Chapter 5

The Kinetics of Toehold-Mediated Four-way Branch Migration[0]

5.1 Abstract

DNA nanotechnology has enabled the implementation of switches, circuits, motors, assembly and amplification using three simple mechanisms: hybridization, three-way branch migration and four-way branch migration. In particular, four-way branch migration, the process by which two double-stranded molecules of DNA that share the same stem sequence simultaneously exchange strands, demonstrates novel capabilities that expand the design-space of what can be engineered using DNA. This mechanism allows molecules to rearrange or insert into a chain without dissociating, and can be initiated conditionally by two matching toeholds (short sequences of DNA that are complementary to single-stranded domains in a target molecule). Because sequences are sequestered by their complement, toehold-mediated four-way branch migration should enable the design of circuits with less crosstalk between strands that are not supposed to interact, as compared with toehold-mediated three-way branch migration. Four-way branch migration has been used to perform directional motion, program molecular walkers and to design efficient molecular probes. We

[0]This work was coauthored by Nadine Dabby, Ho-Lin Chen, Joseph Schaeffer, & Erik Winfree* and is currently in submission [Dabby et al., 2013] with the following contributions: all experiments were performed by N. D.; trajectory simulations were performed by J.S., analysis was performed by N. D., H-L.C., and J.S with supervision by E.W. Experimental design and manuscript was done with input from all authors.

have found that by designing the toeholds involved in a four-way branch migration reaction, we can control the effective reaction rate over at least seven orders of magnitude. We characterize the kinetics of DNA toehold-mediated four-way branch migration using fluorescence spectroscopy, and derive a mechanistic model that can be used in the design of four-way branch migration reactions. The ability to control the kinetics of these reactions will greatly facilitate the programming of dynamic behaviors mediated by four-way branch migration.

5.2 Introduction

DNA enables the construction of dynamic technologies with very simple chemical mechanisms: hybridization and disassociation [Wetmur, 1976, SantaLucia, 1998], strand displacement [Yurke et al., 2000] and four-way branch migration [Panyutin and Hsieh, 1994]. We seek to understand the biophysics that underlies these fundamental mechanisms. By encoding the order of the nucleotides in a sequence, we can control the hydrogen bonding and thus the interaction of DNA strands. Hybridization and disassociation have been well characterized [Wetmur, 1976, SantaLucia, 1998]. Branch migration is the process by which a duplex of DNA exchanges one or two of its strands for new strands with identical sequences; three-way and four-way branch migration are named according to the number of strands involved in the mechanism. Nature uses four-way branch migration powered by proteins to generate genetically varied DNA from one generation to the next.

Dynamic systems of DNA molecules can be controlled by toeholds, the short sequences of DNA that are complementary to single stranded domains in a target molecule [Yurke et al., 2000, Zhang and Winfree, 2009]. We call systems controlled in this manner "toehold-mediated". Many developments in DNA nanotechnology rely on toehold-mediated three-way branch migration to implement switches [Lubrich et al., 2008, Simmel and Yurke, 2002, Yan et al., 2002, Yurke et al., 2000], circuits [Seelig et al., 2006, Yin et al., 2008, Zhang et al., 2007], motors [Gu et al., 2010, Omabegho et al., 2009, Shin and Pierce, 2004, Yin et al., 2008], assembly [Dirks and Pierce, 2004, Lubrich et al., 2009, Yin et al., 2008] and amplification [Dirks and Pierce, 2004, Zhang et al., 2007].

The kinetics and mechanism of three-way branch migration [Green and Tibbetts, 1981, Pa-

nyutin and Hsieh, 1994] and toehold-mediated three-way branch migration [Yurke and Mills, 2003, Zhang and Winfree, 2009] have been thoroughly characterized. The use of toeholds to mediate and control the process of strand displacement allows for sequence specific targeting of DNA fuels to DNA nanodevices that can be cycled through multiple states [Yurke et al., 2000]. The kinetics and mechanism of three-way branch migration has been thoroughly characterized [Green and Tibbetts, 1981, Zhang and Winfree, 2009]. Three-way branch migration reaction rates can be controlled over six orders of magnitude by changing the length [Yurke and Mills, 2003] and strength [Zhang and Winfree, 2009] of the toeholds that initiate the mechanism. One proposed model for toehold-mediated three way branch migration breaks down the mechanism into three steps: the hybridization of the free strand to the toehold domain, the branch migration process, and a final step in which the invading strand completely displaces the incumbent strand to create two separate molecules [Zhang and Winfree, 2009] .

