactants are present. If one or both of the reactants are absent, no irreversible reactions occur. Next, Flux_{ABi} gets consumed in the "produce" step and releases both outputs X and Y. Therefore, taking both react and produce steps together, the reactants B and A have been consumed and the products X and Y have been released. For completeness, we also illustrate how the general unimolecular reaction ($B \rightarrow X$), degradation reaction ($B \rightarrow \phi$), and production reaction ($\phi \rightarrow X$) are implemented (Figures 3.5, 3.6 and 3.7).

The naming scheme we use for the species involved in the reaction pathways in our CRN-to-DNA scheme is both precise and general. By this we mean that, given an arbitrary formal CRN, the naming scheme allows the user to write down the names and molecular specifications for all the species involved in the DNA strand displacement reactions needed to implement the given formal CRN. Moreover, given just the name, the associated molecule can be immediately constructed at the domain level. Essentially, our naming scheme is consistent with a compiler that could be used to generate the DNA implementation for any specified formal CRN.
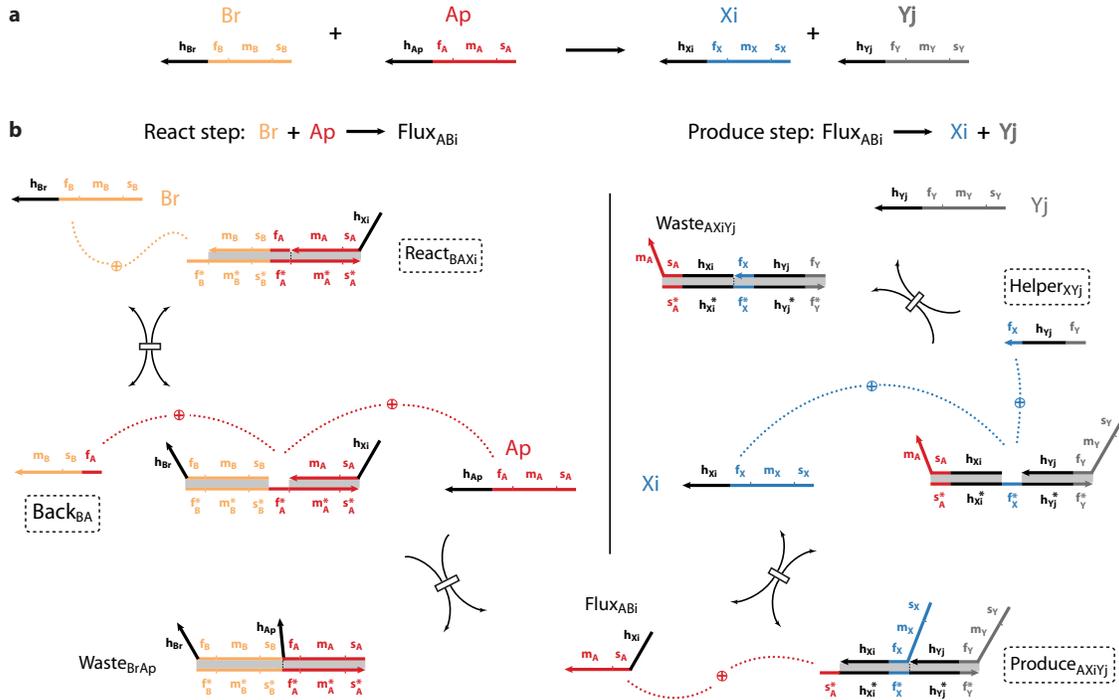
Figure 3.4: a. CRN-to-DNA scheme illustrated with the general bimolecular reaction $B + A \rightarrow X + Y$. Note that the reactants (Br, Ap) and the products (Xi, Yj) have completely independent sequences. The same mechanism can occur with different versions of the formal species B and A. b. Names of fuel species are enclosed in a dashed box. The reaction is implemented in two steps: React and Produce. The React step is mediated by the $\mathrm{React_{BAXi}}$ and $\mathrm{Back_{BA}}$ fuel species. B reacts with $\mathrm{React_{BAXi}}$ to reversibly displace $\mathrm{Back_{BA}}$ by toehold exchange and produces an intermediate species. Note that this process exposes the previously sequestered toehold f_A^*. In case Ap is present, it can react with the intermediate by strand displacement using the toehold f_A to irreversibly displace $\mathrm{Flux_{ABi}}$ and produce $\mathrm{Waste_{BrAp}}$. $\mathrm{Waste_{BrAp}}$ has no free toeholds and is therefore inert. $\mathrm{Flux_{ABi}}$ releases the outputs through the produce step, which is mediated by fuel molecules $\mathrm{Produce_{AXiYj}}$ and $\mathrm{Helper_{XYj}}$. $\mathrm{Flux_{ABi}}$ reacts reversibly with $\mathrm{Produce_{AXiYj}}$ to release the first output Xi and an intermediate species. $\mathrm{Helper_{XYj}}$ reacts irreversibly using the newly exposed toehold f_X^* by strand displacement to release the second output Yj and $\mathrm{Waste_{AXiYj}}$.
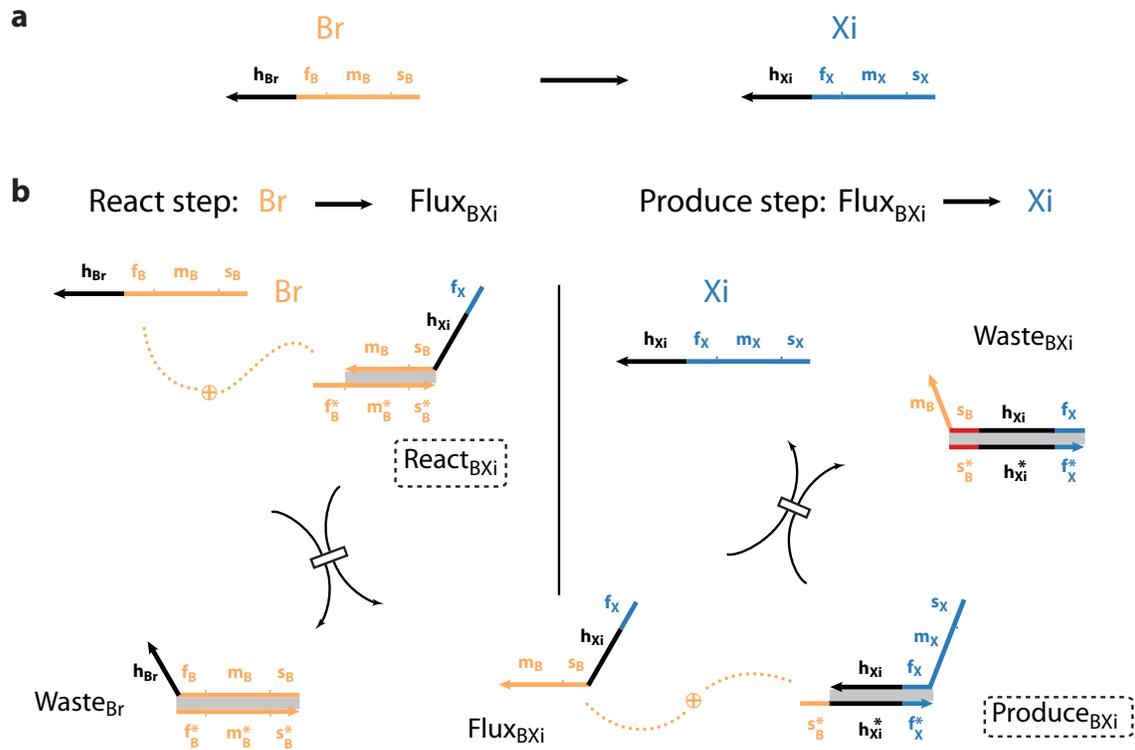
Figure 3.5: Implementation for the reaction B → X. The same mechanism can occur with different versions of B. Names of fuel species are enclosed in a dashed box.

a

Br

$\mathbf{h_{Br}}$ $\mathbf{f_B}$ $\mathbf{m_B}$ $\mathbf{s_B}$

b React step: Br ⟶

$\mathbf{h_{Br}}$ $\mathbf{f_B}$ $\mathbf{m_B}$ $\mathbf{s_B}$

Br

$\mathbf{m_B}$ $\mathbf{s_B}$

React$_{B\varnothing}$

$\mathbf{f_B^*}$ $\mathbf{m_B^*}$ $\mathbf{s_B^*}$

$\mathbf{h_{Br}}$ $\mathbf{f_B}$ $\mathbf{m_B}$ $\mathbf{s_B}$

$\mathbf{m_B}$ $\mathbf{s_B}$

Waste$_{Br}$

$\mathbf{f_B^*}$ $\mathbf{m_B^*}$ $\mathbf{s_B^*}$
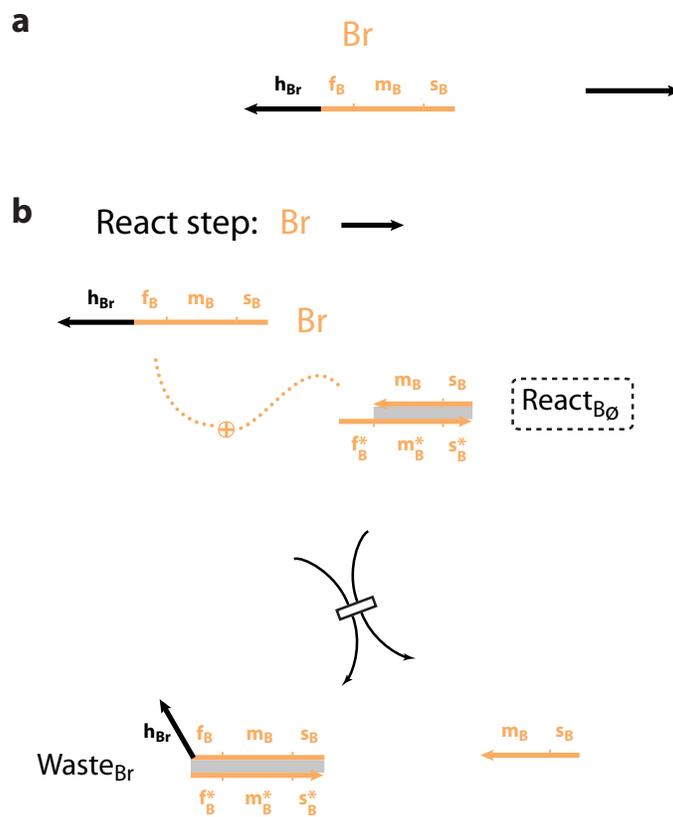
Figure 3.6: Molecular implementation for $B \to \phi$. The same mechanism can occur with different versions of B. Names of fuel species are enclosed in a dashed box.

a

$$\text{Xi}$$

b

React step: GX ⟶ Flux$_{GXi}$ Produce step: Flux$_{GXi}$ ⟶ Xi

Figure 3.7: Molecular implementation for $\phi \to X$. Names of fuel species are enclosed in a dashed box.

3.3 Test case: engineering a strand displacement oscillator

Oscillators have long been an important test of our ability to engineer CRNs with prescribed dynamical behavior (Section 1.2). Therefore, as a challenging test of our CRN-to-DNA technology, we chose to build an oscillator based on DNA strand displacement.

Our target oscillator is the one described in Fig 3.1 (b) (up to scaling rate constants). This oscillator has been widely studied [227–232] and is known as the "rock-paper-scissors" oscillator, "3-way" oscillator, or as the "cyclic Lotka-Volterra" oscillator. In particular, this oscillator has been investigated in the context of ecology, as each species can be thought of as both "predator" and "prey".

Reichenbach et al. [229] study this oscillator in both the deterministic and stochastic regimes. They also present a neat proof that this CRN is indeed an oscillator. The sum of the species A + B + C is trivially conserved. It turns out that, if the rate constants for the three autocatalytic reactions are identical, the product A × B × C is also conserved. Therefore the dynamics is constrained to be on the intersection of the plane A + B + C = constant and the curve A × B × C = constant, but there is no equilibrium on that intersection, resulting in characteristic triangle-like orbits (see Figure 3 of Reichenbach et al. [229]). The same proof is also presented in Lachmann and Sella [227]. Blossey [232] investigates the stability of this oscillator.

We chose this oscillator due to its symmetric, modular nature and its (relatively) small size. It consists of three autocatalytic reactions, each of which may be thought of as a module that can be tested independently. Our molecular implementation of the three autocatalytic modules, $B + A \rightarrow 2B$, $C + B \rightarrow 2C$, and $A + C \rightarrow 2A$ are illustrated in Figures 3.8, 3.9, and 3.10, respectively.
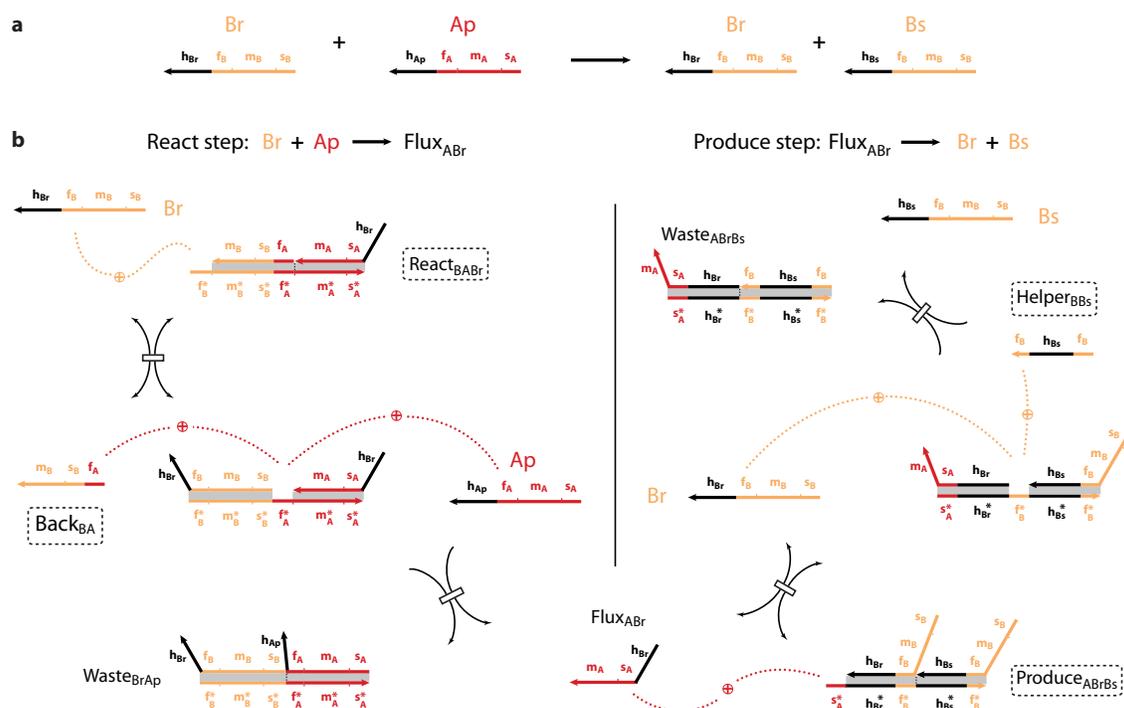
Figure 3.8: Implementation for the reaction $B + A \rightarrow 2B$. The same mechanism can occur with different versions of B and A. Names of fuel species are enclosed in a dashed box.
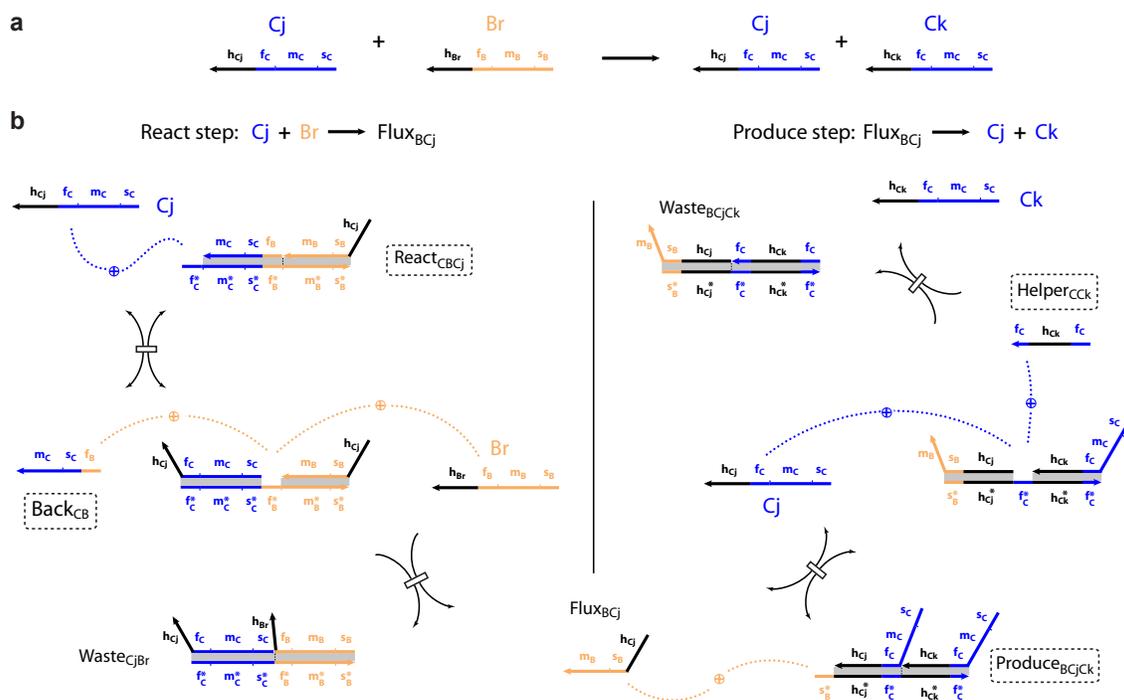
Figure 3.9: Implementation for the reaction $C + B \rightarrow 2C$. The same mechanism can occur with different versions of C and B. Names of fuel species are enclosed in a dashed box.
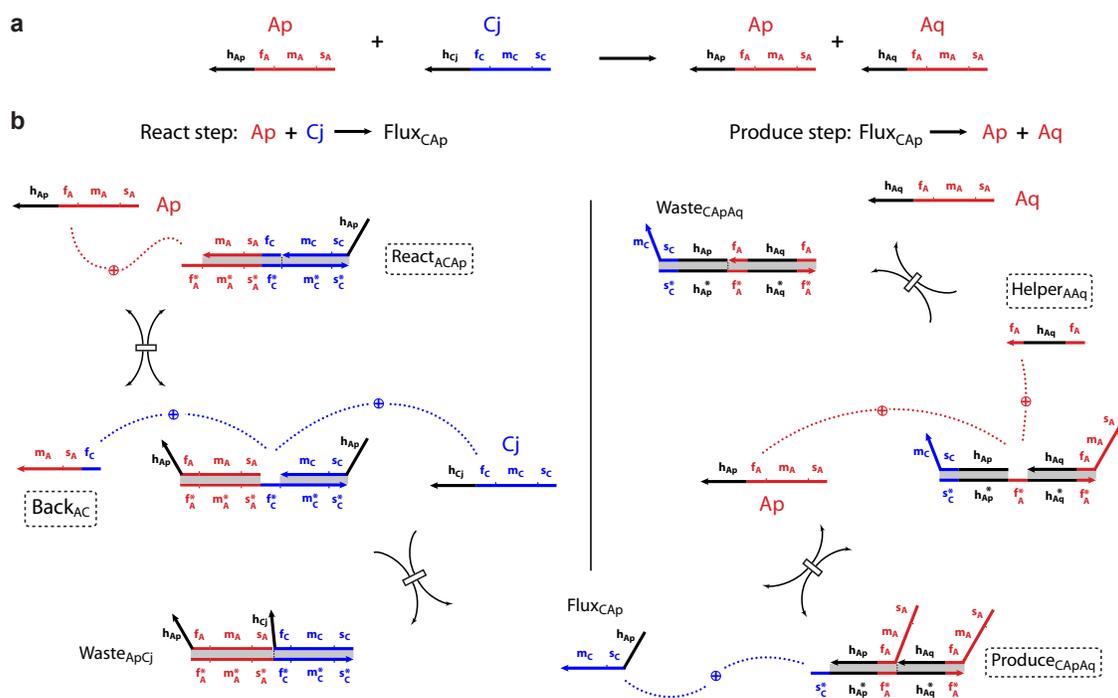
Figure 3.10: Implementation for the reaction $A + C \rightarrow 2A$. The same mechanism can occur with different versions of A and C. Names of fuel species are enclosed in a dashed box.

3.3.1 Modeling the DNA implementation

Fig 3.11 shows simulations from a simple model of the DNA implementation of our oscillatory CRN. Each strand displacement and toehold exchange reaction has been modeled as an effective bimolecular reaction with a rate constant of $k_{sd} = 2 \times 10^5$ /M /s. For clarity, equations 3.1 - 3.4 specify the chemical reaction equations in the strand-displacement level model for the autocatalytic module $B + A \rightarrow 2B$. The model comprises those equations and similar equations for the other two autocatalytic modules. This model assumes that unimolecular reactions (including dissociation of toeholds and branch migration) are effectively instantaneous. All fuel molecules have an initial concentration of 300 nM and are not replenished.

$$B + \text{React}_{BABr} \xleftrightarrow[k_{sd}]{k_{sd}} \text{ReactInt}_{BABr} + \text{Back}_{BA} \tag{3.1}$$

$$\text{ReactInt}_{BABr} + A \xrightarrow{k_{sd}} \text{Flux}_{ABr} + \text{Waste}_{BA} \tag{3.2}$$

$$\text{Flux}_{ABr} + \text{Produce}_{ABrBs} \xleftrightarrow[k_{sd}]{k_{sd}} B + \text{ProduceInt}_{ABrBs} \tag{3.3}$$

$$\text{ProduceInt}_{ABrBs} + \text{Helper}_{BBs} \xrightarrow{k_{sd}} B + \text{Waste}_{ABrBs} \tag{3.4}$$
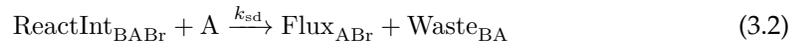
This simple model shows that the signal strands A, B, and C in the DNA implementation demonstrate oscillatory behavior before the fuel species (which power the reactions) get depleted significantly. Therefore, even in 'batch reactor' mode where the fuel species are not being replenished, we should in principle be able to engineer oscillatory dynamics which lasts as long as the fuel species are in significant excess.

3.3.2 Non-idealities in the DNA implementation

The simple model presented above does not include any non-idealities in the DNA implementation. These non-idealities could be of several different kinds, ranging from different strand displacement reactions having very different rate constants to spurious "leak" reactions which compete with desired reaction pathways.

Broadly, there are two classes of non-idealities. The first class refers to those that are a consequence of imperfect molecules, e.g. errors in DNA synthesis or mis-folded complexes. The second class comprises non-idealities that are unavoidable in our CRN-to-DNA scheme even with per-
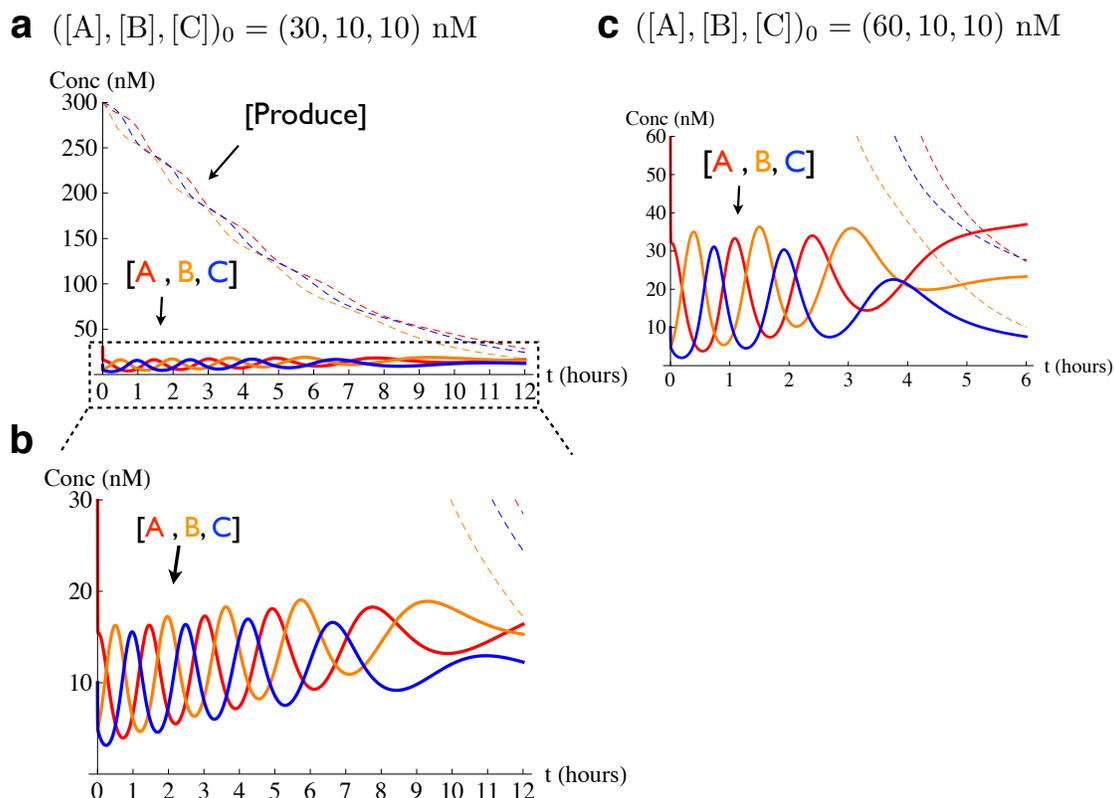
Figure 3.11: Modeling the DNA implementation of the oscillator at the level of individual strand displacement and toehold exchange reactions. Equations 3.1 to 3.4, along with similar equations for the other two modules, specify the model. The fuel species are present at an initial concentration of 300 nM and are not replenished. a. Concentrations of Produce molecules (dashed lines; $\mathrm{Produce_{CApAq}}$ in red, $\mathrm{Produce_{BCjCk}}$ in blue, $\mathrm{Produce_{ABrBs}}$ in orange) and signal strands as a function of time starting with an initial concentration of $([A], [B], [C])_0 = (30, 10, 10)$ nM. b. The plot in (a), zoomed in so that the oscillatory dynamics of the signal strands are visible. c. The plot in (b) with an initial concentration of $([A], [B], [C])_0 = (60, 10, 10)$ nM, all other parameters being the same.

fect molecules, since they arise from the domain-level specification of the molecules or reaction pathways. In particular, leak reactions that arise from "blunt end" (zero-base toehold) strand displacement are good examples of this class (see Figure 3.12). These leak reactions are a direct consequence of the fact that blunt end strand displacement occurs at a non-zero rate.

These non-idealities have consequences for our experiments. First, our initial conditions are imperfect because some fraction of our fuel species are "bad" — they may be mis-folded complexes or DNA strands with synthesis errors on them. One particular manifestation of these imperfect initial conditions is what we measure experimentally as "initial leak" — some fraction of our fuel species release their outputs even when their inputs are not present. Therefore, our initial conditions may involve a smaller concentration of "good" fuel species than we expect, some leaked signal strands or Flux molecules, and some spurious products.

Second, spurious products that form due to leak pathways may be capable of undergoing certain reactions that are legitimate steps in a desired reaction pathway. Figure 3.13 illustrates some reactions of this kind in the case of spurious products that arose from the gradual leak pathways shown in Figure 3.12. Therefore leak reactions may affect dynamical behavior in ways that are more complex than merely the unexpected release of signal strands or Flux strands.

In particular, spurious products can affect reaction stoichiometries. For example, Leaked-React$_{CBr}$ consumes one molecule of Cj and releases no other signal strand; Leaked-Produce$_{BCjCk}$ consumes one molecule of Flux$_{BCj}$ and releases only one molecule of C, since Ck had already "leaked" out (see Figure 3.13). Indeed, this type of mechanism could conceivably explain "incompletion" effects we observe in our experiments - where triggering a particular fuel species (say Produce$_{BCjCk}$) with its input(s) (Flux$_{BCj}$) produces sub-stoichiometric amounts of the output (C).

If we can account for the initial amounts of such leaked products and correspondingly deplete the fuel concentrations (say by fitting those leaked amounts to the data), and include in our model the set of expected reactions in which spurious products may participate, we may be able to mechanistically account for our experiments without any additional assumptions that are fundamentally different from the idealized model. The reactions we may need to include would be of the form outlined below (where L is short for "Leaked").
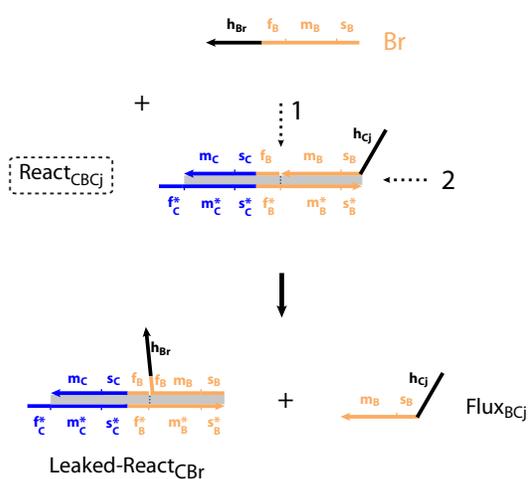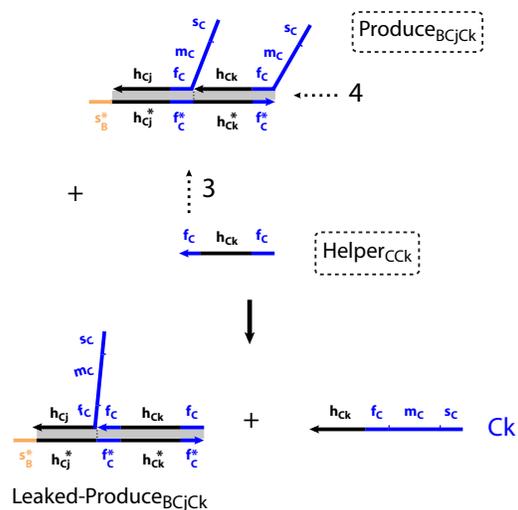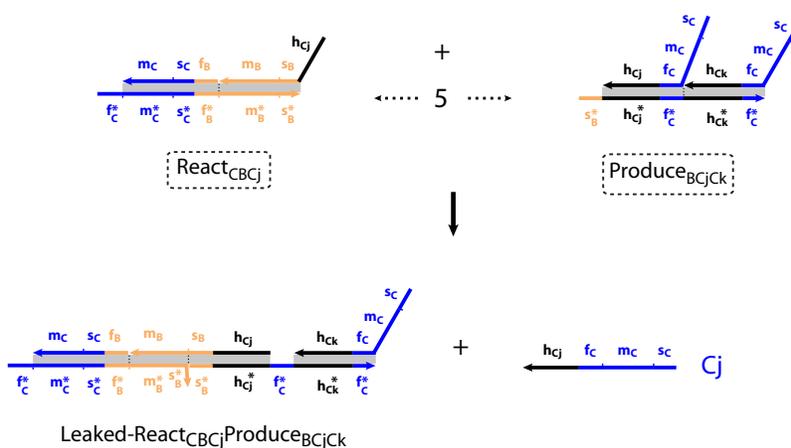
Figure 3.12: Illustrative examples of spurious "leak" pathways that arise due to blunt-end (zero base toehold) strand displacement. These pathways are illustrated in the case of the autocatalytic module $C + B \rightarrow 2C$ but can occur with the other modules as well. Locations of invasion are indicated by numbered dashed arrows. a. The second input (here, Br) can invade at locations 1 (the junction) and 2 (the end of the helix) in the React species. Once strand displacement finishes, the Flux molecule may be released and a spurious species can be formed. b. A similar reaction can happen between the Helper species and the Produce species, releasing the second output of the Produce molecule (here, Ck) and resulting in a spurious species. c. Spontaneous fraying due to thermal fluctuations at the end of the helix in the React molecule may enable the Produce molecule to invade at at location 5. Strand displacement can then result in the release of the first output of the produce gate (here, C_j) and the formation of a spurious species. Notice that all of these spurious species shown here are capable of participating in some reactions that are also a legitimate part of desired reaction pathways (see Figure 3.13).
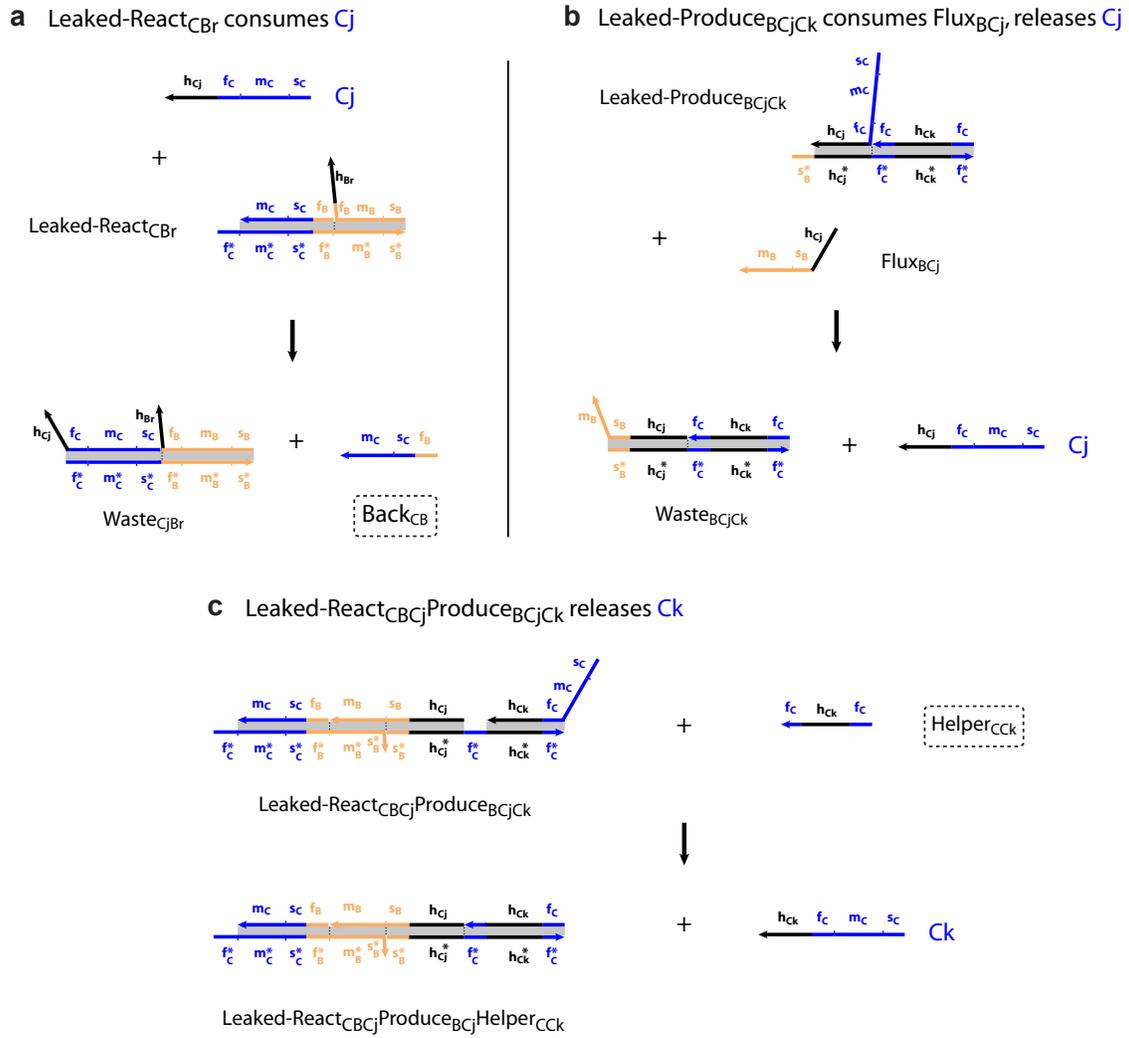
Figure 3.13: Spurious products that are formed due to leak pathways may undergo reactions that are legitimate steps in desired reaction pathways. Therefore leak reactions may affect dynamical behavior in ways that are more complex than merely the unexpected release of signal strands or Flux strands.

$$\text{L-React}_{\text{CBr}} + \text{C} \longrightarrow \text{Waste}_{\text{CjBr}} + \text{Back}_{\text{CB}} \tag{3.5}$$

$$\text{L-Produce}_{\text{BCjCk}} + \text{Flux}_{\text{BCj}} \longrightarrow \text{Waste}_{\text{BCjCk}} + \text{C} \tag{3.6}$$

$$\text{L-React}_{\text{CBCj}} \text{Produce}_{\text{BCjCk}} + \text{Helper}_{\text{CCk}} \longrightarrow \text{C} + \text{L-React}_{\text{CBCj}} \text{Produce}_{\text{BCjCk}} \text{Helper}_{\text{CCk}} \tag{3.7}$$
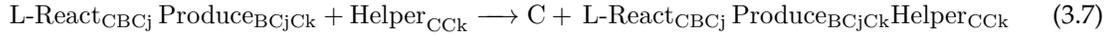
We have explored different augmented versions of the model described in Section 3.3.1 that include some of the non-idealities discussed in this section. We have omitted that exploration in the interest of brevity. However, a rigorous analysis of sensitivity to parameters was not performed. Based on our modeling, we learned that the DNA implementation of our oscillator can tolerate significant variation in individual strand displacement rate constants (factor of 3-10, and possibly more, depending on how ideal other parameters, such as gradual leak rates, are) and bimolecular rate constants for gradual leak pathways (of the kind described in Figure 3.12) as high as 50 - 80 /M /s.

3.3.3 Sequence design challenges

So far, we have discussed the reaction pathways in our molecular implementation in terms of domains (for a definition of "domain", see Figure 3.2). Implicit in the domain level description is the assumption that only complementary domains interact — domains that are designed to be orthogonal are assumed to not interact significantly. Therefore, the ability to engineer strand displacement systems in the lab hinges crucially on sequence design. Sequence design is essentially the process of finding DNA sequences for the domains such that the resulting molecular system stays as faithful to the domain abstraction as possible. While that is a primary requirement, usually there are several other additional requirements, depending on the particular purpose. We now examine our desired criteria and the sequence design challenges towards satisfying those criteria.

First, our strand displacement oscillator presents quite a large sequence design problem. We need to design a total of 177 distinct bases, which presents, naively, 4^{177} possible candidate sequences to choose from. This is a *very* large number — much larger than the total number of positions possible in chess [233]. A large design space presents a serious search challenge: even if we had a working definition for what a "good sequence design" would be, how would we find one?

Second, since we are interested in programming dynamic behavior, we need to be able to control reaction kinetics. We need to ensure that desired strand displacement and toehold exchange rates meet two requirements. First, they need to be as fast as possible relative to the gradual leak pathways, such as blunt end strand displacement rates. Second, they need to be as similar as possible to each other. However, "kinetic sequence design" is an unsolved problem, essentially because the problem of predicting DNA hybridization, branch migration, fraying, and dissociation kinetics from DNA sequence itself is a challenge [148, 153, 192].

We tackle the first requirement by trying to reduce gradual leak rates as much as possible, by a combination of domain-level and sequence-level design. It is worth noting here that digital logic signal restoration methods, which have proved immensely useful in the design of feedforward logic circuits [107], will not apply in our case. Essentially, these methods involve thresholding the leaked products and amplifying the desired products at each layer of the designed circuit. Since our goal is analog dynamic behavior, we cannot co-opt those methods for our purposes.

We address the second requirement by trying our best to control kinetics through thermodynamic proxies, such as toehold strength (Figure 2.2). Therefore, we need our six toeholds (f_A, f_B, f_C, s_A, s_B, s_C) to be roughly equally strong. How strong should they optimally be? For two reasons, we tried to choose toehold energies such that strand displacement rates with those toeholds are in the saturation regime. First, such strong toeholds result in fast strand displacement rates relative to gradual leak rates. Second, if we are in the saturation regime, small variations in toehold energy do not result in a large variation in (relative) strand displacement rates.

Yet, strong toeholds are not without drawbacks. Our toeholds need to be weak enough to ensure that toehold dissociation rates are fast. Fast toehold dissociation rates ensure that (i) unimolecular toehold dissociation reactions will not be rate limiting for desired toehold exchange pathways and (ii) complexes that are not designed to interact with each other but have complementary toehold regions will not be co-localized significantly (see Figure 3.14).

Given this trade-off between strong toeholds for fast kinetics and weak toeholds for fast dissociation of toehold-interactions, it is not clear what the optimal choice is, and this could vary with the desired CRN or dynamical behavior.

Third, several species (the signal strands, Flux, Back, and Helper strands) need to be almost completely single-stranded, with no intra-molecular base pairing, most of the time. This is important because these strands are involved in strand displacement reactions, and secondary structure
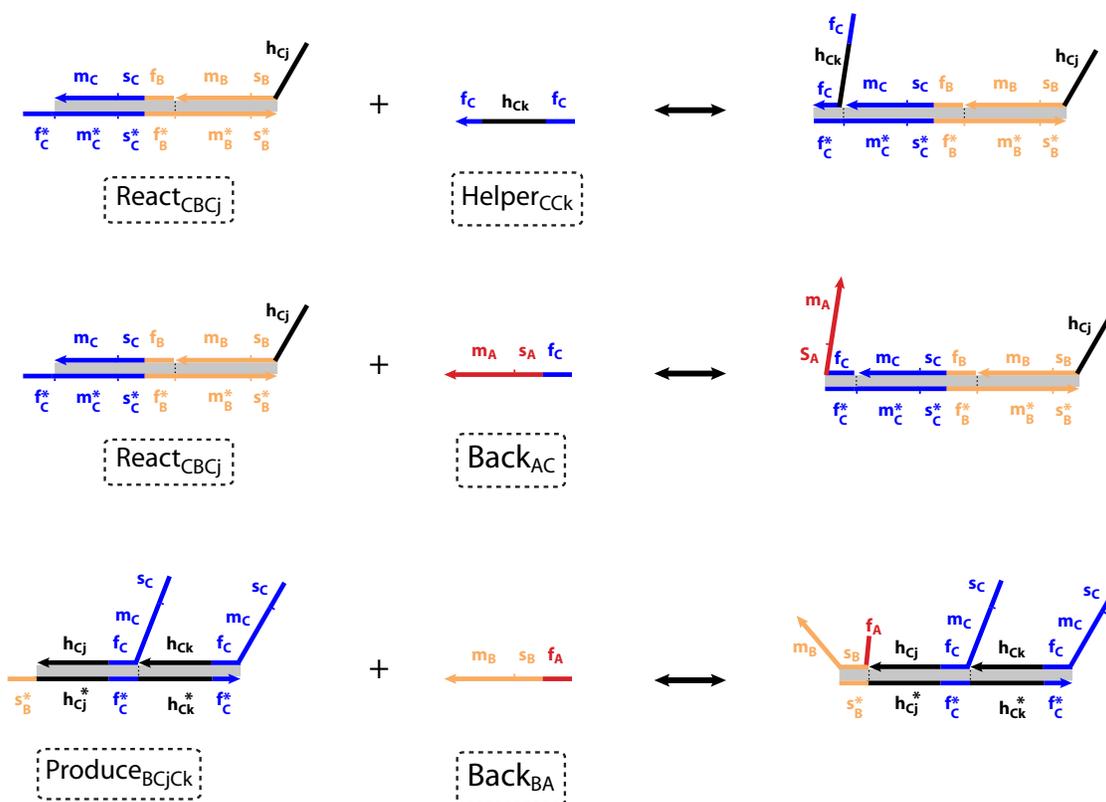
Figure 3.14: Illustrative (but not exhaustive) examples of toehold-only interactions in the molecular implementation of our oscillator. Note that all these interactions (but not all interactions) are between fuel species at high concentration. There are interactions within the same autocatalytic module (e.g. (a)) and across modules (e.g. (b), (c)).

is known to slowdown strand displacement rates [126, 148, 153].

We now recount our journey through multiple sequence designs, focusing on what we learned at each stage and how that informed our design process for subsequent attempts. We would like to stress here that, in particular, our understanding of (i) the challenges described in this subsection and (ii) our journey through sequence space which is to follow have benefitted greatly from hindsight.

For reasons that will soon become clear, the first and second sequence designs have opposite 5′-3′ orientation for all the strands and multi-stranded complexes relative to later designs and all the domain-level diagrams presented in Chapters 3 and 4.

3.4 Sequence design 1

First, we chose all our "top" strands (signal strands, Flux, Back, and Helper strands) to have sequences from the 'ACT' alphabet (no 'G's). This is standard practice in dynamic DNA nanotechnology in order to avoid unintended intramolecular secondary structure in strands that are supposed to be almost entirely single-stranded with no intramolecular base-pairing.

By designing toeholds to have the same number of C-G base pairs, we tried to ensure similar toehold energies. In particular, we chose to do this by designing all toeholds to be of the form "WWWSSSW" where W stands for a "weak" base (A or T) and S stands for a "strong base" (C or G). This rule, combined with the 'ACT' alphabet for top strands, forces all toeholds in the top strands to be "WWWCCCW" and all toeholds in the complementary "bottom" strands to be "WWWGGGW".

To mitigate cross-talk, we designed domains intended for branch migration (m_A, m_B, m_C, and history domains) to be as different as possible at each end. This criterion was enforced because these were the locations where branch migration could initiate, and we wanted to ensure that fleeting binding between toeholds that are not supposed to interact would not result in the initiation of branch migration. In addition, the first and last base of every branch migration domain was constrained to be a strong base, in order to reduce fraying of the helices at the ends. Sequences in the middle of branch migration regions were essentially randomly generated. We call this design "Design 1-PRE", where "PRE" stands for preliminary.

Design 1-PRE was verified by eye with the NUPACK web interface [123] to ensure that the

desired multi-stranded complexes and intermediates were well-formed and that the top strands were (mostly) free of secondary structure.

In our experiments, we encountered two kinds of leak reactions: "initial" and "gradual" leak. These are illustrated in Figure 3.15. Initial leaks involve spontaneous and nearly instantaneous release of the outputs of fuel complexes (such as React and Produce molecules) when a fluorescent reporter for their outputs is present. The fluorescent signal would proceed to completion in as little as 15 minutes.

We found that the initial leak amount scales proportionally with the concentration of the fuel complex but not with the concentration of the reporter for the output (experiments not shown) and this suggested that the initial leak arises due to a fraction of "bad" fuel complexes. At this stage, we did not have a hypothesis for what causes some fraction of the fuel complexes to be "bad". For Design 1-PRE, we measured initial leak to be between 8% and 15% for different auxiliary complexes. Despite our best efforts at purification of complexes, which included use of PAGE-purified strands, ultramers, modified annealing and gel-purification protocols, we could not reduce this initial leak to much less than 10%.

In contrast to initial leaks, gradual leaks are the slow release of output over the course of the experiment (several hours). They are thought to occur through blunt end strand displacement, or strand displacement which begins by invasion at a junction (see Figure 3.12). After preliminary experiments where we observed high gradual leaks, we added 2-nucleotide clamps to the React and Produce molecules (see Figure 3.16) to mitigate some of the pathways shown in Figure 3.12, such as the React-Produce gradual leak pathway in panel (c). We call this design, augmented with 2-nucleotide clamps, Design 1. The clamps did reduce gradual leak, but not substantially. They did not change initial leak. They are included in Design 1 and all subsequent sequence designs, even if they may sometimes not be indicated in domain level diagrams (for convenience). Sequences that comprise Design 1 are provided in Section 5.3.3.

Even with the 2-nucleotide clamps, we observed Produce-Helper gradual leak rates as high as 150 /M /s and React-second input gradual leak rates as high as 50 /M /s in Design 1. A necessary disclaimer about these gradual leak measurements: the leak process that leads to release of output that is measured in these experiments is not truly bimolecular. That is, the gradual leak rate does not scale with the concentration of the reactants exactly as we expect a bimolecular process to. We believe this is because the gradual leak process is likely a conglomeration of multiple pathways,

Figure 3.15: Experiment illustrating leaks in Design 1, with $\text{Produce}_{\text{CApAq}}$-$\text{Helper}_{\text{AAq}}$ leaks as an example. a. The molecules involved. "Rep A", short for "Reporter for A", is a molecule used for fluorescence-based readout of the concentration of A using the mechanism shown in panel b. b. Aq displaces the strand with the quencher (Q), leading to a waste product where the Fluorophore (F) can emit light at its characteristic wavelength. This mechanism works for Ap as well, leading to a quantitative readout of A. c. Experimental setup. All three samples are essentially negative controls for the function of $\text{Produce}_{\text{CApAq}}$, since no Flux_{CAp} is present at any time. Sample 1 shows clean "zero" behavior of Rep A by itself. In Samples 2 and 3, the addition of 100 nM of $\text{Produce}_{\text{CApAq}}$ causes an initial release of A. Note that, if $\text{Produce}_{\text{CApAq}}$ were to be a perfect molecule, there should be no initial spike since no Flux_{CAp} was present. This "initial leak" is higher in Sample 3 because 50 nM of $\text{Helper}_{\text{AAq}}$ was also added, which suggests that $\text{Helper}_{\text{AAq}}$ facilitates this process. After the initial leak, notice the slower but persistent release of A in Sample 3, which we call the "gradual leak" between $\text{Produce}_{\text{CApAq}}$ and $\text{Helper}_{\text{AAq}}$. There is also a much smaller gradual leak between just the $\text{Produce}_{\text{CApAq}}$ and Rep A in Sample 2.

a

m_C s_C f_B m_B s_B

h_{Cj}

f_C^* m_C^* s_C^* f_B^* m_B^* s_B^*

} **first 2 nt of h_{Cj} as a "clamp"**

React$_{CBCj}$

b

s_C s_C

m_C m_C

h_{Cj} f_C h_{Ck} f_C

s_B^* h_{Cj}^* f_C^* h_{Ck}^* f_C^*

} **first 2 nt of m_C as a "clamp"**

Produce$_{BCjCk}$

Figure 3.16: After preliminary experiments with Design 1-PRE demonstrated very high gradual leak rates, we added 2-nucleotide "clamps" to all React and Produce species as illustrated here, to obtain Design 1. These clamps are meant to mitigate some of the gradual leak pathways shown in Figure 3.12, such as the React-Produce gradual leaks in panel (c). The clamps did reduce gradual leak, but not substantially. They are included in Design 1 and all subsequent designs, even if they are not illustrated explicitly in the following diagrams (for convenience).

and we use the bimolecular description as an effective quantitative coarse-graining to describe the process. Taken together, these initial and gradual leaks were too high for our purposes of engineering oscillatory dynamics; we decided to comprehensively revisit our sequence design and verification criteria.

3.5 Choosing algorithms for sequence design and verification

3.5.1 Heuristics for evaluating sequence designs *in silico*

Our experience with Design 1 suggested that we did not really know what we should design for (or against). To understand that better, we developed some heuristic measures which attempted to quantitatively capture various undesired spurious interactions. Using these heuristic measures, we hoped to quantify the "goodness" of candidate designs and sequence design algorithms (by comparing the best designs we could obtain from each algorithm). We now introduce the heuristic measures.

First, we define a "NUPACK interaction score", I (S1, S2), between single-stranded molecules S1 and S2. I (S1, S2) is the percentage of strands that are in any dimer (S1:S1, S1:S2, or S2:S2), as predicted by NUPACK when S1 and S2 are each at a concentration of 1 μM. For this calculation (i) the temperature is set to be 25°C and (ii) the salt concentration to be 1 M Na$^+$, and (iii) complexes

comprising at most 2 strands are considered. Therefore, with some algebra,

$$\mathrm{I}\,(\mathrm{S1}, \mathrm{S2}) = 100 * \left(\frac{[\mathrm{S1} : \mathrm{S1}] + [\mathrm{S1} : \mathrm{S2}] + [\mathrm{S2} : \mathrm{S2}]}{\mathrm{c}} \right), \tag{3.8}$$

where c = 1 μM.

The "Top Strand Interactions (TSI)" score is the sum of interaction scores for every distinct pair of top strands (signal strands, Flux, Back, and Helper strands). That is, if $\{S_i\}$ are the top strands,

$$\mathrm{TSI} := \sum_{(i,j)|i \leq j} \mathrm{I}\,(\mathrm{S_i}, \mathrm{S_j}). \tag{3.9}$$

The "Toehold Occlusion (TO)" score is the sum of I (t^*, S) for every toehold-complement t^* and top strand S, assuming S does not contain toehold t. If S does contain toehold t, I (t^*, S) is replaced by I $(t^*, S_{<t}) + \mathrm{I}\,(t^*, S_{>t})$ where $S_{<t}$ is the subsequence of S, starting at the 5' end, before encountering domain t and $S_{>t}$ is the subsequence after domain t. Strictly, the above definition of $S_{>t}$ could contain an occurrence of t if t occurs multiple times in the original strand S; if that is the case, iteratively apply the same rule. That is,

$$\mathrm{TO} := \sum_{\substack{(t,S) \\ t \,\in\, \mathrm{Toeholds} \\ S \,\in\, \mathrm{Top\ Strands}}} \mathrm{F}\,(t, S), \tag{3.10}$$

where

$$F(t, S) := \begin{cases} I\,(t^*, S) & \text{, if S does not contain } t \\ F(t^*, S_{<t}) + F(t^*, S_{>t}), & \text{if S contains } t. \end{cases}$$

The "Weighted Sum - Branch Migration (WS-BM)" score identifies subsequence matches (not necessarily aligned by position) between distinct branch migration domains, which include the m_A, m_B, m_C and history domains. WS-BM is essentially a weighted count of such subsequence matches. (Note that there are no intended matches between branch migration domains.) A subsequence match of length 5 contributes 1 point, length 6 contributes 2 points, length 7 contributes 4 points, and so on until lengths greater than or equal to 10 contribute 32 points.

If $\{R_i\}$ is the set of branch migration domains,

$$\text{WS-BM} := \sum_{\substack{(R_i, R_j) \\ i < j}} \left(\sum_{s \in \Phi(R_i) \cap \Phi(R_j)} W(s) \right), \tag{3.11}$$

where $\Phi(R_i)$ is the multiset of all subsequences of the sequence R_i and W is the weight function defined by

$$W(s) := \begin{cases} 0 & \text{if } |s| < 5 \\ 2^{|s|-5} & \text{if } |s| \in [5, 10] \\ 32 & \text{if } |s| > 10. \end{cases}$$

The "Maximum Branch Migration subsequence (Max-BM)" score seeks to quantify the same kind of non-ideality in the design but measures the length of the longest subsequence match between distinct branch migration domains, rather than a weighted sum. That is,

$$\text{Max-BM} := \max_{\substack{(R_i, R_j) \\ i < j}} \left(\max_{s \in \Phi(R_i) \cap \Phi(R_j)} |s| \right). \tag{3.12}$$

The "Weighted Sum Inter-Complex (WSIC)" seeks to evaluate unintended subsequence matches between distinct complexes, which includes both single-stranded and multi-stranded complexes (irrespective of the locations of the subsequences within the complexes). In particular, even a subsequence which spans a strand break would be counted as a contiguous subsequence of that complex. A subsequence match of length 6 contributes 1 point, length 7 contributes 2 points, and so on until matches of length 12 and above contribute 7 points.

If $\{C_i\}$ is the set of complexes, then

$$\text{WSIC} := \sum_{\substack{(C_i, C_j) \\ i < j}} \left(\sum_{s \in \Phi(C_i) \cap \Phi(C_j)} W(s) \right), \tag{3.13}$$

where $\Phi(C_i)$ is the multiset of all subsequences of the complex C_i and W is the weight function defined by

$$W(s) := \begin{cases} 0 & \text{if } |s| < 6 \text{ or if } s \text{ is an intended match specified at the domain level} \\ |s| - 5 & \text{if } |s| \in [6, 12] \\ 12 & \text{if } |s| > 12. \end{cases}$$

The "Weighted Sum Inter-Complex Mismatch (WSIC-M)" is identical to the WSIC score, except that it only counts subsequence matches with a 1-base mismatch. To illustrate with a particular example, subsequences "ATAACCA" and "ATATCCA" would be considered a match.

3.5.2 Candidate sequence design methods

Since (i) toehold strengths are critical for controlling reaction kinetics (Figure 2.2) and (ii) excessive spurious interaction between toeholds assumed to be orthogonal at the domain level is undesirable, we decided to design the toeholds first. Our objective was to obtain toeholds strong enough to elicit fast kinetics, with toehold strengths being as close as possible, yet as mutually orthogonal as possible. For this purpose we used StickyDesign [234], a software package used to design "sticky ends" for experimental DNA-tile based self-assembly [235]. Without accounting for flanking bases (which do affect nearest neighbor contributions to the toehold strength), we targeted toehold strengths between 8.0 and 9.0 kcal/mol and toehold cross-talk binding energies of less than or equal to 2.5 kcal/mol.

Once the toeholds were fixed, we tried multiple sequence design algorithms that were available at that time to design the rest of the system. Among these were SpuriousC [124], which minimizes unintended subsequence matches within and between complexes [125]; DomainDesign [126], which focuses on designing domains of user-specified lengths to be as orthogonal as possible to each other based on thermodynamic calculations; and NUPACK [123, 127] which minimizes ensemble-defect [128] to match the user-specified structure for each complex based on thermodynamic calculations.

The best designs we could obtain from each candidate algorithm were compared with each other and Design 1 on the basis of our custom heuristic measures. The results are summarized in Table 3.1. Note that there are three different designs based on NUPACK: "NUPACK" was run without any artificial negative-design constraints; "NUPACK-ND" included some hacks to enforce negative design for minimizing spurious interactions between single strands and toehold complements. These were essentially in the form of artificial target "complexes" between species that were not supposed to interact wherein the desired structure was specified to be without any base pairing between the two species. "NUPACK-ND-BM" included additional such constraints between branch migration domains.

Table 3.1 suggests that, given our toeholds, SpuriousC and DomainDesign gave the best de-

Method	TSI	TO	WS-BM	Max-BM	WSIC	WSIC-M
Design 1	0.8	**466.0**	18	6	**12469**	**8009**
SpuriousC (Design 2)	7.2	21.6	10	6	30	2514
DomainDesign	12.2	19.6	7	6	1239	4016
NUPACK	1.6	56.2	**73**	**10**	**13574**	**8674**
NUPACK-ND	1.4	14.8	**306**	**13**	**23529**	**27578**
NUPACK-ND-BM	2.2	34.0	9	**8**	**25872**	**8429**

Table 3.1: Performance of the best designs from various sequence design methods on our custom heuristic measures. The heuristic measures and sequence design methods used are described in the text. Scores in bold were thought to be unacceptably high relative to other methods; this determination was based on a subjective intuition acquired by repeated design attempts.

signs according to our heuristic measures. Some caveats are in order while interpreting the results. First, our heuristic measures include measures that focus on spurious matches at the level of sequence identity, without a thermodynamic or kinetic evaluation of how physically important those spurious sequence identity matches might be in the test tube. Second, at the time this analysis was performed, NUPACK's second generation algorithms for "test tube design" [129], which perform negative design against competing complexes which could form in a user-specified test tube, had not been released. We have not tested the performance of the second generation NUPACK sequence design algorithms in this analysis.

3.6   Sequence design 2

Based on the analysis in Section 3.5.2, we decided to use SpuriousC for designing our second attempt. We also enforced two other constraints that are included in the optimized SpuriousC design evaluated in Table 3.1. Since fraying due to thermal fluctuations at the ends of helices is thought to contribute to gradual leak pathways of the kind in Figure 3.12 [126, 148, 153], we constrained the ends of helices and junctions to end with two strong (C/G) base pairs, in order to mitigate gradual leak. As a direct consequence of this decision, we had to choose between (i)

a

b
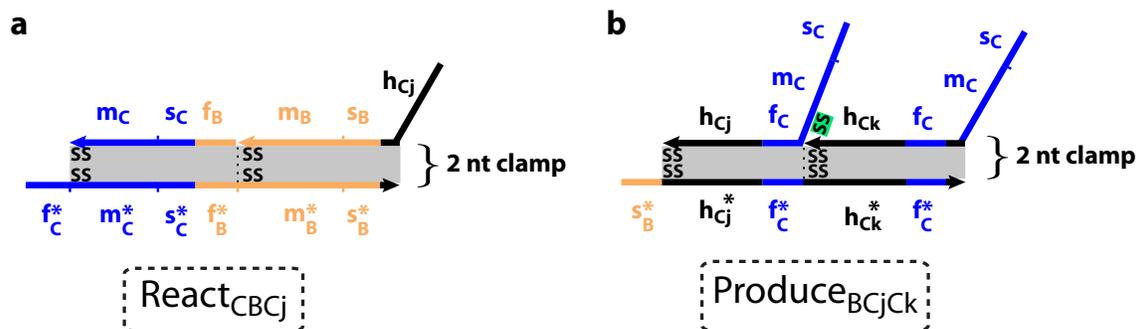
Figure 3.17: To reduce gradual leaks that are thought to be facilitated by thermal fraying of base pairs at junctions or ends of helices, we attempted to close such helices and junctions with two consecutive "strong" (S) base pairs. This strategy is illustrated in the React (a) and Produce (b) molecules. b. In this strategy, there is a trade off between ACT alphabet and preventing branch migration at the junction in the Produce molecules. With the ACT alphabet, back and forth branch migration of 2 nucleotides around the junction is unavoidable since both m_c and h_{ck} will need to begin with "CC". To avoid this, we violated ACT alphabet by fixing the first base of m_C (first of the two highlighted bases in this figure) to be a 'G'. In some situations, e.g. $React_{ACAp}$, the second highlighted base was also a 'G' (optimized by the designer to reduce spurious scores).

allowing 1-2 base pairs to continually branch migrate back and forth on either side of the junction in the Produce species or (ii) violate the 'ACT' alphabet rule by incorporating Guanines (G) at specific places in top strands (see Figure 3.17). Since option (i) could significantly complicate toehold energy calculations, we chose to violate the ATC alphabet.

Experimentally, we found that Design 2 had much less gradual leak — as low as 1 - 10 /M /s. Gradual leaks of this magnitude are illustrated in Figure 3.18 with an example from Design 3. This is "good enough" for oscillatory dynamics according to our modeling. However, Design 2 did not show any reduction in initial leak.

3.7 Sequence design 3

Typically, a domain level design specification can function with either orientation (5'-3') for the DNA strands involved. However, it has been suggested in the literature [217] that certain 5'-3' orientations for the molecules may be preferable in practice because of asymmetries in the distribution of synthesis errors along the 5'-3' axis. In the hope that reversing 5'-3' orientations might change the distribution of synthesis errors to a more favorable one, which may in turn result in
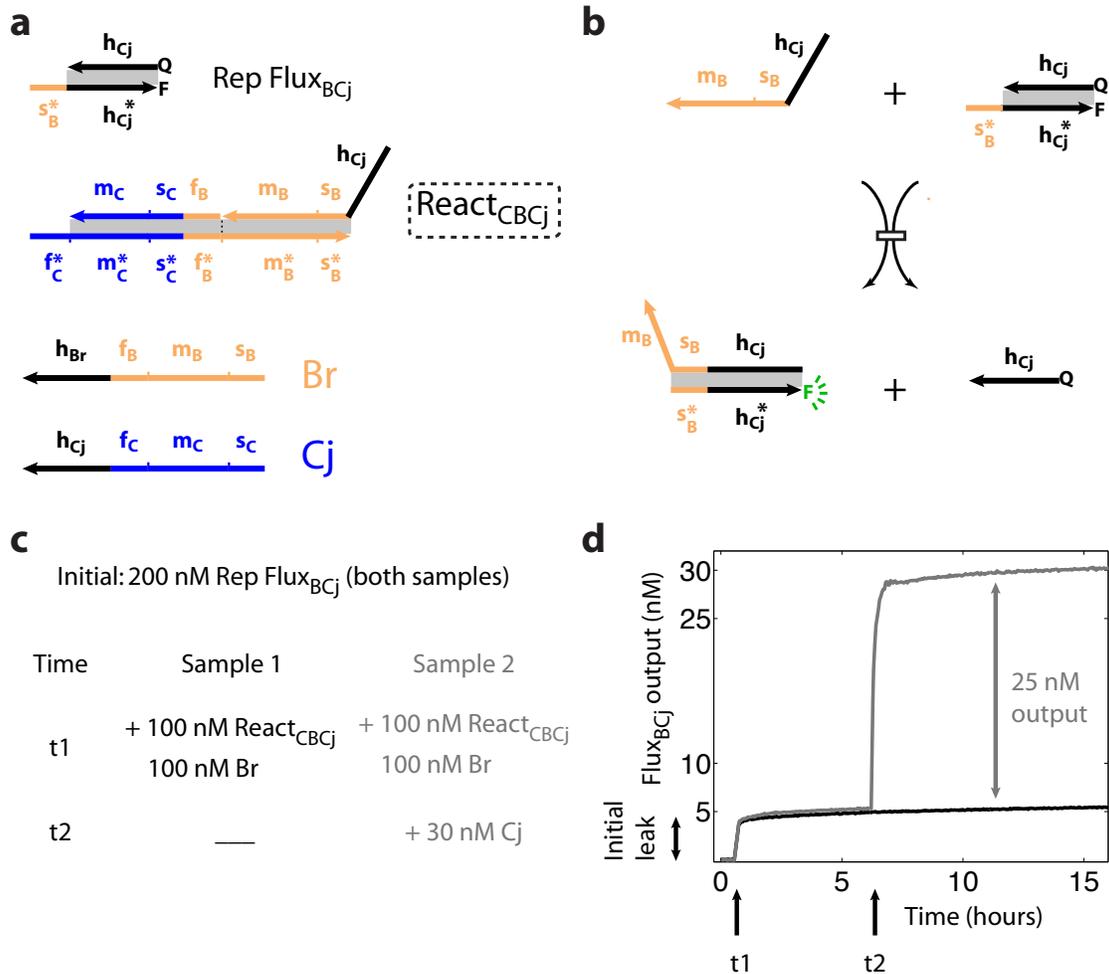
Figure 3.18: Reduced initial and gradual leaks in Design 3, illustrated with $\text{React}_{\text{CBCj}}$-Br leaks. a. The molecules involved. Rep Flux_{BCj} is a reporter for Flux_{BCj}, which works like Rep A in Figure 3.15. For convenience the mechanism is illustrated in (b). c. Experimental setup. Both sample 1 and 2 contain 200 nM Rep Flux_{BCj} initially, and 100 nM $\text{React}_{\text{CBCj}}$ and 100 nM Br are added to both at time t_1. The initial leak is under 5%, which is 2x-3x lower than in Designs 1 and 2. After the initial leak goes to completion, only Sample 2 is triggered with 30 nM of Cj at time t_2. Note the fast triggering in Sample 2 and the much reduced gradual leak in Sample 1 (in general, 5x-10x lower than Design 1; see Figure 3.15 for an example). The apparent incompletion effect is also illustrated. Despite careful quantitation of all species, only 25 nM of Flux_{BCj} is released, which is about 16% less than the full 30 nM expected.

lower initial leak, we decided to try Design 3, which is the same as Design 2 but with 5'-3' orientations reversed.

Since the free energy contributions of individual nearest-neighbor base pair stacks towards double helix stability are not symmetric with respect to 5'-3' orientation, reversing the orientation of our design would perturb the thermodynamics of all our domains, including toeholds. This is an undesirable as it could potentially alter the kinetics of desired strand displacement pathways. In spite of this, we went ahead with testing Design 3.

Experimentally, we found that Design 3 did have much lower initial leaks — reduced to 3-5% of the fuel concentration (see Figure 3.18). This was a big improvement from 10% in Designs 1 and 2. In addition, gradual leaks remained low, except in one particular case, that of $\text{Produce}_{\text{CApAq}}$ and $\text{Helper}_{\text{AAq}}$, where it was very high, approximately 150 /M /s.

Based on careful debugging experiments where we measured the gradual leak with various versions of the $\text{Produce}_{\text{CApAq}}$ and $\text{Helper}_{\text{AAq}}$ molecules, including 1-base mutations, we postulated a remote-toehold [216] style mechanism for the particularly high gradual leak rate (Figure 3.19). This particular gradual leak pathway is tackled in Design 4 (Section 3.8).

While discussing these experimental results with Paul W. K. Rothemund, he suggested that performing these experiments with 0.5 - 1 M Na^+ (as opposed to 12.5 mM Mg^{++}, which was our protocol at that time) may result in lower gradual leak since Mg^{++} is known to stabilize DNA-DNA junctions [194, 236, 237] and might be accelerating this gradual leak pathway by stabilizing invasion of the Helper species at the junction. When we tested the gradual leak pathways in 0.5 M Na^+, we found that there was a reduction in gradual leak across the board by approximately a factor of 2. So, we altered our protocol at this stage to use 0.5 M Na^+ instead of 12.5 mM Mg^{++}. Even though DNA strand displacement kinetics in high sodium (0.5 -1 M Na^+) [145, 146] is not dramatically different from kinetics in 12.5 mM Mg^{++} [147], we experimentally verified that the kinetics of our desired pathways did not slowdown significantly due to the change in salt conditions (experiments not shown).

3.7.1 Kinetics of desired pathways

Apart from investigating leaks, we also experimentally checked whether the desired pathways (such as release of output in appropriate conditions) were occurring with (roughly) the expected kinetics. We found that $\text{React}_{\text{BABr}}$ and $\text{React}_{\text{ACAp}}$ were slow to "trigger" — that is, when both

inputs were present, these molecules were much slower to release their outputs than expected. Compared to $\mathrm{React_{CBCj}}$, where this triggering process takes about 20 minutes (Figure 3.18), these molecules took about 10 hours (slower by a factor of 30).

We re-examined the MFE structures and partition functions of all the complexes involved in NUPACK. We found that hairpins had crept into both these strands, which were supposed to be free of secondary structure, during our last re-design (Figure 3.20). In addition to the MFE structure, we found that the first two bases of the branch migration domain $\mathrm{m_A}$, both G's, were bound almost all the time in some (weak) hairpin or the other. This location is especially critical, as initiation of the first branch migration step is known to be among the slowest unimolecular steps in the strand displacement process and is important in determining kinetics [148]. Given that Ap and Aq are common inputs to both React molecules, we hypothesized that this secondary structure was responsible for the slowdown in triggering both molecules.
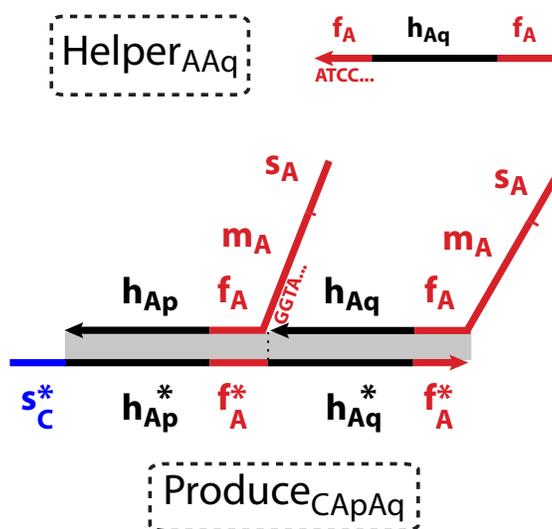
Figure 3.19: Based on experiments measuring gradual leak with single-base changes at the positions illustrated ('ATCC' in $\mathrm{Helper_{AAq}}$ and 'GGTA' in $\mathrm{Produce_{CApAq}}$), these bases contribute to the high gradual leak between $\mathrm{Produce_{CApAq}}$ and $\mathrm{Helper_{AAq}}$. We hypothesize a remote-toehold type mechanism for this gradual leak; the complementarity between 'CC' of $\mathrm{Helper_{AAq}}$ and 'GG' of $\mathrm{Produce_{CApAq}}$ could co-localize the molecules fleetingly to accelerate strand displacement, acting similarly to a strong 2-base pair toehold.
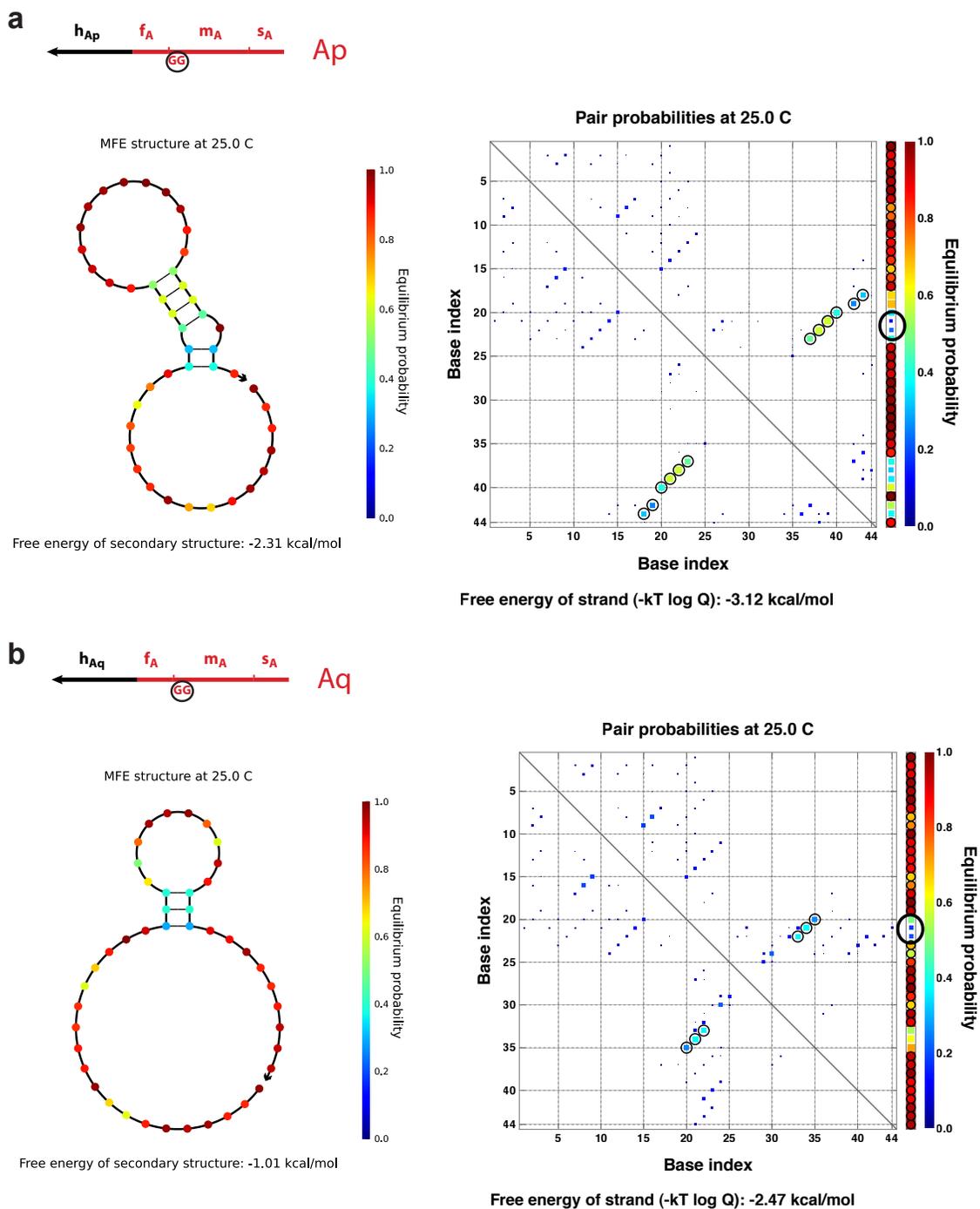
Figure 3.20: NUPACK predicted MFE structure and pair-probabilities matrix for Ap (a) and Aq (b). This secondary structure could slow down desired strand displacement pathways involving these molecules. In particular, the first two bases (GG) of the branch migration region are base paired most of the time and those base pairs occur as a part of several weak hairpins.

3.8   Sequence design 4

In Design 4, we wanted to modify Design 3 to (i) speed up the triggering of $\mathrm{React_{BABr}}$ and $\mathrm{React_{ACAp}}$ to (roughly) match $\mathrm{React_{CBCj}}$ by removing unintended secondary structure in Ap and Aq (Figure 3.20) and (ii) reduce the gradual leak between $\mathrm{Produce_{CApAq}}$ and $\mathrm{Helper_{AAq}}$ by eliminating the remote-toehold style mechanism we postulated in Figure 3.19.

First, we came up with a new heuristic measure for evaluating candidate sequence designs that seeks to quantify the presence of significant secondary structure in strands that are intended to be single-stranded. Let $\mathrm{p_{unpaired}(b, S)}$ be the NUPACK-predicted fraction of a given base b in a given strand S that is unpaired at equilibrium in a solution with only S present (and considering single-stranded complexes only). This measure, Single-Strand Unpaired (SSU), is the minimum of $\mathrm{p_{unpaired}}$ across all bases in all strands that are supposed to be free of secondary structure. That is, if T is the set of all strands that are supposed to be free of secondary structure,

$$\mathrm{SSU} := \min_{S \,\in\, T} \min_{b \,\in\, S} \mathrm{p_{unpaired}}(b, S). \tag{3.14}$$

For Design 3, the SSU was only 0.13, and the base that was the least unpaired was one of the G's in the beginning of the branch migration region $\mathrm{m_A}$. Note that, among all the heuristic measures we have defined, SSU is the only measure where "larger is better".

We decided to modify Design 3 "by hand" to mitigate the problems described above. We identified 8 bases that, if changed, seemed to mitigate one or both of those issues, and identified by trial and error what degrees of freedom those bases had (that is, which options for their new identity did not introduce new secondary structure or malformed complexes as predicted by NUPACK). We changed 3 bases by hand as there seemed to be a clear "best choice" for those 3 bases. For the other 5 bases, we evaluated all candidate designs by brute force on our heuristic measures and chose a design that seemed satisfactory on all those measures. In particular, the SSU for the new design, which we call "Design 4-PRE" increased to 0.76, which was encouraging.

Experimentally, we found (i) a dramatic reduction in the $\mathrm{Produce_{CApAq}}$-$\mathrm{Helper_{AAq}}$ gradual leak, which reduced to approximately 15 /M /s and (ii) a dramatic speedup in the triggering of $\mathrm{React_{BABr}}$, which was now comparable to $\mathrm{React_{CBCj}}$. However, there was no speedup in the triggering of $\mathrm{React_{ACAp}}$, which suggested that another factor was responsible for the slow triggering of $\mathrm{React_{ACAp}}$.

We re-examined everything we knew about these two molecules — the thermodynamics of all the complexes and intermediates, the toehold energies, and so on — and noticed that the toehold exchange reaction in $\mathrm{React_{ACAp}}$ had the two most imbalanced toeholds. The "forward" toehold, f_A, was the weakest and the "backward" toehold, f_C, was the strongest. In addition to that, f_C, being an "internal" toehold, would have an additional (coaxial) stack when bound (see Figure 3.21). As a first order approximation, the imbalance $\Delta\Delta G = \Delta G_{f_C} - \Delta G_{f_A}$, would slowdown the forward toehold exchange rate by a factor of about $10^{\Delta\Delta G}$ [143, 147, 148].

To balance the energies, we tried to remove one or two base pairs from the internal toehold as shown in Figure 3.21, effectively modifying $\mathrm{React_{ACAp}}$ and $\mathrm{Back_{AC}}$ to have a truncated toehold f_C. Experimentally, we were satisfied with the speedup due to a two base pair truncation and called the resulting molecules $\mathrm{React_{ACApi2}}$ and $\mathrm{Back_{ACi2}}$. We call this design, augmented with the 2-base truncations described above, Design 4.

With this modification, all 3 React and all 3 Produce steps in Design 4 had, according to our modeling, low enough gradual leak rates and high enough triggering rates that the DNA implementation could show oscillatory behavior. In the following chapter, we examine our experiments in cascading these molecules to build autocatalytic reactions and a chemical oscillator exclusively based on DNA strand displacement.

We conclude this chapter on debugging molecular networks and sequence design with a note on the "incompletion" effect. We have repeatedly found, across multiple designs, that our fuel species release sub-stoichiometric quantities of output molecules when triggered with a certain nominal concentration of input molecules (see Figures 3.18 and 3.21). Despite careful repeated quantitation of all species involved, and different ways of normalizing our data, we could not quantitatively understand the mechanistic origins of this effect to our satisfaction. This effect has also been reported by Chen et al. [112] in their investigations on implementing another CRN-to-DNA translation scheme [111].

This incompletion effect could arise due to multiple molecular non-idealities, and each of those effects probably partially contributes to it. For example, fuel species with synthesis errors on the output strands could consume their inputs and release non-functional outputs; spurious products formed due to leak pathways could interfere with stoichiometries of desired reaction pathways (Figure 3.13 and discussion thereof). Trying to quantitatively understand this incompletion effect and the extent to which various mechanisms contribute to it is an important direction for future
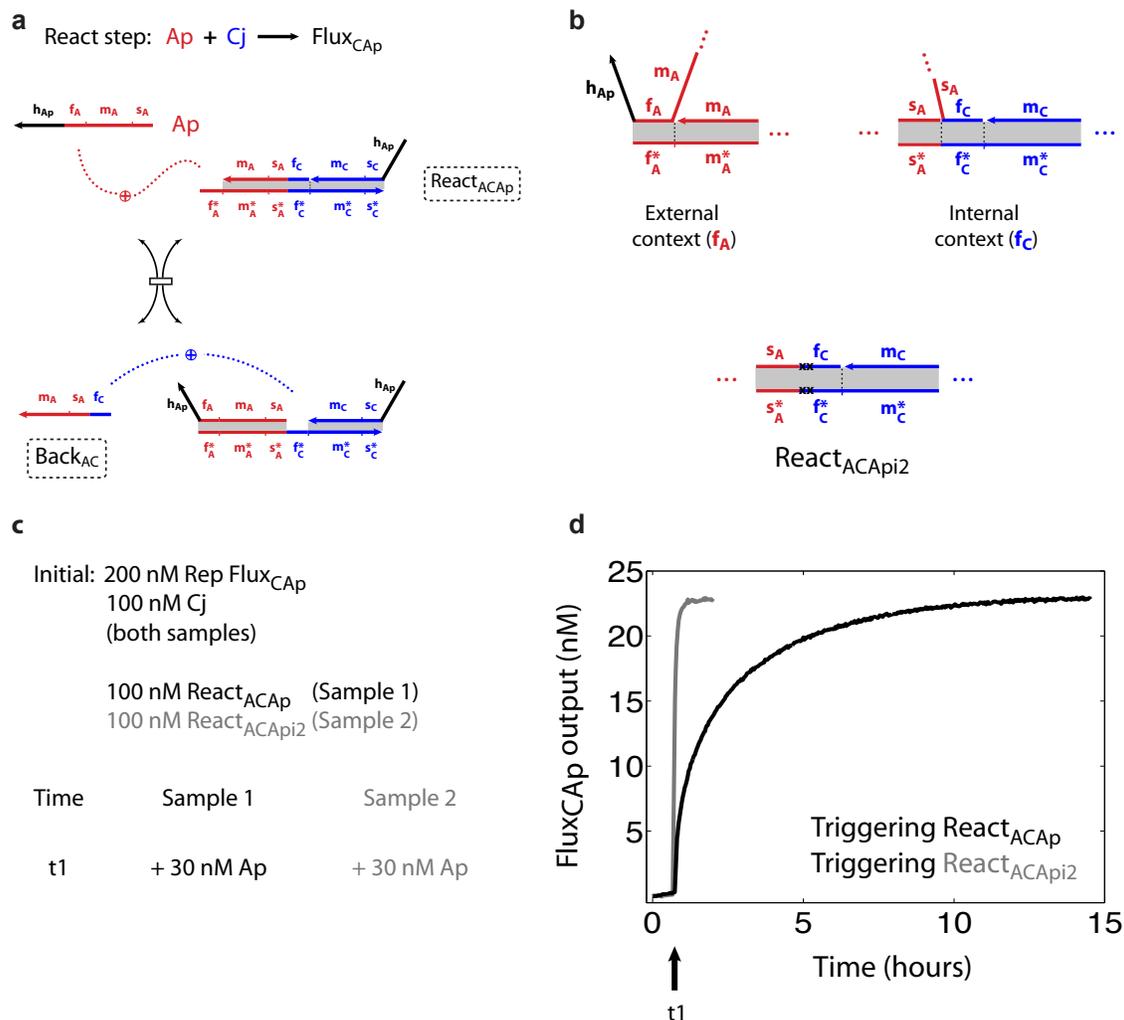
Figure 3.21: Energy imbalance between the external forward toehold (f_A) and the internal backward toehold (f_C) causes slow triggering of $\text{React}_{\text{ACAp}}$. a. Recap of the toehold exchange reaction in triggering $\text{React}_{\text{ACAp}}$. b. Close-up view of the external and internal contexts. Notice that the external toehold, when bound, has only one coaxial stack, while the internal toehold has one on either side. In order to balance the energies, we removed two base pairs from the internal toehold as indicated by truncating the Back_{AC} strand and the bottom strand of $\text{React}_{\text{ACAp}}$ to yield $\text{Back}_{\text{ACi2}}$ and $\text{React}_{\text{ACApi2}}$. c, d. Experiment illustrating that triggering $\text{React}_{\text{ACApi2}}$ is much faster than $\text{React}_{\text{ACAp}}$ (and comparable to the other React molecules). Notice also the apparent incompletion effect: 30 nM of Ap to trigger results in release of about 23 nM of Flux_{CAp}, which is about 23% less than expected.

work.

Supplementary information: All sequence designs and experimental methods are provided in Section 5.3.1.

4   Engineering well-mixed chemical systems with prescribed dynamical behaviors

> *It would not take long to assemble an imposing catalogue of examples illustrating*
> *the resistance of human nature to observations that do not fit into existing theory.*

<div align="right">

- A. T. Winfree [18]

</div>

In Chapter 3, we described our adventures while programming and de-bugging molecular reaction networks. Here, we present our efforts to engineer *in vitro* CRNs with prescribed dynamical behaviors by cascading the molecules and reaction-pathways we optimized in Chapter 3.

We successfully implement three distinct autocatalytic reactions, which we then combine into a *de novo* DNA strand displacement oscillator ("Displacillator"). Unlike biological networks, which use sophisticated evolved molecules (like proteins) to realize such behavior, our test tube realization is the first to demonstrate that Watson-Crick base pairing interactions alone suffice for oscillatory dynamics. Since our design pipeline is general and applicable to any CRN, our experimental demonstration of a *de novo* chemical oscillator could enable the systematic construction of CRNs with other dynamic behaviors.

The scientific contributions described in Chapters 3 and 4 are part of a manuscript currently in preparation:

Niranjan Srinivas, Georg Seelig, Erik Winfree and David Soloveichik, "Programming dynamical behaviors in chemical systems with DNA strand displacement cascades", *in preparation*.

Therefore, the data analysis, modeling and interpretation presented in this chapter should be considered preliminary.

Lastly, I summarize my contributions to this work. When I started working on this project, David Soloveichik, Georg Seelig, and Erik Winfree had already chosen the particular target chem-

ical reaction network (CRN), an oscillator. David had designed the first set of DNA sequences (Design 1 in Chapter 3) and performed preliminary experiments with those. Apart from performing experiments de-bugging Design 1, I came up with all other designs (Designs 2, 3, and 4 in Chapter 3) and performed all other experiments. In that process I formulated and updated the sequence-verification criteria listed in Chapter 3. I also did the modeling described in both Chapter 3 and in this Chapter. Erik, David, and Georg provided valuable feedback throughout these efforts.

4.1   Combining autocatalysis and thresholding into tunable delay elements

As a test of our ability to cascade the individual fuel species from Design 4, we set out to combine the appropriate React and Produce steps into three distinct autocatalytic modules. Our goal was to demonstrate exponential amplification of autocatalyst in each case.

For concreteness, we use the module $C + B \rightarrow 2C$ as a running example in this discussion, but the same holds for the other modules. In preliminary experiments, we found it challenging to control the initial concentrations of autocatalyst (C). When we combined all the fuel species for a given autocatalytic module (React_{CBCj}, Back_{CB} Produce_{BCjCk}, Helper_{CCk}) in the test tube at 100 nM each, along with 50 nM of the signal strand (B) that is consumed by the autocatalyst, the exponential amplification began immediately even without any explicit addition of autocatalyst (C). In fact, the initial rate of amplification (in nM/hr) was quite large and we could not clearly observe the initial "slow" phase of the amplification.

We reasoned that this was due to a high initial leak of autocatalyst (C). This hypothesis is consistent with our measurements of individual initial leaks. We would expect that the approximately 3-5 nM of initially leaked Flux (Flux_{BCj}) would get amplified to about 6 - 10 nM of autocatalyst (C). Combined with the direct release of about 3-5 nM of autocatalyst (C) due to initial leak from the Produce species (Produce_{BCjCk}), this would result in an initial autocatalyst concentration of about 9-15 nM.

In order to control the initial concentrations of the autocatalyst and clearly separate the three processes — initial leak, gradual leak, and exponential amplification — which have different timescales, we combined autocatalysis with thresholding to engineer tunable delay elements. We do not add any autocatalyst and rely on the spurious pathways to provide autocatalyst. The fast

irreversible threshold effectively consumes all the initially leaked autocatalyst before amplification can begin. However, the relatively slow but persistent release of autocatalyst due to gradual leak eventually exhausts all the threshold and kickstarts autocatalytic amplification after that. We may therefore tune the delay before exponential amplification by varying the initial concentration of threshold (Figure 4.1).

In Chapter 3, we had introduced fluorescent reporters for A, B, and C. We use the same molecules as thresholds in the autocatalyst and oscillator experiments. This "threshold readout" allows us to directly observe the consumption of the thresholds Th_A, Th_B, and Th_C. In order to observe both the consumption of the threshold and the autocatalytic amplification simultaneously in the same sample, we combine the threshold readout with a "Helper readout". We achieve this by fluorescently tagging the Helper species with a spectrally distinct fluorophore and placing a quencher on the Produce species, as shown in Figure 4.1. As the autocatalytic amplification proceeds, the Helper species will be consumed to produce the corresponding Waste species and its fluorophore will be quenched. This Helper readout strategy may be used to monitor the progress of any reaction as long as it produces two outputs (and therefore requires a Helper species in the DNA implementation).