Four-way branch migration is the process by which two double-stranded oligonucleotides that share the same stem sequence simultaneously exchange strands. Four-way branch migration has enabled the implementation of a DNA actuator [Zhang and Seelig, 2011]. Toehold-mediated four-way branch migration, initiated by unpaired toeholds that bind together to form an intermediate structure called a Holliday junction, has been used to implement molecular robots [Muscat et al., 2011], directional motors [Venkataraman et al., 2007] and molecular probes [Duose et al., 2012].

Four-way branch migration differs from three-way branch migration in its ability to implement mechanisms via two input toeholds as opposed to only one. The structure of complexes capable of undergoing four-way branch migration subserves an implicit AND function. [Duose et al., 2012] demonstrate the utility of this feature by implementing a molecular probe with a toehold-mediated four-way branch migration design that leaves all waste products double-stranded, or inert. The four-way construction is an efficient alternative to displacing multiple strands on a single probe via toehold-mediated three-way branch migration, because it reduces the number of complexes that must be added. The sequestering of DNA sequences in four-way branch migration systems also result in less cross-talk between single-stranded domains thereby making possible a slew of less noisy designs. The physical process of four-way branch migration enables the insertion or rearrangement of DNA molecules without requiring the complete disassociation of a single strand.

The mechanism is thus capable of allowing the implementation of some novel nanotechnology tools that expand the space of what can be engineered using DNA; such as the ability to implement insertion [Venkataraman et al., 2007] which has not been accomplished to date using three-way branch migration.

The full capabilities of four-way branch migration, and the means for their kinetic control have yet to be fully explored. [Panyutin and Hsieh, 1994, Thompson et al., 1976] have characterized the kinetics of the individual steps within four-way branch migration independent of toehold length, as a function of temperature and ionic conditions. The structural dynamics [McKinney et al., 2002, McKinney et al., 2005, Karymov et al., 2008] and thermodynamics [Seeman and Kallenbach, 1994] of the Holliday junction, have also been studied. As of yet, there has been no characterization of toehold-mediated control of four-way branch migration.

In pursuit of more scalable and robust DNA systems, we seek better control over the kinetics of strand displacement, to minimize leaks and maximize the ability to multiplex inputs into our DNA systems. We experimentally characterize how toehold length and strength can be used to control the speed of toehold-mediated four-way branch migration. We have found that by designing the toeholds involved in four-way branch migration, we can control the effective reaction rate over seven orders of magnitude. We used both thermodynamic and kinetic modeling to arrive at a mechanistic model for toehold-mediated four-way branch migration with four fit parameters. Our model assumes two main phases in the reaction: a bimolecular interaction step in which the two molecules hybridize to form one complex, and a unimolecular phase in which the complex undergoes branch migration. This work aims to improve our general understanding of the mechanism of four-way branch migration by providing a model that predicts how toeholds can be designed to control the kinetics of these reactions.

5.3 System Description

Single-stranded molecules of DNA (henceforth strands) are comprised of concatenated domains, stretches of consecutive nucleotides that act as a unit in binding to complementary stretches of nucleotides. Domains are represented by Latin letters; an asterisk denotes complementary do-

mains, e.g.: x is complementary to x* (Figure 5.1). DNA complexes are composed of two or more noncovalently-bound strands. The experimental reaction undergoes a bimolecular phase during which two double stranded molecules, the "Reporter" and the "Complex", join together to form the "Reporter-Complex" intermediate via hybridization of their toeholds, followed by a unimolecular phase during which branch migration is completed and the intermediate separates into two products, the "m-product" and the "n-product" (Figure 5.1A).

We characterized how toehold length and strength can be used to control the kinetics of a toehold-mediated four-way branch migration. We performed experiments with a series of DNA molecules differing from each other by only a few bases in their toehold regions. The two toeholds in the Complex are labeled **m** (for the 5' toehold on the complex) and **n** (the 3' toehold on the complex), in Figure 5.1A. The toeholds **m** and **n** bind to Reporter toeholds **m*** and **n*** respectively. The binding of one or both toeholds initiates a four-way branch migration between the two complexes; we call this process a toehold-mediated reaction.

For long enough toeholds every molecular collision will lead to complete branch migration, but for shorter toeholds only a fraction of collisions will be effective. The Reporter has toeholds of length **n*** = 6 and **m*** = either 6 or 16 in all experiments (we use a bracket notation to indicate the alternative lengths of each toehold, e.g. **m*** = $\{6; 16\}$). The Reporter is labeled with a fluorophore-quencher pair on opposite strands that are separated after the branch migration reaction with the Complex completes. This allows us to trace the kinetics of the experiment as the fluorescence of the bulk reaction increases over the course of the reaction. The Complex has toeholds of variable lengths (**m** = $\{0; 2; 4; 6; 16\}$ and **n** = $\{0; 2; 4; 6\}$. All sequences can be found in Figure 5.1 and Supplementary Table C.1.

5.4 Experimental Data

We fit experimental data in the toehold-mediated four-way branch migration process to the phenomenological model shown in Figure 5.1A. The rate constant $k_1(m, n)$ denotes the bimolecular rate of intermediate formation between the Reporter and Complex for ultimately successful reactions. The rate constant $k_2(m, n)$ denotes the unimolecular rate for completing the branch

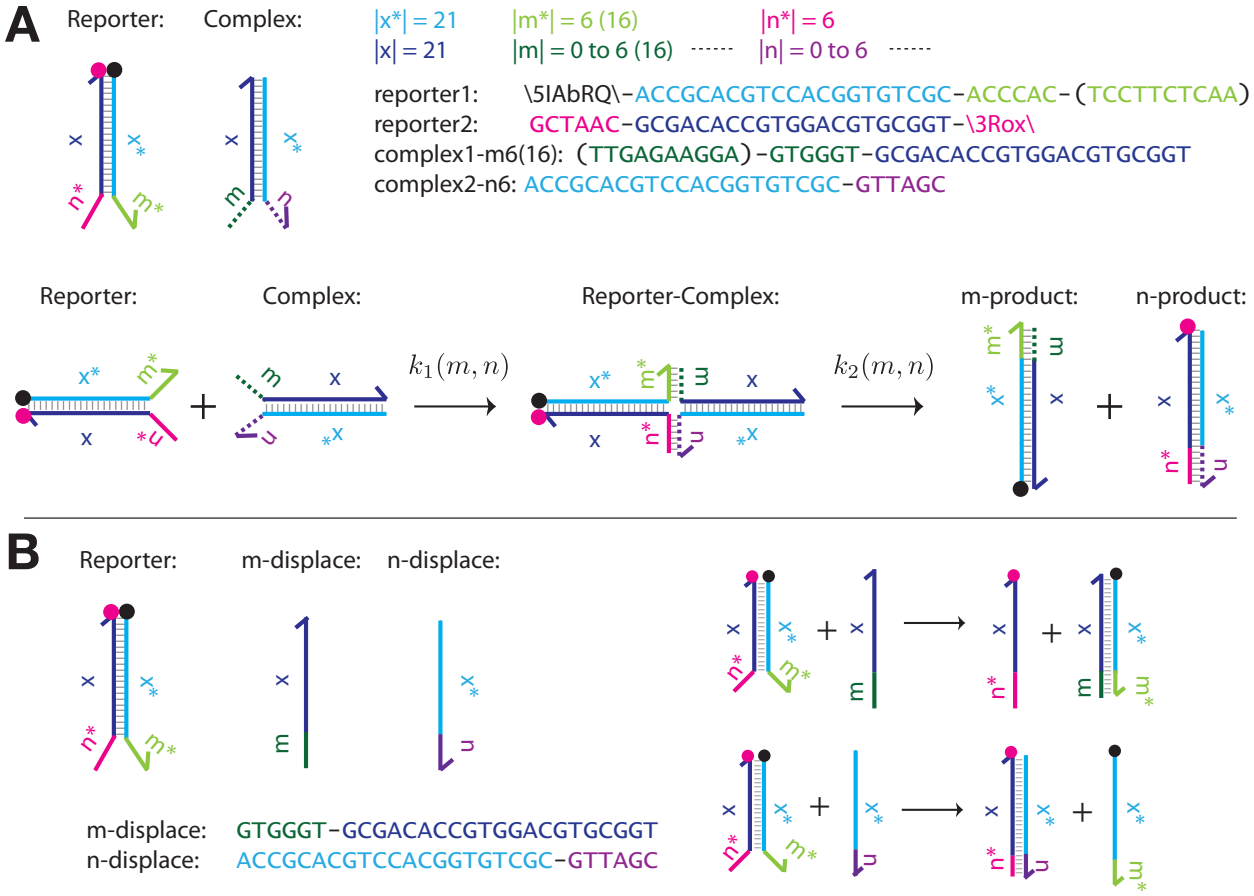


Figure 5.1: (A) A typical fluorescence kinetics experiment contains a Complex with toeholds (m, n), and a Reporter (labeled with a fluorophore and quencher pair on opposite strands) mixed together in solution. As the Complex and and Reporter exchange strands, the fluorophore and quencher pair on the Reporter are separated from each other yielding an increased fluorescence signal in the solution. Sequences for DNA strands are color-coded by domain, \5IAbRQ\ indicates a 5' Iowa Black red quencher modification, \3Rox\ indicates a 3' ROX fluorophore modification. (B)At the end of the experiment another strand of DNA is added into the solution in order to fully displace all unreacted quencher strands on the Reporter. As above, sequences are color-coded by domain.