A CRN model modified to allow non-integer product stoichiometry ("modified-CRN" model; equations 4.1 to 4.3) semi-quantitatively captures most features of the data, including the temporal correspondence between the exhaustion of threshold and the beginning of exponential amplification. Since the modified-CRN model does not include any fuel species, the quantity plotted in Figure 4.1 lieu of the concentration of $Helper_{CCk}$ consumed is the equivalent quantity at each time T, which is given by $\int_0^T k_{CBCC} \cdot C(t) \cdot B(t)\, dt$.

$$C + Th_C \xrightarrow{k_{ThC}} \phi, \qquad \text{rate} = k_{ThC} \cdot C \cdot Th_C \qquad (4.1)$$

$$\phi \xrightarrow{k_{leakC}} C, \qquad \text{rate} = k_{leakC} \qquad (4.2)$$

$$C + B \xrightarrow{k_{CBCC}} s_C \cdot C, \qquad \text{rate} = k_{CBCC} \cdot C \cdot B \qquad (4.3)$$
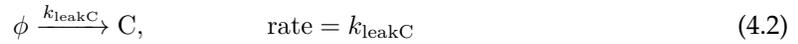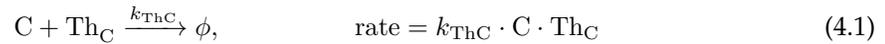
The rate constants k_{leakC} and k_{CBCC} are fitted, while $k_{ThC} = 1.2 \times 10^6$ /M /s has been measured experimentally (see Table 4.3 and Section 4.3.3). The stoichiometry s_C of the output C is also a fitted parameter, since we do not expect our DNA implementation to have perfect stoichiometry for each reaction (see Section 3.3.2). The initial leak was modeled by fitting the initial concentration (C_0)

a

Cj + Br ⇌ Cj + Ck

b

React step: Cj + Br ⟶ Flux$_{BCj}$ Produce step: Flux$_{BCj}$ ⟶ Cj + Ck

Cj

React$_{CBCj}$

Back$_{CB}$

Br

Waste$_{CjBr}$

Waste$_{BCjCk}$ Ck

Helper$_{CCk}$

ROX

Cj

Flux$_{BCj}$

Produce$_{BCjCk}$

c

Cj + Th$_C$

Waste$_{Cj}$ + A488

d Initial: 100 nM Fuels except Produce$_{BCjCk}$ (all samples)
50 nM Br

	Sample 1	Sample 2	Sample 3
Th$_C$ (nM)	10	11	13

t = 0   + 100 nM Produce$_{BCjCk}$ to all samples

e Data

Th$_C$ remaining (nM) — initial / gradual

Helper$_{CCk}$ consumed (nM) — init. grad.

Time (hours)   0 30 60 90 120
initial [Th$_C$] = 10, 11, 13 nM

f modified-CRN model

Th$_C$ remaining (nM)

Helper$_{CCk}$ consumed (nM)

Time (hours)   0 30 60 90 120
initial [Th$_C$] = 10, 11, 13 nM

Figure 4.1: Exponential kinetics of the autocatalytic reaction $C + B \rightarrow 2C$ with delays tuned by threshold concentrations. a, b. Molecular implementation of $C + B \rightarrow 2C$ using our general CRN to DNA scheme. The Helper species is tagged with a fluorophore which gets quenched when it is consumed during the reaction cascade to form Waste. c. Thresholding mechanism. The reporter for C from Chapter 3 is used here as a threshold. It contains a fluorophore spectrally distinct from the Helper which enables simultaneous readout of threshold consumption. d. Experimental setup. Initially, all samples contain 100 nM of the React$_{CBCj}$, Back$_{CB}$ and Helper$_{CCk}$ species (i.e. all fuel species except for Produce$_{BCjCk}$). In addition, all samples contain 50 nM of Br. No C is explicitly added. A threshold species, Th$_C$, is present at 10 nM (sample 1), 11 nM (sample 2), or 13 nM (sample 3). At $t = 0$, 100 nM of Produce$_{BCjCk}$ is added to each sample. e. Experimental inferences of Th$_C$ remaining (top) and Helper$_{CC}$ consumed (bottom). Approximately 8 nM of C is released due to initial leak. This corresponds to the initial drop in the concentration of Th$_C$ (due to consumption of initially leaked C) and the initial spike in the consumption of Helper$_{CC}$ (because Helper$_{CC}$ was consumed in the process of initial leak). After the initial leak, C is released at a slow but constant rate due to the gradual leak. When no more threshold Th$_C$ remains, exponential amplification of C by auto-catalysis takes over. Note the temporal correspondence (indicated by dashed vertical lines) between the times at which Th$_C$ is exhausted and the times at which exponential amplification begins. f. Predictions of a semi-quantitative CRN-level model comprising reactions 4.1 to 4.3.

of autocatalyst (C), which was also assumed to be the amount of Helper (Helper_CCk) consumed in the initial leak process. The best fit values were

$$(\text{C}_0, \text{k}_\text{CBCC}, \text{k}_\text{leakC}, s_\text{C}) = (6.6 \text{ nM}, 2.2 \times 10^4 \text{ /M /s}, 0.05 \text{ nM/hr}, 1.5).$$

Note that $s_\text{C} = 1.5$ suggests lower than expected output stoichiometry, which is consistent with our understanding based on leak and triggering experiments on individual fuel species (Chapter 3).

The biggest discrepancy between the semi-quantitative model and the experimental data lies in the shape of the threshold concentration curves. The model predicts that the threshold concentration is depleted at a constant rate due to gradual leak, whereas the experimental data shows a more complex variation in the rate of threshold depletion. A mechanistic model of the DNA implementation seems necessary for understanding this feature of the data.

Figure 4.2 demonstrates exponential amplification with tunable delay based on the other two autocatalytic modules. Having succeeded in cascading the individual fuel species into three autocatalytic modules, we attempted to combine all three autocatalytic modules in one test tube to implement a *de novo* oscillator based on DNA strand displacement.

a

$$\phi \xrightarrow{\text{slow}} C$$
$$C + Th_C \xrightarrow{\text{fast}} \phi$$
$$C + B \longrightarrow 2C$$

initial [Th$_C$] = 10, 11, 13 nM

b

$$\phi \xrightarrow{\text{slow}} A$$
$$A + Th_A \xrightarrow{\text{fast}} \phi$$
$$A + C \longrightarrow 2A$$

initial [Th$_A$] = 10, 11, 13 nM

c

$$\phi \xrightarrow{\text{slow}} B$$
$$B + Th_B \xrightarrow{\text{fast}} \phi$$
$$B + A \longrightarrow 2B$$

initial [Th$_B$] = 10, 11, 13 nM
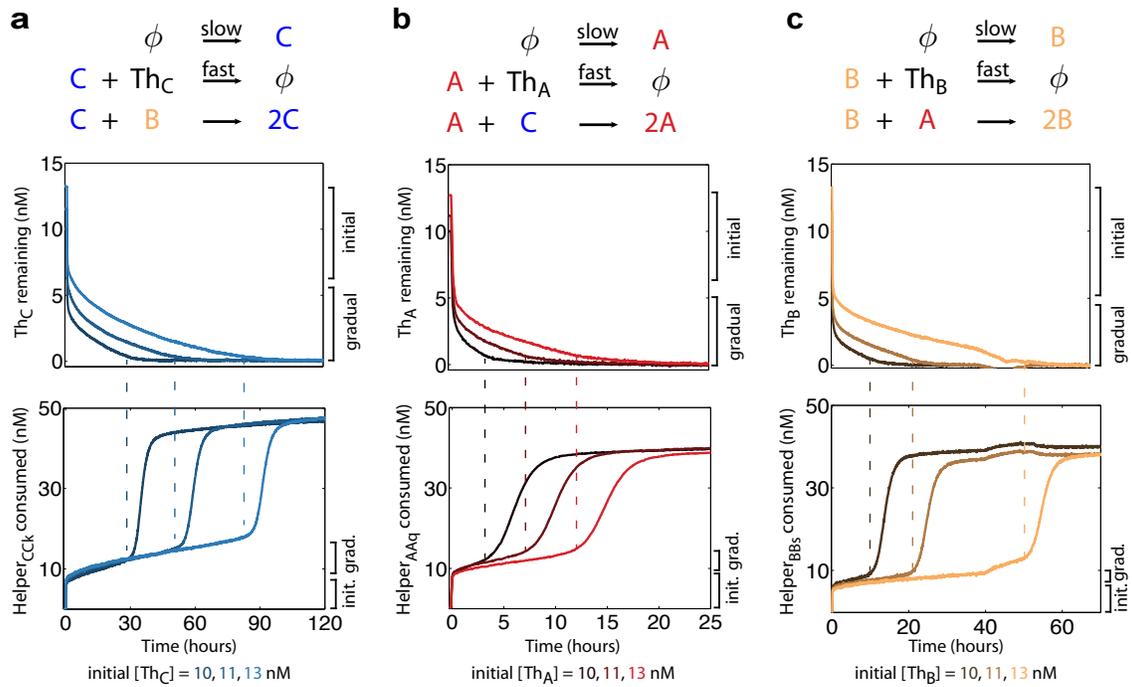
Figure 4.2: Exponential kinetics of all three autocatalytic reactions with delays tuned by threshold concentrations. These modules use the same kinds of threshold and Helper readouts but with different pairs of fluorophores.

4.2 Displacillator: a *de novo* strand displacement oscillator

In trying to combine our autocatalytic modules into a strand displacement oscillator, we encountered several challenges. First, we had to achieve initial values for the concentrations of (A, B, C) that would be suitable for obtaining and observing oscillations. The initial concentrations needed to be different enough that we would obtain observably large oscillations, while still being much smaller than the concentration of our fuel species in order to ensure multiple cycles of oscillation before the fuel species are exhausted. Second, in our initial experiments, we observed strongly damped oscillations. We hypothesized that the damping arose from each autocatalytic module having lower than ideal stoichiometry due to molecular non-idealities (Section 4.1 and Chapter 3).

4.2.1 Counteracting damping: Catalytic helper mechanism

One way of counteracting damping and "tuning up" our Displacillator would be to engineer an alternative mechanism for the Helper step where, in addition to the second output of the Produce species, the Flux strand is also released. If that happens, the Flux strand can continue to interact with more Produce species and release more outputs, thereby effectively increasing output stoichiometry. Inspired by Zhang et al. [143]'s entropy driven catalyst, we came up with the "catalytic Helper" mechanism (see Figure 4.3) to engineer such an alternative Helper pathway. For clarity we note here that in our case, the entropic benefit is a bonus and not the only reason our desired pathway is thermodynamically favorable.

A catalytic Helper (CatHelper) species is a modified version of the corresponding Helper species that releases not just the second output, but also the Flux strand (by toehold exchange). The released Flux strand is subsequently free to interact with more Produce species and release more outputs, thereby increasing the output stoichiometry of the desired CRN. By tuning the concentration of CatHelper species relative to Helper species, we are able to control the probability of each pathway, thereby providing a flexible knob for increasing the output stoichiometry of our reactions to compensate for the incompletion effect. From now on, the term "total-Helper" for a given autocatalytic module refers to the sum of CatHelper and Helper species for that module.
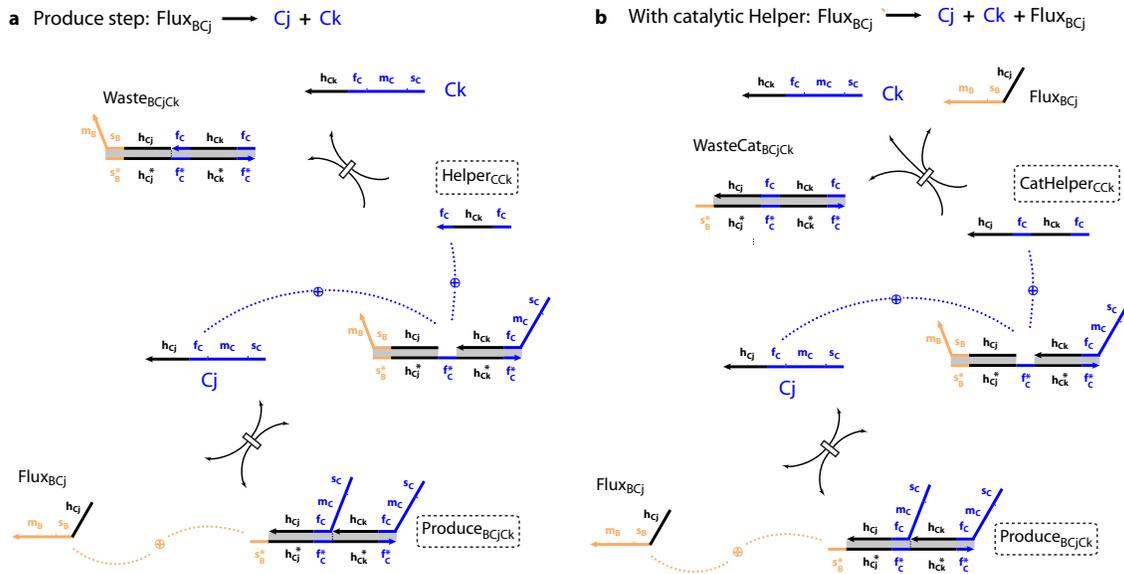
Figure 4.3: Catalytic Helper mechanism. a. Standard produce step for the reaction $C + B \rightarrow 2C$ facilitated by the traditional "Helper" species. b. Produce step for the reaction $C + B \rightarrow 2C$ using the "catalytic Helper" species. The CatHelper strand is nothing but the Helper strand extended at the 5' end with the history domain of the first output of the Produce species (here, h_{Cj}). Apart from releasing the second output (here, Ck) the catalytic Helper also displaces the Flux strand by toehold exchange, which is then free to interact with more Produce species to release more outputs, thereby effectively "tuning up" the output stoichiometry of the desired CRN. The term "total-Helper" for a given autocatalytic module refers to the sum of catalytic and standard Helper species for that module.

4.2.2 Optimized Displacillator experiments

We describe our optimized experimental protocol for Displacillator experiments in Figure 4.4. We attempt to control the initial concentrations of A, B, and C by using thresholds to irreversibly consume initial leaks, and subsequently adding the desired amounts of A, B, and C by hand to kickstart oscillation. It is challenging to exactly balance the initial concentrations of thresholds such that both our objectives, the consumption of the initial leak and the consequent extinction of thresholds, are achieved in reasonable time. This is illustrated in Figure 4.4, which shows that at the time of the second addition, about 7 nM of Th_A is leftover (see Sample 1 in panel (b)). We do not, at this time, quantitatively understand this transient phase where initial leaks, desired reactions, and thresholding all occur at the same time; nor do we fully understand the effect of the leftover thresholds on the observed dynamical behavior. Therefore we do not yet know how to optimally set the initial concentrations of thresholds. These experiments have been optimized by trial and error.

Similarly, we tuned the relative concentrations of Helper and CatHelper (ratio of 3:1) in order to optimize the number of oscillatory cycles observed. These ratios for the three modules may likely permit further optimization if aided by quantitative mechanistic modeling.

We use two copies of each sample: one to observe the consumption of Helper and CatHelper species ("Helper readout") and the other to observe the consumption of thresholds ("threshold readout"). The first sample uses Helper and CatHelper species that are fluorescently tagged and plain versions of thresholds, which do not have fluorophores or quenchers. In the other copy, plain versions of Helper and CatHelper species are used along with fluorescently tagged thresholds. Produce species in both samples are labeled with a quencher on the bottom strand.

Experiments where the Displacillator is started from multiple initial conditions are presented in Figure 4.5. The rate of consumption of the different Helper species, which measures the instantaneous rate (in nM/hr) at which each autocatalytic module is proceeding (apart from a small contribution from gradual leak), clearly shows that the modules are "firing" one after another in the expected order.

a

Initial: 100 nM React and Back species (both samples)

Sample 1   Sample 2

10 nM Th_A, Th_B, Th_C 10 nM $\text{Th}_A{}^\dagger$, $\text{Th}_B{}^\dagger$, $\text{Th}_C{}^\dagger$

75 nM Helpers AAq†, BBs†, CCk† 75 nM Helpers AAq, BBs, CCk

25 nM PL-CatHelpers AAq†, BBs†, CCk† 25 nM CatHelpers AAq, BBs, CCk

Addition 1: + 100 nM Produce species to both samples

Addition 2: + (0, 10, 13) nM of (Ap, Br, Cj) to both samples

b
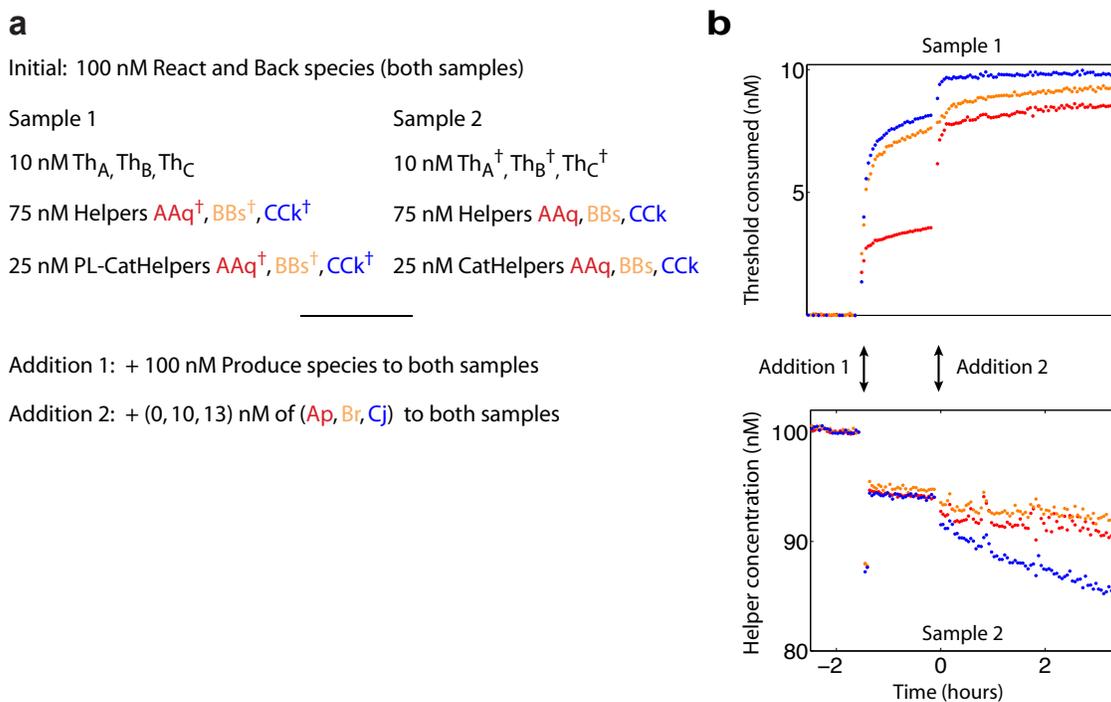
Figure 4.4: Experimental setup for Displacillator experiments. Two samples are used for each experiment: Sample 1 uses "plain" versions of Helper and CatHelper (indicated with a †), which do not contain fluorophores, and thresholds Th_A, Th_B, and Th_C with fluorophores. Sample 2 uses plain versions of thresholds (indicated with a †), which do not contain fluorophores, and Helper and CatHelper species with fluorophores. Otherwise Samples 1 and 2 are identical. (In particular, Produce species in both samples are labeled with a quencher on the bottom strand.) Helper:CatHelper ratio of 3:1 was arrived at by experimental tuning. a. All fuel species except Produce species are present initially, along with 10 nM of each threshold. After measuring the background, a "soup" containing 100 nM of each Produce species is added to each sample. This causes initial leak of signal strands A, B, and C, which is reflected in (i) the initial drop in total-Helper concentration (Sample 1) because these species are consumed and quenched during initial leak and (ii) the initial spike in consumption of thresholds (Sample 1) as the leaked signal strands are consumed. Once the initial leak stabilizes, we add (A, B, C) to kickstart oscillation with the initial conditions we desire.
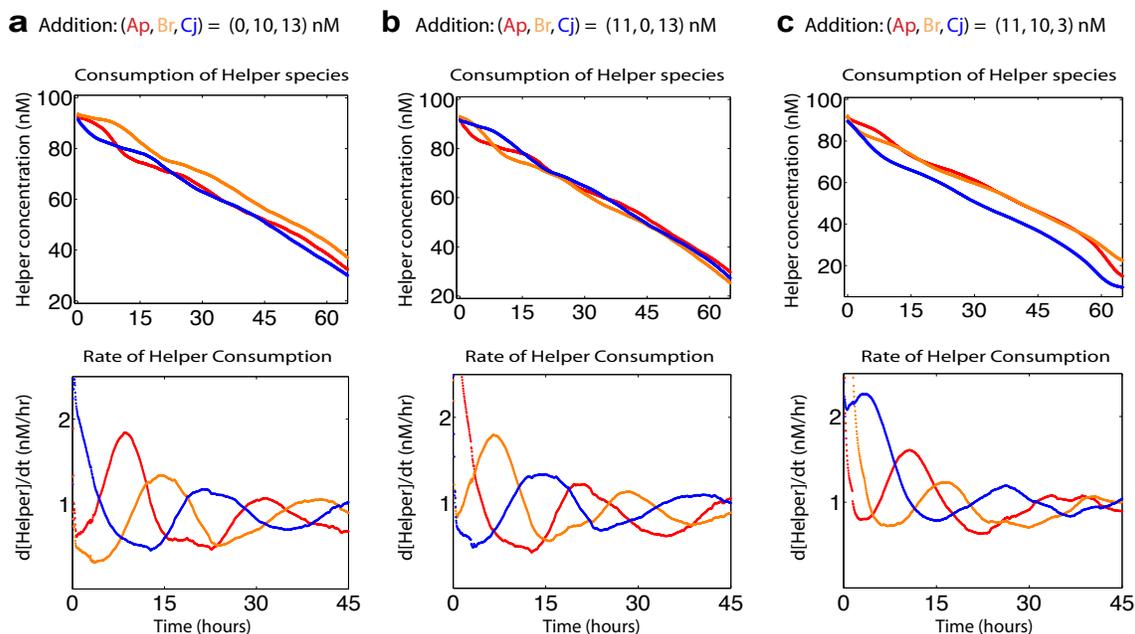
Figure 4.5: Displacillator experiments from multiple initial conditions. Total-Helper (Helper + CatHelper) concentrations are presented in the top row; time derivatives of those measurements are presented in the bottom row. Time derivatives of Total-Helper measurements indicate instantaneous rate (in nM/hr) at which each autocatalytic module is proceeding in the test tube, apart from consumption of Total-Helper due to gradual leak. The sequential 'firing' of the autocatalytic modules in the expected order (as long as the fuel species are at significant concentration) indicates oscillatory behavior.

4.2.3 Inferring signal strand concentrations

Ideally, we would like to directly measure the concentrations of the signal strands A, B, and C. Unfortunately, such direct measurement is difficult since irreversibly consuming any of these strands for reading out the concentration would alter the dynamical behavior of our system. Tagging the target species with fluorophores has been shown to enable direct monitoring of dynamic circuits without irreversible consumption [238]. However, in our design this approach necessitates placement of fluorophores in the middle of the first toehold of each signal strand (f_A, f_B, and f_C). Since modifying toeholds could significantly perturb strand displacement and toehold exchange rates, we chose to infer the concentrations of A, B, and C from measurements of Helper concentrations. We now discuss two approaches for doing so.

4.2.3.1 Ideal stoichiometry approach

The CRN we seek to implement (ignoring rate constants) is

$$B + A \longrightarrow 2\,B \tag{4.4}$$

$$C + B \longrightarrow 2\,C \tag{4.5}$$

$$A + C \longrightarrow 2\,A. \tag{4.6}$$
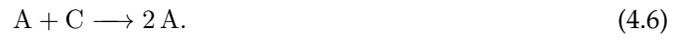
With two assumptions, there is an easy way of inferring signal strand concentrations from Helper measurements. The first assumption is that our DNA implementation is faithful to the stoichoimetry specified by the formal CRN in equations 4.4 - 4.6. The second assumption is that the contribution of gradual leaks (such as Produce-Helper gradual leaks) to total-Helper consumption is negligible. For the purposes of this section, $\mathrm{Helper}_i(t)$ for an autocatalytic module i is defined as the total-Helper concentration for that module at time t.

With the above two assumptions, we can define

$$\mathrm{Flow_{BABB}(t)} := \mathrm{Helper_{BBs}}(0) - \mathrm{Helper_{BBs}}(t) \tag{4.7}$$

$$\mathrm{Flow_{CBCC}(t)} := \mathrm{Helper_{CCk}}(0) - \mathrm{Helper_{CCk}}(t) \tag{4.8}$$

$$\mathrm{Flow_{ACAA}(t)} := \mathrm{Helper_{AAq}}(0) - \mathrm{Helper_{AAq}}(t) \tag{4.9}$$

to be the progress through each reaction upto time t (in nM). In other words, $\mathrm{Flow_{BABB}}(t)$ measures "how much" of reaction 4.4 has taken place upto time t. Consequently, one can infer the

a Addition: $(A_p, B_r, C_j) = (0, 10, 13)$ nM

b Addition: $(A_p, B_r, C_j) = (11, 0, 13)$ nM

c Addition: $(A_p, B_r, C_j) = (11, 10, 3)$ nM
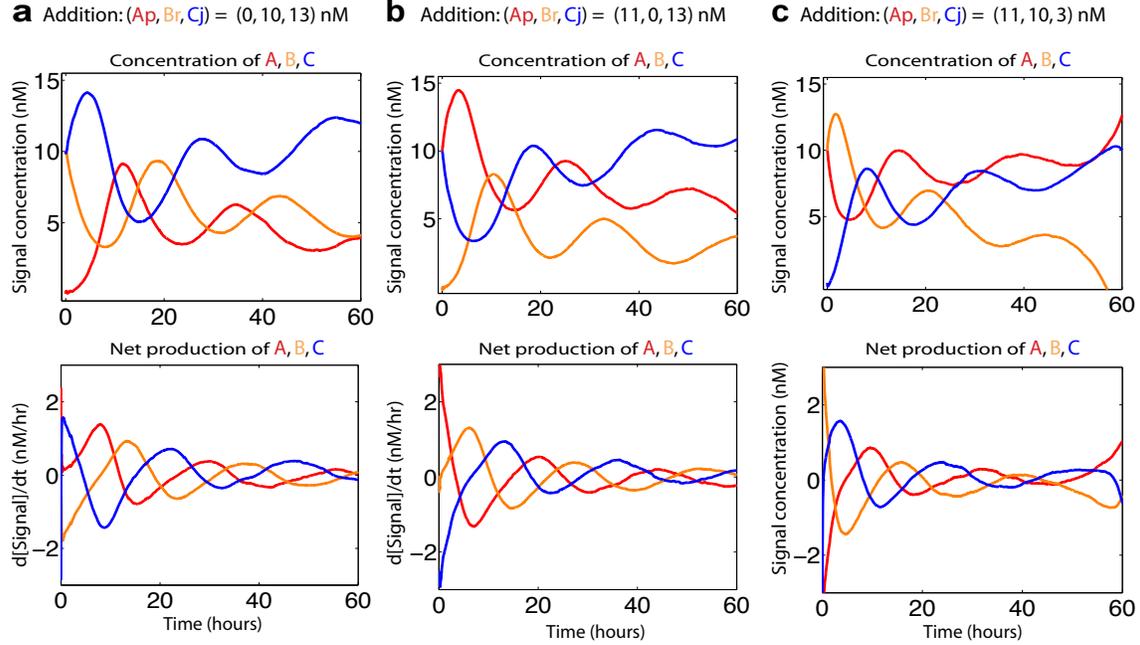
Figure 4.6: Inferences of A(t), B(t) and C(t) concentrations with the ideal stoichiometry approach outlined in Section 4.2.3.1. Time derivatives, which signify net production rate of each signal strand, are also presented.

concentrations of A, B, and C as a function of time as follows:

$$A(t) = A(0) - \text{Flow}_{\text{BABB}}(t) + \text{Flow}_{\text{ACAA}}(t) \tag{4.10}$$

$$B(t) = B(0) - \text{Flow}_{\text{CBCC}}(t) + \text{Flow}_{\text{BABB}}(t) \tag{4.11}$$

$$C(t) = C(0) - \text{Flow}_{\text{ACAA}}(t) + \text{Flow}_{\text{CBCC}}(t). \tag{4.12}$$

Inferences of A(t), B(t), and C(t) concentrations under this ideal stoichiometry approach are provided in Figure 4.6. Here initial concentrations A(0), B(0) and C(0) are calculated from (i) the amounts added to kickstart the oscillations and (ii) the threshold concentrations remaining at the time of addition.

In this method, the assumption of ideal stoichiometries is far more crucial than the choice to neglect gradual leak contributions. It would be possible to easily correct definitions 4.7- 4.9 for gradual leaks, say with independently measured gradual leak parameters for a certain model of gradual leak. However, the stoichiometry assumptions are hard to justify, for two reasons. First,

we know from the experiments and discussions in Chapter 3 that individual React and Produce steps yield sub-stoichiometric outputs. Second, although we have compensated for such incompletion effects by using the catalytic Helper mechanism, it is not clear at all that we have achieved exact compensation and restored ideal stoichiometry for our reactions.

4.2.3.2 Phenomenological model

To avoid the rigid stoichiometry assumptions in Section 4.2.3.1, which are hard to justify, we present here an alternative phenomenological model where reactant and product stoichiometries are fit parameters. This model, presented in Equations 4.13 - 4.18, attempts to infer signal concentrations in an internally consistent way. For the purposes of this section, $\text{Helper}_i(t)$ for an autocatalytic module i is defined as the total-Helper concentration for that module at time t.

$$\text{r}_{\text{BA}} \cdot (\text{B} + \text{A}) \xrightarrow{k_{\text{BABB}}} \text{p}_{\text{BA}} \cdot \text{B}, \qquad \text{rate} = k_{\text{BABB}} \cdot \text{B} \cdot \text{A} \tag{4.13}$$

$$\text{r}_{\text{CB}} \cdot (\text{C} + \text{B}) \xrightarrow{k_{\text{CBCC}}} \text{p}_{\text{CB}} \cdot \text{C}, \qquad \text{rate} = k_{\text{CBCC}} \cdot \text{C} \cdot \text{B} \tag{4.14}$$

$$\text{r}_{\text{AC}} \cdot (\text{A} + \text{C}) \xrightarrow{k_{\text{ACAA}}} \text{p}_{\text{AC}} \cdot \text{A}, \qquad \text{rate} = k_{\text{ACAA}} \cdot \text{A} \cdot \text{C} \tag{4.15}$$

$$\phi \xrightarrow{k_{\text{leakA}}} \text{A}, \qquad \text{rate} = k_{\text{leakA}} \tag{4.16}$$

$$\phi \xrightarrow{k_{\text{leakB}}} \text{B}, \qquad \text{rate} = k_{\text{leakB}} \tag{4.17}$$

$$\phi \xrightarrow{k_{\text{leakC}}} \text{C}, \qquad \text{rate} = k_{\text{leakC}} \tag{4.18}$$

Rate constants $k_{\text{BABB}}, k_{\text{CBCC}}, k_{\text{ACAA}}$ are in /M /s; rate constants $k_{\text{leakA}}, k_{\text{leakB}}, k_{\text{leakC}}$ are in M /s. r_i and p_i determine effective stoichiometries for the reactants and products, respectively. Along with the initial conditions $(\text{A}(0), \text{B}(0), \text{C}(0))$, which are fitted for each experiment, these comprise the parameters (Θ) of the phenomenological model.

How do we interpret the parameters r_i and p_i? For a given autocatalytic module, r_i is interpreted as the average number of reactants consumed per unit consumption of total-Helper species for that module. Similarly, p_i is interpreted as the average number of products released per unit consumption of total-Helper species for that module.

Interpreting the data. We now outline how, given a parameter set Θ and the data set D containing measurements of total-Helper concentrations, we infer concentrations A(t), B(t), and C(t).

We define the leak-corrected consumption of total-Helper as below:

$$\text{Flow}^{\Theta,\text{D}}_{\text{BABB}}(t) := \text{Helper}_{\text{BBs}}(0) - \text{Helper}_{\text{BBs}}(t) - k_{\text{leakB}} \cdot t \tag{4.19}$$

$$\text{Flow}^{\Theta,\text{D}}_{\text{CBCC}}(t) := \text{Helper}_{\text{CCk}}(0) - \text{Helper}_{\text{CCk}}(t) - k_{\text{leakC}} \cdot t \tag{4.20}$$

$$\text{Flow}^{\Theta,\text{D}}_{\text{ACAA}}(t) := \text{Helper}_{\text{AAq}}(0) - \text{Helper}_{\text{AAq}}(t) - k_{\text{leakA}} \cdot t. \tag{4.21}$$

This leak correction involves an implicit assumption: that gradual leaks of signal strands arise primarily from the Produce-Helper gradual leak mechanism. This is a good approximation because the Produce-Helper gradual leak arises from two fuel species which are both at high concentration, in contrast to the React-second input gradual leak where only the React species is at high concentration. In addition, for Design 4, the React-second input gradual leaks have a rate constant of about 1 - 5 /M /s, whereas Produce-Helper gradual leaks have a relatively larger rate constant between 5 - 15 /M /s.

By taking time derivatives of equations 4.19 - 4.21, we can infer instantaneous rates of total-Helper consumption through each autocatalytic module.

$$\text{FlowRate}^{\Theta,\text{D}}_{\text{BABB}}(t) := -\frac{d\text{Helper}_{\text{BBs}}(t)}{dt} - k_{\text{leakB}} \tag{4.22}$$

$$\text{FlowRate}^{\Theta,\text{D}}_{\text{CBCC}}(t) := -\frac{d\text{Helper}_{\text{CCk}}(t)}{dt} - k_{\text{leakC}} \tag{4.23}$$

$$\text{FlowRate}^{\Theta,\text{D}}_{\text{ACAA}}(t) := -\frac{d\text{Helper}_{\text{AAq}}(t)}{dt} - k_{\text{leakA}}. \tag{4.24}$$

The interpretation of r_i and p_i as average number of reactants consumed and products released per unit consumption of total-Helper provides a clean way of inferring concentrations of signal strands from measurements of total-Helper concentrations, given a set of model parameters Θ.

$$\text{A}^{\Theta,\text{D}}(t) = \text{A}(0) - r_{\text{BA}} \cdot \text{Flow}^{\Theta,\text{D}}_{\text{BABB}}(t) + (p_{\text{AC}} - r_{\text{AC}}) \cdot \text{Flow}^{\Theta,\text{D}}_{\text{ACAA}}(t) + k_{\text{leakA}} \cdot t \tag{4.25}$$

$$\text{B}^{\Theta,\text{D}}(t) = \text{B}(0) - r_{\text{CB}} \cdot \text{Flow}^{\Theta,\text{D}}_{\text{CBCC}}(t) + (p_{\text{BA}} - r_{\text{BA}}) \cdot \text{Flow}^{\Theta,\text{D}}_{\text{BABB}}(t) + k_{\text{leakB}} \cdot t \tag{4.26}$$

$$\text{C}^{\Theta,\text{D}}(t) = \text{C}(0) - r_{\text{AC}} \cdot \text{Flow}^{\Theta,\text{D}}_{\text{ACAA}}(t) + (p_{\text{CB}} - r_{\text{CB}}) \cdot \text{Flow}^{\Theta,\text{D}}_{\text{CBCC}}(t) + k_{\text{leakC}} \cdot t. \tag{4.27}$$

Model predictions. Given the reactions and rates (4.13 - 4.18), we may generate model predic-

tions by numerically simulating the following mass action ODEs. We define

$$\text{F}^{\Theta}_{\text{BABB}}(t) := k_{\text{BABB}} \cdot \text{B}(t) \cdot \text{A}(t) \tag{4.28}$$

$$\text{F}^{\Theta}_{\text{CBCC}}(t) := k_{\text{CBCC}} \cdot \text{C}(t) \cdot \text{B}(t) \tag{4.29}$$

$$\text{F}^{\Theta}_{\text{ACAA}}(t) := k_{\text{ACAA}} \cdot \text{A}(t) \cdot \text{C}(t). \tag{4.30}$$

The mass action ODEs are

$$\left(\frac{d\text{A}}{dt}\right)^{\Theta} = -r_{\text{BA}} \cdot \text{F}^{\Theta}_{\text{BABB}}(t) + (p_{\text{AC}} - r_{\text{AC}}) \cdot \text{F}^{\Theta}_{\text{ACAA}}(t) + k_{\text{leakA}} \tag{4.31}$$

$$\left(\frac{d\text{B}}{dt}\right)^{\Theta} = -r_{\text{CB}} \cdot \text{F}^{\Theta}_{\text{CBCC}}(t) + (p_{\text{BA}} - r_{\text{BA}}) \cdot \text{F}^{\Theta}_{\text{BABB}}(t) + k_{\text{leakB}} \tag{4.32}$$

$$\left(\frac{d\text{C}}{dt}\right)^{\Theta} = -r_{\text{AC}} \cdot \text{F}^{\Theta}_{\text{ACAA}}(t) + (p_{\text{CB}} - r_{\text{CB}}) \cdot \text{F}^{\Theta}_{\text{CBCC}}(t) + k_{\text{leakC}}. \tag{4.33}$$

Given the initial concentrations of A, B, and C specified in Θ, we may numerically solve the mass action ODEs above to generate model predictions for A(t), B(t), and C(t). Further, the predicted curves for A(t), B(t), and C(t) imply a prediction of the instantaneous rates (in nM/hr) through each autocatalytic module. There is a subtlety involved in this calculation. The instantaneous rate (in nM/hr) of consumption of the *reactants* of each autocatalytic module is calculated as specified in equations 4.28 - 4.30. However, since we measure total-Helper concentrations, we would like to calculate the instantaneous rate of consumption of *total-Helper species*, rather than the reactants, for each autocatalytic module. Since r_i is interpreted as the average number of reactants consumed per unit total-Helper consumption, all we need to do is to divide equations 4.28 - 4.30 by the corresponding r_i; this gives us the instantaneous rate of consumption of total-Helper species for each module, according to our model.

Parameter fitting. We simultaneously fit (i) the model predicted signal concentrations ($\text{A}^{\Theta}(t)$, $\text{B}^{\Theta}(t)$, $\text{C}^{\Theta}(t)$) to the signal concentrations inferred from data ($\text{A}^{\Theta,\text{D}}(t)$, $\text{B}^{\Theta,\text{D}}(t)$, $\text{C}^{\Theta,\text{D}}(t)$) and (ii) the model predicted instantaneous rate of total-Helper consumption for each autocatalytic module (calculated as $\text{F}^{\Theta}_{\text{BABB}}(t)/r_{\text{BA}}$ and so on) to the same quantity inferred from data using equations 4.22 - 4.24.

We briefly address the fact that the two fitting criteria, (i) and (ii) above, are in different units (nM and nM/hr respectively). As a quick preliminary solution to this issue, we just add the two

Parameter	Best-fit value
(r_{BA}, r_{CB}, r_{AC})	$(1.2, 0.8, 0.8)$
(p_{BA}, p_{CB}, p_{AC})	$(2.0, 1.4, 1.9)$
$(k_{leakB}, k_{leakC}, k_{leakA})$	$(2, 2, 6)$ in /M /s
$(k_{BABB}, k_{CBCC}, k_{ACAA})$	$(1.2 \times 10^4, 7.9 \times 10^3, 6.4 \times 10^3)$ in /M /s

Table 4.1: Best-fit values for parameters from the phenomenological model. The leak rate constants $k_{leakB}, k_{leakC}, k_{leakA}$ are provided in terms of the equivalent bimolecular Produce-Helper leak rate constants they imply, assuming Produce and Helper species are at 100 nM each.

error functions corresponding to (i) and (ii) above to obtain the objective function which is subsequently minimized by nonlinear least squares. Essentially, this procedure involves an implicit weighting arising from the different units. Currently, the optimal error value has a ratio of approximately 20 : 1 in favor of the first component. Further analysis with a more principled choice of weights may result in better fits.

Interpreting stoichiometric best fit parameters. The best fit values (Θ_{opt}) for the parameters are listed in Table 4.1. The initial concentrations of (A, B, C) for each experiment, which were fitted around expected values, are omitted for brevity. We now attempt to interpret the stoichiometric best fit parameters.

Since r_i is the average number of reactants consumed per unit of total-Helper consumption, $r_i < 1$ implies that more than one unit of total-Helper is consumed, on average, per unit consumption of reactants. This indicates the effectiveness of the catalytic Helper mechanism. While at first glance $r_{BA} = 1.2$ may be surprising, we point out that in this preliminary analysis we do not have estimates for the error bars associated with these parameters; therefore, it is hard to interpret the best fit parameters at face value.

Since p_i is the average number of products released per unit of total-Helper consumption, $p_i < 2$ implies sub-stoichiometric release of products per unit of total-Helper consumed. This could arise from incompletion effects in the produce step, possibly due to mechanisms similar to that of Figure 3.13 (b).

Lastly, p_i/r_i, the average number of products released per unit consumption of reactants, is

approximately (1.7, 1.6, 2.2) for the three modules.

Results from the phenomenological model. The results of the data analysis with the phenomenological model are provided in Figures 4.7 and 4.8. In particular, the phase and velocity plots (panels (e) and (h) of Figure 4.8) clearly illustrate the triangular orbit which is characteristic of the Displacillator.
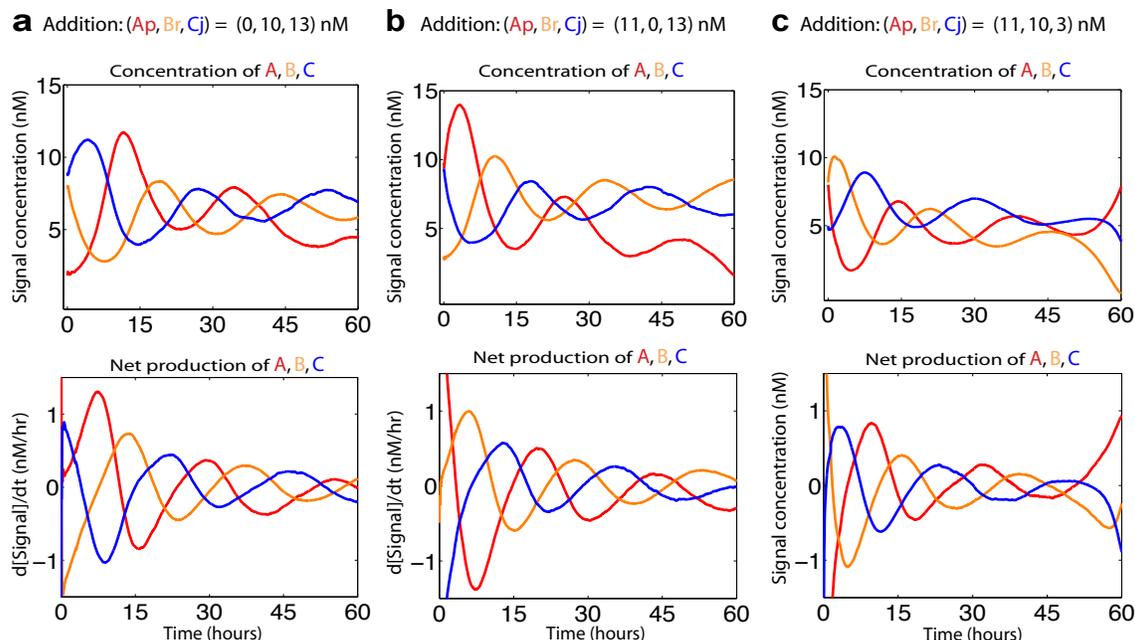
Figure 4.7: a. Concentrations of A, B, and C inferred from Helper concentration measurements based on the phenomenological model in Section 4.2.3.2. b. Time derivatives of A, B, and C. The order of the peaks is as expected in experiments with all three initial conditions.

4.3 Mechanistic model of the Displacillator

Given that the Displacillator has been engineered at the level of molecular pathways, where each strand displacement or toehold exchange reaction can be quantitatively understood, it would be very satisfying to be able to explain the dynamical behavior of the system directly in terms of those designed elementary reactions. In particular, our focus in this section is to use a "mechanistic model" to semi-quantitatively explain the temporal features of the data (such as period of oscillation) in terms of elementary reactions.
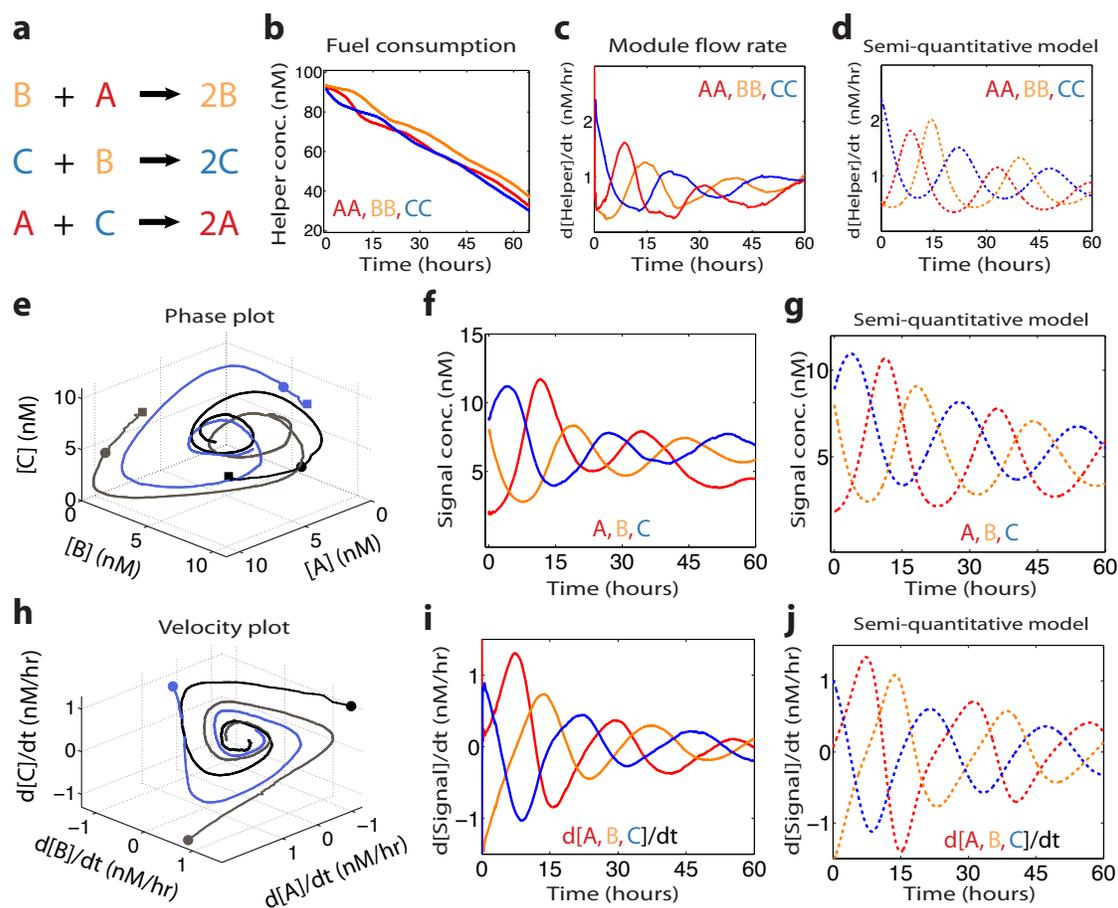
Figure 4.8: Experimental demonstration of the Displacillator, a *de novo* oscillator based on DNA strand-displacement. All solid lines (including phase and velocity plots) are measurements or model-based inferences from experimental data; model predictions are indicated by dashed lines. The model used is the semi-quantitative phenomenological model presented in Section 4.2.3.2. a. The 3 autocatalytic modules. b. Measurement of fuel consumption (concentrations of Helper species) in one experiment. c. Instantaneous reaction rates (nM/hr) for each autocatalytic reaction, inferred from measurements in (b). d. Model predictions corresponding to the module flow rates plotted in (c). e. Phase plot comprising the three different trajectories, starting from 3 different initial points (indicated by squares), shown in Figure 4.7. f. Inferred A, B, C trajectories for the experiment shown in (b), corresponding to the pale blue trajectory in (e). g. Model predictions corresponding to the data in (f). h. Velocity plots for the three trajectories shown in (e), which follow a roughly triangular orbit as expected for this oscillator. Solid circles correspond to the solid circles in (e) to indicate relationship to initial points. i. Velocity trajectories as a function of time for the experiment shown in (b), corresponding to the pale blue trajectory in (h). j. Model prediction corresponding to the data shown in (i).

4.3.1 Mechanistic model

We now present a mechanistic model of the Displacillator that models each elementary strand displacement and toehold exchange reaction. The mechanistic model comprises the following reactions for the autocatalytic module $B + A \rightarrow 2B$ and similar reactions for the other two modules:

$$B + \text{React}_{\text{BABr}} \xrightleftharpoons[k_{\text{back}}^{\text{BABr}}]{k_{\text{fwd1}}^{\text{BABr}}} \text{ReactInt}_{\text{BABr}} + \text{Back}_{\text{BA}} \tag{4.34}$$

$$\text{ReactInt}_{\text{BABr}} + A \xrightarrow{k_{\text{fwd2}}^{\text{BABr}}} \text{Flux}_{\text{ABr}} + \text{Waste}_{\text{BA}} \tag{4.35}$$

$$\text{Flux}_{\text{ABr}} + \text{Produce}_{\text{ABrBs}} \xrightleftharpoons[k_{\text{back}}^{\text{ABrBs}}]{k_{\text{fwd1}}^{\text{ABrBs}}} B + \text{ProduceInt}_{\text{ABrBs}} \tag{4.36}$$

$$\text{ProduceInt}_{\text{ABrBs}} + \text{Helper}_{\text{BBs}} \xrightarrow{k_{\text{fwd2}}^{\text{ABrBs}}} B + \text{Waste}_{\text{ABrBs}} \tag{4.37}$$
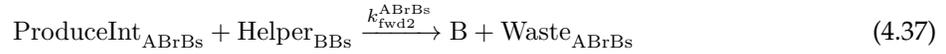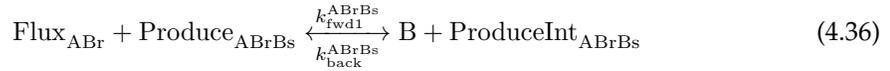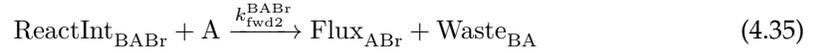
In addition, we include the two main gradual leak pathways

$$\text{Produce}_{\text{ABrBs}} + \text{Helper}_{\text{BBs}} \xrightarrow{k_{\text{leak}}^{\text{ABrBs}}} B \tag{4.38}$$

$$\text{React}_{\text{BABr}} + A \xrightarrow{k_{\text{leak}}^{\text{BABr}}} \text{Flux}_{\text{ABr}}. \tag{4.39}$$
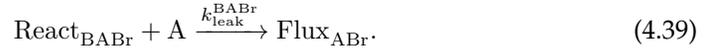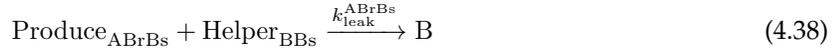
For simplicity, this model does not account for spurious products (of the kind in Figure 3.13) that could form through these leak pathways and subsequently participate in additional reactions. Versions of this model that hope to quantitatively match data must likely incorporate such pathways. Further, since incompletion effects are not included in the model, the catalytic-Helper mechanism is omitted as well. Lastly, we have assumed that the initial leaks have been consumed entirely by thresholds, and that the thresholds have been entirely exhausted. So, for each experiment, we set initial conditions to be what we would expect assuming that thresholding is instantaneous.

How do we choose the rate constants for the mechanistic model? In the course of optimizing the performance of the Displacillator, we experimentally characterized 18 bimolecular rate constants for the intended strand displacement and toehold exchange reactions involving fuel and intermediate species, along with the rate constants for the consumption of signal strands by thresholds. These include *all* designed strand displacement and toehold exchange reactions *except* those involved in the catalytic Helper pathway, which have not been characterized. Along with the measured rate constants, we use rate constants estimated from gradual leak experiments for
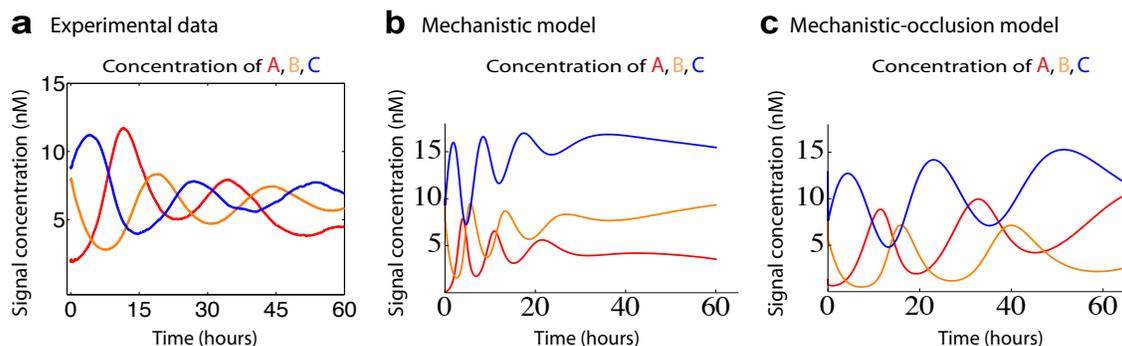
Figure 4.9: Attempts to explain the temporal dynamics of the Displacillator in terms of elementary strand displacement and toehold exchange reactions whose rate constants have been characterized experimentally (see Section 4.3.3). The mechanistic model predicts much faster oscillations than observed in experiment (panel b). When augmented with the toehold occlusion effect to obtain the mechanistic-occlusion model, semi-quantitative agreement with the periods of oscillation observed in experiment is obtained (c).

gradual leak pathways ($k_{leak}^{ABrBs} \sim k_{leak}^{BCjCk} \sim 5$; $k_{leak}^{CApAq} \sim 15$; $k_{leak}^{BABr} \sim k_{leak}^{ACApi2} \sim k_{leak}^{CBCj} \sim 1$; all in /M /s).

Figure 4.9 presents predictions of the mechanistic model and experimental data for one run of the Displacillator. The mechanistic model predicts much faster oscillations than experimentally observed; the periods of oscillations are strikingly different.

4.3.2 Mechanistic-occlusion model

We wondered if the discrepancy in timescales could be explained by toehold occlusion (Figure 3.14). The mechanistic model assumes that toehold dissociation, like all unimolecular steps, is effectively instantaneous. However, each of our toehold and toehold-complements are present in multiple species, including fuel species at high concentration. Therefore, if toehold dissociation steps are not fast enough to be assumed instantaneous, two complexes that are supposed to interact fleetingly through toehold hybridization might actually be co-localized for significant time intervals. When they are co-localized, the toeholds involved would be unavailable for desired reactions, which could cause a slowdown relative to the mechanistic model. This phenomenon has been reported elsewhere in the literature [107].

We enumerated, by hand, all such toehold hybridization interactions specified at the domain level between two species where at least one is a fuel species at high concentration. For species

involved in the autocatalytic module $B + A \rightarrow 2B$, these interactions are listed below. A generic variable O_i is used in place of symbols for each distinct pairwise complex. For simplicity, the hybridization rate k_{on} and dissociation rate k_{off} are assumed to be identical for each interaction.

The following are toehold occlusion interactions arising within the module $B + A \rightarrow 2B$.

$$\text{Helper}_{\text{BBs}} + \text{React}_{\text{BABr}} \xrightleftharpoons[k_{off}]{k_{on}} O_1 \tag{4.40}$$

$$\text{Produce}_{\text{ABrBs}} + A \xrightleftharpoons[k_{off}]{k_{on}} O_2 \tag{4.41}$$

$$\text{ProduceInt}_{\text{ABrBs}} + \text{Helper}_{\text{BBs}} \xrightleftharpoons[k_{off}]{k_{on}} O_3 \tag{4.42}$$

$$\text{ProduceInt}_{\text{ABrBs}} + B \xrightleftharpoons[k_{off}]{k_{on}} O_4. \tag{4.43}$$
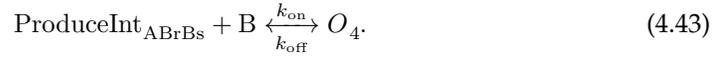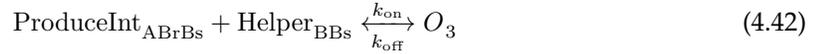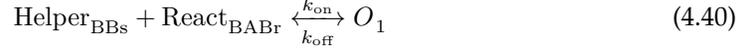
The following are toehold occlusion interactions arising between one species from the module $B + A \rightarrow 2B$ and another species from a different module.

$$\text{Back}_{\text{CB}} + \text{React}_{\text{BABr}} \xrightleftharpoons[k_{off}]{k_{on}} O_5 \tag{4.44}$$

$$\text{React}_{\text{ACApi2}} + \text{Back}_{\text{BA}} \xrightleftharpoons[k_{off}]{k_{on}} O_6 \tag{4.45}$$

$$\text{ProduceInt}_{\text{CApAq}} + \text{Back}_{\text{BA}} \xrightleftharpoons[k_{off}]{k_{on}} O_7 \tag{4.46}$$

$$\text{ProduceInt}_{\text{ABrBs}} + \text{Back}_{\text{CB}} \xrightleftharpoons[k_{off}]{k_{on}} O_8 \tag{4.47}$$

$$\text{Produce}_{\text{BCjCk}} + \text{Back}_{\text{BA}} \xrightleftharpoons[k_{off}]{k_{on}} O_9 \tag{4.48}$$

$$\text{Produce}_{\text{ABrBs}} + \text{Back}_{\text{ACi2}} \xrightleftharpoons[k_{off}]{k_{on}} O_{10} \tag{4.49}$$

$$\text{Produce}_{\text{BCjCk}} + \text{Produce}_{\text{ABrBs}} \xrightleftharpoons[k_{off}]{k_{on}} O_{11} \tag{4.50}$$

$$\text{Produce}_{\text{CApAq}} + \text{Produce}_{\text{ABrBs}} \xrightleftharpoons[k_{off}]{k_{on}} O_{12} \tag{4.51}$$

$$\text{Produce}_{\text{BCjCk}} + \text{ProduceInt}_{\text{ABrBs}} \xrightleftharpoons[k_{off}]{k_{on}} O_{13} \tag{4.52}$$

$$\text{ProduceInt}_{\text{CApAq}} + \text{Produce}_{\text{ABrBs}} \xrightleftharpoons[k_{off}]{k_{on}} O_{14}. \tag{4.53}$$

We augmented the mechanistic model with these reactions and similar ones for the other autocatalytic modules to obtain the "mechanistic-occlusion model". Fixing $k_{on} = 2 * 10^6$ /M /s for convenience, we found that the predictions of the mechanistic-occlusion model are quite sensitive

to the dissociation parameter k_{off}. For k_{off} = 0.1 /s, for example, we could obtain good semi-quantitative agreement with the temporal dynamics of the data (see Figure 4.9). These rates imply a toehold binding ΔG of about 10 kcal/mol, which is within 1.0 - 2.0 kcal/mol of our estimated toehold energies.

4.3.3 Characterizing individual strand displacement and toehold exchange rates

As mentioned earlier, we experimentally characterized 18 bimolecular rate constants for the intended strand displacement and toehold exchange reactions involving fuel and intermediate species, along with the rate constants for the consumption of signal strands by thresholds. These include *all* designed strand displacement and toehold exchange reactions *except* those involved in the catalytic Helper pathway, which have not been characterized. These rate constants were measured in independent experiments where only the relevant species were present (see Section 5.3.2). The results are summarized in Tables 4.2 and 4.3.

Overall, the rates for different strand displacement and toehold exchange steps vary by as much as a factor of 30. Yet certain patterns seem to emerge. For the React species, k_{back} is higher than k_{fwd2} in all cases, at least by a factor of 2. For the Produce species the opposite is true by at least a factor of 3. This comparison is important because, for both the React and Produce steps, k_{back} and k_{fwd2} represent reactions that compete directly with each other at the intermediate stage of execution.

We do not fully understand, at this stage, the significance of the rate constants we have measured in influencing the kinetics of each module and the Displacillator. Informally, the model predicted plots in Figure 4.9 can vary quite dramatically if some rate constants, such as k_{fwd1} for the React species, are varied. It is possible that a formal analysis of sensitivity to parameters is required to better understand these effects.

4.4 Sequence design and *in silico* verification

In this thesis, we have demonstrated experimentally that DNA strand displacement is a general technology for systematically implementing prescribed dynamical behaviors in a test tube. In this process, we have developed heuristics for sequence design and verification, biophysical understanding of desired and spurious strand displacement pathways and experimental tools to assay

Complex	k_{fwd1}	k_{back}	k_{fwd2}
$\text{React}_{\text{ACApi2}}$	2.7×10^5	1.1×10^6	1.4×10^5
$\text{React}_{\text{BABr}}$	1.8×10^5	6.2×10^5	2.7×10^5
$\text{React}_{\text{CBCj}}$	8.6×10^4	9.8×10^5	3.0×10^5
$\text{Produce}_{\text{CApAq}}$	2.1×10^5	2.2×10^5	1.2×10^6
$\text{Produce}_{\text{ABrBs}}$	6.0×10^5	4.6×10^5	1.5×10^6
$\text{Produce}_{\text{BCjCk}}$	1.6×10^6	2.4×10^5	2.6×10^6

Table 4.2: Measured rate constants (all in /M /s) for designed strand displacement and toehold exchange reactions in the Displacillator. The reactions corresponding to the notation for rate constants are specified in Equations 4.34 - 4.37. Note that rate constants involving the catalytic Helper pathway have not been characterized.

Threshold	Measured rate constant (/M /s)
k_{ThA}	7.4×10^5
k_{ThB}	1.7×10^6
k_{ThC}	1.2×10^6

Table 4.3: Measured rate constants (all in /M /s) for the consumption of signal species (Ap, Br, and Cj) by thresholds (Th_{A}, Th_{B}, and Th_{C}, respectively) .

and debug *in vitro* chemical networks. Here we summarize what we learned about sequence design and *in silico* verification while engineering experimental DNA strand displacement systems based on our CRN-to-DNA scheme. We believe that while these design rules are likely particularly relevant to our CRN-to-DNA scheme, the general principles may apply to any DNA strand displacement system.

First, given the importance of toehold strengths in determining strand displacement and toehold exchange kinetics, we recommend that toeholds be designed first, using software that allows the design of short strands that are iso-energetic as well as orthogonal (such as StickyDesign [234]). In particular, it might be helpful to consider the different local contexts for each toehold (internal, external, with or without overhangs or dangles).

Once the toeholds are designed, we suggest designing the rest of the system using software that can accommodate negative design. This is because, given that all intended complexes and intermediates are well-formed, negative design to prevent spurious interactions may need more emphasis. Also, closing helices and junctions with strong C-G base pairs, wherever possible, will likely minimize gradual leak rates.

For both design steps above, we strongly recommend using 'ACT' alphabet for the "top" strands. Violating ACT alphabet should be done with caution, if at all, since multiple problems can arise both due to unintended secondary structure in single-strands and unexpected leak pathways arising out of a short but strong (2 or 3 G-C base pairs) spurious match.

Once multiple candidate sequence designs have been obtained, an *in silico* verification step where all candidate designs are evaluated on heuristic metrics of the kind we discuss in Chapter 3 may help identify designs with the fewest potential problems. In general, we believe that this verification step is likely more important than the choice of design algorithm or criteria.

Lastly, once strong candidate designs are identified, it might be a good idea to test the corresponding designs with 5'-3' orientations reversed. This is because, while we believe that initial leaks are likely caused by synthesis errors, we do not have confidence in our ability to predict the optimal 5'-3' orientation. Therefore, in case a design is found where both orientations satisfy the verification step, it might be informative to try both orientations experimentally. It is important to remember at this stage that reversing 5'-3' orientations may perturb the thermodynamics of toeholds and other critical regions; those checks will need to be repeated for the reversed designs.

We conclude this section with a necessary caveat. The design rules summarized here have not

yet been rigorously tested. We are currently collaborating with James Parkin and Erik Winfree on streamlining software tools that incorporate these design rules and testing new "from scratch" sequence designs for our Displacillator. This effort is currently in progress and preliminary results are encouraging.

4.5   Challenges in scaling up CRN-to-DNA approaches

Scaling up the CRN-to-DNA approach investigated in this thesis for engineering much larger and more complex CRNs is an attractive but difficult task. Here, we briefly discuss the multiple challenges that would need to be addressed to make progress towards that goal.

In our efforts, we chose to design orthogonal toeholds. That design principle has the advantages of minimizing cross-talk between fuel species belonging to different reactions and the problem of toehold occlusion. However, there are a finite number of orthogonal toeholds available at any (reasonable) toehold length. So, as the desired CRN gets larger, we may be constrained to re-use toeholds, which will exacerbate cross talk and toehold occlusion.

Since the strand displacement system is implemented in a well-mixed solution, the number of pair-wise interactions possible in the test tube scales quadratically with the size of the DNA implementation. Therefore, even if we achieve perfect orthogonality for all our toeholds, fleeting binding of different complexes due to "toehold-only interactions" will cause a significant fraction of toeholds to be occluded at any given time.

Both the problems discussed above may be mitigated by lowering the concentrations of the fuel species. However, that will necessarily scale down the concentration range of the intended dynamical behavior. In addition, the intended dynamical behavior would occur at much lower rates and take much longer to observe, which might pose an experimental challenge.

Implementing CRNs tethered to a surface, such as that of DNA origami, has been suggested as a possible solution [239–242] to some of the challenges described above, since limiting interactions to species that are co-localized can reduce spurious interactions and allow domains to be re-used multiple times.

However, there remain three general obstacles that are common to both CRNs in solution and CRNs on surface. First, kinetic sequence design is an unsolved problem, since predicting DNA hybridization, branch migration, fraying, and dissociation kinetics from DNA sequence itself is a

challenge [148, 153, 192]. However, without scientific understanding and algorithms that permit hybridization and strand displacement kinetics to be predicted from DNA sequence within a factor of 20 (at the very least), it is hard to imagine that large CRNs with prescribed dynamic behavior can be reliably engineered. This is because multiple rounds of re-design will likely be required just to adjust rate parameters.

Second, scaling up will require (i) better mechanistic understanding of initial and gradual leaks, so that they can be mitigated further and (ii) design principles to modify the domain-level specifications of current CRN-to-DNA schemes for increased fault tolerance and robust performance in the face of molecular non-idealities.

Third, enabling sustained dynamic behavior will require mechanisms for continuous "power supply". By this, we mean that either macroscopic or molecular solutions would need to be found for the problem of replenishing fuel species and removing the buildup of waste molecules.

The challenges we have outlined here lead to a natural question. Given the "real-world" constraints involved in engineering CRNs with our CRN-to-DNA technology (such as spurious pathways, limited number of orthogonal toehold sequences, quadratic scaling of toehold occlusion, and so on), what sub-class of all formal CRNs can we reasonably hope to engineer in a test tube? At this time, we are not sure what the answer to this question is. However, with further work on understanding the experimental constraints and their implications for our design pipeline, answers to this question will begin to emerge.

5 Appendix

5.1 Appendix to Chapter 1

Table 5.1 lists the references indexed in Figure 1.2.

Index	Reference
R1	Kallenbach et al. [156]
R2	Chen and Seeman [243]
R3	Yurke et al. [145]
R4	Yurke and Mills [146]
R5	Shin and Pierce [244]
R6	Shih et al. [245]
R7	Seelig et al. [105]
R8	Zhang et al. [143]
R9	Yin et al. [160]
R10	Rothemund [132]
R11	Rothemund et al. [142]
R12	Han et al. [246]
R13	Wickham et al. [247]
R14	Wei et al. [134]
R15	Ke et al. [248]
R16	Teichmann et al. [249]
R17	Gerling et al. [250]
R18	Qian and Winfree [107]
R19	Qian et al. [251]
R20	Wickham et al. [252]
R21	Chen et al. [112]
R22	Srinivas [253]
R23	Zhang et al. [254]

Table 5.1: References in Figure 1.2.

5.2 Appendix to Chapter 2.

5.2.1 Introduction.
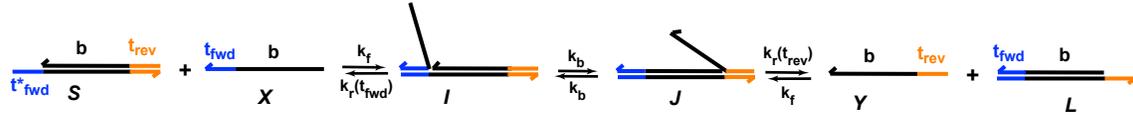
Figure 5.1: The phenomenological model of Zhang and Winfree [147] for reversible toehold exchange. For simplicity, they assume the hybridization rate constants for both toeholds to be the same (k_f). k_f and k_b are fitted to the data.

The phenomenological model of Zhang and Winfree [147] directly treats reversible toehold exchange (Figure 5.1), which is a generalization of strand displacement. In the main text, we restricted their model to irreversible strand displacement without any fundamental changes.

5.2.2 Intuitive Energy Landscape model.

Zero toehold case. The IEL for the zero toehold case ($h = 0$) is illustrated in Figure 5.2. The main modification is the addition of a fraying step (state A to A′, at the cost of one base pair stack - $|\Delta G_{bp}|$), which permits a collision leading to the formation of a base pair between the substrate-incumbent duplex and the invader. Such a collision essentially results in the displacement of one incumbent base by the invader (state B′). The remaining states are identical to the case with a toehold at least 1-base long, except that the branch migration domain is shorter by one base (length $b - 1$ rather than b). Since the substrate-incumbent duplex needs to fray at either end to permit strand displacement from that end, the collision rate includes a multiplicative factor of $2 \times e^{-|\Delta G_{bp}|/RT}$, which accounts for the average fraction of duplexes frayed at either end. So, the complexes join at a rate $k_{bi} \times u \times (2 \times e^{-|\Delta G_{bp}|/RT})$ rather than just $k_{bi} \times u$, where u is the concentration.

We now approximate $k_{eff}(0)$ in terms of $k_{eff}(1)$, which will prove useful once we derive expressions for $k_{eff}(h)$ for $h > 0$. As we pointed out earlier, once the first base pair forms between the frayed substrate-incumbent duplex and the invader, the situation is identical to the $h = 1$ case with the toehold just bound, except with a branch migration domain of length $b - 1$ rather than b.

Assuming b is long enough that $b - 1 \approx b$,

$$k_{\text{eff}}(0) \approx (2\, e^{-|\Delta G_{\text{bp}}|/RT})\, k_{\text{eff}}(1) \ . \tag{5.1}$$

Choosing k_{uni}. We choose k_{uni} such that the rate of dissociation of the last base pair of the toehold is approximately equal to the rate of fraying of every other base pair. So, we set

$$k_{\text{uni}}\, e^{-|\Delta G_{\text{bp}}|/RT} \approx k_{\text{bi}}\, e^{(-|\Delta G_{\text{bp}}| + |\Delta G_{\text{assoc}}|)/RT}\, u_0, \tag{5.2}$$

which yields

$$k_{\text{uni}} \approx k_{\text{bi}}\, e^{|\Delta G_{\text{assoc}}|/RT}\, u_0 \tag{5.3}$$

where $u_0 = 1\,\text{M}$ is the reference concentration.

Modeling k_{eff}. Conceptually, we may split the kinetics of strand displacement into two parts: the concentration-dependent, bimolecular part involving collision and formation of the first base pair of the toehold (state A to B) and the unimolecular part comprising the formation of the remaining base pairs of the toehold (states B to C) and branch migration (states C to F). The unimolecular part is modeled as a random walk on the free energy landscape starting at B, and ending either at A (no displacement) or at E (successful displacement). In the low-concentration regime, the unimolecular part finishes (one way or the other) much faster than the rate of collision; so, we model the effective rate constant of strand displacement, k_{eff}, as

$$k_{\text{eff}} = k_v \times p \tag{5.4}$$

where k_v is the rate constant for the formation of the first toehold base pair and $p = \mathbb{P}(\text{reaching E before A starting at B})$ is the probability of successful displacement starting at B. Since the IEL assumes a collision rate of $k_{\text{bi}} \times u$, $k_v = k_{\text{bi}}$.

Calculating p. First, given the rates of the individual transitions as specified in the kinetic model, the probabilities of transition from each state to its neighbors can be calculated. Let $p_{i,j}$ be the probability of going from state i to state j. Then, we choose

$$p_{i,i+1} = \frac{k_{i,i+1}}{k_{i,i+1} + k_{i,i-1}} \ , \tag{5.5}$$

$$p_{i,i-1} = \frac{k_{i,i-1}}{k_{i,i+1} + k_{i,i-1}} \ . \tag{5.6}$$

Since we are interested in the probability that the random walk finishes at one end rather than the
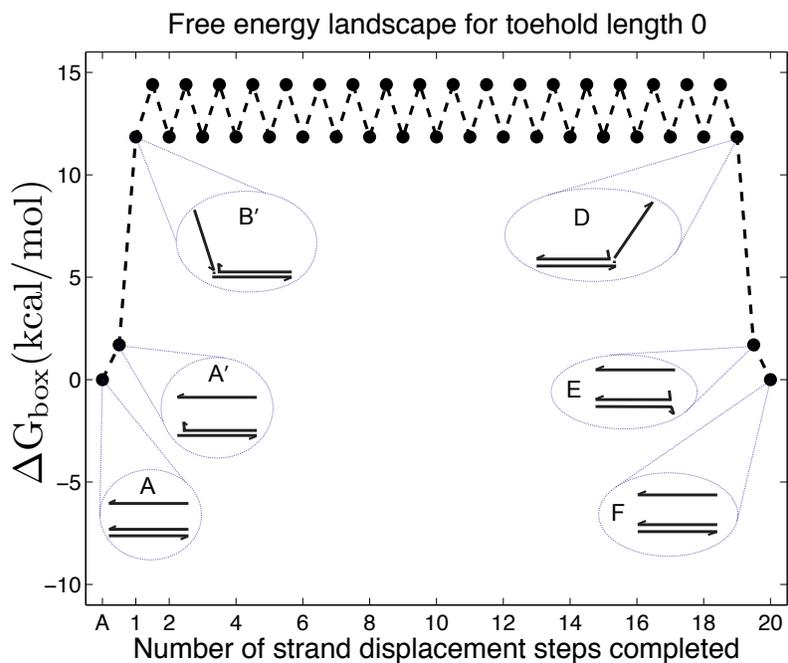
Figure 5.2: IEL free energy landscape at 25 °C for a 0-base toehold. First, the invader and the substrate-incumbent complex are unconstrained by each other (A). Then, the substrate-incumbent complex frays at one end (A′). This fraying step, at the cost of one base pair stack ($|\Delta G_{\mathrm{bp}}|$), is necessary to enable a collision leading to a base pair, since no toehold is present. Once such a collision occurs, the first base of the incumbent has effectively been displaced by the invader (B′). The remaining states are identical to the positive toehold case.

other, we use absorbing boundaries:

$$p_{i,i} = 1 \text{ for } i = \text{A, E (A}' \text{ for 0 toehold) .} \tag{5.7}$$

Given these transition probabilities, p can be calculated analytically using the method of Rudolph [173] for calculating absorption probabilities for a 1-dimensional random walk with absorbing boundaries.
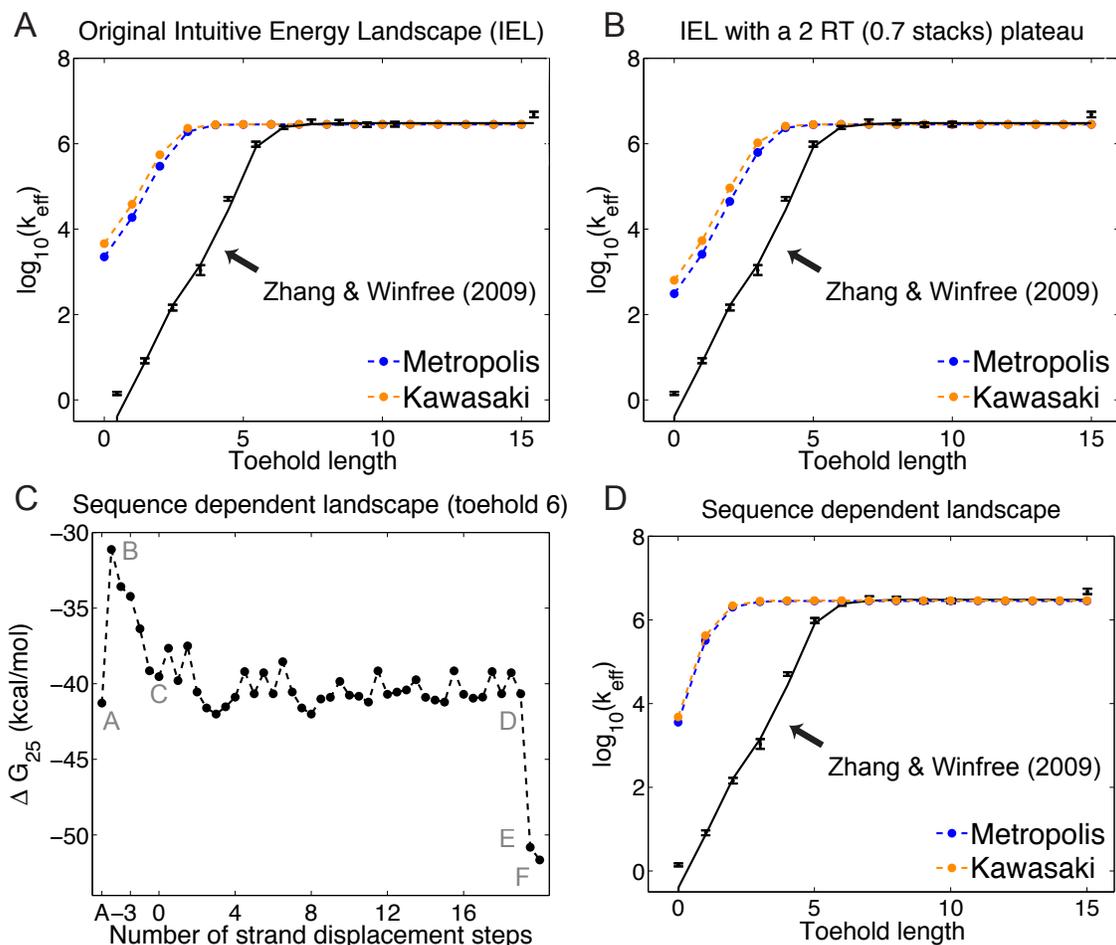
Figure 5.3: Strand displacement kinetics predicted by the IEL under varying conditions, for both Metropolis and Kawasaki unimolecular rate models. (A) IEL (2.6, 0) (B) IEL (2.6, 1.2) (D) for the sequence dependent free energy landscape predicted by the NUPACK energy model with Dangles = "Some" (shown in (C)). None of these variations is consistent with experimental data [147]. (C) States A–F pertain to Figure 2.3 in the main text. The sequence dependent landscape is much "rougher" than the IEL, but this "roughness" goes no further in accounting for the data (D).

Alternate unimolecular rate model. We also considered the Kawasaki [174] method for assigning unimolecular transition rates in the IEL:

$$k_{ij} = k_{\text{uni}} \times \text{e}^{-(\Delta G_{\text{box}}(j) - \Delta G_{\text{box}}(i))/2RT} \tag{5.8}$$

$$k_{ji} = k_{\text{uni}} \times \text{e}^{-(\Delta G_{\text{box}}(i) - \Delta G_{\text{box}}(j))/2RT} \tag{5.9}$$

As with the Metropolis rate model, this choice guarantees detailed balance and thus that the system eventually converges to a Boltzmann distribution:

$$\text{Prob}(i) = \frac{1}{Q}\text{e}^{-\frac{\Delta G_{\text{box}}(i)}{RT}} \quad \text{where} \quad Q = \sum_i \text{e}^{-\frac{\Delta G_{\text{box}}(i)}{RT}}.$$

Once again, we choose k_{uni} such that the rate of dissociation of the last base pair of the toehold is approximately equal to the rate of fraying of every other base pair. For the Kawasaki method, we have

$$k_{\text{uni}} \, e^{-|\Delta G_{\text{bp}}|/2RT} \approx k_{\text{bi}} \, e^{(-|\Delta G_{\text{bp}}| + |\Delta G_{\text{assoc}}|)/RT} \, u_0, \tag{5.10}$$

which yields

$$k_{\text{uni}} \approx k_{\text{bi}} \times e^{(-|\Delta G_{\text{bp}}|/2 + |\Delta G_{\text{assoc}}|)/RT} \, u_0 \,. \tag{5.11}$$

With $k_{\text{uni}} \approx 1.8 \times 10^7$ /s as calculated above, the Kawasaki method consistently predicts a slightly smaller acceleration in k_{eff} with toehold length than the Metropolis [172] method (see Figure 5.3); therefore our conclusions about constraints on IEL parameters that are necessary to match experimental data are robust to this choice. In the analysis below, the Metropolis method is assumed throughout.

Analytical approximations for $k_{\text{eff}}(h)$. The IEL is simple enough to yield itself to an approximate analysis that permits intuitive understanding of the model. For $h > 1$, we may approximate $k_{\text{eff}}(h)$ as:

$$k_{\text{eff}}(h) \approx k_{\text{bi}} \times p_{\text{zip}} \times p_{\text{bm}|\text{toe}}(h) \tag{5.12}$$

where p_{zip} is the probability that the full toehold "zips up" once the first base pair is formed, $k_{\text{bi}} \times p_{\text{zip}}$ is the bimolecular rate constant for all the base pairs in the toehold to form and $p_{\text{bm}|\text{toe}}(h)$ is the probability that, once the toehold has zipped up, the incumbent is displaced before the toehold dissociates.

For $h = 1$, there are no other bases in the toehold to zip up, so $k_{\text{eff}}(1)$ may simply be approxi-

mated as

$$k_{\text{eff}}(1) \approx k_{\text{bi}} \times p_{\text{bm|toe}}(1). \tag{5.13}$$

We shall now estimate $p_{\text{bm|toe}}(h)$. Let k_{first} be the rate at which the first base of the incumbent is displaced by the invader, once the toehold is bound. Let $k_r(h)$ be the rate at which the toehold of length h dissociates from the state in which it is fully bound (i.e. state C of Figure 2.3 in the main text). We estimate k_{first} and $k_r(h)$ later. The probability of displacing the first base of the incumbent before the toehold dissociates is simply $k_{\text{first}}/(k_{\text{first}} + k_r(h))$.

For simplicity, we assume that once $(b-1)$ bases of the incumbent have been displaced, the last base always dissociates. This approximation allows us to think about branch migration as a flat random walk. Our approximation is reasonable because the probability that the invader dissociates without completing strand displacement after displacing $(b-1)$ bases of the incumbent can only be 0.5, in the worst case (for a 0-base toehold, with $k_{\text{uni}} >> k_{\text{bi}}$), and hence introduces negligible error on a logarithmic scale. Moreover, if branch migration is indeed significantly slow relative to fraying, we would expect that the last few bases of the incumbent would fray, leading to the dissociation of the incumbent even before branch migration completes. Once the first base of the incumbent has been displaced, the probability of displacing the remaining bases of the incumbent before going back to the toehold-only-bound state is $1/(b-1)$, according to the gambler's ruin analysis (see Section 14.2 of Feller [169]). With a probability of $(b-2)/(b-1)$, we return to the toehold bound state - from where, by definition, displacement succeeds with a probability of $p_{\text{bm|toe}}(h)$.

Therefore, we have $p_{\text{bm|toe}}(h) \approx$

$$\left(\frac{k_{\text{first}}}{k_{\text{first}} + k_r(h)} \right) \left[\frac{1}{b-1} + \frac{b-2}{b-1} \times p_{\text{bm|toe}}(h) \right] \tag{5.14}$$

which yields

$$p_{\text{bm|toe}}(h) \approx \frac{k_{\text{first}}}{k_{\text{first}} + (b-1)k_r(h)} . \tag{5.15}$$

Therefore, equation 5.12 yields

$$k_{\text{eff}}(h) \approx k_{\text{bi}} \times p_{\text{zip}} \times \frac{k_{\text{first}}}{k_{\text{first}} + (b-1)k_r(h)} . \tag{5.16}$$

With some algebra, equation 5.16 yields

$$k_{\text{eff}}(h) \approx \frac{k_{\text{bi}} \times p_{\text{zip}}}{1 + (b-1)\frac{k_r(h)}{k_{\text{first}}}} \tag{5.17}$$

for $h > 1$.

Analyzing equation 5.17 provides intuition about the dependence of $k_{\text{eff}}(h)$ on h. For long toeholds, i.e. in the saturation limit, $(b-1) \frac{k_r(h)}{k_{\text{first}}} << 1$; this yields the saturation value

$$k_{\text{eff}}(\infty) \approx k_{\text{bi}} \times p_{\text{zip}}. \tag{5.18}$$

For short toeholds, i.e in the "sloping" part of the $k_{\text{eff}}(h)$ curve, $(b-1) \frac{k_r(h)}{k_{\text{first}}} >> 1$ and this yields

$$k_{\text{eff}}(h) \approx \frac{k_{\text{bi}} \, p_{\text{zip}} \, k_{\text{first}}}{(b-1) \, k_r(h)}. \tag{5.19}$$

We now estimate p_{zip}. For $h > 1$, we assume, for simplicity, that the formation of a second base pair of the toehold will guarantee zippering. Then, p_{zip} may be approximated as

$$p_{\text{zip}} \approx \frac{k_{\text{uni}}}{k_{\text{uni}} + k_r(1)} \, . \tag{5.20}$$

Note that $k_r(1)$ is just the rate for the simple bimolecular dissociation step, and so we have

$$k_r(1) \approx k_{\text{bi}} \, \lambda \tag{5.21}$$

where λ is defined as

$$\lambda = e^{-(|\Delta G_{\text{bp}}| - \Delta G_{\text{assoc}})/RT} \, u_0 \, . \tag{5.22}$$

We now estimate $k_r(h)$ for $h > 1$. Once a toehold of length h is fully bound, $(h-1)$ uphill fraying steps and 1 dissociation step need to occur for the toehold to dissociate. Therefore,

$$k_r(h) \approx k_{\text{fray}}(h-1) \times (1 - p_{\text{zip}}) \tag{5.23}$$

where $k_{\text{fray}}(h-1)$ is the rate at which $(h-1)$ base pairs of the toehold fray, starting from the toehold fully bound state, to leave the invader bound by 1 base of the toehold (state C to B in Figure 2.3 in the main text).

Clearly,

$$k_{\text{fray}}(h-1) = k_{\text{uni}} \, e^{-(h-1) \, |\Delta G_{\text{bp}}|/RT}. \tag{5.24}$$

From equations 5.20, 5.23 and 5.24, we get

$$k_r(h) \approx e^{-(h-1) \, |\Delta G_{\text{bp}}|/RT} \times \frac{1}{1/k_{\text{uni}} + 1/k_r(1)} \, . \tag{5.25}$$

We also recall that in our particular parameterization we chose k_{uni} such that $k_r(1) \approx k_{\text{fray}}(1)$, although we do not use or need that condition in this analysis.

Lastly,

$$k_{\text{first}} = \frac{1}{2} \times k_{\text{uni}} \times e^{-\Delta G_{s+p}/RT} \tag{5.26}$$

where ΔG_{s+p} is defined as

$$\Delta G_{s+p} = \Delta G_s + \Delta G_p \tag{5.27}$$

and ΔG_s is the sawtooth amplitude and ΔG_p is the plateau height.

Substituting for $k_r(h)$ from equation 5.25, k_{first} from equation 5.26 and p_{zip} from equation 5.20 into equation 5.19, we may approximate $k_{\text{eff}}(h)$ for $h > 1$ in the pre-saturation regime. We also assume that b is long enough that $b - 1 \approx b$. With some algebra, we get

$$k_{\text{eff}}(h) \approx \frac{k_{\text{uni}} \; e^{-(\Delta G_{s+p} - h|\Delta G_{\text{bp}}|)/RT} \; \eta}{2 \; b} \tag{5.28}$$

where η is defined as

$$\eta = \frac{1}{e^{\Delta G_{\text{assoc}}/RT} \; u_0}. \tag{5.29}$$

We now approximate $k_{\text{eff}}(1)$ using equation 5.13. Using equations 5.15, 5.21 and 5.26, we get

$$k_{\text{eff}}(1) \approx \frac{\rho \; k_{\text{bi}}}{\rho + 2 \; b \; \lambda \; e^{\Delta G_{s+p}/RT}} \tag{5.30}$$

where ρ is defined as

$$\rho = k_{\text{uni}}/k_{\text{bi}}. \tag{5.31}$$

Having already assumed that $b - 1 \approx b$, we may approximate $k_{\text{eff}}(0)$ in terms of $k_{\text{eff}}(1)$ using equation 5.1. Along with equation 5.30, this yields

$$k_{\text{eff}}(0) \approx \frac{2 \; e^{-|\Delta G_{\text{bp}}|/RT} \; \rho \; k_{\text{bi}}}{\rho + 2 \; b \; \lambda \; e^{\Delta G_{s+p}/RT}} \; . \tag{5.32}$$

Although equations 5.30 and 5.32 seem different in form from equation 5.28, that is only because equation 5.28 assumes the pre-saturation ("sloping") regime, whereas equations 5.30 and 5.32 are more general. For reasonable values of ρ and long enough b, such as the default IEL values ($\rho \approx 25, b = 20$), we may use the approximation

$$\rho + 2 \; b \; \lambda \; e^{\Delta G_{s+p}/RT} \approx 2 \; b \; \lambda \; e^{\Delta G_{s+p}/RT}. \tag{5.33}$$

This yields

$$k_{\text{eff}}(1) \approx \frac{k_{\text{uni}} \; e^{-(\Delta G_{s+p} - |\Delta G_{\text{bp}}|)/RT} \; \eta}{2 \; b} \tag{5.34}$$

and

$$k_{\text{eff}}(0) \approx \frac{k_{\text{uni}} \ e^{-\Delta G_{s+p}/RT} \ \eta}{b} \qquad (5.35)$$

which are consistent in form with equation 5.28, but for the factor of two for $h = 0$.