Fitted Four-way Unimolecular Rates		
toehold length	average (sec^{-1})	standard deviation
k_2^{open}	4.4×10^{-4}	4.5×10^{-5}
k_2^{closed}	1.5×10^{-3}	4.7×10^{-4}

Table 5.1: Unimolecular rates k_2^{open} and k_2^{closed} determined by mean squared error fitting to experimental traces. We assume that the bimolecular rate constant k_1 is $k_f = 3.0 \times 10^6$ M^{-1} sec^{-1}.

migration process that results in the m-product and n-product.

We first attempted to isolate $k_2(m, n)$ by examining the kinetics of toehold combinations in which the unimolecular reaction was the rate-limiting step. We used the toehold combinations $(m = 16, n = 0)$, $(m = 16, n = 2)$, $(m = 16, n = 4)$, $(m = 16, n = 6)$ to this end, and assumed that the bimolecular step when a 16-base toehold is present is limited by the rate of hybridization $k_f = 3.0 \times 10^6$ with a negligible backward reaction. This value for k_f was measured for DNA strands in a study on toehold-mediated three-way branch migration [Zhang and Winfree, 2009]. Although three-way branch migration and four-way branch migration are different molecular mechanisms, we presume that in both cases the initial hybridization step is the same, and that the differences in mechanism only affect the probability that a given molecular interaction results in successful branch migration completion.

We discovered that for $m = 16$, all toehold pairs in which $n \geq 2$ achieved the same unimolecular rate constant, within experimental error. There are two cases for k_2: the closed case (wherein $m \geq 2$ and $n \geq 2$), and the open case (wherein $m = 0$ or $n = 0$). The open unimolecular rate, k_2^{open} (4.4×10^{-4} sec^{-1}), was almost an order of magnitude slower than the closed rate, k_2^{closed} (1.5×10^{-3} sec^{-1}). See Figure 5.2A for $(m = 16, n = 2)$ data, and Table C.2 for rates and standard deviations. For individual rates, concentrations used and experimental traces, see Table S2, Table S3 and Figure S1. After finding experimental values for k_2^{open} and k_2^{closed}, we used these values to determine k_1 from experimental traces for each combinatorial pair of (m, n). Each experiment was run at least three times, with at least two concentrations. Concentrations were selected to ensure that the rate of the overall reaction was limited by the bimolecular step rather than the unimolecular step (Table S4).

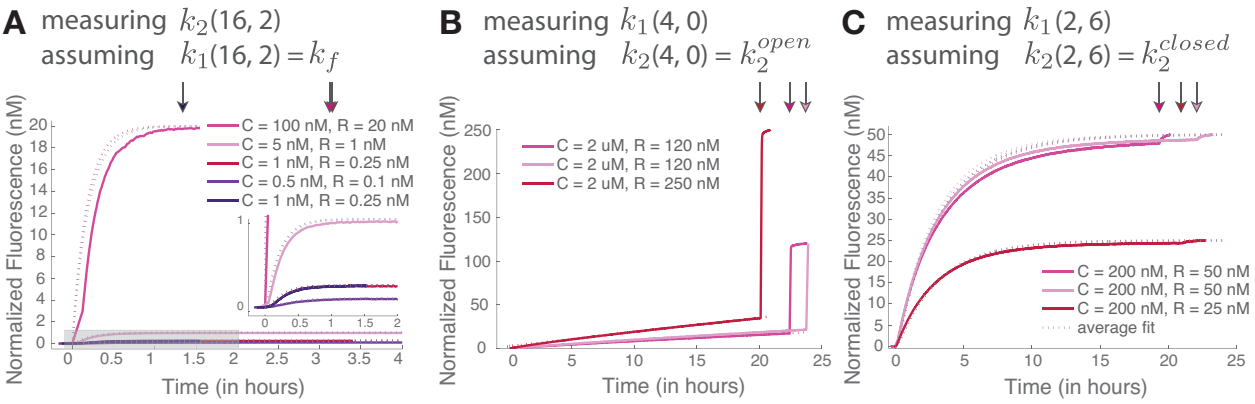


Figure 5.2: The kinetics of the reactions between Reporter and Complexes with kinetics in the (A) fast (B) slow and (C) medium-speed regimes. At the conclusion of each experiment n-displace is added in excess (as indicated by arrows). (A) Traces of $(m = 16, n = 2)$ with zoom-in of low concentration data (shaded region). (B) Traces of $(m = 4, n = 0)$. (C) Traces of $(m = 2, n = 6)$. Solid lines indicate experimental traces, dotted colored lines depict simulation of individual traces, and grey dotted lines show fits of the average fit rate at each concentration.