We define

$$\mathcal{A}_{h_2,h_1} = \log_{10}(k_{\text{eff}}(h_1)) - \log_{10}(k_{\text{eff}}(h_2)) \qquad (5.36)$$

to be the orders of magnitude acceleration in k_{eff} as toehold length increases from h_1 to h_2. We now use equations 5.18 and 5.32 to study the dependence of $\mathcal{A}_{\infty,0}$ on ρ, ΔG_{s+p} and b. With some algebra, equations 5.18 and 5.32 yield

$$\mathcal{A}_{\infty,0} \approx \mu + \log_{10}\left(\frac{\rho + 2\,b\,\lambda\,e^{\Delta G_{s+p}/RT}}{\rho + \lambda}\right) \qquad (5.37)$$

where μ is defined as

$$\mu = \log_{10}\left(e^{|\Delta G_{\text{bp}}|/RT}/2\right) \approx 0.9\ . \qquad (5.38)$$

Dependence on $\rho = \mathbf{k_{uni}}/\mathbf{k_{bi}}$. To analyze the effect of changing ρ on $\mathcal{A}_{\infty,0}$, let us choose a priori reasonable values for the other parameters: $b = 20$, $\Delta G_s = 1.5$ stacks (2.6 kcal/mol) and $\Delta G_p = 0$. Then, equation 5.37 becomes

$$\mathcal{A}_{\infty,0} \approx 0.9 + \log_{10}\left(\frac{\rho + 4000}{\rho + 1.4}\right) \qquad (5.39)$$

whose behavior is qualitatively similar to the numerical curve plotted in Figure 5.4. Moreover, the difference between $\mathcal{A}_{\infty,0}$ at $\rho = \infty$ and $\rho = 0$ as predicted by equation 5.39 is approximately 3.5, which is consistent with the total variation in $\mathcal{A}_{15,0}$ as numerically plotted in Figure 5.4.

Dependence on $\mathbf{\Delta G_{s+p}}$. For the default k_{uni} and k_{bi} values used in the IEL, $\rho/(\rho + \lambda)$ may be neglected in equation 5.37 to yield

$$\mathcal{A}_{\infty,0} \approx \mu + \log_{10}\left(\frac{2\,b\,e^{\Delta G_{s+p}/RT}\,\lambda}{\rho + \lambda}\right)\ . \qquad (5.40)$$

This yields an approximately linear dependence on ΔG_{s+p}:

$$\mathcal{A}_{\infty,0} \approx c_1 + c_2\,\Delta G_{s+p} \qquad (5.41)$$

where

$$c_1 = \mu + \log_{10}\left(\frac{2\,b\,\lambda}{\rho + \lambda}\right) \qquad (5.42)$$
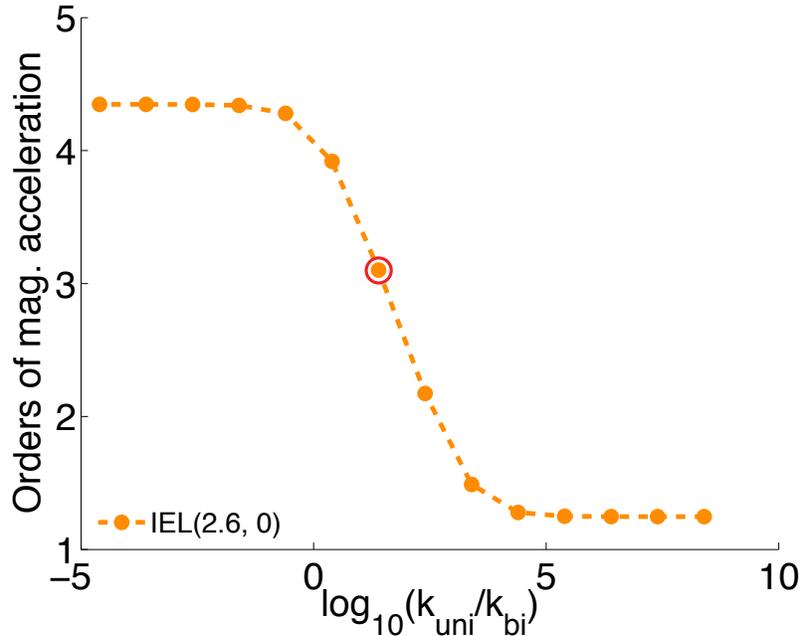
Figure 5.4: The influence of $k_{\mathrm{uni}}/k_{\mathrm{bi}}$ on the orders of magnitude acceleration in k_{eff} between toehold lengths 15 and 0 ($\mathcal{A}_{15,0}$) predicted by IEL(2.6, 0). The red circle indicates the default value of $k_{\mathrm{uni}}/k_{\mathrm{bi}}$. The IEL exhibits sigmoidal response to $\log_{10}(k_{\mathrm{uni}}/k_{\mathrm{bi}})$ and even implausibly low values of $k_{\mathrm{uni}}/k_{\mathrm{bi}}$ do not result in acceleration that matches the experimentally observed value of 6.5 orders of magnitude.

and

$$c_2 = \frac{1}{RT \log_e(10)}. \tag{5.43}$$
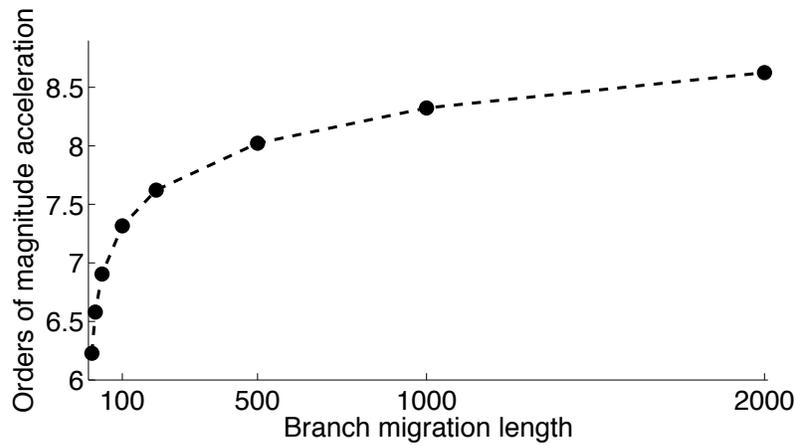
Figure 5.5: Orders of magnitude acceleration in k_{eff} between toehold lengths 15 and 0 ($\mathcal{A}_{15,0}$), as a function of the length of the branch migration domain, as predicted by IEL(5.3, 2.0).

Dependence on b. From equation 5.40, we have

$$\mathcal{A}_{\infty,0} \approx \log_{10}(b) + c_3 \tag{5.44}$$

where

$$c_3 = \mu + \log_{10}\left(\frac{2\,e^{\Delta G_{s+p}/RT}\,\lambda}{\rho + \lambda}\right)\quad. \tag{5.45}$$

The logarithmic dependence of $\mathcal{A}_{\infty,0}$ on b is qualitatively similar to the numerical curve of $\mathcal{A}_{15,0}$ plotted in Figure 5.5.

Estimating branch migration step times. In order to compare with experimentally inferred values [165, 166], we define average branch migration step time (t_{bm}) as the average time taken for a single step of branch migration in either direction. For the IEL and AEL, t_{bm} is calculated as

$$t_{\mathrm{bm}} = \frac{1}{k_{\mathrm{bm}}} \tag{5.46}$$

where

$$k_{\mathrm{bm}} = k_{\mathrm{uni}} \times e^{-\Delta G_s/RT}. \tag{5.47}$$

5.2.3 Augmented Energy Landscape model

The IEL assumes the initial attachment of all toeholds occurs at the same rate. It is possible to consider alternatives, such as an attachment rate that is linear in toehold length. At the cost of some of its intuitive simplicity, the IEL can be augmented to include such a linear dependence of the binding rate on toehold length. We outline the Augmented Energy Landscape model (AEL) below. The state space of the model is identical to the IEL, although we now include a multiplicity factor in the free energy of states with a partially formed toehold. This accounts for the fact that there are several ways in which to form an incomplete toehold. Note that the AEL represents the 0-base toehold system in an identical fashion to the IEL.

Free-energy model. The AEL free-energy model is identical to that of the IEL except for states in which the toehold is partially formed. We adjust the free energy of these states by a factor

$$\Delta G_{\mathrm{mult}}(n) = -RT \ln(h - n + 1), \tag{5.48}$$

where n is the number of base pairs present and h is the total toehold length. This factor accounts for the multiplicity of states with n base pairs (assuming the toehold base pairs form a continuous helix).

Rate model. The AEL has an identical rate model to the IEL, except for transitions involving states in which the toehold is partially formed. The transition from n to $n-1$ base pairs with the toehold could occur by fraying a base pair at either end of the partially formed duplex. We therefore assign these processes a rate

$$k_{n,n-1} = 2k_{\text{uni}}\, \mathrm{e}^{-|\Delta G_{\text{bp}}|/RT} \ . \tag{5.49}$$

The requirement of detailed balance fixes the rates in the opposite direction:

$$k_{n-1,n} = 2k_{\text{uni}}\frac{h+1-n}{h+2-n}. \tag{5.50}$$

Attachment to the toehold is assumed to occur in h equally likely ways: the rate of attachment is then

$$k_{0,1} = h\, k_{\text{bi}}\, u_0\, \mathrm{e}^{-\Delta G_{\text{volume}}/RT}, \tag{5.51}$$

where $u_0 = 1\,\text{M}$ is the reference concentration. Once again, detailed balance fixes the reverse rate.

$$k_{1,0} = k_{\text{bi}}\, u_0\, \mathrm{e}^{(-|\Delta G_{\text{bp}}|+\Delta G_{\text{assoc}})/RT} \ . \tag{5.52}$$

The remaining consideration is the relation between k_{uni} and k_{bi}. Once again, we assume that breaking of the final base pair in the toehold occurs at the same rate as other base pairs are disrupted, giving

$$k_{\text{uni}} = k_{\text{bi}}u_0\, \mathrm{e}^{\Delta G_{\text{assoc}}/RT} \ . \tag{5.53}$$

The AEL is simulated in a manner exactly analogous to Multistrand (see Supplementary Section S4), using a Gillespie algorithm [73] and utilizing "first step" mode in which the strands are initialized in the state with 1 base pair, and the system is evolved to see if displacement is successful. The success probability is then multiplied by the initial binding rate to find the overall reaction rate. The probability is estimated using at least 10^4 successful trajectories for each case.

Performance of the AEL. Overall, these modifications to the IEL result in a self-consistent model with an initial binding rate that is linear in the length of the toehold. As with the IEL, the AEL has free parameters k_{bi}, ΔG_s and ΔG_p, and the overall acceleration with increased toehold length is largely determined by $\Delta G_s + \Delta G_p$. We fix $\Delta G_p = 2.0\,\text{kcal/mol}$ in accordance with experiment (Figure 7 of the main text), and adjust k_{bi} and ΔG_s to fit the data of Zhang and Winfree [147]. We find that $k_{\text{bi}} = 3.30 \times 10^5\,\text{/M/s}$ and $\Delta G_s = 3.60\,\text{kcal/mol}$ (implying $k_{\text{uni}} = 8.17 \times 10^6\,\text{/s}$)

Toehold length	$k_{\text{eff}}^{\text{AEL}}$ (/M/s)	$k_{\text{eff}}^{\text{expt}}$ (/M/s)
0	1.4	1.4
1	12	8.2
2	214	144
3	3.7×10^3	1.1×10^3
4	6.2×10^4	5.1×10^4
5	6.7×10^5	9.6×10^5
6	1.7×10^6	2.4×10^6
15	4.8×10^6	4.8×10^6

Table 5.2: Comparison of AEL predictions and experimental results from Zhang and Winfree [147] for the rate of displacement as a function of toehold length.

give a good fit to experiment, as summarised in Table 5.2.

5.2.4 Secondary structure kinetics model

Multistrand [153] essentially employs a Gillespie algorithm [73] for generating statistically correct trajectories of a stochastic Markov process. Code implementing the Multistrand model is available for public download (http://www.dna.caltech.edu/Multistrand/).

Choice and timing of next transition. Suppose the box is in state i. Then, the next state m is chosen randomly from the states j which are adjacent to i (i.e. they differ from i by only one base pair), weighted by the rate of transition to each.

$$\mathbb{P}(\text{state } m \text{ is chosen}) = \frac{k_{im}}{\sum_j k_{ij}} \tag{5.54}$$

The time taken for the transition from i to m to occur (Δt) is chosen randomly from an exponential distribution with rate parameter λ, where λ is the total rate of transitioning from the current state i. That is,

$$\mathbb{P}(\text{time taken is } \Delta t) = \lambda e^{-\lambda \Delta t} \tag{5.55}$$

where $\lambda = \sum_j k_{ij}$.

First step mode. We will describe first step mode for a general reaction of the form:

$$A + B \xrightarrow{k_{\text{eff}}} C + D \tag{5.56}$$

For a strand displacement reaction, which involves intermediate steps, this model assumes a low-concentration regime where the bimolecular step occurs on a much longer timescale than the uni-

molecular steps; equation 5.56 may then be used to accurately capture the dynamics.

The first step simulation mode begins with the bimolecular join step where A and B collide and form a base pair. The secondary structures of A and B are obtained by Boltzmann sampling the non-pseudoknotted secondary structure space for each molecule. If the bimolecular reaction rates are slow enough for the reactants to be in equilibrium, this sampling is valid. Once the secondary structures of A and B have been sampled, one of the available join steps is chosen at random and the simulation proceeds. As more trajectories are run, different initial secondary structures for A and B and different join steps are explored.

Note that we are not directly simulating the bimolecular join steps, whose rates are proportional to the simulated concentration (and are hence much slower than the unimolecular steps). This allows Multistrand to focus on the trajectories where a collision does occur, rather than spending most of the time simulating unimolecular reactions while waiting for the rare bimolecular reaction.

As simulation of a trajectory proceeds, two distinct end states are tracked: the molecules falling apart into the reactants (one of the A + B configurations), or forming the products (one of the C + D configurations). Each trajectory simulated may be classified as one that failed (if the former happens) or one that reacted (if the latter happens).

Our simulations yield the following data: first passage times for trajectories that reacted ($\Delta t_{\text{react}}^i$), first passage times for trajectories that failed (Δt_{fail}^i), the number of trajectories that reacted (N_{react}) and failed (N_{fail}), and the estimated average rate of collision (k_{coll} in /M/s) of the reactants A and B. For each trajectory, the rate of collision is calculated as $N_{\text{first}} \times k_{\text{bi}} \times u$, where N_{first} is the number of possible first base pairs between the sampled secondary structures of A and B for that trajectory and u is the simulated concentration. k_{coll} is estimated as the mean of the rates of collision for the trajectories simulated.

The following model is used to analyze simulation data. We assume that molecules A and B collide to either form a reactive molecule that will yield the products C and D (in our case, successful displacement) or a nonreactive molecule that will fall apart into the reactants A and B in some time (in our case, unsuccessful displacement).

$$A + B \xrightarrow{k_1} AB \xrightarrow{k_2} C + D \tag{5.57}$$

$$\text{A} + \text{B} \underset{k_2'}{\overset{k_1'}{\rightleftharpoons}} \text{AB}' \tag{5.58}$$

Our model (equations 5.57, 5.58) is fitted as follows.

$$k_1 = \frac{N_{\text{react}}}{N_{\text{react}} + N_{\text{fail}}} \times k_{\text{coll}} \tag{5.59}$$

$$k_1' = \frac{N_{\text{fail}}}{N_{\text{react}} + N_{\text{fail}}} \times k_{\text{coll}} \tag{5.60}$$

$$k_2 = \frac{1}{\mathrm{E}[\Delta t_{\text{react}}^i]} \tag{5.61}$$

$$k_2' = \frac{1}{\mathrm{E}[\Delta t_{\text{fail}}^i]} \tag{5.62}$$

Assuming equation 5.56 is valid, k_{eff} may be predicted [153] based on our model as follows.

$$k_{\text{eff}} = \frac{1}{\Delta t_{\text{correct}}} \times \frac{1}{u} \tag{5.63}$$

where $\Delta t_{\text{correct}}$ is the expected time taken for a successful reaction to occur. $\Delta t_{\text{correct}}$ is calculated from the expected time for a failed collision to fall apart into the reactants (Δt_{fail}), and the expected time for a reactive collision to produce the products (Δt_{react}). Δt_{fail} and Δt_{react} depend on the expected time for any collision to occur (Δt_{coll}). These quantities are calculated as follows.

$$\Delta t_{\text{correct}} = \Delta t_{\text{fail}} \times \frac{k_1'}{k_1} + \Delta t_{\text{react}} \tag{5.64}$$

$$\Delta t_{\text{fail}} = \Delta t_{\text{coll}} + \frac{1}{k_2'} \tag{5.65}$$

$$\Delta t_{\text{react}} = \Delta t_{\text{coll}} + \frac{1}{k_2} \tag{5.66}$$

$$\Delta t_{\text{coll}} = \frac{1}{(k_1 + k_1') \times u} \tag{5.67}$$

In the low-concentration regime, the resolution of the three-stranded complex (resulting in successful displacement of the incumbent or dissociation of the invader) may be assumed to be effectively instantaneous compared to the rate of the bimolecular collision step. That is, we may assume $k_{\text{coll}} \times u << k_2, k_2'$. Indeed, we make this assumption since we are inferring a bimolecular rate constant (equation 5.56). With that assumption, the general formulation (equation 5.63) may be reduced to

$$k_{\text{eff}} = k_{\text{coll}} \times p \tag{5.68}$$

where p is the probability that the collision results in successful displacement of the incumbent.

Simulation details. We simulated the "average strength toehold" experimental system of Zhang and Winfree [147], measuring strand displacement rates as a function of toehold length
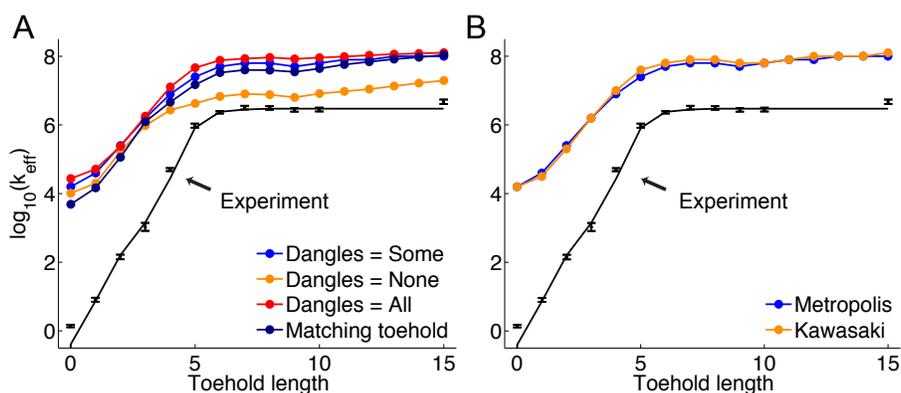
Figure 5.6: Multistrand simulations at 25 °C with different choices: (A) (i) in treating free energy contributions due to dangles [178] (options "Some"(default), "None" and "All" in the NUPACK [123] energy model [120]) and (ii) with substrate overhangs only as long as the toehold on the invader - i.e. matching length bottom toeholds on the substrate, rather than the full 15 base overhang used by Zhang and Winfree [147] (B) different ways of assigning absolute transition rates for unimolecular steps while satisfying detailed balance. Note that none of the variations are able to account for the experimental data points (in black) from Zhang and Winfree [147]; solid black line is their phenomenological model. Standard errors for Multistrand simulations are under 1% (not shown).

Strand	Sequence
Substrate	5'- *GAAGTGACATGGAGA* CGTAGGGTATTGAATGAGGG -3'
Incumbent	5'- CCCTCATTCAATACCCTACG -3'
Invader	5'- CCCTCATTCAATACCCTACG *TCTCCATGTCACTTC* -3'

Table 5.3: Sequences used in Multistrand simulations of strand displacement, with toeholds in italics. For toehold lengths less than 15, the toehold of the invader is truncated to the appropriate length, measured from the 5' end. For simulations with a matching length substrate overhang, the toehold of the substrate was also truncated to match the toehold of the invader.

at 25 °C. We omitted the downstream step used for experimental detection purposes, and deleted the extra domain in the incumbent which was used only in that step. The sequences we used are provided in Table 5.3. Simulations were performed in first step mode.

Multistrand variations. The experimental system of Zhang and Winfree [147] employs a substrate strand with a 15-base overhang. Depending on the length of the invading toehold, a subset of this overhang is complementary to the toehold. The fact that the substrate overhang is longer than the toehold it binds to could conceivably have two effects: (i) stabilizing the first toehold base pair between the invader and the substrate through a dangle free energy contribution and/or (ii) allowing unexpected pathways of displacement through a larger set of possible first base pairs. Multistrand simulations with a matching length substrate overhang (truncated to match the length of the invading toehold) are closer to experiment by only 0.6 orders of magnitude (Figure 5.6(A)).
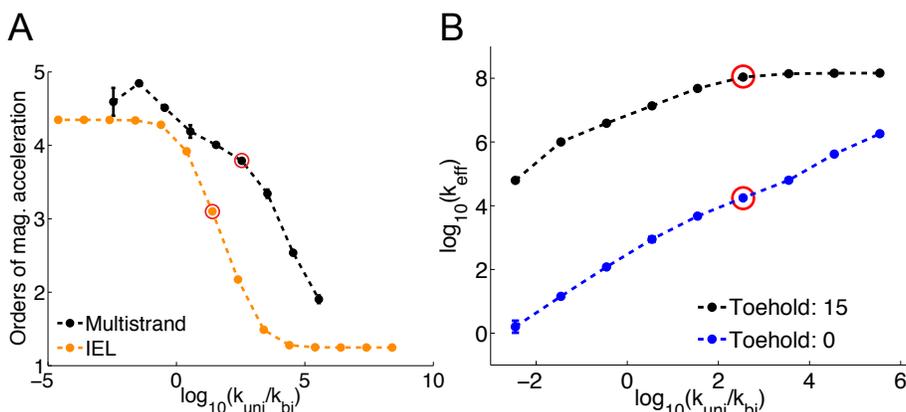
Figure 5.7: The dependence of Multistrand predictions on k_{uni}/k_{bi} (red circles indicate default value of k_{uni}/k_{bi}). Error bars are $3*$SE long where SE is the standard error. (A) Orders of magnitude acceleration in k_{eff} between toehold lengths 15 and 0 ($\mathcal{A}_{15,0}$), as predicted by Multistrand, as a function of $\log_{10}(k_{uni}/k_{bi})$. IEL(2.6, 0) predictions are shown for comparison. Even implausibly low values of k_{uni}/k_{bi} do not result in acceleration that matches the experimentally observed value of 6.5 orders of magnitude. (B) $\log_{10}(k_{eff})$ vs $\log_{10}(k_{uni}/k_{bi})$ for toehold lengths 15 and 0. The surprising non-monotonicity in (A) at the lowest value of k_{uni}/k_{bi} is observed to arise from the disproportionately large decrease in $\log_{10}(k_{eff})$ for toehold length 15 in (B). We hypothesize that this is likely a sequence dependent effect.

This suggests that possible effects (i) and (ii) are not large enough to explain the discrepancy between Multistrand predictions and experimental measurements of strand displacement kinetics.

Because the energy models used by Multistrand [153], NUPACK [123], Mfold [184], and Vienna RNA [183] do not have a consensus method for handling dangle energy terms, we ran Multistrand simulations with each of the three options (Figure 5.6(A)). For each dangles option, k_{bi} and k_{uni} were separately calibrated to the same data (hybridization, zippering) used for calibrating the Metropolis method, as described in Schaeffer [153]. Only minor differences were observed.

We also performed Multistrand simulations using the Kawasaki [174] method for assigning unimolecular transition rates, for which k_{bi} and k_{uni} were also recalibrated; again this yielded nearly identical results (Figure 5.6(B)).

Dependence on the ratio k_{uni}/k_{bi}. When the invader is bound to the substrate by just one base of the toehold, it can either dissociate, leading to unsuccessful displacement, or form another base pair of the toehold and proceed towards zippering. Since dissociation is a bimolecular process, its rate is influenced by k_{bi}, while the rate of the unimolecular zippering process is influenced by k_{uni}. Decreasing k_{uni}/k_{bi} increases the rate of the former relative to the latter. Although this is true for both short and long toeholds, short toeholds reach the state where the invader is bound to the sub-

strate by just one base more often than long toeholds. So, decreasing k_{uni}/k_{bi} disproportionately reduces the displacement rate of short toeholds and increases the orders of magnitude acceleration due to toehold length predicted by Multistrand (Figure 5.7).

5.2.5 Measuring relative stability of strand displacement intermediates

DNA sequence design. DNA oligonucleotide sequences were designed by modifying sequences from Zhang and Winfree [147] by hand to get orthogonal domains d and e. Secondary structures of oligonucleotides and complexes were verified using the NUPACK web server [123] to be as intended.

Buffer conditions. DNA oligonucleotides were stored at 4 °C in TE buffer (10 mM Tris.HCl pH balanced to 8.0, with 1 mM EDTA.Na$_2$, purchased as 100x stock from Sigma-Aldrich). Prior to experimentation, TE buffer containing 62.5 mM MgCl$_2$ was added at a ratio of 1:4 to the sample, resulting in a final MgCl$_2$ concentration of 12.5 mM, out of which 1 mM is bound to EDTA. This buffer was used to prepare and store all oligonucleotide complexes, and to conduct all temperature dependent absorbance experiments, and will be referred to as "TE/Mg^{2+}" buffer.

Annealing and purification of complexes. All DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). Oligonucleotides of length 60 bases or less were ordered with HPLC purification, while those longer than 60 bases were ordered with IE-HPLC purification. Concentrations were estimated from absorbance at 260 nm (measured using a Thermo Scientific NanoDrop cuvette-free spectrophotometer) using calculated extinction coefficients [219].

All complexes Xi:Yj were prepared by annealing Xi (at an approximate concentration of 25 µM) with Yj (at 20% excess) in TE/Mg^{2+} buffer. All annealing operations were performed in an Eppendorf Mastercycler Gradient thermocycler. The samples were cooled at a constant rate from 95 °C to 20 °C in 90 min.

All complexes were purified by nondenaturing (ND) polyacrylamide gel electrophoresis (PAGE) by running the samples on 12% gel at 150 V for approximately 6h. The acrylamide (19:1 acrylamide:bis) was diluted from 40% acrylamide stock purchased from Ambion. ND loading dye containing xylene cyanol FF in 50% glycerol was added to all samples, resulting in a final gycerol concentration of 10% by volume. The appropriate bands were cut out and eluted in 1 mL of TE/Mg^{2+} buffer for 2 days. Purified complexes were quantitated from absorbance at 260 nm (measured using a Thermo Scientific NanoDrop cuvette-free spectrophotometer) using calculated extinction coefficients [219]. Typical yields ranged from 40% to 60%.
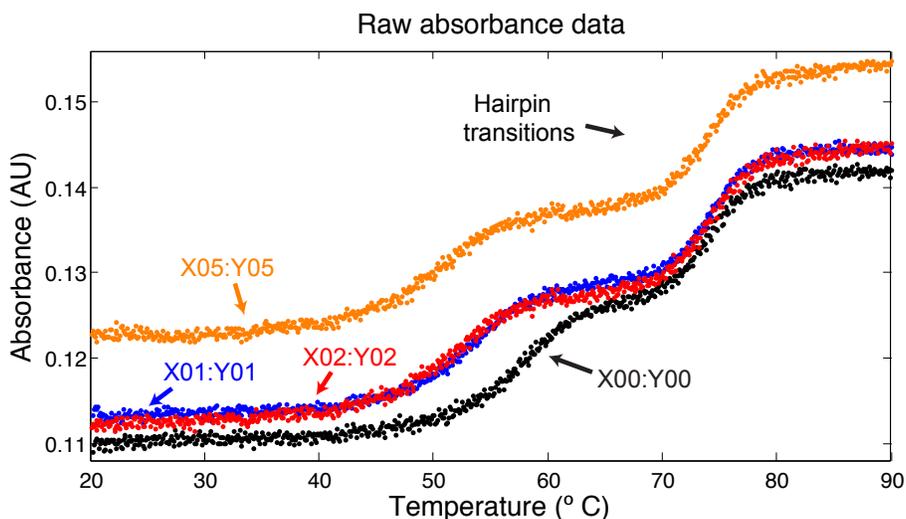
Figure 5.8: Raw absorbance data (at 260 nm), while annealing, at a concentration of 200 nM. Measurements were taken every 0.1 °C between 20 °C and 90 °C. The lower temperature transition is the (bimolecular) formation of the complex, while the higher temperature transition is the (unimolecular) formation of the hairpin. Data acquired by annealing and melting are essentially superimposable.

Temperature-dependent absorbance experiment protocols. Temperature-dependent absorbance experiments were performed using a Model 14 UV-Vis spectrophotometer, equipped with a water bath temperature controller, from AVIV Biomedical, Lakewood, NJ. UV absorbance at 260 nm between 20 °C and 90 °C was measured with a 1 nm bandwidth using 1.6 mL quartz cuvettes. The temperature step was set at 0.1 °C/min with a 0.1 °C dead-band and an equilibration time of 0.25 minutes. All cuvettes were thoroughly cleaned before each experiment: each cuvette was washed 15 times in distilled water, once in 70% ethanol, another five times in distilled water, and finally once more in 70% ethanol.

One temperature-dependent absorbance experiment consisted of: (i) heating from 20 °C to 90 °C, before taking any measurements; (ii) annealing from 90 °C to 20 °C while measuring absorbance every 0.1 °C; (iii) holding for 2h at 20 °C; (iv) melting from 20 °C to 90 °C while measuring absorbance every 0.1 °C. All heating and annealing steps in an experiment were at the temperature step mentioned above (0.1 °C/min). An example of raw temperature-dependent absorbance data at 200 nM obtained while annealing (step ii) is provided in Figure 5.8.

For each complex, one sample was prepared at each of four different concentrations. For each of those samples, two runs of the temperature-dependent absorbance experiment described above were performed.

Concentration (nM)	Upper normalization range ($^\circ$C)
100	[61, 63]
150	[63, 65]
200	[64, 66]
300	[65, 67]
400	[65, 67]
500	[65, 67]

Table 5.4: Melt fraction for each complex is calculated from smoothed absorbance data by normalizing the absorbance in the [20, 35] $^\circ$C range to 0 and the absorbance in the concentration-dependent upper normalization range, specified in this Table, to 1. Our results are robust to this choice; this was verified by repeating the analysis with [65, 67] $^\circ$C as the upper normalization range across all concentrations.

Two state model. We analyze the temperature-dependent absorbance data using a two-state model [191]: each molecule is assumed to be either in the fully bound state (`Xi:Yj`) or the fully dissociated state (`Xi + Yj`).

The raw absorbance data was smoothed by a moving average of 30 points (corresponding to a temperature interval of 3 $^\circ$C). The "melt fraction" or fraction of complex dissociated at temperature T ($f(T)$) was calculated by normalizing the average absorbance of the bound state (between [20 $^\circ$C, 35 $^\circ$C]) to 0 and that of the dissociated state (between a concentration-dependent upper normalization range - see Table 5.4) to 1. Note that the upper normalization range at a given concentration is the same for all complexes. Our results are robust to the choice of upper normalization range; this was verified by repeating the analysis with [65, 67] $^\circ$C as the upper normalization range across all concentrations.

Given the initial concentration c of the complex `Xi:Yj`, the melt fraction $f(T)$ at temperature T in the two-state model can be calculated from $(\Delta H^\circ, \Delta S^\circ)$ as follows. Consider the reaction `Xi + Yj` $\rightleftarrows$ `Xi:Yj`, at temperature T. Let us assume that the initial concentration c of `Xi:Yj` dissociates to `Xi` and `Yj` at concentration x each. Then, `Xi:Yj` is at concentration $c - x$. We know that the equilibrium constant $K_{\mathrm{eq}}(T)$ is related to $f(T)$ as

$$K_{\mathrm{eq}}(T) = \frac{c - x}{x^2} = \frac{1 - f(T)}{cf(T)^2} \tag{5.69}$$

Solving the quadratic equation for $f(T) \geq 0$, we get

$$f(T) = \frac{-1 + \sqrt{1 + 4cK_{\mathrm{eq}}(T)}}{2cK_{\mathrm{eq}}(T)} \tag{5.70}$$

Since $K_{\mathrm{eq}}(T) = \exp(-\frac{\Delta G^\circ(T)}{RT})$, we may predict the entire temperature-dependent melt fraction curve by varying T appropriately.
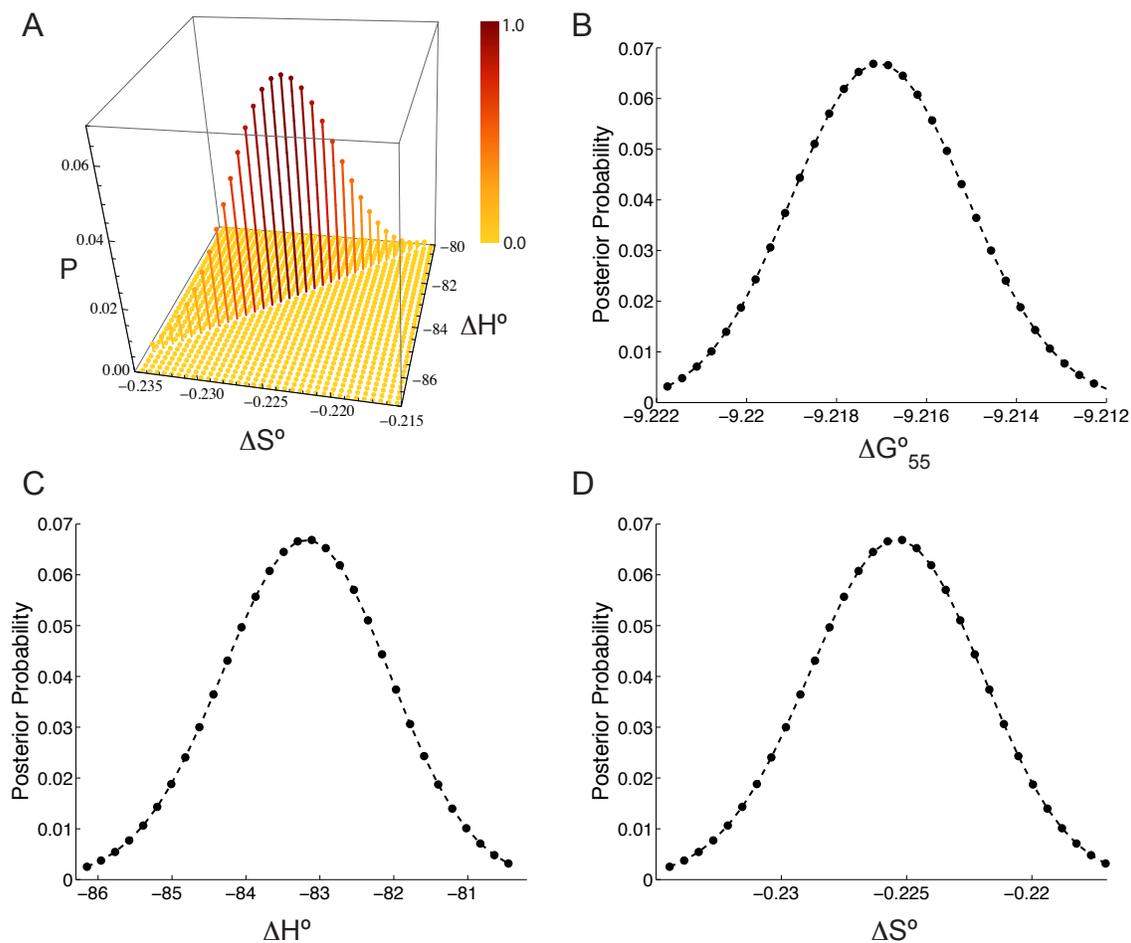
Figure 5.9: Example posterior probability distributions obtained by Bayesian analysis over (A) $(\Delta H^\circ, \Delta S^\circ)$ space and marginals over (B) ΔG° at 55 °C, (C) ΔH° and (D) ΔS° for complex X10:Y10. All ΔG°_{55} and ΔH° values are in kcal/mol while ΔS° values are in kcal/K/mol. Note that the 99% confidence interval is much more narrow for ΔG°_{55} compared to ΔH° and ΔS°.

For each complex, we infer $(\Delta H^\circ, \Delta S^\circ)$ (and hence ΔG_{25}°, ΔG_{55}°) by fitting the predicted melt fraction curves to smoothed and normalized absorbance data across different concentrations. By comparing the free energies of different complexes, we can infer the contribution of the poly-T overhangs. We do this in two ways: a Bayesian analysis and a descriptive "leave-one-concentration-out" fit.

Bayesian analysis. We essentially discretize the $(\Delta H^\circ, \Delta S^\circ)$ space into a grid and calculate the likelihood that our experimental data for each complex (all data traces at four concentrations) arose from each candidate pair in the discretization, assuming an independent Gaussian noise model. Normalizing the likelihood yields the posterior distribution for ΔH°, ΔS° and ΔG° (e.g. Figure 5.9). We calculate posterior means and 99% confidence intervals, under the assumptions of the two-state model and our Bayesian framework.

In other words, given candidate values of the standard enthalpy and entropy of formation, $(\Delta H_0^\circ, \Delta S_0^\circ)$, for a particular complex, smoothed and normalized absorbance data D_{T_i, c_j} at a certain temperature T_i and concentration c_j is assumed to be related to the predicted melt fraction f_{T_i, c_j} as follows:

$$D_{T_i, c_j} = f_{T_i, c_j} + \xi_{T_i, c_j} \tag{5.71}$$

where

$$\xi_{T_i, c_j} \sim \mathcal{N}(0, \sigma_{T_i, c_j}^2)$$

is independent additive Gaussian noise. That is, ξ_{T_i, c_j} and ξ_{T_l, c_k} are assumed to be independent if $T_i \neq T_l$ or $c_j \neq c_k$. σ_{T_i, c_j}^2 is calculated as the sample variance of smoothed and normalized absorbance data points at concentration c_j in a neighborhood around T_i (three points on either side of T_i).

With these assumptions, the likelihood of observing the data given the estimate $(\Delta H_0^\circ, \Delta S_0^\circ)$ is simply

$$L(\Delta H_0^\circ, \Delta S_0^\circ) = \prod_{T_i, c_j} \phi_{f_{T_i, c_j}, \sigma_{T_i, c_j}^2}(D_{T_i, c_j}) \tag{5.72}$$

where ϕ_{μ, σ^2} is the probability density function of the Gaussian distribution with mean μ and variance σ^2.

Starting with a uniform prior over the $(\Delta H^\circ, \Delta S^\circ)$ grid, the posterior probability distribution
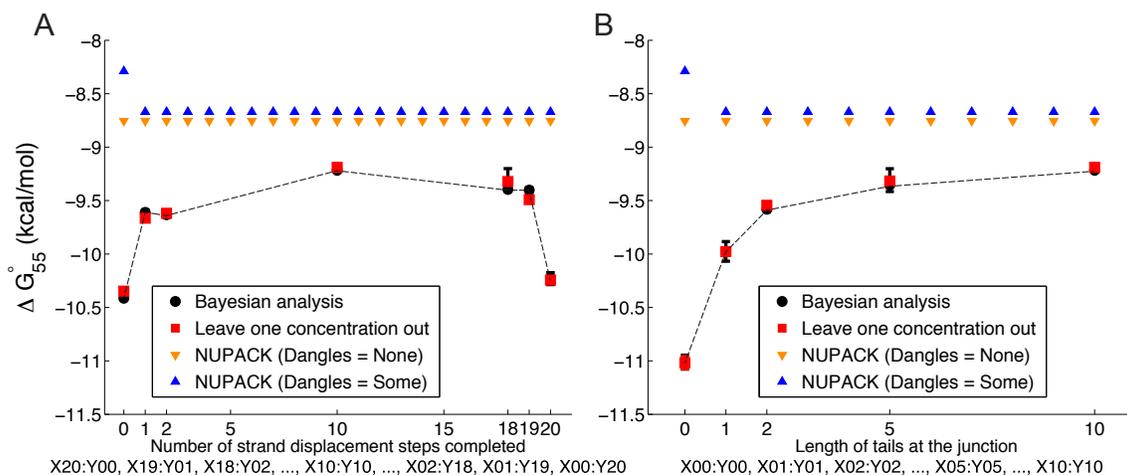
Figure 5.10: ΔG° of formation (at $55\ ^\circ$C) of complexes in the *strand displacement snapshot* study (A) or the *local overhang* study (B). Error bars in black indicate Bayesian posterior means and 99% confidence intervals. Error bars in red indicate means and standard deviations of leave-one-concentration-out least square fits. NUPACK predictions with dangles options "some" and "none" are provided for comparison.

is proportional to the likelihood (this standard result from Bayesian statistics is justified later). So, normalizing the likelihood of observing our data, we can calculate the posterior distribution:

$$P(\Delta H_0^\circ, \Delta S_0^\circ) = \frac{L(\Delta H_0^\circ, \Delta S_0^\circ)}{\displaystyle\sum_{\Delta H_i^\circ, \Delta S_j^\circ} L(\Delta H_i^\circ, \Delta S_j^\circ)} \tag{5.73}$$

In practice, we first perform a coarse discretization of $(\Delta H^\circ, \Delta S^\circ)$ space in order to identify the region containing non-zero values of the posterior probability; we then perform a fine discretization of that region and evaluate the posterior probability over it.

Once we have the posterior probability over $(\Delta H^\circ, \Delta S^\circ)$ space, we find the smallest region containing 99% of the probability, and then evaluate marginal posterior probability distributions for ΔH°, ΔS°, ΔG_{25}° and ΔG_{55}° (Figure 5.9). The 99% confidence intervals are relatively much narrower for ΔG_{55}° than ΔH° or ΔS°. This shows that our data permits accurate comparison of the stability of our complexes through ΔG°, but cannot easily separate the enthalpic and entropic contributions. Also note that error bars and 99% confidence intervals are much narrower for ΔG_{55}° (Figure 5.10) compared to ΔG_{25}° (Figure 2.8 in the main text). This is to be expected because the former temperature is closer to the experimental melting temperature of our complexes. With the assumptions in the two-state model and our Bayesian framework, we report posterior means and 99% confidence intervals for quantities of interest (Table 5.6, Figure 5.10).

Relationship between posterior probability and likelihood. We now recall that with a uniform prior, the posterior probability distribution is proportional to the likelihood. For a more detailed introduction, see Gelman et al. [220]. Suppose θ is a vector of parameters we want to infer, and that we have data $\mathcal{D}$ which is informative about θ. Then, we know

$$\mathbb{P}(\theta, \mathcal{D}) = \mathbb{P}(\theta) \times \mathbb{P}(\mathcal{D}|\theta) = \mathbb{P}(\mathcal{D}) \times \mathbb{P}(\theta|\mathcal{D}) \tag{5.74}$$

Therefore the posterior distribution $\mathbb{P}(\theta|\mathcal{D})$ is obtained by

$$\mathbb{P}(\theta|\mathcal{D}) = \frac{\mathbb{P}(\theta) \times \mathbb{P}(\mathcal{D}|\theta)}{\mathbb{P}(\mathcal{D})} \tag{5.75}$$

Here, $\mathbb{P}(\theta)$ is constant because we start with a uniform prior. $\mathbb{P}(\mathcal{D}) = \sum_\theta \mathbb{P}(\theta) \times \mathbb{P}(\mathcal{D}|\theta)$ is also independent of θ. $\mathbb{P}(\mathcal{D}|\theta)$ is nothing but the likelihood. Hence, with a uniform prior, the posterior distribution is proportional to the likelihood.

Leave-one-concentration-out analysis. This is a simple and descriptive way of analyzing the data, which essentially serves as a sanity check. Data from each complex is analyzed separately to infer the free energy of formation of that complex. We measured temperature-dependent absorbance data at four concentrations. Here, we sequentially leave out data from one concentration at a time, thus generating four datasets, each containing data from three concentrations. For each dataset, we perform a simultaneous nonlinear least squares fit (using the Levenberg-Marquardt algorithm, implemented by a built-in MATLAB function) of the predicted melt fraction curves to the smoothed and normalized absorbance data across all three concentrations present in the dataset. This procedure generates four estimates of $(\Delta H^\circ, \Delta S^\circ)$ of formation for each complex, one for each leave-one-concentration-out dataset. We then calculate ΔG_{25}° and ΔG_{55}° for each of those four estimates and report the mean and standard deviation, for each complex (Table 5.5).

5.2.6   Coarse-grained molecular modeling

A Coarse-grained molecular model. OxDNA and its interaction potentials have been described in detail by Ouldridge [154]. Code implementing OxDNA is available for public download (http://dna.physics.ox.ac.uk/). OxDNA represents DNA as a string of nucleotides, where each nucleotide (sugar, phosphate and base group) is a rigid body with interaction sites for backbone, stacking and hydrogen-bonding interactions. The potential energy of the system can be decom-

Complex	$\Delta G^\circ_{25,\,avg}$	$\Delta G^\circ_{25,\,sd}$	$\Delta G^\circ_{55,\,avg}$	$\Delta G^\circ_{55,\,sd}$
X20:Y00	-18.00	0.15	-10.35	0.023
X19:Y01	-16.48	0.14	-9.66	0.013
X18:Y02	-16.15	0.11	-9.62	0.031
X10:Y10	-16.03	0.14	-9.19	0.033
X02:Y18	-16.36	0.08	-9.32	0.029
X01:Y19	-16.44	0.16	-9.49	0.014
X00:Y20	-17.96	0.09	-10.24	0.027
X00:Y00	-19.17	0.31	-11.02	0.058
X01:Y01	-17.78	0.06	-9.98	0.017
X02:Y02	-16.92	0.27	-9.54	0.008
X05:Y05	-16.22	0.28	-9.32	0.028

Table 5.5: Leave-one-concentration-out mean and standard deviation for ΔG° at 25 °C and 55 °C, for each complex. All values in kcal/mol.