Combinatorial pairs of (m, n) fell into one of three regimes: slow (Figure 3B), medium (Figure 3C), and fast (Figure 3A). Figure 5.2 shows data and ODE simulations for one set of experiments in each of these regimes. Slow reactions were conducted at high concentrations and failed to achieve a 20% completion rate within 24 hours (Figure 5.2B). Medium-speed reactions were conducted at concentrations on the order of 25 nM and reached completion within 24 hours (Figure 5.2C). Fast reactions were conducted at low concentrations of 5 nM or less, and resulted in greater error, as there were fewer data points to fit (Figure 5.2A). Over these 16 combinatorial pairs of toeholds, we see $k_1^{fit}(m, n)$ rates ranging from less than 0.033 M^{-1} sec^{-1} to 6.9×10^5 M^{-1} sec^{-1} (Figure 3). The data suggest that four-way branch migration can be controlled over at least seven orders of magnitude by utilizing toeholds to mediate the reaction. Further, as was observed with toehold-mediated three-way branch migration, there is a rough correlation between (total) toehold length and the effective rate constant. (All rates and standard deviations can be found in Table 5.2, see Figures S2-S7 for all traces.) The speed of the reactions limits our confidence on the extreme ends: slow reactions were too slow to accurately measure using our methods (these rates may be slower than what we could measure), and fast reactions offered too few data points to accurately measure completion time (these rates may be faster than what we could measure).

Fitted Four-way Bimolecular Rates				
toehold length	$\Delta G^{\circ}_{k_1}$ (kcal mol^{-1})	$k_1^{fit}(m,n)$ (M^{-1}sec^{-1})	$k_1^{calc}(m,n)$ (M^{-1}sec^{-1})	$k_1^{sim}(m,n)$ (M^{-1}sec^{-1})
$m=0, n=0$	2.41	0.034 ± 0.0071	1.1×10^{-5}	≤ 0.099
$m=2, n=0$	0.57	0.047 ± 0.015	2.4×10^{-4}	≤ 0.35
$m=2, n=2$	-0.21	0.10 ± 0.0063	0.045	≤ 0.61
$m=0, n=2$	-1.34	0.033 ± 0.0023	5.9×10^{-3}	≤ 0.36
$m=2, n=4$	-1.82	0.93 ± 0.088	0.68	2.29 ± 1.3
$m=0, n=4$	-2.95	0.039 ± 0.0032	0.089	≤ 0.51
$m=4, n=0$	-4.02	0.97 ± 0.23	0.54	0.65 ± 0.645
$m=4, n=2$	-4.80	56 ± 1.4	100	92 ± 10.45
$m=2, n=6$	-5.64	490 ± 34	430	185 ± 117
$m=6, n=0$	-6.24	58 ± 6.8	23	7.30 ± 2.475
$m=4, n=4$	-6.41	770 ± 88	1.6×10^3	$5.20 \times 10^3 \pm 640$
$m=0, n=6$	-6.77	5.0 ± 0.028	56	24 ± 4
$m=6, n=2$	-7.02	$9.4 \times 10^3 \pm 3.5 \times 10^3$	4.4×10^3	$2.59 \times 10^3 \pm 640$
$m=6, n=4$	-8.63	$7.0 \times 10^4 \pm 3.0 \times 10^4$	6.4×10^4	$1.59 \times 10^5 \pm 1.925 \times 10^4$
$m=4, n=6$	-10.23	$2.8 \times 10^5 \pm 1.7 \times 10^5$	7.4×10^5	$1.66 \times 10^5 \pm 2.055 \times 10^4$
$m=6, n=6$	-12.45	$6.9 \times 10^5 \pm 3.5 \times 10^5$	2.8×10^6	$4.63 \times 10^5 \pm 2.09 \times 10^4$

Table 5.2: Bimolecular rates k_1^{fit} determined by mean squared error fitting to experimental traces, with comparison to numerical models. We assume the unimolecular rate is $k_2^{open} = 4.4 \times 10^{-4}$ sec^{-1} in the case where only one toehold is present on the intermediate complex, and $k_2^{closed} = 1.5 \times 10^{-3}$ sec^{-1} when two toeholds are present on the intermediate complex. The unimolecular rate was not utilized in cases where the experiments did not reach a 20% completion level within 24 hours, as we fit these rates linearly. Calculated reaction rates, k_1^{calc}, are predicted by the mechanistic model (see Section 5). Simulated reaction rates, k_1^{sim}, were calculated from between 200 and 10^6 Multistrand trajectories and scaled uniformly to best fit the data as described in Section 4 and the Materials and Methods section. Experimental reaction rates are shown with the standard deviation of independent fits to three or more experimental trajectories, taken with at least two different reactant concentrations. Simulated reaction rates are shown with an estimate of the standard error of the mean. Values that are upper-bounded indicate reactions that did not have a single successful trajectory over all simulations.