Complex	E[ΔH°]	ΔH°_{CI}	E[ΔS°]	ΔS°_{CI}	E[ΔG°_{25}]	$\Delta G^\circ_{25,\,CI}$
X20:Y00	-93.2	[-97.8, -88.8]	-0.252	[-0.266, -0.239]	-17.99	[-18.42, -17.56]
X19:Y01	-84.9	[-89.1, -81.0]	-0.229	[-0.242, -0.217]	-16.49	[-16.85, -16.16]
X18:Y02	-80.9	[-84.8, -77.3]	-0.217	[-0.229, -0.206]	-16.15	[-16.49, -15.84]
X10:Y10	-83.2	[-86.2, -80.4]	-0.226	[-0.234, -0.217]	-15.98	[-16.25, -15.73]
X02:Y18	-83.9	[-92.8, -80.1]	-0.227	[-0.255, -0.215]	-16.21	[-16.84, -15.87]
X01:Y19	-87.5	[-91.6, -83.5]	-0.238	[-0.250, -0.226]	-16.54	[-16.99, -16.19]
X00:Y20	-97.0	[-99.8, -89.8]	-0.264	[-0.273, -0.242]	-18.16	[-18.47, -17.54]
X00:Y00	-100.3	[-104.6, -96.2]	-0.272	[-0.285, -0.260]	-19.18	[-19.63, -18.74]
X01:Y01	-95.8	[-103.5, -88.1]	-0.262	[-0.285, -0.238]	-17.82	[-18.49, -17.16]
X02:Y02	-89.9	[-94.3, -85.6]	-0.245	[-0.258, -0.232]	-16.92	[-17.30, -16.55]
X05:Y05	-83.9	[-92.0, -78.7]	-0.227	[-0.252, -0.211]	-16.17	[-16.78, -15.75]

Table 5.6: Bayesian posterior means and 99% confidence intervals for ΔH°, ΔS° and ΔG°_{25} for each complex. ΔH° and ΔG°_{25} values are in kcal/mol while ΔS° is in kcal/K/mol.
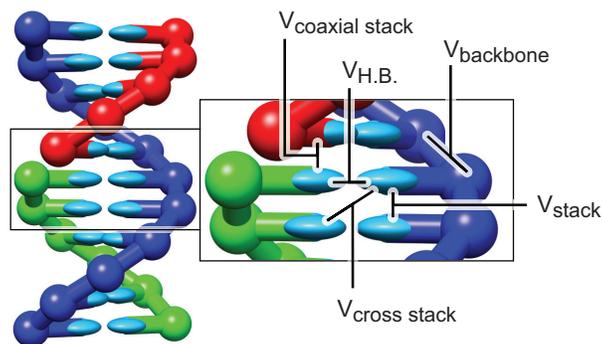
Figure 5.11: A model DNA duplex, with stabilising interactions depicted schematically. The backbone sites are shown as spheres, the bases as ellipsoids. Backbone colouring indicates strand identity. All nucleotides also interact with repulsive excluded volume interactions. The coaxial stacking interaction acts like a stacking interaction between bases that are not immediate neighbours along the backbone of a strand.

posed as

$$V = \sum_{\langle ij \rangle} \left(V_{\text{b.b.}} + V_{\text{stack}} + V'_{\text{exc}} \right) +$$

$$\sum_{i,j \notin \langle ij \rangle} \left(V_{\text{HB}} + V_{\text{cr.st.}} + V_{\text{exc}} + V_{\text{cx.st.}} \right), \tag{5.76}$$

where the first sum is taken over all nucleotides that are nearest neighbors on the same strand and the second sum comprises all remaining pairs. The interactions between nucleotides are schematically shown in Figure 5.11. The backbone potential $V_{\text{b.b.}}$ is an isotropic spring that imposes a finite maximum distance between backbone sites of neighbours, mimicking the covalent bonds along the strand. The hydrogen bonding (V_{HB}), cross stacking ($V_{\text{cr.st.}}$), coaxial stacking ($V_{\text{cx.st.}}$) and stacking interactions (V_{stack}) are anisotropic and explicitly depend on the relative orientations of the nucleotides as well as the distance between the relevant interaction sites. This orientational dependence captures the planarity of bases, and drives the formation of helical duplexes. The coaxial stacking term is designed to capture stacking interactions between bases that are not immediate neighbours along the backbone of a strand. Bases and backbones also have excluded volume interactions V_{exc} or V'_{exc}.

Hydrogen-bonding interactions are only possible between complementary (A-T and C-G) base-pairs, but no other sequence-dependence is included in oxDNA. Consequently, the interactions were fitted to reproduce melting temperatures of 'average' oligonucleotides, obtained by averag-

ing over the parameters of SantaLucia's nearest-neighbour model [118], and the structural and mechanical properties of double- and single-stranded DNA.

oxDNA does not have any explicit electrostatic interactions. It was fitted to reproduce DNA behavior at salt concentration $[\mathrm{Na^+}] = 0.5\,\mathrm{M}$, where the electrostatic properties are strongly screened, and it may be reasonable to incorporate them into a short-ranged excluded volume. Possible issues related to salt concentrations are discussed in the main text. It should be noted that the oxDNA neglects several features of the DNA structure and interactions due to the high level of coarse-graining. Specifically, the double helix in the model is symmetrical rather than the grooves between the backbone sites having different sizes (i.e., major and minor grooving), and all four nucleotides have the same structure.

Simulation Techniques. The thermodynamic properties of the molecular model are obtained by averaging over the Boltzmann distribution

$$\rho(\mathbf{r}^N, \mathbf{p}^N, \mathbf{q}^N, \mathbf{L}^N) \quad \propto \quad \mathrm{e}^{(-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N, \mathbf{q}^N, \mathbf{L}^N))}. \tag{5.77}$$

Here $\mathcal{H}$ is the system Hamiltonian, which is a function of positional and angular particle coordinates $\mathbf{r}^N$ and $\mathbf{q}^N$ and their generalized momenta $\mathbf{p}^N$ and $\mathbf{L}^N$. As the terms containing $\mathbf{p}^N$ and $\mathbf{L}^N$ in $\mathcal{H}$ are separable and can be analytically integrated out, the probability of a certain configuration is proportional to a Boltzmann factor for its potential energy, $\exp(-\beta V(\mathbf{r}^N, \mathbf{q}^N))$. Obtaining kinetic properties requires an additional choice of dynamics. The Virtual Move Monte Carlo (VMMC), and Langevin Dynamics (LD) algorithms used for thermodynamic and kinetic properties respectively are outlined in the following sections.

Virtual Move Monte Carlo. Monte Carlo techniques involve randomly generating trial moves of a system consisting of a set of particles, and accepting those moves with probabilities that ensure the simulation samples from the distribution given in equation 5.77. The widely used Metropolis Monte Carlo algorithm (MMC) [172] attempts random moves of single particles, and accepts those moves with a probability

$$P_{\mathrm{acc}}(\mu \to \nu) = \min\{1, \exp(-\beta(V^\nu - V^\mu))\}, \tag{5.78}$$

where μ and ν represent initial and final states respectively. Although simple to implement, MMC can struggle to equilibrate strongly interacting systems as moves of single particles tend to generate large increases in energy, and are therefore rejected.
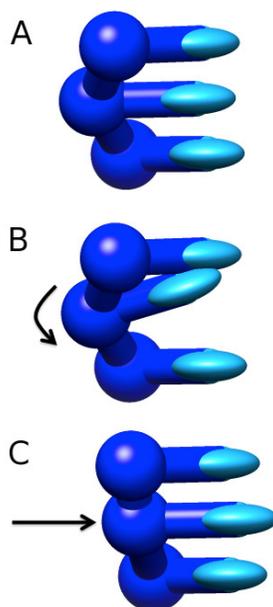
Figure 5.12: Examples of initial 'single particle' seed moves attempted in our implementation of the VMMC algorithm. (A) Initial state. (B) New state obtained from (A) after rotating the central nucleotide about its backbone site. (C) New state obtained from (A) after translating the central nucleotide. In each of the new states, the flanking nucleotides act as a reference: they have exactly the same position and orientation as in (A). Once an initial 'seed' move such as (B) or (C) has been selected, clusters are grown from this seed in the manner outlined by Whitelam et al. [200].

The VMMC algorithm [199, 200] overcomes this problem by generating clusters that depend on energy changes resulting from attempted moves (we use the variant presented in the appendix of Whitelam et al. [200]). A random single particle move is chosen, and energy changes due to that move are calculated with all neighbours of the seed particle. Those particles for which the energy is increased are probabilistically added to the cluster, and move together with the seed particle. The process is then repeated with neighbours of the newly added particles, until no more new links are formed. Due to the statistical biases introduced during cluster building, a more complex acceptance factor than equation 5.78 is required in order to satisfy detailed balance.

In the context of the molecular model studied in this work, a 'single particle' is a nucleotide, and the attempted moves are translation of a whole nucleotide and rotation of a nucleotide about its backbone site. These moves are illustrated in Figure 5.12.

Umbrella sampling. Despite the efficiency of VMMC, obtaining accurate statistics for the free-energy landscape of displacement and the stability of a duplex can be difficult. Equilibration can be accelerated by flattening free-energy barriers with an artificial biasing weight $W(\mathbf{r}^N, \mathbf{q}^N)$

[202]. A lower barrier means transitions occur more quickly, and equilibration is accelerated. The thermodynamic expectation of any variable A follows from the biased sample obtained as

$$\langle A \rangle = \frac{\int \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{q}^N A(\mathbf{r}^N, \mathbf{q}^N) \mathrm{e}^{-\beta V(\mathbf{r}^N, \mathbf{q}^N)}}{\int \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{q}^N \mathrm{e}^{-\beta V(\mathbf{r}^N, \mathbf{q}^N)}}$$

$$= \frac{\int \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{q}^N \frac{A(\mathbf{r}^N, \mathbf{q}^N)}{W(\mathbf{r}^N, \mathbf{q}^N)} W(\mathbf{r}^N, \mathbf{q}^N) \mathrm{e}^{-\beta V(\mathbf{r}^N, \mathbf{q}^N)}}{\int \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{q}^N \frac{1}{W(\mathbf{r}^N, \mathbf{q}^N)} W(\mathbf{r}^N, \mathbf{q}^N) \mathrm{e}^{-\beta V(\mathbf{r}^N, \mathbf{q}^N)}}$$

$$= \frac{\langle A(\mathbf{r}^N, \mathbf{q}^N) / W(\mathbf{r}^N, \mathbf{q}^N) \rangle_W}{\langle 1 / W(\mathbf{r}^N, \mathbf{q}^N) \rangle_W}. \tag{5.79}$$

Here $\langle \rangle_W$ indicates the expectation found by sampling from the biased distribution $W(\mathbf{r}^N, \mathbf{q}^N) \exp(-\beta V(\mathbf{r}^N, \mathbf{q}^N))$. For simulations in this paper an initial $W(\mathbf{r}^N, \mathbf{q}^N)$ was chosen from experience, and improved by hand as required.

Langevin Dynamics. Langevin Dynamics is an approach for generating diffusive motion of coarse-grained models with implicit solvent. The principle is that the solvent exerts both random forces and dissipative drag on the solute, and that the two are related by a fluctuation-dissipation relation to ensure that the steady-state distribution is given by equation 5.77. Newton's equations, with the addition of these solvent-mediated forces, can then be integrated to give dynamical trajectories. In this work, we use the quaternion-based algorithm of Davidchack et al. [201] as an efficient methodology for simulating rigid bodies.

To use this algorithm, it is necessary to specify a friction tensor relating the drag forces experienced by a particle to its momenta. For simplicity, we assume each nucleotide interacts with the solvent in a spherically symmetric manner, meaning that the task is reduced to identifying linear and rotational damping coefficients, γ and Γ. We choose values of $\gamma = 0.59 \, \mathrm{ps}^{-1}$ and $\Gamma = 1.76 \, \mathrm{ps}^{-1}$. These values are around one to two orders of magnitude smaller than would be inferred for a objects the size of nucleotides in a fluid with the viscosity of water. Lower friction coefficients accelerate dynamics, which would have been prohibitively slow otherwise, but still reproduce diffusive motion. It is reasonable to assume that, for the comparison of relative rates of similar processes, this choice will not be overly important given the number of approximations already inherent in the oxDNA. Indeed, simulations with $\gamma = 5.9 \, \mathrm{ps}^{-1}$ and $\Gamma = 17.6 \, \mathrm{ps}^{-1}$ showed no measurable difference in the probability of successful displacement for a 3-base toehold once attached. Furthermore, the sequence of states visited during unbiased VMMC simulations provides another (less rigorous) approximation to dynamics. These simulations give a somewhat lower success rate

of displacement from a toehold-bound state than LD. The net effect, however, is simply a shifting of the start of the plateau by around half a base pair, and the qualitative arguments are still valid. These results suggest the findings presented here are not strongly sensitive to the choice of dynamics. Simulations reported in this work were performed with a time step of 5.13 fs. As shown by Ouldridge [154], such a time step reproduces the average energies of a duplex and the hairpin-folding kinetics obtained with much smaller integration time steps.

Forward flux sampling. Forward flux sampling (FFS) allows the calculation of the flux between two local minima of free energy, and also sampling from the ensemble of trajectories that link the two minima [203, 204]. The term 'flux' from (meta)stable state A to state B has the following definition.

> Given an infinitely long simulation in which many transitions are observed, the flux of trajectories from A to B is $\Phi_{AB} = N_{AB}/(\tau f_A)$, where N_{AB} is the number of times the simulation leaves A and then reaches B, τ is the total time simulated and f_A is the fraction of the total time simulated for which state A has been more recently visited than state B.

The concept of flux is therefore a generalization of a transition rate for processes that are not instantaneous: it incorporates the time spent in intermediate states between A and B. In our simulations of displacement, state A is the initial configuration of incumbent bound to substrate, with invading strand unattached, and B is the state with the invading strand bound to the substrate with the incumbent strand unattached. The three-stranded complexes involved in branch migration are intermediate states.

In the experimental work of Zhang and Winfree [147], bimolecular rate constants for strand displacement are measured using bulk systems at low concentrations ($\sim$ nM). In our simulations, we use three strands in a periodic cell of volume 1.67×10^{-20} L. The periodic boundary conditions allow us to mimic a bulk system, and the volume used corresponds to a concentration of approximately 100 μM for each strand. If the simulation time spent in intermediate states were negligible, fluxes could then be taken directly as instantaneous reaction rates, allowing bimolecular rate constants to be inferred and compared to experiments. At such a high concentration as 100 μM, however, we find that the time scale for resolving a three-stranded complex (when either
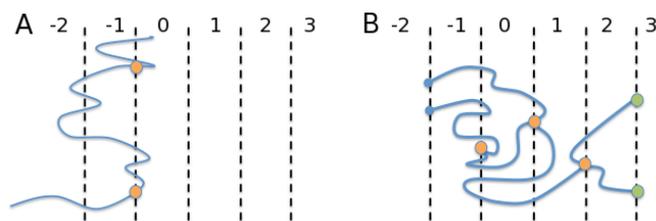
Figure 5.13: Schematic illustration of FFS. The space is divided into different values of the order parameter Q by interfaces, with Q values indicated by numbers in the figure. We wish to estimate the flux from $Q = -2$ to $Q = Q_{\max} = 3$. (A) First stage, estimating the flux across the interface λ_{-1}^0 between $Q = -1$ and $Q = 0$. Circles indicate crossings of the interface that contribute to the flux; only positive crossings are counted, and only the first crossing since leaving $Q = -2$. (B) Subsequent stages, estimating the probability of reaching interface λ_{Q-1}^Q from λ_{Q-1}^{Q-2}. In this example, two trajectories are launched from each point at the prior interface. These trajectories end either by successfully reaching λ_{Q-1}^Q, or by failing and returning to λ_{-2}^{-1}.

displacement is completed, or the invading strand detaches) can be a significant contribution to the overall displacement reaction time. It is reasonable to assume that the time required to resolve the three-stranded complex does not scale with simulation concentration, and therefore should not contribute to rate constants at the low concentrations typical of experiment. To make a fair comparison to experiment, therefore, we must not include time spent in three-stranded complexes in our estimate of the flux. Mathematically, this corresponds to redefining f_A so that it doesn't include time spent in three-stranded intermediates. The measurements of flux thus recorded are reported in the main text as transition rates, as the assumption of instantaneous transitions for a dilute solution is implicit in the measurement process.

FFS is illustrated schematically in Figure 5.13. We first discuss FFS generally, before specifying the implementation for our system. We require an order parameter Q that measures the extent of the reaction, such that non-intersecting interfaces λ_{Q-1}^Q can be drawn between consecutive values of Q. We define the lowest value of Q as $Q = -2$ because the procedure for interfaces λ_{Q-1}^Q with $Q > 0$ is distinct from that for $Q \leq 0$. Initially, simulations are performed that begin in the lowest value of Q and the flux of trajectories crossing the surface λ_{-1}^0 (for the first time since leaving $Q = -2$) is measured. Note that the simulation is not restarted when trajectories cross λ_{-1}^0.

The total flux of trajectories from $Q = -2$ to the alternative minima ($Q = Q_{\max}$) is then calculated as the flux across λ_{-1}^0 from $Q = -2$, multiplied by the conditional probability that these trajectories reach $Q = Q_{\max}$ before returning to $Q = -2$, $P(\lambda_{Q_{\max}-1}^{Q_{\max}}|\lambda_{-1}^0)$. This probability can be factorized into the product of the probabilities of trajectories starting from the interface λ_{Q-1}^Q

reaching the interface λ_Q^{Q+1} before returning to $Q = -2$

$$P(\lambda_{Q_{\max}-1}^{Q_{\max}}|\lambda_{-1}^0) = \Pi_{Q=1}^{Q_{\max}} P(\lambda_{Q-1}^Q|\lambda_{Q-2}^{Q-1}). \tag{5.80}$$

The simulation then proceeds by randomly loading microstates which correspond to the crossing of λ_{-1}^0, and using these as initial points from which to estimate $P(\lambda_0^1|\lambda_{-1}^0)$. The process is then iterated for successive interfaces, allowing the estimation of flux and the construction of trajectories sampled from the distribution of transition pathways.

Simulation implementation.

Kinetics of displacement. Three strands of DNA (substrate, incumbent and invading strand, with sequences as given in the Table 5.7) were simulated in a periodic cubic cell of 1.67×10^{-20} l, at a temperature of $T = 25°$C, using the LD algorithm with FFS. Strands were initialized with the incumbent bound to the substrate and the invading strand separate, and equilibrated for $0.513\,\mu$s. Given the diffusion constant of model DNA ($\sim 10^{-9}\,\mathrm{m^2 s^{-1}}$ for our system with $\gamma = 0.59\,\mathrm{ps^{-1}}$ and $\Gamma = 1.76\,\mathrm{ps^{-1}}$), this is approximately the time required for the reactant DNA molecules to diffuse by the length of the simulation cell. It is also far longer than the decorrelation time of the potential energy of the reactants (around $1\,$ns).

To simplify the sampling, we used an approach in which only the expected base pairs between the incumbent and the substrate or the invading strand and the substrate are given non-zero binding strength - we shall refer to these base pairs as *correct base pairs*. In a full system, incorrect base pairs can potentially contribute to association processes, typically through the formation of metastable misbonded structures that then relax into the intended configuration [192]. The metastability of these misbonded structures would make them difficult to incorporate into the FFS simulations of displacement. As the sequences used by Zhang and Winfree [147] were deliberately designed to minimize the occurrence of misbonded configurations, we expect that this simplification should have minor systematic effects on our measurements. We also tested this assumption by measuring the rate at which two strands, corresponding to a full invader with a toehold of six bases and the 15-base toehold domain of the substrate, formed the intended 6-base-pair toehold. In one case we considered only native interstrand interactions, in the other we allowed all possible complementary base pairs. We found that allowing non-native base pairs increased association rates by approximately 50%, a small effect given the range of relative displacement rates considered in this work. More details are provided in the following sections.

Duplex	Sequence (5′–3′)
Substrate	GACATGGAGA*CGTAGGGTATTGAATGAGGG*
Incumbent	*TCCCTCATTCAATACCCTACG*
Invader	*CCCTCATTCAATACCCTACG*[TCTCCAT]

Table 5.7: Sequences used in simulations of strand displacement. Bases in italics constitute the primary branch migration domain. The bases in square brackets indicate those that are added to create a toehold.

The order parameter Q used for FFS up to the attachment of the invading strand to the substrate is the same for all toehold lengths, and the definition is given in Table 5.8. For this parameter, a potential correct base pair between the invading and substrate strands is counted as *nearly formed* if

- The separation of hydrogen bonding sites is ≤ 0.85 nm.

- The hydrogen-bonding potential consists of a separation dependent factor multiplied by a number of modulating angular factors. At most one of these factors that contributes multiplicatively to the hydrogen-bonding energy is zero.

- The hydrogen-bonding energy is less negative than -2.98 kcal mol^{-1}.

In this context, a 'correct' base pair is a base pair that is expected in the final substrate/invading strand duplex. Physically, these conditions mean that the bases are close and fairly well aligned, but not forming a strong base pair.

For higher values ($Q > 4$), the order parameter definition depends on toehold length. For shorter toeholds, a greater number of interfaces are needed to measure the flux accurately. Successive values of Q correspond to increasing numbers of correct base pairs (with a hydrogen-bonding energy more negative than -2.98 kcal mol^{-1}) between the invading strand and the substrate. Finally, the maximum value of Q corresponds to the invading strand having the maximum number of base pairs with the substrate, and all correct base pairs for the incumbent and substrate strand having a distance of at least 2.56 nm between hydrogen-bonding sites. The explicit definitions of $Q > 4$ for each toehold are given in Table 5.9.

For each toehold length, 100 independent simulations were performed to measure the initial flux. In each simulation, 10 states at the interface λ_{-1}^0 were collected, giving 1000 in total (a typical number in FFS simulations [203, 204]). Crossings of the λ_{-1}^0 interface were saved with a 10 % probability, meaning that ~ 100 crossings were observed in total for each independent simulation. Saving states with only a 10 % probability ensures that the 1000 states collected are more statisti-

Order parameter Q	Separation d/nm	Nearly-formed bp n	Formed bp b
$Q = -2$	$d > 5.11$	0	0
$Q = -1$	$5.11 \geq d > 3.42$	0	0
$Q = 0$	$3.42 \geq d > 2.56$	0	0
$Q = 1$	$2.56 \geq d > 1.71$	0	0
$Q = 2$	$1.71 \geq d > 0.85$	0	0
$Q = 3$	$d \leq 0.85$	0	0
$Q = 4$	$d \leq 0.85$	$n \geq 1$	0

Table 5.8: Order parameter definitions for early stages of FFS simulations of displacement, up to the stage of attachment of the invading strand to the substrate. The separation d is the minimum distance between hydrogen-bonding sites over all potential correct base pairs in the invading and substrate strands.

Q	0,1	2	3	4
5	$b = 1$	$b = 1$	$1 \leq b < 3$	$1 \leq b < 4$
6	$2 \leq b < 4$	$2 \leq b < 5$	$3 \leq b < 8$	$4 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$
7	$4 \leq b < 10$	$5 \leq b < 9$	$8 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$	$Q = Q_{\mathrm{max}}$
8	$10 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$	$9 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$	$Q = Q_{\mathrm{max}}$	
9	$Q = Q_{\mathrm{max}}$	$Q = Q_{\mathrm{max}}$		

Q	5	6	7
5	$1 \leq b < 5$	$1 \leq b < 6$	$1 \leq b < 7$
6	$5 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$	$6 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$	$7 \leq b \,\&\, Q \neq Q_{\mathrm{max}}$
7	$Q = Q_{\mathrm{max}}$	$Q = Q_{\mathrm{max}}$	$Q = Q_{\mathrm{max}}$

Table 5.9: Order parameter definitions for the later stages of FFS simulations of displacement. b is the number of base pairs between substrate and invading strand with hydrogen-bonding energy more negative than $-2.98\,\mathrm{kcal\,mol^{-1}}$. $Q = Q_{\mathrm{max}}$ is defined as b having its maximum value (toehold length plus branch migration domain length) and all correct base pairs for the incumbent and substrate strand having a distance of at least $2.56\,\mathrm{nm}$ between hydrogen-bonding sites.

cally independent than otherwise. For each subsequent interface, a large number of trajectories were launched and those that reached the next interface before returning to $Q = -2$ were saved. The number of attempts and successes for each interface are given in Table 5.10.

Uncertainties in the measurements can be estimated in the following manner. As 100 independent simulations were performed for each toehold length to measure the initial flux, the standard error can be estimated in the usual way. Assuming we have obtained a representative set of states at each interface, the later stages can be modelled as Bernoulli trials – the probability of success measured after N attempts has a variance of $p(1-p)/N$, where p is the true probability of success. The measured p can then be used to estimate the standard error on p for each stage. The errors for individual stages can then be added in quadrature, and overall errors are given alongside the overall fluxes in Table 5.10. Note that this estimation of errors assumes that the set of states at each interface is a representative sample of the true distribution. As such, this error estimate is a lower bound on the true uncertainty. In particular, the assumption will be weakest for the 0-base toehold, where the number of completely independent reactive trajectories is small. However, given that the relative rates for different toehold lengths in the exponential regime are approximately consistent with what would be expected from the known thermodynamics of oxDNA, and that the overall difference between the shortest and longest toeholds is larger than six orders of magnitude, errors due to measurement uncertainty are unlikely to affect our conclusions.

In some simulations involving the longer toeholds, complete (and long-lived) binding to the toehold occurred during the initial measurements of the flux across λ^0_{-1}. In a conventional implementation of FFS [203, 204], the time spent in these states would be included in the estimate of the flux. As stated in the main text, however, for comparison to dilute systems we wish to ignore time spent in three-stranded complexes. Therefore these simulations were ended early, and the time spent bound to the toehold was disregarded. These binding events are not useless, however: the frequency with which they occurred can be compared to the predictions of FFS. The number of unsolicited binding events (9 in total for $\sim 0.63\,\mathrm{ms}$ of total simulation time for toeholds of length 4, 5, 6 and 7) is consistent with the predictions of FFS ($\sim 15\,\mathrm{ms}^{-1}$), suggesting that the implementation of FFS is reliable.

Simulations of association allowing non-native base pairs. We measured the rate at which two strands, corresponding to a full invader with a toehold of six bases and the 15-base toehold domain of the substrate (see Table 5.7) formed the intended 6-base-pair toehold, both with and with-

Target interface	Toehold length			
	0	1	2	3
λ^0_{-1}	9581 (181 µs)	10400 (201 µs)	10532 (191 µs)	9851 (175 µs)
λ^1_0	10000 / 4031	10000 / 4406	10000 / 4187	10000 / 4264
λ^2_1	10000 / 4607	10000 / 4716	10000 / 4588	10000 / 4627
λ^3_2	10000 / 2304	10000 / 2468	10000 / 2635	10000 / 2721
λ^4_3	10000 / 1415	10000 / 1973	10000 / 2436	10000 / 2405
λ^5_4	50000 / 179	30000 / 745	25000 / 728	20000 / 765
λ^6_5	30000 / 526	15000 / 1877	3000 / 942	4500 / 1036
λ^7_6	60000 / 289	50000 / 315	16001 / 462	2000 / 261
λ^8_7	10000 / 1025	5000 / 325	2500 / 319	200 / 50
λ^9_8	300 / 149	300 / 131	300 / 92	
flux and error / s^{-1}	4.97×10^{-3} 6.5×10^{-4}	0.292 3.4×10^{-3}	7.06 0.89	210 31

Target interface	Toehold length			
	4	5	6	7
λ^0_{-1}	10179 (168 µs)	9859 (163 µs)	9617 (158 µs)	9148 (143 µs)
λ^1_0	10000 / 4316	10000 / 4497	10000 / 4229	10000 / 4327
λ^2_1	10000 / 4632	10000 / 4778	10000 / 4642	10000 / 4768
λ^3_2	10000 / 2732	10000 / 2920	10000 / 2999	10000 / 2966
λ^4_3	10000 / 2575	10000 / 2785	10000 / 3151	10000 / 3227
λ^5_4	20000 / 797	10000 / 408	10000 / 490	10000 / 490
λ^6_5	3000 / 767	3000 / 967	1000 / 289	1000 / 290
λ^7_6	200 / 51	199 / 177	100 / 100	100 / 100
flux and error / s^{-1}	2.22×10^3 300	1.24×10^4 840	1.61×10^4 1.2×10^3	1.81×10^4 1.4×10^3

Table 5.10: Trajectories attempted/successful at each stage of FFS for all toehold lengths. For λ^0_{-1}, values are given for the total number of trajectories crossing λ^0_{-1} (for the first time since leaving $Q = -2$), and the total simulation time taken. The interfce corresponding to the formation of the first base pair is highlighted in yellow, and the interface corresponding to the formation of the full toehold in green (when this is distinct from the first bp).

out non-native interactions between the strands (intrastrand hairpins were allowed in both cases). The order parameters used, which are very similar to those used in the initial stages of the displacement simulations, and the results, are given in Table 5.11. 20 initial flux simulations were initialized for 0.86μ s each before sampling was started, and simulations that reached $Q = 7$ were restarted to avoid including the time taken to resolve a misbonded complex in the flux calculation (as such a time would be negligible compared to diffusional time scales at the low concentrations relevant to experiment). Errors are estimated analogously to those for displacement.

As can be seen from Table 5.11, although the rate of formation of structures with six base pairs is much faster when non-native interactions are present, many of these structures involve misbonds that subsequently melt. The overall result is that non-native interactions accelerate binding by only 50%, suggesting that the systematic effect of ignoring non-native base pairs on the relative rates of displacement for different toehold lengths is small.

Free energy profiles of displacement. Free energy profiles of displacement were sampled using VMMC simulations of the three strands at a temperature $T = 25°$C. In this work, the trial moves used to generate VMMC clusters were:

- Rotation of a nucleotide about its backbone site, with the axis chosen from a uniform random distribution and the angle from a normal distribution with mean of zero and a standard deviation of 0.12 radians.

- Translation of a nucleotide with the direction chosen from a uniform random distribution and the distance from a normal distribution with mean of zero and a standard deviation of $1.02\,Å$.

These trial moves are illustrated in Figure 5.12. The variances are chosen from experience to provide efficient sampling. Umbrella sampling was performed using a biasing of the system according to the number of base pairs between the substrate and the incumbent strand, and the substrate and the invading strand. The umbrella biasing potential forbade complete detachment of any strand. States were recorded in histograms according to the intact base pairs between invading and substrate strands closest to the 3′ end of the substrate. Further, states were divided between coaxially stacked and unstacked macrostates according to the configurations of the nucleotides between the final base to which the invading strand was paired and the first base to which the incumbent was bound. If any of these substrate bases was not stacked with its neighbour, the

Order parameter Q	Definition	Native interactions only	Non-native interactions permitted
$Q = -2$	$d > 5.11$		
$Q = -1$	$5.11 \geq d > 3.42$	Crossings of λ^0_{-1} (simulation time / μs)	
$Q = 0$	$3.42 \geq d > 2.56$	8195 (69.3)	8145 (66.9)
		Attempts/successes at reaching λ^Q_{Q-1} from λ^{Q-1}_{Q-2}	
$Q = 1$	$2.56 \geq d > 1.71$	10000/4327	10000/4416
$Q = 2$	$1.71 \geq d > 0.85$	10000/5186	10000/5167
$Q = 3$	$0.85 \geq d \;\&\; n = 0$	10000/5333	10000/5152
$Q = 4$	$n \geq 1 \;\&\; b = 0 \;\&\; c = 0$	10000/968	10000/5333
$Q = 5$	$b = 1 \;\&\; c = 0$	12000/852	12000/2094
$Q = 6$	$b \geq 2 \;\&\; 6 > c \geq 1$	7500/3468	12000/3213
$Q = 7$	$c \geq 6 \;[\&\; c_{\mathrm{nat}} < c\,]$	2500/1879	5850/2251
$Q = 8$	$c \geq 6 \;\&\; c_{\mathrm{nat}} = c$		277/104
		Overall flux (and error) / s^{-1}	
		3.83×10^4 (1.57×10^3)	5.15×10^4 (3.28×10^3)

Table 5.11: Order parameter definitions FFS simulations of association for an invader with a 6-base toehold and the 15-base toehold domain of the substrate, and simulation results. The separation d is the minimum distance between hydrogen-bonding sites over *all* pairs of bases in the invading and substrate strands. n is the number of nearly formed base pairs, as defined in the text. b is the number of base pairs with energy more negative than -1.43 kcal/mol, and c the number of base pairs with energy more negative than -1.79 kcal/mol. c_{nat} is the number of those base pairs which are native (i.e., the intended toehold base pairs). In the native-only simulations, $c_{\mathrm{nat}} = c$ by definition, and $Q = 7$ is the maximal value in this case.

Molecule	Sequence
Hairpin	5′-(TTTTTTTTTT) [GTACATCTGAAG] TTTTTT [CTTCAGATGTAC] TACCGT{AG}-3′
ssDNA	5′- {CT}ACGGTA(TTTTTTTTTT) -3′

Table 5.12: Sequences used in simulations of the hybridization of a strand to the single-stranded overhang of a hairpin to measure the effect of ssDNA overhangs on melting equilibria. Bases in brackets "()" represent the overhangs, which are not present in every simulation. Bases enclosed by "{}" are only used in simulations of the 8-base hybridization. Square brackets "[]" enclose the stem of the sequence.

state was deemed to be coaxially unstacked. For the purpose of these simulations, an interaction energy (hydrogen-bonding or stacking energy) more negative than $-0.60\,\text{kcal}\,\text{mol}^{-1}$ was counted as an instance of hydrogen-bonding or stacking, as appropriate. Final data was obtained from 10 simulations of 4×10^{10} attempted VMMC moves.

Duplex formation. To measure the effect of ssDNA overhangs on melting equilibria, simulations of the hybridization of a strand to the single-stranded overhang of a hairpin were performed. Dangling poly(dT) overhangs of various lengths were included to indicate the destabilizing effect of excess ssDNA at a junction. Hairpin stems of length 12, with a loop of length 6, were used. Complementary sections of length 6 and 8 were used to allow accurate estimates of the destabilization at 25°C and 55°C. The sequences used are provided in Table 5.12.

In this case, all complementary base pairs were allowed to form hydrogen bonds. VMMC simulations were performed in periodic simulation cells of volume 1.67×10^{-20}l at a range of temperatures in the vicinity of 25°C and 55°C. Histogram reweighting [221] was used to infer the results at the desired temperatures. 4 simulations of 4×10^{10} steps each were performed in a periodic cubic cell of volume 3.96×10^{20} l for each system, with attempted moves identical to those used for the displacement landscape (except that rotations were drawn from a distribution with a standard deviation of 0.2 radians and displacements from a distribution with standard deviation 1.7 Å) . Umbrella sampling as a function of the number of interstrand base pairs was used to accelerate sampling. The ratio of bound to unbound states in simulations, ϕ, was recorded (any state with at least one interstrand hydrogen-bond more negative than $-0.60\,\text{kcal}\,\text{mol}^{-1}$ was counted as bound). The free-energy of formation follows as $\Delta G = -RT \ln \phi$: values for various combinations of ssDNA overhangs are reported in Table 2.2. For each system, free energies inferred from the 4 separate simulations were consistent to within $\sim 0.1\,\text{kcal}\,\text{mol}^{-1}$.

5.2.7 Notes on 1D Landscape Models

A sequence-dependent free energy landscape for RNA strand displacement. Figure 5.14 shows the sequence-dependent free energy landscape for strand displacement with a 10-base toehold at 25°C predicted by *efn2* [207, 208]. From this we infer that their model accounts for junction-context only when the junction is part of an 'interior multiloop'; during branch migration, the overhangs at the junction form an 'exterior loop', for which it appears that the coaxial stacking terms are not calculated.

Comparing IEL predictions to a prior phenomenological model. We present below the formulae derived by Zhang and Winfree [147] using their phenomenological model of reversible toehold exchange, restricted to the irreversible strand displacement case. We use their equations 1 and 2, in the limit where the 0-length "reverse toehold" dissociation rate is arbitrarily large, as further described in their Figure 8. Denoting their effective bimolecular rate constant for strand displacement by $k_{\mathrm{eff}}^{\mathrm{ZW}}(h)$, we have

$$k_{\mathrm{eff}}^{\mathrm{ZW}}(h) \approx \frac{k_f \, k_b}{k_{\mathrm{rev}}(h) + k_b} \tag{5.81}$$

where

$$k_{\mathrm{rev}}(h) \approx \frac{2 \, k_f}{b} e^{-|\Delta G^\circ(h)|/RT} \tag{5.82}$$

and

$$k_b \approx \frac{400}{b^2}. \tag{5.83}$$

$k_f \approx 3 \times 10^6$ /M/s is their fitted rate constant for toehold hybridization, k_b is their fitted rate constant for crossing the "half-way point" of branch migration, $k_{\mathrm{rev}}(h)$ is the calculated unimolecular rate constant for toehold dissociation, b is the length of the branch migration domain, and $|\Delta G^\circ(h)|$ is the absolute free energy of binding between the toehold and its complement.

We compared the dependence of strand displacement kinetics on branch migration domain length as predicted by IEL(5.3, 2.0) and the phenomenological model of Zhang and Winfree [147] (Figure 5.15). Numerically, predictions of their phenomenological model are quite consistent with the IEL's predictions in the cases we examined.
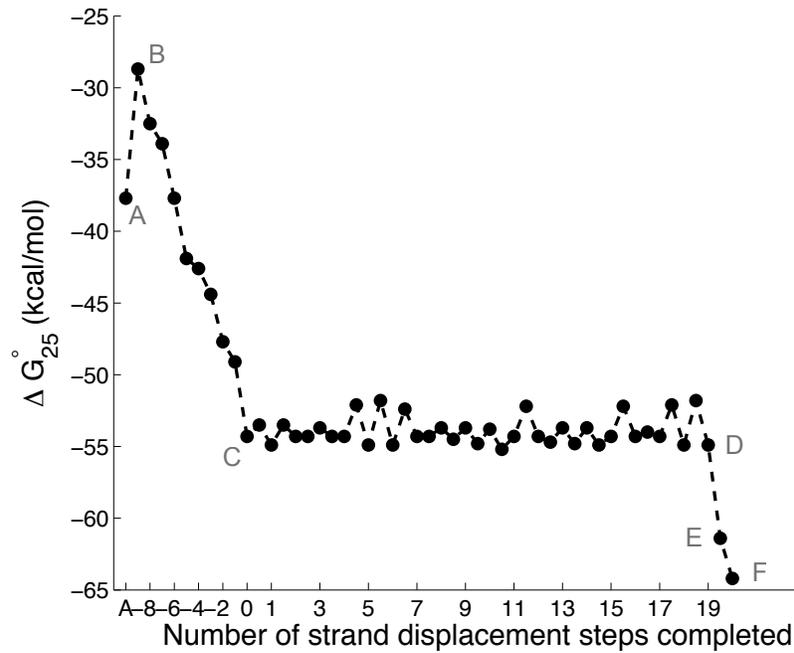
Figure 5.14: The sequence-dependent free energy landscape of strand displacement for a 10-base toehold at 25°C predicted by *efn2* for RNA molecules. States A–F correspond to those in the IEL analysis, Figure 3 of the main text.
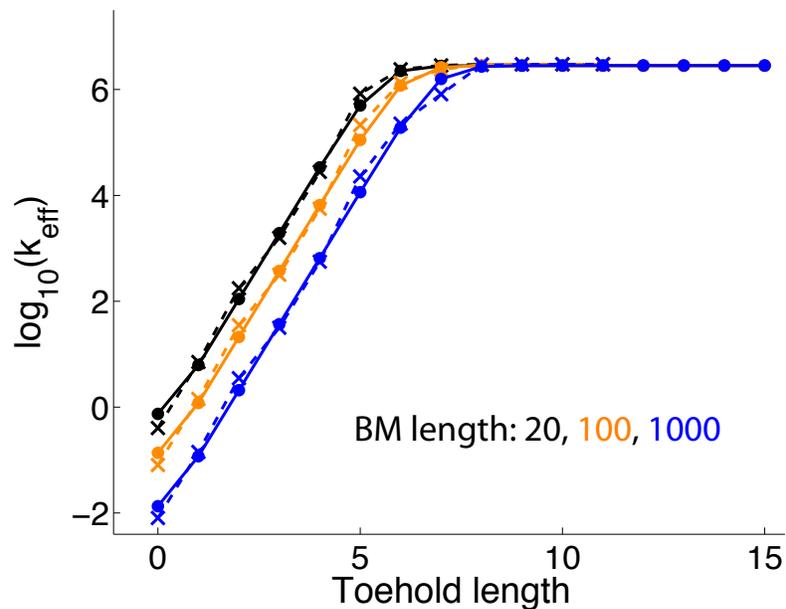
Figure 5.15: Predicted dependence of k_{eff} on toehold length for various lengths of the branch migration domain. IEL(5.3, 2.0)'s predictions are marked by filled circles and solid lines, while predictions of the phenomenological model of Zhang and Winfree [147] are indicated by crosses and dashed lines.

5.3 Appendix to Chapters 3 and 4

5.3.1 Materials and Methods

DNA oligonucleotides. All DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT). Oligonucleotides of length less than 60 bases were ordered with HPLC purification, while those 60 bases or longer were ordered with IE-HPLC purification. Where applicable, fluorophores and quenchers were attached by IDT as well.

Buffer conditions. All DNA oligonucleotides were stored at 4 °C in TE buffer (10 mM Tris.HCl pH balanced to 8.0, with 1 mM EDTA.Na$_2$, purchased as 100x stock from Sigma-Aldrich). We define two buffer conditions. First, "TE/Mg^{++}", which was prepared by adding TE buffer containing 62.5 mM MgCl$_2$ in a ratio of 1:4 to the sample, thereby achieving a final MgCl$_2$ concentration of 12.5 mM, out of which 1 mM is bound to EDTA. Second, "TE/Na$^+$", which was prepared by adding the appropriate quantity of dry NaCl salt (purchased 99% pure from EM Science, lot number 43076317) to a given volume of TE buffer to achieve a final NaCl concentration of 0.5 M. All buffer solutions were pH adjusted to 8.0 and filtered with a 1 micron filter (Nalgene rapid-flow).

All spectrofluorimetry experiments with Designs 3 and 4 reported here were performed in TE/Na$^+$ buffer. Spectrofluorimetry experiments with earlier Designs 1 and 2 were performed in TE/Mg^{++} buffer.

Quantitation of single strands. Single strands were quantitated from absorbance at 260 nm (measured using a Thermo Scientific NanoDrop cuvette-free spectrophotometer) using calculated extinction coefficients [219]. After thorough vortexing to ensure homogeneity in concentration, 3 samples of 2 μL each were typically used to measure absorbance. Two readings were taken from each sample and all data points were averaged. Typically, readings were within 2-5% of each other.

Annealing protocol. All annealing steps in this study were identical and were performed with an Eppendorf Mastercycler Gradient thermocycler. The samples were first heated up to 95 °C and then slowly cooled to 20 °C at the constant rate of 1 °C/min.

Annealing Reporters. All reporter complexes were annealed with a 20% excess of top strand (which is labeled with quencher in each case). For experiments with Designs 3 and 4, reporters were annealed in TE/Na$^+$ buffer. For experiments with Designs 1 and 2, reporters were annealed in TE/Mg^{++} buffer.

Reporter complexes were not gel purified after annealing. Reporter complexes were annealed

to have a bottom strand concentration of 5 μM, which also determines nominal concentration of reporter since the top strand was added in excess. This procedure was chosen because of two reasons. First, accurate quantification of bottom strands leads to accurate estimates of the concentration of reporter complex. This is important since the total concentration of reporter complex is used for normalization of spectrofluorimetry data. In contrast, quantifying the concentration of multistranded complexes is relatively less accurate because of larger errors in estimating extinction coefficients. Second, since reporter top strands have no toehold domains and are modified with quenchers, the excess addition of top strands ensures that all bottom strands form complexes, even with somewhat imperfect stoichiometry. This mitigates the chances of any active single-stranded DNA being present and ensures a stable fluorescence baseline.

Annealing and purification of multistranded fuel species. Each multistranded fuel species (React and Produce species) were prepared as follows. First, 5 nanomoles of the bottom strand was annealed with 6 nanomoles (20% excess) of each of the top strands in TE/Mg^{++} buffer (at an approximate concentration of 25 μM).

After annealing, complexes were purified by nondenaturing (ND) polyacrylamide gel electrophoresis (PAGE) by running the samples on 12% gel at 150 V for approximately 6 hours. The purpose of the purification was (i) to remove the excess top strands that were added and (ii) remove multimers of the desired complexes that form due to the high concentrations in the annealing step. The acrylamide (19:1 acrylamide:bis) was diluted from 40% acrylamide stock purchased from Ambion. ND loading dye containing xylene cyanol FF in 50% glycerol was added to the React species, resulting in a final glycerol concentration of 10% by volume. Since the Produce species contain a quencher, 50% glycerol was added directly (rather than the ND loading dye) to achieve a final gycerol concentration of 10% by volume.

For Designs 3 and 4, the appropriate bands were cut out and eluted in 1 mL of TE/Na$^+$ buffer for 18-24 hours. For Designs 1 and 2, the elution was done in TE/Mg^{++} buffer instead.

Dialysis of multistranded fuel species. For Designs 3 and 4, since the experiments were to be performed in TE/Na$^+$ buffer, a further reduction (approximately 2500 fold) in Mg^{++} concentration was achieved using 2 rounds of dialysis. Each round of dialysis is expected to achieve a reduction of approximately 50 fold, since 1 ml of purified multistranded fuel species was dialyzed with approximately 50 ml of TE/Na$^+$ buffer for 2 hours using a 2 ml Thermo Scientific Slide-A-Lyzer MINI dialysis device with a 10K MWCO membrane.

Quantitation of multistranded fuel species. The procedure for quantitating multistranded fuel species is essentially identical to the procedure for single strands, except for the calculation of extinction coefficients, which involves corrections for hyperchromicity [219]. We expect the inferred concentrations to be less accurate because of larger uncertainties in the estimated extinction coefficients. Typical yields after purification ranged from 40% to 60%.

Experimental protocols for spectrofluorimetry. Spectrufluorimetry experiments were performed using a commercial SPEX Fluorolog-3 from Horiba, equipped with a water bath temperature controller. All spectrofluorimetry experiments were performed at 25 °C unless otherwise mentioned. Synthetic round-top quartz cuvettes (119-004F) from Hellma, with a total volume 1.6 ml, were used.

Prior to each experiment, all cuvettes were cleaned thoroughly. Each cuvette was washed 10 times in Milli-Q water, twice in 70% ethanol, and finally another 5 times in Milli-Q water. After washing, cuvettes were dried by gently tapping them on a Kimtech Science wipe placed on paper towels for cushioning. They were subsequently left to air-dry for about one hour. Cuvette caps were washed once thoroughly with Milli-Q water, once with 70% ethanol, and once again with Milli-Q water. They were then dried with Kimtech Science wipe and left to air-dry for an hour. After adding the sample, the exterior of the cuvette was washed with the same procedure as the caps.

For experiments involving the ROX fluorophore, excitation was at 584 nm, while emissions were at 602 nm. For experiments involving the Alexa-488 fluorophore, excitation was at 492 nm, while emissions were at 517 nm. For experiments involving the Alexa-647 fluorophore, excitation was at 650 nm, while emissions were at 670 nm. Band pass value of 2 nm was used for both excitation and emission monochrometers for all experiments except those measuring individual strand displacement and toehold exchange rate constants, for which a bandpass of 4 nm was used. All experiments were done with integration time of 10 seconds for each data point.

For experiments involving one fluorophore, measurements were taken every minute; for those with multiple fluorophores, measurement interval increased proportionally because each excitation/emission channel was allotted 1 minute for measurement.

Reporter	Measured rate constant (/M /s)
k_{RepA}	7.4×10^5
k_{RepB}	1.7×10^6
k_{RepC}	1.2×10^6
$k_{\mathrm{RepFluxABr}}$	2.2×10^6
$k_{\mathrm{RepFluxBCj}}$	3.1×10^6
$k_{\mathrm{RepFluxCAp}}$	7.7×10^5
$k_{\mathrm{RepBackBA}}$	6.8×10^5
$k_{\mathrm{RepBackCB}}$	1.0×10^6
$k_{\mathrm{RepBackACi2}}$	1.1×10^5

Table 5.13: Measured rate constants (all in /M /s) for reporters for signal strands, flux strands, and back strands, respectively. Molecular design diagrams at the sequence level are shown in Figures 5.24 - 5.31. Note that since Rep A, Rep B, and Rep C have also been used as thresholds in the autocatalytic delay element and Displacillator experiments, rates for those steps are identical to the threshold rates listed in Table 4.3.

5.3.2 Measuring individual strand displacement and toehold exchange rates

Figures 5.33 - 5.50 provide molecular design diagrams at the sequence level for each desired pathway for which the rate constant has been measured. The measured values for rate constants of those pathways are provided in Table 4.2. Molecular design diagrams at the sequence level for "reporter" steps are provided in Figures 5.24 - 5.32. Rate constants were also characterized for these reporter steps; these are presented in Table 5.13.

5.3.3 Sequences from Designs 1, 2, 3 and 4

DNA Sequences from Design 1 (Table 5.14), Design 2 (Table 5.15), Design 3 (Table 5.16) and Design 4 (Tables 5.17 and 5.18) are provided. Molecular design diagrams at the sequence level are provided for Design 4 in Figures 5.16 - 5.23. Table 5.19 provides sequences for reporter species used for characterizing the rates of individual strand displacement and toehold exchange reactions in-

volved in desired pathways for Design 4.
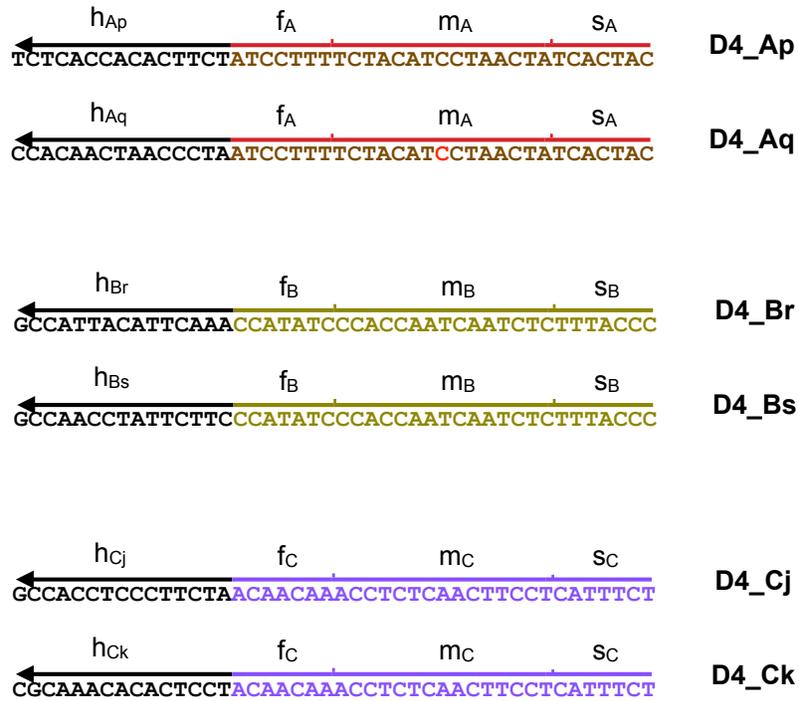
Figure 5.16: Molecular design diagrams at the sequence level for signal strands Ap, Aq, Br, Bs, Cj, and Ck in Design 4.

Strand name	Sequence
D1_React_BOT_CBCj	AGTGGGTTAGTAGAGAGTTGTTAGTGGGAAATGGGAATGTTGTGAGGAATGAGAGGGTAT
D1_Back_CB	CTCATTCCTCACAACATTCCCATTTCCCA
D1_Flux_BCj	CTAACAACTCTCTACTAACCCACTTCATACCTTATCC
D1_Produce_BOT_BCjCk	AGAGGGTATGAAGGTGTAAGAAGGAGGGTATGGATAAGGTATGAAGTGGGTTA
D1_Cj	CTTCATACCTTATCCATACCCTCTCATTCCTCACAACATTCCCA
D1_Ck	CCTTCTTACACCTTCATACCCTCTCATTCCTCACAACATTCCCA
D1_Helper_CCk	ATACCCTCCTTCTTACACCTTCATACCCT
D1_React_BOT_BABr	TGAGGGTTAGAGGTTTGAAGAGTGAGGGATTTGGGTTAGTAGAGAGTTGTTAGTGGGAAA
D1_Back_BA	CTAACAACTCTCTACTAACCCAAATCCCT
D1_Flux_ABr	CACTCTTCAAACCTCTAACCCTCATTCAAATCTCACC
D1_Produce_BOT_ABrBs	AGTGGGAAAGGAGAGAATGAATGGTGGGAAAGGTGAGATTTGAATGAGGGTTA
D1_Br	CATTCAAATCTCACCTTTCCCACTAACAACTCTCTACTAACCCA
D1_Bs	CCATTCATTCTCTCCTTTCCCACTAACAACTCTCTACTAACCCA
D1_Helper_BBs	TTTCCCACCATTCATTCTCTCCTTTCCCA
D1_React_BOT_ACAp	TGTGGGAATGTTGTGAGGAATGAGAGGGTATAGGGTTAGAGGTTTGAAGAGTGAGGGATT
D1_Back_AC	CACTCTTCAAACCTCTAACCCTATACCCT
D1_Flux_CAp	CTCATTCCTCACAACATTCCCACACAATACTATCATC
D1_Produce_BOT_CApAq	TGAGGGATTGTGTTTGAGTTTAGGAGGGATTGATGATAGTATTGTGTGGGAAT
D1_Ap	CACAATACTATCATCAATCCCTCACTCTTCAAACCTCTAACCCT
D1_Aq	CCTAAACTCAAACACAATCCCTCACTCTTCAAACCTCTAACCCT
D1_Helper_AAq	AATCCCTCCTAAACTCAAACACAATCCCT

Table 5.14: DNA sequences from Design 1.

Strand name	Sequence
D2_React_BOT_CBCj	TCGGGTAAAGAGATTGATTGGTGGGATATGGAGAAATGAGGAAGTTGAGAGGCTTGTTGT
D2_Back_CB	GCCTCTCAACTTCCTCATTTCTCCATATC
D2_Flux_BCj	CCACCAATCAATCTCTTTACCCGACACCTCCCTTCTA
D2_Produce_BOT_BCjCk	GCTTGTTGTAGGAGTGTGTTTGCGTTGTTGTTAGAAGGGAGGTGTCGGGTAAA
D2_Cj	GACACCTCCCTTCTAACAACAAGCCTCTCAACTTCCTCATTTCT
D2_Ck	CGCAAACACACTCCTACAACAAGCCTCTCAACTTCCTCATTTCT
D2_Helper_CCk	ACAACAACGCAAACACACTCCTACAACAA
D2_React_BOT_BABr	CCGTAGTGATAGTTAGTATGTACCAAAGGATGGGTAAAGAGATTGATTGGTGGGATATGG
D2_Back_BA	CCACCAATCAATCTCTTTACCCATCCTTT
D2_Flux_ABr	GGTACATACTAACTATCACTACGGCATTACATTCAAA
D2_Produce_BOT_ABrBs	GGGATATGGGAAGAATAGGTTGCCGATATGGTTTGAATGTAATGCCGTAGTGA
D2_Br	GGCATTACATTCAAACCATATCCCACCAATCAATCTCTTTACCC
D2_Bs	GGCAACCTATTCTTCCCATATCCCACCAATCAATCTCTTTACCC
D2_Helper_BBs	CCATATCGGCAACCTATTCTTCCCATATC
D2_React_BOT_ACAp	CCAGAAATGAGGAAGTTGAGAGGCTTGTTGTGTAGTGATAGTTAGTATGTACCAAAGGAT
D2_Back_AC	GGTACATACTAACTATCACTACACAACAA
D2_Flux_CAp	GCCTCTCAACTTCCTCATTTCTGGTCACCACACTTCT
D2_Produce_BOT_CApAq	CCAAAGGATTAGGGTTAGTTGTGGAAAGGATAGAAGTGTGGTGACCAGAAATG
D2_Ap	GGTCACCACACTTCTATCCTTTGGTACATACTAACTATCACTAC
D2_Aq	CCACAACTAACCCTAATCCTTTGGTACATACTAACTATCACTAC
D2_Helper_AAq	ATCCTTTCCACAACTAACCCTAATCCTTT

Table 5.15: DNA sequences from Design 2.

Strand name	Sequence
D3_React_BOT_CBCj	TGTTGTTCGGAGAGTTGAAGGAGTAAAGAGGTATAGGGTGGTTAGTTAGAGAAATGGGCT
D3_Back_CB	CTATACCTCTTTACTCCTTCAACTCTCCG
D3_Flux_BCj	ATCTTCCCTCCACAGCCCATTTCTCTAACTAACCACC
D3_Produce_BOT_BCjCk	AAATGGGCTGTGGAGGGAAGATTGTTGTTGCGTTTGTGTGAGGATGTTGTTCG
D3_Cj	TCTTTACTCCTTCAACTCTCCGAACAACAATCTTCCCTCCACAG
D3_Ck	TCTTTACTCCTTCAACTCTCCGAACAACATCCTCACACAAACGC
D3_Helper_CCk	AACAACATCCTCACACAAACGCAACAACA
D3_React_BOT_BABr	GGTATAGGGTGGTTAGTTAGAGAAATGGGTAGGAAACCATGTATGATTGATAGTGATGCC
D3_Back_BA	TTTCCTACCCATTTCTCTAACTAACCACC
D3_Flux_ABr	AAACTTACATTACGGCATCACTATCAATCATACATGG
D3_Produce_BOT_ABrBs	AGTGATGCCGTAATGTAAGTTTGGTATAGCCGTTGGATAAGAAGGGTATAGGG
D3_Br	CCCATTTCTCTAACTAACCACCCTATACCAAACTTACATTACGG
D3_Bs	CCCATTTCTCTAACTAACCACCCTATACCCTTCTTATCCAACGG
D3_Helper_BBs	CTATACCCTTCTTATCCAACGGCTATACC
D3_React_BOT_ACAp	TAGGAAACCATGTATGATTGATAGTGATGTGTTGTTCGGAGAGTTGAAGGAGTAAAGACC
D3_Back_AC	AACAACACATCACTATCAATCATACATGG
D3_Flux_CAp	TCTTCACACCACTGGTCTTTACTCCTTCAACTCTCCG
D3_Produce_BOT_CApAq	GTAAAGACCAGTGGTGTGAAGATAGGAAAGGTGTTGATTGGGATTAGGAAACC
D3_Ap	CATCACTATCAATCATACATGGTTTCCTATCTTCACACCACTGG
D3_Aq	CATCACTATCAATCATACATGGTTTCCTAATCCCAATCAACACC
D3_Helper_AAq	TTTCCTAATCCCAATCAACACCTTTCCTA

Table 5.16: DNA sequences from Design 3.

Strand name	Sequence
D4_React_BOT_CBCj	TGTTGTTTGGAGAGTTGAAGGAGTAAAGAGGTATAGGGTGGTTAGTTAGAGAAATGGGCG
D4_Back_CB	CTATACCTCTTTACTCCTTCAACTCTCCA
D4_Flux_BCj	ATCTTCCCTCCACCGCCCATTTCTCTAACTAACCACC
D4_Produce_BOT_BCjCk	AAATGGGCGGTGGAGGGAAGATTGTTGTTGCGTTTGTGTGAGGATGTTGTTTG/3IAbRQSp/
D4_Cj	TCTTTACTCCTTCAACTCTCCAAACAACAATCTTCCCTCCACCG
D4_Ck	TCTTTACTCCTTCAACTCTCCAAACAACATCCTCACACAAACGC
D4_Helper_CCk	/56-ROXN/AACAACATCCTCACACAAACGCAACAACA
D4_Cat_Helper_CCk	/56-ROXN/AACAACATCCTCACACAAACGCAACAACAATCTTCCCTCCACCG
D4_React_BOT_BABr	GGTATAGGGTGGTTAGTTAGAGAAATGGGTAGGAAAAGATGTAGGATTGATAGTGATGCG
D4_Back_BA	TTTCCTACCCATTTCTCTAACTAACCACC
D4_Flux_ABr	AAACTTACATTACCGCATCACTATCAATCCTACATCT
D4_Produce_BOT_ABrBs	AGTGATGCGGTAATGTAAGTTTGGTATAGCGGTTGGATAAGAAGGGTATAGGG/3IAbRQSp/
D4_Br	CCCATTTCTCTAACTAACCACCCTATACCAAACTTACATTACCG
D4_Bs	CCCATTTCTCTAACTAACCACCCTATACCCTTCTTATCCAACCG
D4_Helper_BBs	/5Alex647N/CTATACCCTTCTTATCCAACCGCTATACC
D4_Cat_Helper_BBs	/5Alex647N/CTATACCCTTCTTATCCAACCGCTATACCAAACTTACATTACCG
D4_React_BOT_ACApi2	TAGGAAAAGATGTAGGATTGATAGTGATGTTGTTTGGAGAGTTGAAGGAGTAAAGAAG
D4_Back_ACi2	AACAACATCACTATCAATCCTACATCT
D4_Flux_CAp	TCTTCACACCACTCTTCTTTACTCCTTCAACTCTCCA
D4_Produce_BOT_CApAq	GTAAAGAAGAGTGGTGTGAAGATAGGAAAGGTGTTGATTGGGATTAGGAAAAG/3IABkFQ/
D4_Ap	CATCACTATCAATCCTACATCTTTTCCTATCTTCACACCACTCT
D4_Aq	CATCACTATCAATCCTACATCTTTTCCTAATCCCAATCAACACC
D4_Helper_AAq	/5Alex488N/TTTCCTAATCCCAATCAACACCTTTCCTA
D4_Cat_Helper_AAq	/5Alex488N/TTTCCTAATCCCAATCAACACCTTTCCTATCTTCACACCACTCT

Table 5.17: DNA sequences from Design 4.

Strand name	Sequence
D4_Rep_BOT_C	TGTTGTTTGGAGAGTTGAAGGAGTAAAGA/3AlexF488N/
D4_Rep_TOP_B	/5IAbRQ/CCCATTTCTCTAACTAACCACC
D4_Rep_BOT_B	GGTATAGGGTGGTTAGTTAGAGAAATGGG/3Rox_N/
D4_Rep_TOP_A	/5IAbRQ/CATCACTATCAATCCTACATCT
D4_Rep_BOT_A	TAGGAAAAGATGTAGGATTGATAGTGATG/3AlexF647N/
D4_Helper_CCk†	AACAACATCCTCACACAAACGCAACAACA
D4_Cat_Helper_CCk†	AACAACATCCTCACACAAACGCAACAACAATCTTCCCTCCACCG
D4_Helper_BBs†	CTATACCCTTCTTATCCAACCGCTATACC
D4_Cat_Helper_BBs†	CTATACCCTTCTTATCCAACCGCTATACCAAACTTACATTACCG
D4_Helper_AAq†	TTTCCTAATCCCAATCAACACCTTTCCTA
D4_Cat_Helper_AAq	TTTCCTAATCCCAATCAACACCTTTCCTATCTTCACACCACTCT
D4_Rep_TOP_C†	TCTTTACTCCTTCAACTCTCCA
D4_Rep_BOT_C†	TGTTGTTTGGAGAGTTGAAGGAGTAAAGA
D4_Rep_TOP_B†	CCCATTTCTCTAACTAACCACC
D4_Rep_BOT_B†	GGTATAGGGTGGTTAGTTAGAGAAATGGG
D4_Rep_TOP_A†	CATCACTATCAATCCTACATCT
D4_Rep_BOT_A†	TAGGAAAAGATGTAGGATTGATAGTGATG
D4_Rep_TOP_C	/5IABkFQ/TCTTTACTCCTTCAACTCTCCA
D4_QUE_Helper_CCk	TGTTGTTGCGTTTGTGTGAGGATGTTGTTTG/3IAbRQSp/
D4_QUE_Helper_BBs	GGTATAGCGGTTGGATAAGAAGGGTATAGGG/3IAbRQSp/
D4_QUE_Helper_AAq	TAGGAAAGGTGTTGATTGGGATTAGGAAAAG/3IABkFQ/

Table 5.18: Additional DNA sequences from Design 4. † signifies "plain version", without fluorophores or quenchers attached. Complexes Rep A, Rep B and Rep C, comprising the corresponding top and bottom strands, also act as thresholds.

Strand name	Sequence
D4_Rep_Back_CB_Top	TCTTTACTCCTTCAACTCTCCA/3IAbRQSp/
D4_Rep_Back_CB_Bot	/56-ROXN/TGGAGAGTTGAAGGAGTAAAGAGGTATAG
D4_Rep_Back_BA_Top	CCCATTTCTCTAACTAACCACC/3IAbRQSp/
D4_Rep_Back_BA_Bot	/56-ROXN/GGTGGTTAGTTAGAGAAATGGGTAGGAAA
D4_Rep_Back_ACi2_Top	CATCACTATCAATCCTACATCT/3IAbRQSp/
D4_Rep_Back_ACi2_Bot	/56-ROXN/AGATGTAGGATTGATAGTGATGTTGTT
D4_Rep_Flux_ABr_Top	/5IAbRQ/CAAACTTACATTACCG
D4_Rep_Flux_ABr_Bot	AGTGATGCGGTAATGTAAGTTTG/3Rox_N/
D4_Rep_Flux_BCj_Top	/5IAbRQ/CATCTTCCCTCCACCG
D4_Rep_Flux_BCj_Bot	AAATGGGCGGTGGAGGGAAGATG/3Rox_N
D4_Rep_Flux_BCj_Top	/5IAbRQ/CATCTTCCCTCCACCG
D4_Rep_Flux_BCj_Bot	AAATGGGCGGTGGAGGGAAGATG/3Rox_N
D4_Rep_Flux_CAp_Top	/5IAbRQ/CTCTTCACACCACTCT
D4_Rep_Flux_CAp_Bot	GTAAAGAAGAGTGGTGTGAAGAG/3Rox_N

Table 5.19: DNA sequences used for characterizing individual rate constants in Design 4.

Figure 5.17: Molecular design diagrams at the sequence level for the fuel species that mediate the React steps in Design 4.

Figure 5.18: Molecular design diagrams at the sequence level for the Flux species in Design 4.

Figure 5.19: Molecular design diagrams at the sequence level for the fuel species that mediate the Produce step for the module B + A → 2B in Design 4. Fluorophores and quenchers are as indicated. "A647" stands for the fluorophore "Alexa 647".

Figure 5.20: Molecular design diagrams at the sequence level for the fuel species that mediate the Produce step for the module $C + B \rightarrow 2C$ in Design 4. Fluorophores and quenchers are as indicated.

D4_Cat_Help_AAq

D4_Help_AAq

D4_Produce_CApAq

Figure 5.21: Molecular design diagrams at the sequence level for the fuel species that mediate the Produce step for the module A + C → 2A in Design 4. Fluorophores and quenchers are as indicated. "A488" stands for the fluorophore "Alexa 488".

D4_Que_Help_AAq

D4_Que_Help_BBs

D4_Que_Help_CCk

Figure 5.22: Molecular design diagrams at the sequence level for the species that are used for normalizing Helper readout. These species are added at the end of the experiment for quenching the remaining Helper species in Design 4. Quenchers are as indicated.

Figure 5.23: Molecular design diagrams at the sequence level for the reporter species used for reading out the concentrations of signal strands (left) and Flux strands (right) in Design 4. Rep A, B, and C are also used as thresholds in autocatalytic delay element and Displacillator experiments. Fluorophores and quenchers are as indicated. "A488" and "A647" stand for the fluorophores "Alexa 488" and "Alexa 647", respectively.

Figure 5.24: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.25: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.26: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.27: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.28: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.29: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.30: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.31: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.
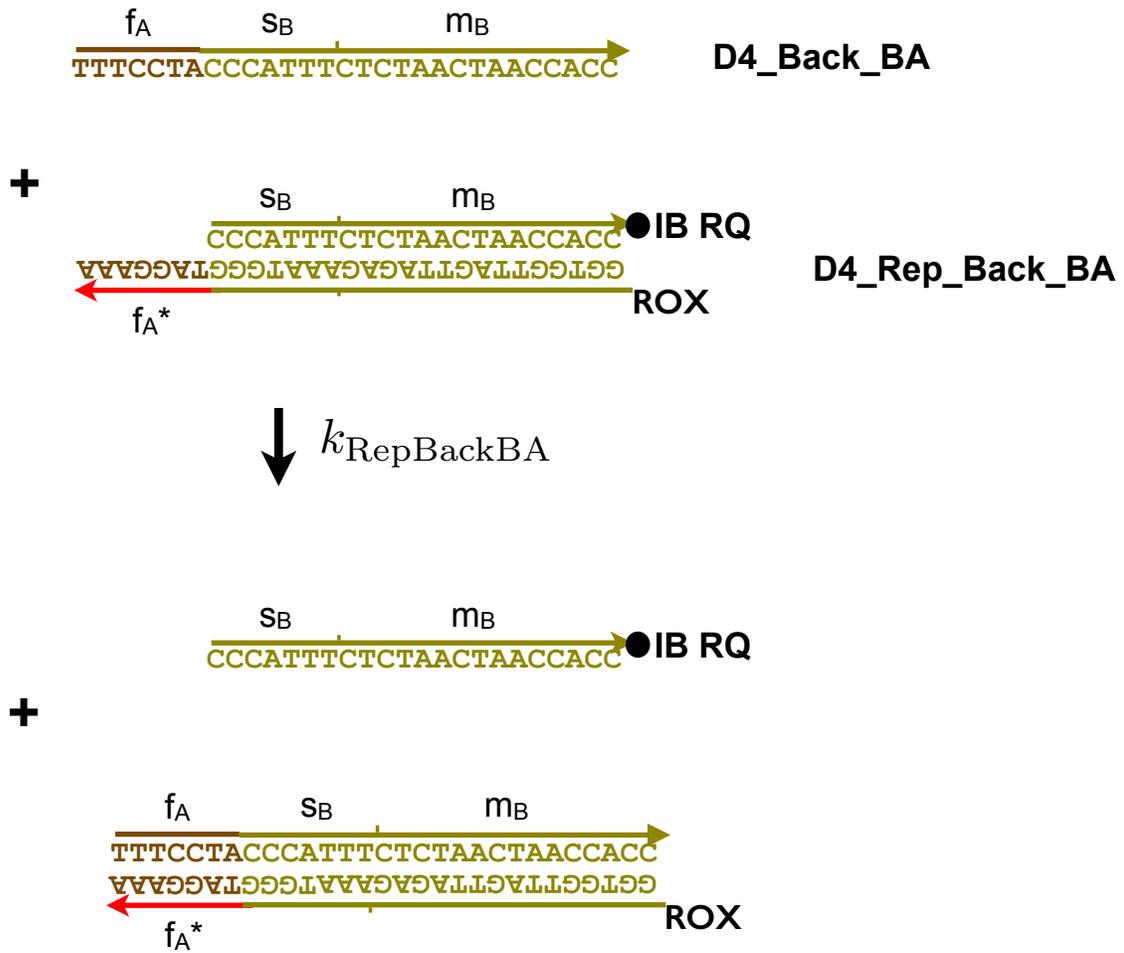
Figure 5.32: Molecular design diagram at the sequence level for reporters. The measured value for the bimolecular rate constant for this pathway is specified in Table 5.13.

Figure 5.33: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.34: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
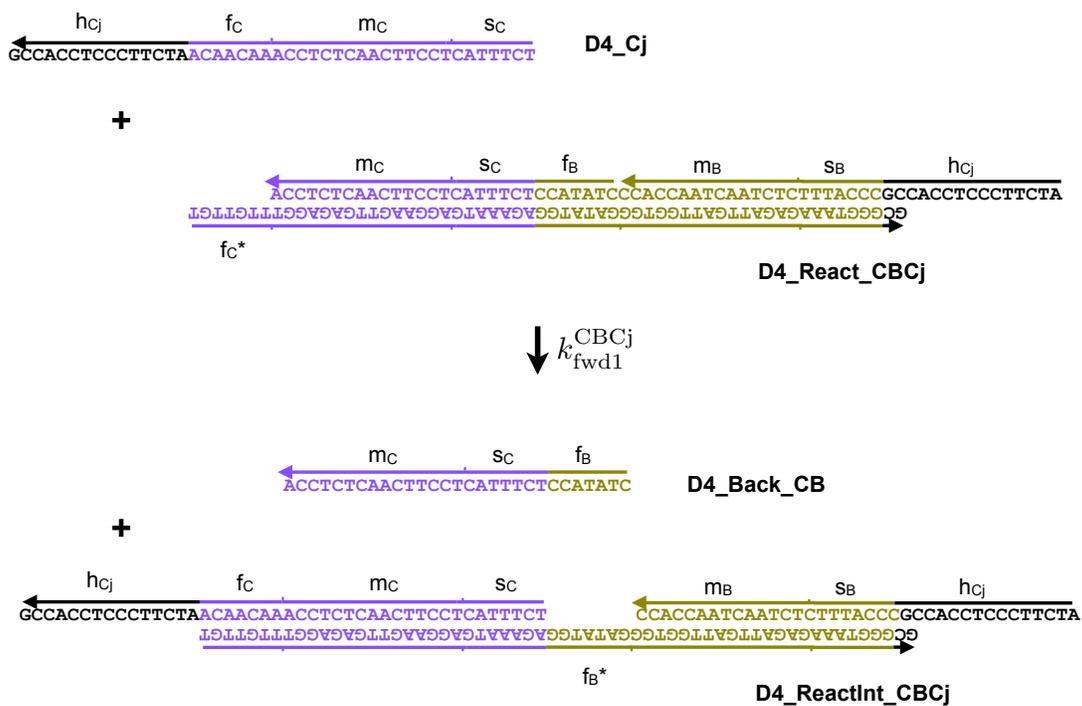
Figure 5.35: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.36: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
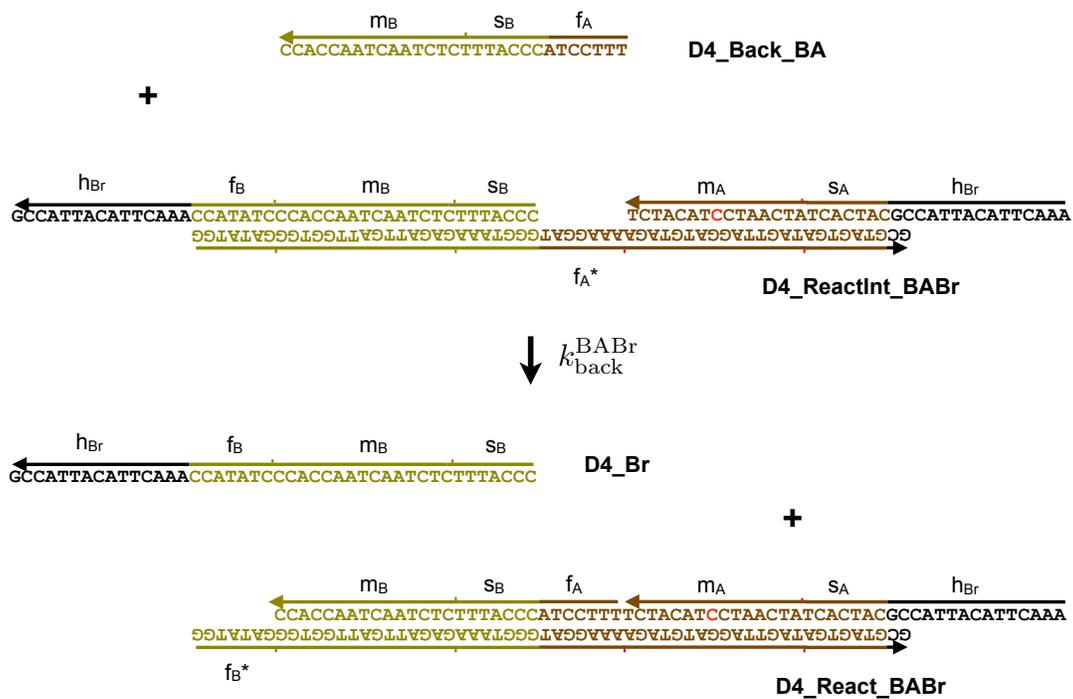
Figure 5.37: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.38: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
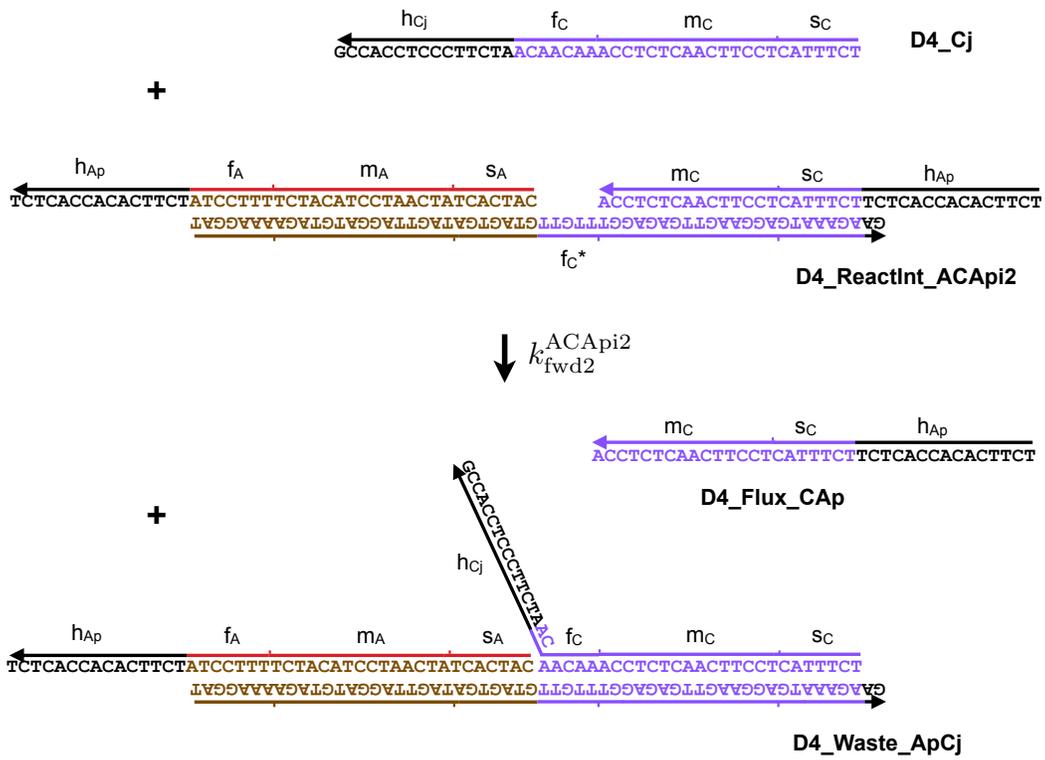
Figure 5.39: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
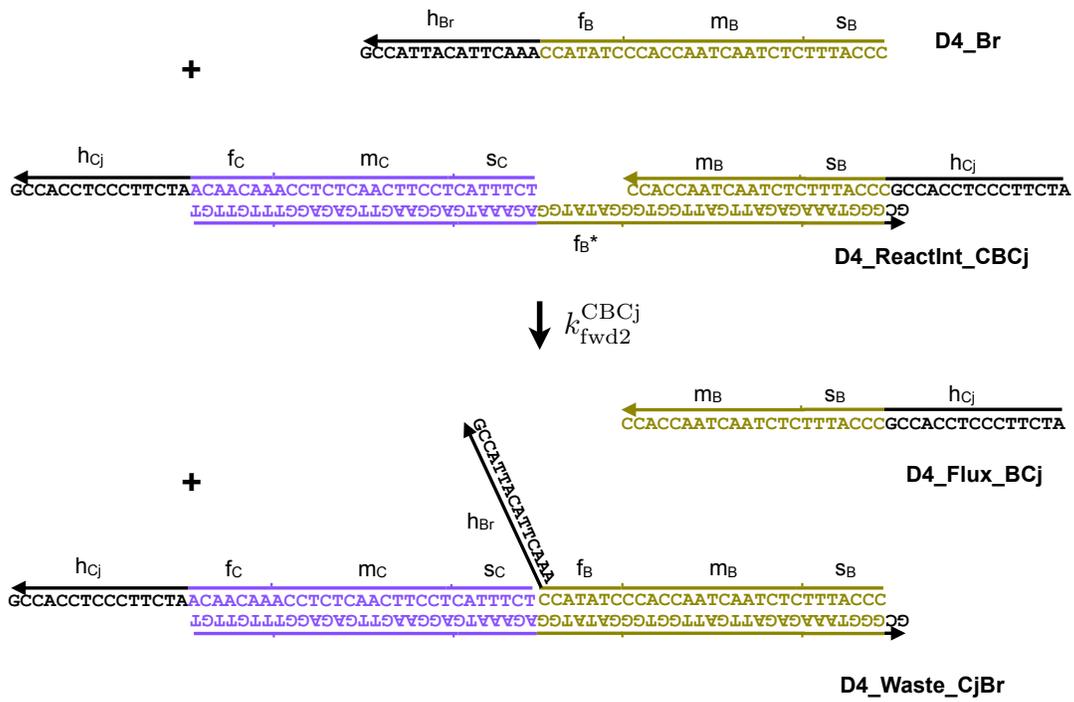
Figure 5.40: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.41: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.42: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.43: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.44: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.45: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
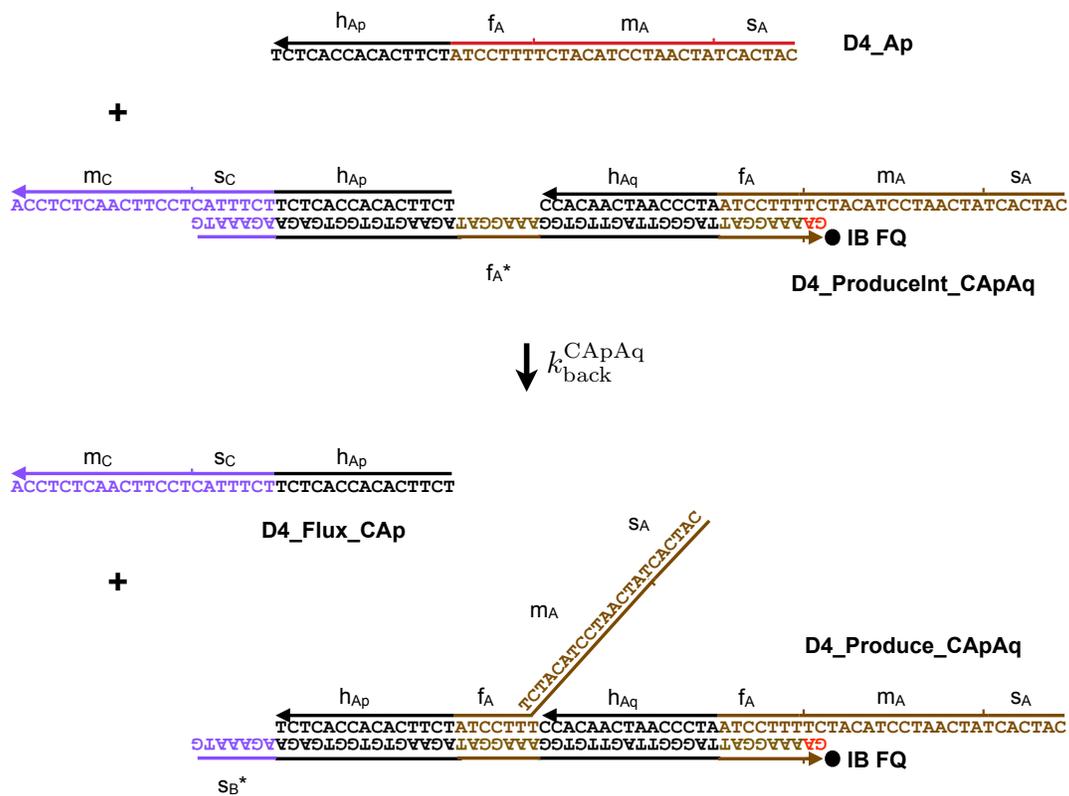
Figure 5.46: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
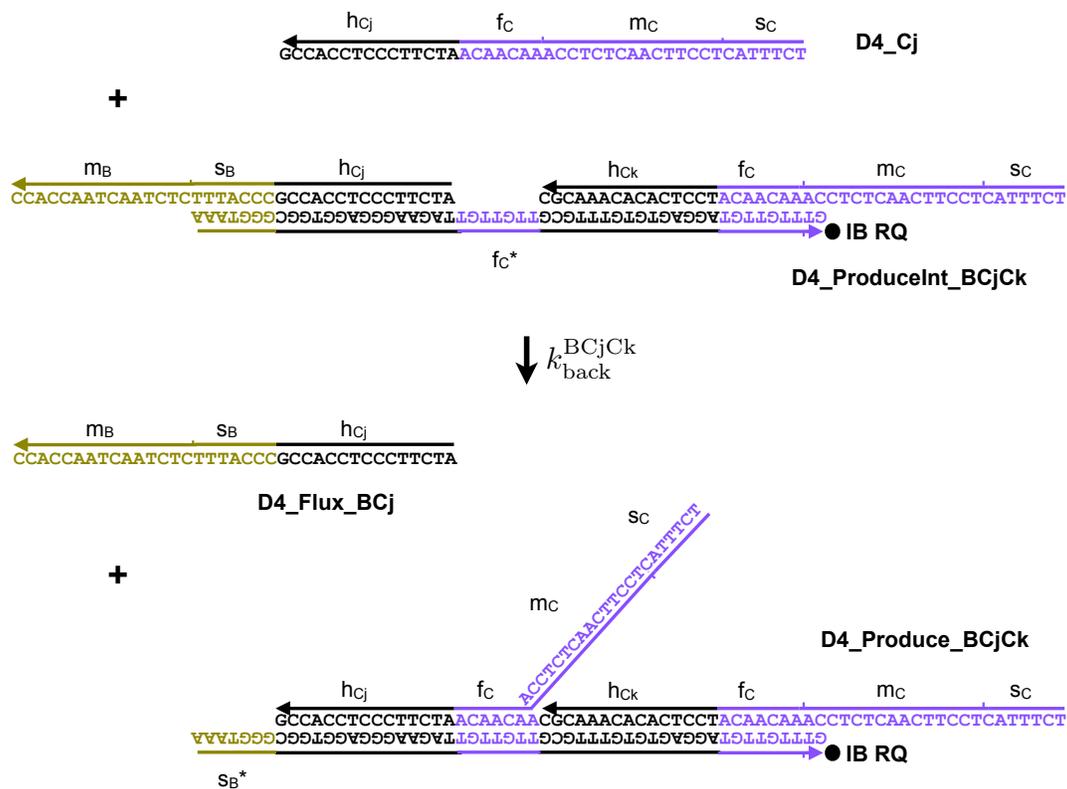
Figure 5.47: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.
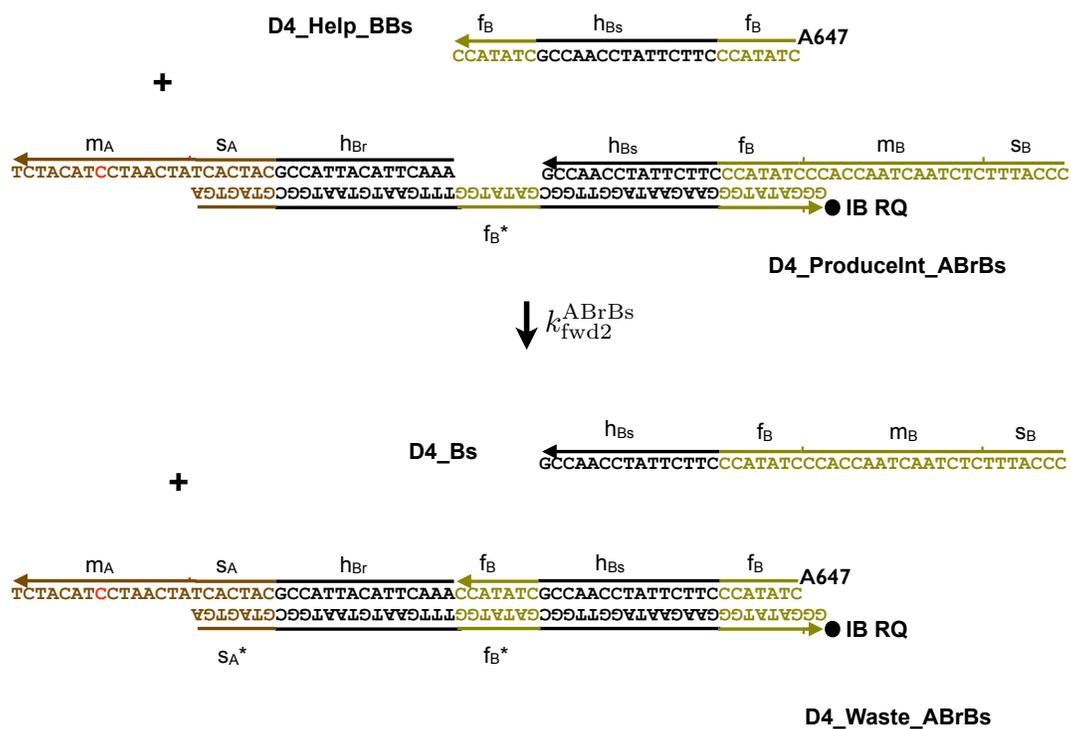
Figure 5.48: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.49: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Figure 5.50: Molecular design diagram at the sequence level for desired reaction pathways. The measured value for the bimolecular rate constant for this pathway is specified in Table 4.2.