Each experiment was concluded with the addition of 20 μL of a displacement strand to the reaction solution for a final concentration of 6.3 μM of displacement strand (this amount is at least $50\times$ the concentration of reporter in each cuvette, and results in a dilution factor of 1.3% for the 1500 μL solution). We were able to use the maximum fluorescence of a fully consumed Reporter to normalize the fluorescence data such that the maximum signal is equivalent to a reaction completion level of 100%. We noted a lack of completion in the faster reactions with very low concentrations. In the $(m = 4, n = 6)$, $(m = 6, n = 2)$, $(m = 6, n = 4)$, and $(m = 6, n = 6)$ experiments, we set the 100% completion level to the maximum fluorescence signal that was recorded before the displacement strand was added. We traced this effect back to the age of each complex and observed that older complexes reached lower levels of completion. Across all batches, the variance of fitted rates is not correlated with completion level, independent of the Reporter's age. We are confident that our fits of these traces to the lowered completion level does not affect the results within the standard deviation reported (see Figures S4-S7).

We observed a disparity in experimental reaction rates between the m and n toeholds. For example, $(m = 0, n = 4)$ has a k_1^{fit} of 0.039 $\text{M}^{-1}\text{sec}^{-1}$, while $(m = 4, n = 0)$ has a k_1^{fit} of 0.97 $\text{M}^{-1}\text{sec}^{-1}$. This may result from sequence differences, toehold interference or small unanticipated secondary structures in the toeholds. Even though both the m and n toeholds were designed to have equal GC content, this changes as the sequences are truncated (Table S1).

The kinetics of the displacement reaction when using the n-displace strand were 2 orders of magnitude slower than when using the m-displace strand (see Table S5 for concentrations used, Figure S8 for experimental traces). These results are consistent with our observation that the same toehold lengths can differ in their rates by more than an order of magnitude when present on the m-toehold versus the n-toehold. Negligible rate differences of the displacement reaction in the presence or absence of the other toehold gives us confidence that the presence of both full-length toeholds on the Reporter did not interfere with the experiments (Figure S8).

We gauged the significance of a reverse reaction in which additional toeholds were present on the opposite side of the m-product and n-product. We tested the case where the m-product and n-product have one complementary toehold, y, of three bases in length (Figure S9 illustrates the experimental set-up and Tables S6 and S7 list the sequences and concentrations used). In this case

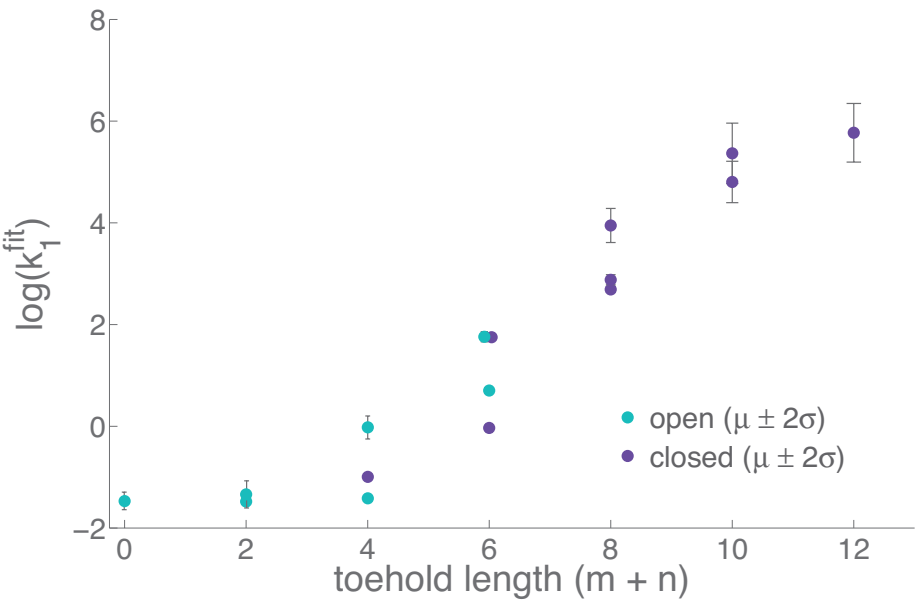


Figure 5.3: Plot of toehold length $(m + n)$ versus $\log(k_1^{fit})$, showing the correlation between the sum of toeholds m and n to the experimentally fit mean k_1 rate constants. Dots correspond to open (blue) and closed (purple) loop reactions. Error bars show two standard deviations from the mean.