Bibliography

[1] S. Weinberg, *Scientist: Four golden lessons*. Nature, 2003, vol. 426, no. 389.

[2] J. von Neumann, *The Theory of self reproducing automata*, A. Burks, Ed. Univ. of Illinois Press, 1966.

[3] A. M. Turing, "The chemical basis of morphogenesis," *Bull. of Math. Biol.*, vol. 52, no. 1/2, pp. 153–197, 1990.

[4] W. S. Klug, M. R. Cummings, and C. A. Spencer, *Concepts of Genetics*, 8th ed., 2005.

[5] D. de Solla Price, "Gears from the greeks. the antikythera mechanism: A calendar computer from ca. 80 b. c." *Transactions of the American Philosophical Society*, vol. 64, no. 7, pp. 1–70, 1974.

[6] T. Freeth, Y. Bitsakis, X. Moussas, J. H. Seiradakis, A. Tselikas, H. Mangou, M. Zafeiropoulou, R. Hadland, D. Bate, A. Ramsey, M. Allen, A. Crawley, P. Hockley, T. Malzbender, D. Gelb, W. Ambrisco, and M. G. Edmunds, "Decoding the ancient greek astronomical calculator known as the antikythera mechanism," *Nature*, vol. 444, pp. 587–591, 2006.

[7] T. Freeth, A. Jones, J. M. Steele, and Y. Bitsakis, "Calendars with olympiad display and eclipse prediction on the antikythera mechanism," *Nature*, vol. 454, pp. 614–617, 2008.

[8] D. Y. Zhang and G. Seelig, "Dynamic DNA nanotechnology using strand-displacement reactions." *Nature Chem.*, vol. 3, no. 2, pp. 103–113, 2011.

[9] Y. Krishnan and F. C. Simmel, "Nucleic acid based molecular devices," *Angew. Chem. Int. Ed.*, vol. 50, no. 14, pp. 3124–3156, 2011.

[10] A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan, "Challenges and opportunities for structural DNA nanotechnology," *Nat. Nano.*, vol. 6, no. 12, pp. 763–772, 2011.

[11] V. Linko and H. Dietz, "The enabled state of DNA nanotechnology," *Curr. Opin. Biotechnol.*, vol. 24, pp. 555–561, 2013.

[12] F. Zhang, J. Nangreave, Y. Liu, and H. Yan, "Structural DNA nanotechnology: State of the art and future perspective," *Journal of the American Chemical Society*, vol. 136, pp. 11 198–11 211, 2014.

[13] W. W. Grabow and L. Jaeger, "RNA self-assembly and RNA nanotechnology," *Accounts of Chemical Research*, vol. 47, pp. 1871–1880, 2014.

[14] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, pp. 114–117, April 1965.

[15] R. Carlson, "The changing economics of DNA synthesis," *Nature Biotechnology*, vol. 27, no. 12, pp. 1091–1094, 2009.

[16] T. Ando, "High-speed atomic force microscopy coming of age," *Nanotechnology*, vol. 23, no. 062001, 2012.

[17] W. C. Bray, "A periodic reaction in homogeneous solution and its relation to catalysis," *J. Am. Chem. Soc*, vol. 43, no. 6, pp. 1262–1267, 1921.

[18] A. T. Winfree, "The prehistory of the Belousov-Zhabotinsky oscillator," *Journal of Chemical Education*, vol. 61, no. 8, pp. 661–663, 1984.

[19] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell," *Current Opinion in Cell Biology*, vol. 15, pp. 221–231, 2003.

[20] (2009, March). [Online]. Available: http://commons.wikimedia.org/wiki/File:Peptide_syn.png

[21] J. B. Chang and J. E. Ferrell Jr., "Mitotic trigger waves and the spatial coordination of the Xenopus cell cycle," *Nature*, vol. 500, pp. 603–607, 2013.

[22] A. M. Zhabotinsky, "A history of chemical oscillations and waves," *Chaos*, vol. 1, no. 4, pp. 379–386, 1991.

[23] I. R. Epstein and K. Showalter, "Nonlinear chemical dynamics: Oscillations, patterns, and chaos," *J. Phys. Chem.*, vol. 100, pp. 13 132–13 147, 1996.

[24] I. R. Epstein and J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*.   Oxford Univ Press, London, 1998.

[25] F. Sagues and I. R. Epstein, "Nonlinear chemical dynamics," *Dalton Trans.*, pp. 1201–1217, 2003.

[26] B. P. Belousov, *Sb. Ref. Radiats. Med. za 1958, Medgiz, Moscow*, vol. 1, p. 145, 1959 (in Russian).

[27] A. M. Zhabotinsky, *Biofizika*, vol. 9, p. 306, 1964 (in Russian).

[28] E. N. Harvey, *A history of luminescence: From the earliest times until 1900*.   American Philosophical Society, 1957, vol. 44.

[29] G. Fechner, *Schweigg J. Chem. Phys*, vol. 53, p. 129, 1828.

[30] W. Ostwald, *Phys. Zeitsch.*, vol. 8, pp. 87–88, 1899.

[31] H. L. Heathcote, *Z. Phys. Chem. (English: Journal of Physical Chemistry)*, vol. 37, p. 368, 1901.

[32] G. Bredig and J. Weinmayr, *Z. Phys. Chem. (English: Journal of Physical Chemistry)*, vol. 42, p. 601, 1903.

[33] E. S. Hedges, *Nature*, vol. 128, p. 398, 1931.

[34] J. Keizer, P. A. Rock, and S. W. Lin, *J. Am. Chem. Soc*, vol. 101, p. 5637, 1979.

[35] O. M. R. F. O. Rice, "The thermal decomposition of hydrogen peroxide," *J. Phys. Chem.*, vol. 31, no. 9, pp. 1352–1356, 1927.

[36] R. J. Field, E. Körös, and R. M. Noyes, *J. Am. Chem. Soc*, vol. 94, p. 8649, 1972.

[37] D. Edelson, R. J. Field, and R. M. Noyes, *Int. J. Chem. Kinet.*, vol. 7, p. 417, 1975.

[38] L. Treindl and R. M. Noyes, *J. Phys. Chem.*, vol. 97, p. 11354, 1993.

[39] G. Laurenczy and M. T. Beck, *J. Phys. Chem.*, vol. 98, p. 5188, 1994.

[40] P. D. Kepper, K. Kustin, and I. R. Epstein, *J. Am. Chem. Soc*, vol. 103, p. 2133, 1981.

[41] M. Orbán and I. R. Epstein, *J. Am. Chem. Soc*, vol. 107, no. 2302, 1985.

[42] J. H. Jensen, *J. Am. Chem. Soc*, vol. 105, p. 2639, 1983.

[43] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, 2000.

[44] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, 2000.

[45] M. R. Atkinson, M. A. Savageau, J. T. Myers, and A. J. Ninfa, "Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli," *Cell*, vol. 113, 2003.

[46] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S. Lee, and J. C. Liao, "A synthetic gene-metabolic oscillator," *Nature*, vol. 435, 2005.

[47] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," *Nature*, vol. 456, 2008.

[48] D. Endy, "Foundations for engineering biology," *Nature*, vol. 438, no. 7067, pp. 449–453, 2005.

[49] A. L. Slusarczyk, A. Lin, and R. Weiss, "Foundations for the design and implementation of synthetic genetic circuits," *Nat. Rev. Genet.*, vol. 13, pp. 406–420, 2012.

[50] J. Ackermann, B. Wlotzka, and J. S. McCaskill, "In vitro DNA-based predator-prey system with oscillatory kinetics," *Bull. of Math. Biol.*, vol. 60, pp. 329–353, 1998.

[51] B. Wlotzka and J. S. McCaskill, "A molecular predator and its prey: coupled isothermal amplification of nucleic acids," *Chemistry & Biology*, vol. 4, no. 1, pp. 25–33, 1997.

[52] G. F. Joyce, "Forty years of in vitro evolution," *Angew. Chem. Int. Ed.*, vol. 46, pp. 6420–6436, 2007.

[53] J. Kim and E. Winfree, "Synthetic in vitro transcriptional oscillators," *Mol. Syst. Biol.*, vol. 7, no. 1, p. 465, 2011.

[54] J. Kim, J. J. Hopfield, and E. Winfree, "Neural network computation by *in vitro* transcriptional circuits," *Advances in Neural Information Processing Systems (NIPS)*, vol. 17, pp. 681–688, 2004.

[55] J. Kim, "In vitro synthetic transcriptional networks," Ph.D. dissertation, California Institute of Technology, 2007.

[56] J. Kim, K. S. White, and E. Winfree, "Construction of an *in vitro* bistable circuit from synthetic transcriptional switches," *Mol. Syst. Biol.*, vol. 2, no. 68, 2006.

[57] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Natl. Acad. Sci. U.S.A*, vol. 81, pp. 3088–3092, 1984.

[58] P. Subsoontorn, J. Kim, and E. Winfree, "Ensemble Bayesian analysis of bistability in a synthetic transcriptional switch," *ACS Synthetic Biology*, vol. 1, pp. 299–316, 2012.

[59] K. Montagne, R. Plasson, Y. Sakai, T. Fujii, and Y. Rondelez, "Programming an in vitro DNA oscillator using a molecular networking strategy," *Mol. Syst. Biol.*, vol. 7, no. 466, 2011.

[60] A. Baccouche, K. Montagne, A. Padirac, T. Fujii, and Y. Rondelez, "Dynamic DNA-toolbox reaction circuits: A walkthrough," *Methods*, vol. 67, pp. 234–249, 2014.

[61] A. Padirac, T. Fujii, and Y. Rondelez, "Bottom-up construction of in vitro switchable memories," *Proc. Natl. Acad. Sci. U.S.A*, vol. E3213, 2012.

[62] T. Fujii and Y. Rondelez, "Predator prey molecular ecosystems," *ACS Nano*, vol. 7, no. 1, pp. 27–34, 2013.

[63] H. W. H. van Roekel, L. H. H. Meijer, S. Masroor, Z. C. F. Garza, A. Estévez-Torres, Y. Rondelez, A. Zagaris, M. A. Peletier, P. A. J. Hilbers, and T. F. A. de Greef, "Automated design of programmable enzyme-driven DNA circuits," *ACS Synthetic Biology*, vol. DOI: 10.1021/sb500300d, 2014.

[64] N. Aubert, C. Mosca, T. Fujii, M. Hagiya, and Y. Rondelez, "Computer-assisted design for scaling up systems based on DNA reaction networks," *J. R. Soc. Interface*, vol. 11, no. 20131167, 2014.

[65] A. Padirac, T. Fujii, and Y. Rondelez, "Nucleic acids for the rational design of reaction circuits," *Curr. Opin. Biotechnol.*, vol. 24, pp. 575–580, 2013.

[66] A. J. Genot, T. Fujii, and Y. Rondelez, "In vitro regulatory models for systems biology," *Biotechnology Advances*, vol. 31, no. 6, pp. 789–796, 2013.

[67] E. Franco, J. Kim, and F. C. Simmel, "Transcriptional oscillators," in *Multiscale Analysis and Nonlinear Dynamics: From Genes to the Brain*. Wiley-VCH Verlag GmbH & Co. KGaA, 2013.

[68] M. Weitz and F. C. Simmel, "Synthetic in vitro transcription circuits," *Transcription*, vol. 3, no. 2, pp. 87–91, doi: 10.4161/trns.19734 2012.

[69] J. Kim, "Synthetic networks," in *Automation in Proteomics and Genomics*. John Wiley & Sons, Ltd., New York, 2009, pp. 251–271.

[70] D. T. Gillespie, "The chemical Langevin equation," *J. Chem. Phys.*, vol. 113, no. 1, pp. 297–306, 2000.

[71] D. T. Gillespie, "The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction," *J. Phys. Chem. A*, vol. 106, pp. 5063–5071, 2002.

[72] D. A. Mcquarrie, "Stochastic approach to chemical kinetics," *J. Appl. Prob.*, vol. 4, pp. 413–478, 1967.

[73] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2361, 1977.

[74] N. G. van Kampen, *Stochastic processes in physics and chemistry*, revised edition ed. Elsevier, 1997.

[75] I. Oppenheim, K. E. Shuler, and G. H. Weiss, "Stochastic and deterministic formulation of chemical rate equations," *J. Chem. Phys.*, vol. 50, no. 460, p. doi: 10.1063/1.1670820, 1969.

[76] T. G. Kurtz, "The relationship between stochastic and deterministic models for chemical reactions," *J. Chem. Phys.*, vol. 57, no. 7, pp. 2976–2978, 1972.

[77] J. Esparza and M. Nielsen, "Decidability issues for Petri Nets - a survey," *J. Info. Proc. and Cybernet.*, vol. 3, pp. 143–160, 1994.

[78] R. M. Karp and R. E. Miller, "Parallel program schemata," *J. Comput. Sys. Sci.*, vol. 3, no. 4, pp. 147–195, 1969.

[79] J. Aspnes and E. Ruppert, *An Introduction to Population Protocols*, ser. Middleware for Network Eccentric and Mobile Applications, B. Garbinato, H. Miranda, and L. Rodrigues, Eds. Springer-Verlag, 2009.

[80] M. Cook, D. Soloveichik, E. Winfree, and J. Bruck, *Programmability of Chemical Reaction Networks*, A. Condon, D. Harel, J. N. Kok, A. Salomaa, and E. Winfree, Eds. Springer Berlin Heidelberg, 2009.

[81] D. Soloveichik, M. Cook, E. Winfree, and J. Bruck, "Computation with finite stochastic chemical reaction networks," *Nat Comput*, vol. 7, no. DOI 10.1007/s11047-008-9067-y, pp. 615–633, 2008.

[82] M. Minsky, *Computation: Finite and Infinite machines*. Prentice-Hall, NJ, 1967.

[83] D. Angluin, J. Aspnes, and D. Eisenstat, "Stably computable predicates are semilinear," in *PODC 2006: Proceedings of the twenty-fifth annual ACM symposium on principles of distributed computing*, New York, NY, USA, 2006, pp. 292–299.

[84] P. L. Houston, *Chemical Kinetics and Reaction Dynamics*, 1st ed. McGraw-Hill, 2001.

[85] M. Feinberg, "Chemical reaction network structure and the stability of complex isothermal reactors - i. the deficiency zeor and deficiency one theorems," *Chem. Engg. Sci.*, vol. 42, no. 10, pp. 2229–2268, 1987.

[86] M. Feinberg, "Lectures on chemical reaction networks," *Notes of lectures given at the Mathematics Research Center, University of Wisconsin*, 1979.

[87] M. Korzuhin, *Oscillatory Processes in Biological and Chemical Systems*. Nauka, Moscow (in Russian), 1967.

[88] W. Klonowski, "Simplifying principles for chemical and enzyme reaction kinetics," *Biophysical chemistry*, vol. 18, pp. 73–87, 1983.

[89] E. N. Kerner, "Universal formats for nonlinear ordinary differential systems," *J. Math. Phys.*, vol. 22, pp. 1366–1371, 1981.

[90] K. Kowalski, "Universal formats for nonlinear dynamical systems," *Chem. Phys. Lett.*, vol. 209, no. 1, pp. 167–170, 1993.

[91] T. Wilhelm, "Chemical systems consisting only of elementary steps – a paradigma for nonlinear behavior," *J. Math. Chem.*, vol. 27, pp. 71–88, 2000.

[92] N. Samardzija, L. D. Greller, and E. Wasserman, "Nonlinear chemical kinetic schemes derived from mechanical and electrical dynamical systems," *J. Chem. Phys.*, vol. 90, no. 2296, p. doi: 10.1063/1.455970, 1989.

[93] H. T. Davis, *Introduction to nonilnear differential and integral equations*. Dover, NY, 1962.

[94] N. Minorsky, *Nonlinear Oscillations*. Kreiger, NY, 1974.

[95] E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, vol. 20, p. 130, 1963.

[96] O. E. Rössler, "An equation for continuous chaos," *Phys. Lett. A*, vol. 57, p. 397, 1976.

[97] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, pp. 115–133, 1943.

[98] J. von Neumann, *Automata Studies*. Princeton Univ. Press, Princeton, 1956, pp. 43–98.

[99] A. Hjelmfelt, E. D. Weinberger, and J. Ross, "Chemical implementation of neural networks and Turing machines," *Proc. Natl. Acad. Sci. U.S.A*, vol. 88, pp. 10 983–10 987, 1991.

[100] A. Hjelmfelt, E. D. Weinberger, and J. Ross, "Chemical implementation of finite-state machiens," *Proc. Natl. Acad. Sci. U.S.A*, vol. 89, pp. 383–387, 1992.

[101] A. H. amd J. Ross, "Chemical implementation and thermodynamics of collective neural networks," *Proc. Natl. Acad. Sci. U.S.A*, vol. 89, pp. 388–391, 1992.

[102] V. Hárs and J. Tóth, *On the inverse problem of reaction kinetics*, ser. Qualitative Theory of Differential Equations, M. Farkas, Ed. Colloquia Mathematica Societatis János Bolyai, (Szeged, Hungary, 1979), 1981, vol. 30.

[103] H. W. Engl, C. Flamm, P. Kʹugler, J. Lu, S. Mʹuller, and P. Schuster, "Inverse problems in systems biology," *Inverse Problems*, vol. 25, no. 123014, p. 51, 2009.

[104] M. N. Stojanovic, S. Semova, D. Kolpashchikov, J. Macdonald, C. Morgan, and D. Stefanovic, "Deoxyribozyme-based ligase logic gates and their initial circuits," *J. Am. Chem. Soc.*, vol. 127, no. 19, pp. 6914–6915, 2005.

[105] G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree, "Enzyme-free nucleic acid logic circuits," *Science*, vol. 314, no. 5805, pp. 1585–1588, 2006.

[106] J. Elbaz, O. Lioubashevski, F. Wang, F. Remacle, R. D. Levine, and I. Willner, "DNA computing circuits using libraries of DNAzyme subunits," *Nat. Nano.*, vol. 5, no. 6, pp. 417–422, 2010.

[107] L. Qian and E. Winfree, "Scaling up digital circuit computation with DNA strand displacement cascades," *Science*, vol. 332, no. 6034, pp. 1196–1201, 2011.

[108] A. J. Genot, J. Bath, and A. J. Turberfield, "Reversible logic circuits made of DNA," *J. Am. Chem. Soc*, vol. 133, no. 50, pp. 20 080–20 083, 2011.

[109] W. Li, Y. Yang, H. Yan, and Y. Liu, "3-input majoriy logic gate and multiple input logic circuit based on DNA strand displacement," *Nano Letters*, vol. 13, pp. 2980–2988, 2013.

[110] D. Soloveichik, G. Seelig, and E. Winfree, "DNA as a universal substrate for chemical kinetics," *Proc. Natl. Acad. Sci. U.S.A*, vol. 107, no. 12, pp. 5393–5398, 2010.

[111] L. Cardelli, "Two domain DNA strand displacement," *Mathematical Structures in Computer Science*, vol. 23, no. 02, pp. 247–271, 2013.

[112] Y.-J. Chen, N. Dalchau, N. Srinivas, A. Phillips, L. Cardelli, D. Soloveichik, and G. Seelig, "Programmable chemical controllers made from DNA," *Nature Nanotech.*, vol. 8, p. 755, 2013.

[113] R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu, "Synthetic analog computation in living cells," *Nature*, vol. 497, no. 7451, pp. 619–623, 2013.

[114] R. Sarpeshkar, "Analog synthetic biology," *Phil. Trans. R. Soc. A*, vol. 372:20130110, 2014.

[115] K. E. Drexler, "Molecular engineering:an approach to the development of general capabilities for molecular manipulation," *Proc. Natl. Acad. Sci. U.S.A*, vol. 78, no. 9, pp. 5275–5278, 1981.

[116] L. Cardelli, *Abstract Machines of Systems Biology*. Transactions on Computational Systems Biology III, LNBI, Springer., 2005, vol. 3737, pp. 145–168.

[117] J. SantaLucia, Jr., H. T. Allawi, and P. A. Seneviratne, "Improved nearest-neighbor parameters for predicting DNA duplex stability," *Biochemistry*, vol. 35, no. 11, pp. 3555–3562, 1996.

[118] J. SantaLucia, Jr., "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics." *Proc. Natl. Acad. Sci. U.S.A*, vol. 17, no. 95(4), pp. 1460–5, 1998.

[119] J. SantaLucia, Jr. and D. Hicks, "The thermodynamics of DNA structural motifs," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 33, pp. 415–40, 2004.

[120] R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree, and N. A. Pierce, "Thermodynamic analysis of interacting nucleic acid strands," *SIAM Review*, vol. 49, no. 1, pp. 65–88, 2007.

[121] P. J. Hagerman, "Flexibility of DNA," *Annu. Rev. Biophys. Biophys. Chem.*, vol. 17, pp. 265–286, 1988.

[122] R. M. Dirks and N. A. Pierce, "A partition function algorithm for nucleic acid secondary structure including pseudoknots," *J Comput. Chem.*, vol. 24, no. 13, pp. 1664–1677, 2003.

[123] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce., "NUPACK: analysis and design of nucleic acid systems." *J Comput. Chem.*, vol. 32, pp. 170–173, 2011.

[124] E. Winfree, "Unpublished software for designing DNA sequences based on sequence symmetry minimization."

[125] N. Seeman, "De novo design of sequences for nucleic acid structural engineering," *J. Biomol. Struct. Dyn.*, vol. 8, no. 3, pp. 573–81, 1990.

[126] D. Y. Zhang, "Toward domain-based sequence design for DNA strand displacement reactions," *DNA Computing and Molecular Programming*, pp. 162–175, 2011.

[127] R. M. Dirks, M. Lin, E. Winfree, and N. A. Pierce, "Paradigms for computational nucleic acid design," *Nucl. Acids Res.*, vol. 32, no. 4, pp. 1392–1403, 2004.

[128] J. N. Zadeh, B. R. Wolfe, and N. A. Pierce, "Nucleic acid sequence design via efficient ensemble defect optimization," *J Comput. Chem.*, vol. 32, pp. 439–452, 2011.

[129] B. R. Wolfe and N. A. Pierce, "Sequence design for a test tube of interacting nucleic acid strands," *ACS Synthetic Biology*, vol. doi: 10.1021/sb5002196, 2014.

[130] N. C. Seeman, "Nucleic acid junctions and lattices," *J. Theor. Biol.*, vol. 99, pp. 237–247, 1982.

[131] J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. Sha, P. E. Constantinou, S. L. Ginell, C. Mao, and N. C. Seeman, "From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal," *Nature*, vol. 461, p. 74, 77 2009.

[132] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," *Nature*, vol. 440, no. 7082, pp. 297–302, 2006.

[133] H. Dietz and W. M. Shih, "Folding DNA into twisted and curved nanoscale shapes," *Science*, vol. 325, no. 5941, pp. 725–730, 2009.

[134] B. Wei, M. Dai, and P. Yin, "Complex shapes self-assembled from single-stranded DNA tiles," *Nature*, vol. 485, pp. 623–627, 2012.

[135] N. C. Seeman, "An overview of structural DNA nanotechnology," *Mol. Biotech.*, vol. 37, no. 3, pp. 246–257, 2007.

[136] C. Geary, P. W. K. Rothemund, and E. S. Andersen, "A single-stranded architecture for co-transcriptional folding of RNA nanostructures," *Science*, vol. 345, no. 6198, pp. 799–804, 2014.

[137] C. Geary and E. S. Andersen, "Design principles for single-stranded RNA origami structures," *DNA Computing and Molecular Programming*, 2014.

[138] P. Guo, "The emerging field of RNA nanotechnology," *Nat. Nanotechnol.*, vol. 5, pp. 833–842, 2010.

[139] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, 1994.

[140] D. Doty, "Theory of algorithmic self-assembly," *Communications of the ACM*, vol. 55, no. 12, pp. 78–88, 2012.

[141] M. N. Stojanovic and D. Stefanovic, "A deoxyribozyme-based molecular automaton," *Nat. Biotech.*, vol. 21, no. 9, pp. 1069–74, Sep 2003.

[142] P. W. K. Rothemund, N. Papadakis, and E. Winfree, "Algorithmic self-assembly of DNA Sierpinski triangles," *PLoS Biol*, vol. 2, no. 12, p. e424, Dec 2004.

[143] D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree, "Engineering entropy-driven reactions and networks catlayzed by DNA," *Science*, vol. 318, pp. 1121–1125, 2007.

[144] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce, "Programming biomolecular self-assembly pathways," *Nature*, vol. 451, no. 7176, pp. 318–22, Jan 2008.

[145] B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. Neumann, "A DNA-fueled molecular machine made of DNA," *Nature*, vol. 406, pp. 605–608, 2000.

[146] B. Yurke and A. Mills, "Using DNA to power nanostructures," *Genet. Prog. Evol. Mach.*, vol. 4, pp. 111–122, 2003.

[147] D. Zhang and E. Winfree, "Control of DNA strand displacement kinetics using toehold exchange," *J. Am. Chem. Soc.*, vol. 131, no. 47, pp. 17 303–17 314, 2009.

[148] N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, and E. Winfree, "On the biophysics and kinetics of toehold-mediated DNA strand displacement," *Nucl. Acids Res.*, vol. 41, no. 22, pp. 10 641–10 658, 2013.

[149] X. Chen, N. Briggs, J. R. McLain, and A. D. Ellington, "Stacking nonenzymatic circuits for high signal gain," *Proc. Natl. Acad. Sci. U.S.A*, vol. 110, no. 14, pp. 5386–5391, 2013.

[150] A. Phillips and L. Cardelli, "A programming language for composable DNA circuits," *J. R. Soc. Interface*, vol. 6, pp. S419–36, 2009.

[151] M. Lakin and A. Phillips, "Modelling, simulating and verifying turing-powerful strand displacement systems," *DNA Computing and Molecular Programming LNCS*, vol. 6937, pp. 130–144, 2011.

[152] M. Lakin, S. Youssef, F. Polo, S. Emmott, and A. Phillips, "Visual DSD: a design and analysis tool for DNA strand displacement systems," *Bioinformatics*, vol. 27, no. 22, pp. 3211–3213, 2011.

[153] J. M. Schaeffer, "Stochastic simulation of the kinetics of multiple interacting nucleic acid strands," Ph.D. dissertation, California Institute of Technology, 2013.

[154] T. E. Ouldridge, "Coarse-grained modelling of DNA and DNA nanotechnology," Ph.D. dissertation, University of Oxford, 2011.

[155] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Structural, mechanical and thermodynamic properties of a coarse-grained DNA model," *J. Chem. Phys.*, vol. 134, no. 8, p. 085101, 2011.

[156] N. R. Kallenbach, R.-I. Ma, and N. C. Seeman, "An immobile nucleic acid junction constructed from oligonucleotides," *Nature*, vol. 305, no. 5937, pp. 829–831, 1983.

[157] E. Winfree, F. R. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional DNA crystals," *Nature*, vol. 394, p. 539, 1998.

[158] Y. He, T. Ye, M. Su, C. Zhang, A. Ribbe, W. Jiang, and C. Mao, "Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra," *Nature*, vol. 452, pp. 198–201, 2008.

[159] R. Penchovsky and R. R. Breaker, "Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes," *Nat. Biotech.*, vol. 23, pp. 1424–1433, 2005.

[160] P. Yin, H. M. Choi, C. R. Calvert, and N. A. Pierce, "Programming biomolecular self-assembly pathways," *Nature*, vol. 451, pp. 318–323, 2008.

[161] R. M. Dirks and N. A. Pierce, "Triggered amplification by hybridization chain reaction," *Proc. Natl. Acad. Sci. U.S.A*, vol. 101, no. 43, pp. 15 275–15 278, 2004.

[162] R. Pei, S. K. Taylor, D. Stefanovic, S. Rudchenko, T. E. Mitchell, and M. N. Stojanovic, "Behavior of polycatalytic assemblies in a substrate-displaying matrix," *J. Am. Chem. Soc.*, vol. 128, no. 39, pp. 12 693–12 699, 2006.

[163] S. J. Green, J. Bath, and A. J. Turberfield, "Coordinated chemomechanical cycles: a mechanism for autonomous molecular motion," *Phys. Rev. Lett.*, vol. 101, no. 23, p. 238101, 2008.

[164] J. Bath and A. J. Turberfield, "DNA nanomachines," *Nat. Nanotechnol.*, vol. 2, pp. 275–284, 2007.

[165] C. M. Radding, K. L. Beattie, W. K. Holloman, and R. C. Wiegand, "Uptake of homologous single-stranded fragments by superhelical DNA," *J. Mol. Biol.*, vol. 116, pp. 825–839, 1977.

[166] C. Green and C. Tibbetts, "Reassociation rate limited displacement of DNA strands by branch migration," *Nucl. Acids Res.*, vol. 9, no. 8, pp. 1905–1918, 1981.

[167] L. E. Morrison and L. M. Stols, "Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution," *Biochemistry*, pp. 3095–3104, 1993.

[168] J. G. Wetmur and N. Davidson, "Kinetics of renaturation of DNA," *J. Mol. Biol.*, vol. 31, pp. 349–370, 1968.

[169] W. Feller, *An introduction to probability theory and its applications*. New York: Wiley and Sons, 1968, vol. 1.

[170] L. P. Reynaldo, A. V. Vologodskii, B. P. Neri, and V. I. Lyamichev, "The kinetics of oligonucleotide replacements," *J. Mol. Biol.*, vol. 297, pp. 511–520, 2000.

[171] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "DNA nanotweezers studied with a coarse-grained model of DNA," *Phys. Rev. Lett.*, vol. 104, p. 178101, 2010.

[172] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculation by fast computing machines." *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.

[173] G. Rudolph, "The fundamental matrix of the general random walk with absorbing boundaries." Collaborative Research Center "Computational Intelligence" CI-75, University of Dortmund, Tech. Rep., 1999.

[174] K. Kawasaki, "Diffusion constants near the critical point for time-dependent Ising models." *Phys Rev*, vol. 145, pp. 224–230, 1966.

[175] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster, "RNA folding at elementary step resolution." *RNA*, vol. 6, pp. 325–338, 2000.

[176] N. von Ahsen, C. T. Wittwer, and E. Schutz, "Oligonucleotide melting temperatures under PCR conditions: Nearest-neighbor corrections for Mg2+, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas," *Clin. Chem.*, vol. 47, no. 11, pp. 1956–1961, 2001.

[177] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, "Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations," *Biochemistry*, vol. 47, no. 19, pp. 5336–5353, 2008.

[178] S. Bommarito, N. Peyret, and J. SantaLucia, Jr., "Thermodynamic parameters for DNA sequences with dangling ends," *Nucl. Acids Res.*, vol. 28, no. 9, pp. 1929–1934, 2000.

[179] N. Peyret, "Prediction of nucleic acid hybridization: parameters and algorithms," Ph.D. dissertation, Wayne State University, 2000.

[180] E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, "Stacked-unstacked equilibrium at the nick site of DNA," *J. Mol. Biol.*, vol. 342, no. 3, pp. 775 – 785, 2004.

[181] D. V. Pyshnyi and E. M. Ivanova, "Thermodynamic parameters of coaxial stacking on stacking hybridization of oligodeoxyribonucleotides," *Russ. Chem. B+*, vol. 51, pp. 1145–1155, 2002.

[182] V. A. Vasiliskov, D. V. Prokopenko, and A. D. Mirzabekov, "Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips," *Nucl. Acids Res.*, vol. 29, no. 11, pp. 2303–2313, 2001.

[183] I. L. Hofacker, "Vienna RNA secondary structure server," *Nucl. Acids Res.*, vol. 31, no. 13, pp. 3429–3431, 2003.

[184] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucl. Acids Res.*, vol. 31, no. 13, pp. 3406–3415, 2003.

[185] M. E. Craig, D. M. Crothers, and P. Doty, "Relaxation kinetics of dimer formation by self-complementary oligonucleotides," *J. Mol. Biol.*, vol. 62, pp. 383–401, 1971.

[186] D. Pörschke, O. C. Uhlenbeck, and F. H. Martin, "Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs," *Biopolymers*, vol. 12, no. 6, pp. 1313–1335, 1973.

[187] D. Pörschke, "A direct measurement of the unzippering rate of a nucleic acid double helix," *Biophysical Chemistry*, vol. 2, no. 2, pp. 97–101, 1974.

[188] H. Kuhn, V. V. Demidov, J. M. Coull, M. J. Fiandaca, B. D. Gildea, and M. D. Frank-Kamenetskii, "Hybridization of DNA and PNA molecular beacons to single-stranded and double-stranded DNA targets," *J. Am. Chem. Soc.*, vol. 124, no. 6, pp. 1097–1103, 2002.

[189] A. Tsourkas, M. A. Behlke, S. D. Rose, and G. Bao, "Hybridization kinetics and thermodynamics of molecular beacons," *Nucl. Acids Res.*, vol. 31, no. 4, pp. 1319–1330, 2003.

[190] Y. Gao, L. K. Wolf, and R. M. Georgiadis, "Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison," *Nucl. Acids Res.*, vol. 34, no. 11, pp. 3370–3377, 2006.

[191] D. D. Albergo, L. A. Marky, K. J. Breslauer, and D. H. Turner, "Thermodynamics of (dG-dC)$_3$ double-helix formation in water and deuterium oxide," *Biochemistry*, vol. 20, no. 6, pp. 1409–1413, 1981.

[192] T. E. Ouldridge, P. Šulc, F. Romano, J. P. K. Doye, and A. A. Louis, "DNA hybridization kinetics: zippering, internal displacement and sequence dependence," *Nucl. Acids Res.*, vol. 41, no. 19, pp. 8886–8895, 2013.

[193] R. Elber and S. Kirmizialtin, "Molecular machines," *Curr. Opin. Struct. Biol.*, vol. doi:10.1016/j.sbi.2012.12.002, 2013.

[194] I. G. Panyutin and P. Hsieh, "The kinetics of spontaneous DNA branch migration," *Proc. Natl. Acad. Sci. U.S.A*, vol. 91, pp. 2021–2025, 1994.

[195] C. Joo, S. A. McKinney, D. M. Lilley, and T. Ha, "Exploring rare conformational species and ionic effects in DNA Holliday Junctions using single-molecule spectroscopy," *J. Mol. Biol.*, vol. 342, pp. 739–751, 2004.

[196] T. E. Ouldridge, R. L. Hoare, A. A. Louis, J. P. K. Doye, J. Bath, and A. J. Turberfield, "Optimizing DNA nanotechnology through coarse-grained modelling: A two-footed DNA walker," *ACS Nano*, in press.

[197] F. Romano, A. Hudson, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, "The effect of topology on the structure and free energy landscape of DNA kissing complexes," *J. Chem. Phys.*, vol. 136, p. 215102, 2012.

[198] F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, "Coarse-grained simulations of DNA overstretching," *J. Chem. Phys.*, vol. 138, p. 085101, 2013.

[199] S. Whitelam and P. L. Geissler, "Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles," *J. Chem. Phys.*, vol. 127, p. 154101, 2007.

[200] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, "The role of collective motion in examples of coarsening and self-assembly," *Soft Matter*, vol. 5, no. 6, pp. 1521–1262, 2009.

[201] R. L. Davidchack, R. Handel, and M. V. Tretyakov, "Langevin thermostat for rigid body dynamics," *J. Chem. Phys.*, vol. 130, no. 23, p. 234101, 2009.

[202] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *J. Comp. Phys.*, vol. 23, pp. 187–199, 1977.

[203] R. J. Allen, P. B. Warren, and P. R. ten Wolde, "Sampling rare switching events in biochemical networks," *Phys. Rev. Lett.*, vol. 94, no. 1, p. 018104, 2005.

[204] R. J. Allen, C. Valeriani, and P. R. ten Wolde, "Forward flux sampling for rare event simulations," *J. Phys.: Condens. Matter*, vol. 21, no. 46, p. 463102, 2009.

[205] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix," *Nucl. Acids Res.*, vol. 34, no. 2, pp. 564–574, 2006.

[206] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker, "Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding," *Proc. Natl. Acad. Sci. U.S.A*, vol. 91, pp. 9218–9222, 1994.

[207] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure," *J. Mol. Biol.*, vol. 288, no. 5, pp. 911–940, 1999.

[208] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner, "Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure," *Proc. Natl. Acad. Sci. U.S.A*, vol. 101, no. 19, pp. 7287–7292, 2004.

[209] H. M. T. Choi, J. Y. Chang, L. A. Trinh, J. E. Padilla, S. E. Fraser, and N. A. Pierce, "Programmable in situ amplification for multiplexed imaging of mRNA expression," *Nat. Biotech.*, vol. 28, pp. 1208–1212, 2010.

[210] J. M. Callura, D. J. Dwyer, F. J. Issacs, C. R. Cantor, and J. J. Collins, "Tracking, tuning, and terminating microbial physiology using synthetic riboregulators," *Proc. Natl. Acad. Sci. U.S.A*, vol. 107, no. 36, pp. 15 898–15 903, 2010.

[211] J. G. Wetmur, "Hybridization and renaturation kinetics of nucleic acids," *Annu. Rev. Biophys. Bioeng.*, vol. 5, pp. 337–361, 1976.

[212] D. Pörschke, "Elementary steps of base recognition and helix-coil transitions in nucleic acids." *Mol. Biol., Biochem. and Biophys.*, vol. 24, pp. 191–218, 1977.

[213] W. F. Lima, "Implication of RNA structure on antisense oligonucleotide hybridization kinetics," *Biochemistry*, vol. 31, no. 48, p. 12055, 1992.

[214] M. Homann, W. Nedbal, and G. Sczakiel, "Dissociation of long-chain duplex RNA can occur via strand displacement in vitro: Biological implications," *Nucl. Acids Res.*, vol. 24, no. 22, pp. 4395–4400, 1996.

[215] X. Chen, "Expanding the rule set of DNA circuitry with associative toehold activation," *Journal of the American Chemical Society*, vol. 134, no. 1, pp. 263–271, 2011.

[216] A. J. Genot, D. Y. Zhang, J. Bath, and A. J. Turberfield, "Remote toehold: A mechanism for flexible control of DNA hybridization kinetics," *J. Am. Chem. Soc.*, vol. 133, no. 7, pp. 2177–2182, 2011.

[217] D. Y. Zhang and E. Winfree, "Robustness and modularity properties of a non-covalent DNA catalytic reaction," *Nucleic acids research*, vol. 38, no. 12, pp. 4182–4197, 2010.

[218] D. Pörschke, "Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction," *Biophysical chemistry*, vol. 2, no. 2, pp. 83–96, 1974.

[219] J. D. Puglisi and I. T. Jr., "Absorbance melting curves of RNA," *Methods Enzymol.*, vol. 180, pp. 304–325, 1989.

[220] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. CRC Press, 2004.

[221] A. M. Ferrenberg and R. H. Swendsen, "New Monte Carlo technique for studying phase transitions," *Phys. Rev. Lett.*, vol. 61, pp. 2635–2638, 1988.

[222] A. Arkin and J. Ross, "Computational functions in biochemical reaction networks," *Biophys J*, vol. 67, pp. 560–578, 1994.

[223] D. Wolf and A. Arkin, "Motifs, modules and games in bacteria," *Curr Opin Microbiol*, vol. 6, pp. 125–134, 2003.

[224] L. Cardelli, "Strand algebras for DNA computing," *Nat Comput*, vol. 10, pp. 407–428, 2011.

[225] L. Qian, D. Soloveichik, and E. Winfree, "Efficient Turing-universal computation with DNA polymers," *DNA Computing and Molecular Programming*, pp. 123–140, 2011.

[226] E. Chiniforooshan, D. Doty, L. Kari, and S. Seki', "Scalable, time-responsive, digital, energy-efficient molecular circuits using DNA strand displacement," *DNA Computing and Molecular Programming LNCS*, vol. 6518, pp. 25–36, 2011.

[227] M. Lachmann and G. Sella, "The computationally complete ant colony: Global coordination in a system with no hierarchy," *Lecture Notes in Computer Science (LNCS)*, vol. 929, pp. 784–800, 1995.

[228] S. K. Bayramov, "New theoretical schemes of the simplest chemical oscillators," *Biochemistry (Moscov)*, vol. 70, no. 12, pp. 1377–1384, 2005.

[229] T. Reichenbach, M. Mobilia, and E. Frey, "Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model," *Phys. Rev. E*, vol. 74, no. 051907, 2006.

[230] L. Cardelli, "Artificial biochemistry," *Algorithmic Bioprocesses*, pp. 429–462, 2009.

[231] M. R. Lakin, S. Youssef, L. Cardelli, and A. Phillips, "Abstractions for DNA circuit design," *J. R. Soc. Interface*, vol. 9, pp. 470–486, 2012.

[232] R. Blossey, "Stability of the 3-way oscillator," courtesy L. Cardelli.

[233] L. V. Allis, "Searching for solutions in games and artificial intelligence," Ph.D. dissertation, Maastricht University, 1994.

[234] C. G. Evans and E. Winfree, "DNA sticky end design and assignment for robust algorithmic self-assembly." *DNA Computing and Molecular Programming LNCS*, vol. 8141, pp. 61–75, 2013.

[235] C. Evans, "Crystals that count! physical principles and experimental investigations of DNA tile self-assembly," Ph.D. dissertation, California Institute of Technology, 2014.

[236] I. G. Panyutin, I. Biswas, and P. Hsieh, "A pivotal role for the structure of the holliday junction in DNA branch migration." *The EMBO Journal*, vol. 14, no. 8, pp. 1819–1826, 1995.

[237] S. A. McKinney, A. D. J. Freeman, D. M. J. Lilley, and T. Ha, "Observing spontaneous branch migration of holliday junctions one step at a time," *Proc. Natl. Acad. Sci. U.S.A*, vol. 102, no. 16, pp. 5715–5720, 2005.

[238] A. Padirac, T. Fujii, and Y. Rondelez, "Quencher-free multiplexed monitoring of DNA reaction circuits," *Nucl. Acids Res.*, vol. 40, no. 15, p. e118, 2012.

[239] H. Chandran, N. Gopalkrishnan, A. Phillips, and J. Reif, "Localized hybridization circuits," *DNA Computing and Molecular Programming LNCS*, vol. 6937, pp. 64–83, 2011.

[240] R. A. Muscat, K. Strauss, L. Ceze, and G. Seelig, "DNA-based molecular architecture with spatially localized components," *Proceedings of the 40th Annual international symposium on computer architecture (ISCA)*, pp. 177–188, 2013.

[241] F. Dannenberg, M. Kwiatkowska, C. Thachuk, and A. J. Turberfield, "DNA walker circuits: Computational potential, design, and verification," *DNA Computing and Molecular Programming LNCS*, vol. 8141, pp. 31–45, 2013.

[242] L. Qian and E. Winfree, "Parallel and scalable computation and spatial dynamics with DNA-based chemical reaction networks on a surface," *DNA Computing and Molecular Programming LNCS*, vol. 8727, pp. 114–131, 2014.

[243] J. Chen and N. C. Seeman, "Synthesis from DNA of a molecule with the connectivity of a cube," *Nature*, vol. 350, no. 6319, pp. 631–633, 1991.

[244] J.-S. Shin and N. A. Pierce, "A synthetic DNA walker for molecular transport," *J. Am. Chem. Soc.*, vol. 126, no. 35, pp. 10 834–10 835, 2004.

[245] W. M. Shih, J. D. Quispe, and G. F. Joyce, "A 1.7-kilobase single-stranded dna that folds into a nanoscale octahedron," *Nature*, vol. 427, no. 6975, pp. 618–621, 2004.

[246] D. Han, S. Pal, Y. Liu, and H. Yan, "Folding and cutting DNA into reconfigurable topological nanostructures," *Nat. Nanotechnol.*, vol. 5, pp. 712–717, 2010.

[247] S. F. J. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, and A. J. Turberfield, "Direct observation of stepwise movement of a synthetic molecular transporter," *Nat. Nanotechnol.*, vol. 6, pp. 166–169, 2011.

[248] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, "Three-dimensional structures self-assembled from DNA bricks," *Science*, vol. 338, no. 6111, pp. 1177–1183, 2012.

[249] M. Teichmann, E. Kopperger, and F. C. Simmel, "Robustness of localized DNA strand displacement cascades," *ACS Nano*, vol. 26, no. 8, pp. 8487–96, 2014.

[250] T. Gerling, K. Wagenbauer, A. Neuner, and H. Dietz, "Dynamic DNA devices and assemblies formed by shape-complementary, non-basepairing 3D components," *Science*, vol. 347, pp. 1446–1452, 2015.

[251] L. Qian, E. Winfree, and J. Bruck, "Neural network computation with DNA strand displacement cascades," *Nature*, vol. 475, pp. 368–372, 2011.

[252] S. F. J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama, and A. J. Turberfield, "A DNA-based molecular motor that can navigate a network of tracks," *Nat. Nano.*, vol. 7, pp. 169–173, 2012.

[253] N. Srinivas, "Programming chemical kinetics: engineering dynamic reaction networks with DNA strand displacement," Ph.D. dissertation, California Institute of Technology, 2015.

[254] D. Y. Zhang, R. F. Hariadi, H. M. T. Choi, and E. Winfree, "Integrating DNA strand-displacement circuitry with DNA tile self-assembly," *Nature Comm.*, vol. 4, no. 1965, 2013.