the reactions most likely to go backward are $(m = 0, n = 0)$, $(m = 0, n = 2)$, $(m = 2, n = 0)$. The reverse reactions in the smallest toehold cases were negligible over the course of three days (Figure S10 and Table S8). Since these were the most energetically favorable of all of the reverse reactions and all experiments were conducted over a period of less than 24 hours, we are confident that reverse reactions should have no measurable effect on estimated rate constants.

The experimental data (Table 2) corroborates the hypothesis that toehold length can be used to control the rate of toehold-mediated four-way branch migration (Figure 5.3). However, the data also indicates that for a given sum of toeholds **m** and **n**, there can be a one to two order of magnitude spread of bimolecular reaction rates. Following the observation that toehold free energy is a better predictor of bimolecular reaction rates than toehold length [Yurke and Mills, 2003,Zhang and Winfree, 2009] in toehold-mediated three-way branch migration, we reasoned that similar arguments could better explain the behavior of toehold-mediated four-way branch migration. This required careful consideration of the relevant energy landscape and reaction mechanism steps.

5.5 Energy Landscapes and Elementary Step Simulations

The analysis of toehold-mediated four-way branch migration is complicated by the variety of ways that the two toeholds can initiate (or fail to initiate) the binding process, and by the multiple steps in which base pairs break or are formed during each branch migration step. These features can, in principle, be accounted for by secondary structure models that explicitly track each change in base pairing, while incorporating known thermodynamic and kinetic behaviors of DNA.

DNA secondary structure is the base pairing information within a set of DNA strands. We can assign an energy to a particular configuration of the molecules using the well-known nearest neighbor energy model [Mathews et al., 2004, SantaLucia Jr and Hicks, 2004]. This model breaks down the secondary structure configuration into local components known as loops, which are defined by single stranded regions and their neighboring base pairs. There are several different categories of loops, such as stacks (i.e., two neighboring base pairs with no intervening single stranded regions), hairpins, bulges, and multiloops. We typically consider states in the energy landscape to be adjacent if they differ by exactly one base pairing. Even though the total number of adjacent states to a given state is at most quadratic in the total length of the strands, the entire state space is typically at least exponential in that length. In order to modify the energy model to handle multi-stranded systems, an additional energy term is required to account for the entropic initiation cost of bringing two strands together [Dirks et al., 2007]. For each complex we have an additional energy contribution associated with the entropy of the volume [Schaeffer, 2012]. In this paper, we refer to the specific version of the energy model that we use as "the NUPACK model".

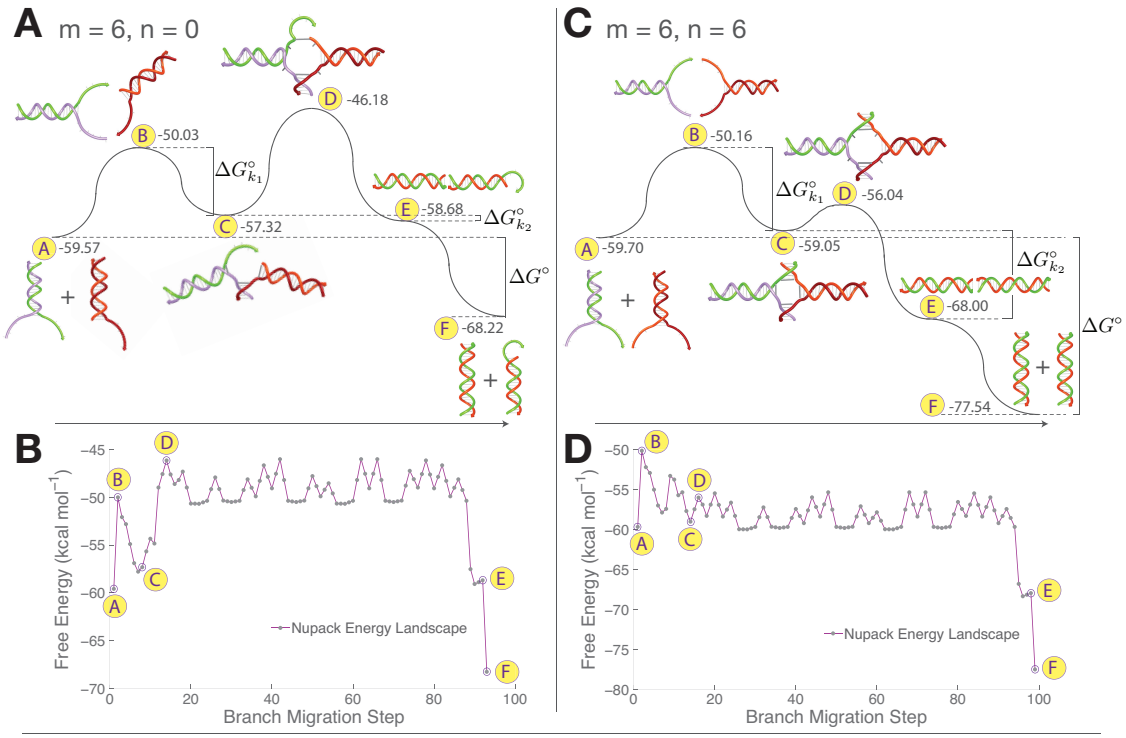
Secondary structure kinetics models can be defined as a continuous time Markov process over a secondary structure state space where states are considered adjacent if they differ by exactly one base pair [Flamm et al., 2000], and the elementary step is the formation or breaking of a single base pair. These kinetics models define the rate of transition between adjacent states using what is usually called a rate method. For example, the Metropolis rate method defines the rate of energetically favorable steps as being 1 (in arbitrary time units) and the rate of unfavorable steps as the negative exponential of the energy difference between the two states [Metropolis et al., 1953], thus ensuring that the equilibrium probabilities are consistent with the Boltzmann distribution using the

thermodynamic energies. However, thermodynamics does not determine kinetics: the rates for any step and its reverse can be arbitrarily scaled by the same amount without affecting the equilibrium distribution. Using this principle, the simulated kinetics can be calibrated to experimental data by uniformly scaling all steps.

To simulate toehold-mediated four-way branch migration, we used the Multistrand software, which simulates elementary-step secondary structure kinetics of a finite number of DNA molecules interacting within a finite volume, effectively performing random walks on the energy landscape of the system [Schaeffer, 2012]. Employing Multistrand's "first step" mode, every simulation was started with a bimolecular base-pair formation step that occurs between the Reporter and the Complex and was followed until one of two distinct end states was reached: the Reporter and Complex falling apart, or the molecules reacting into the m-product and n-product. Each trajectory ends in one of two states (non-reactive or reactive). A full simulation provides data on what percentage of initial interactions react to completion and the rate at which reactive or nonreactive collisions occur. That data is then used to calculate the simulated bimolecular reaction rate, k_1^{sim}, between the two complexes [Schaeffer, 2012].

All rates in the Multistrand kinetic model have scaling factors so that the simulated time approximates real time. These scaling factors were calibrated to match duplex formation and three-way branch migration experiments using two fitting parameters [Schaeffer, 2012]. Given the degree to which kinetics is underdetermined from thermodynamics, we expect the reported times from the simulator to require an additional scaling factor in order to approximate real time for this experimental system (see Materials and Methods). We found that the simulated rates for the complexes, k_1^{sim}, correspond well with experimental fits after being scaled uniformly by a factor of 20 (data shown in Table 2). The secondary structure energy landscape can be used as the basis for understanding toehold-mediated four-way branch migration, but because it is exponentially large, it does not provide a simple understanding suitable for analytic explanations.

We present an intuitive scheme for a reaction pathway in Figure 5.4A and C, where we have abstracted away the entire secondary structure state space into six key states: **(A)** The Reporter and Complex separate within a volume, **(B)** the Reporter and Complex co-localized with no base-pairs between them, **(C)** the Reporter and Complex bound by all available toeholds, **(D)** the initation



(A) $\Delta G_A = \Delta G^\circ_{reporter} + \Delta G^\circ_{complex} + 2 \times \left(RT \ln \frac{V}{V_0} \right)$

(B) $\Delta G_B = \Delta G^\circ_{reporter} + \Delta G^\circ_{complex} + 3 \times \left(RT \ln \frac{V}{V_0} \right)$

(C) $\Delta G_C = \Delta G^\circ_{reporter-complex} + 3 \times \left(RT \ln \frac{V}{V_0} \right)$

(D) $\Delta G_D = \Delta G^\circ_{reporter-complex^\ddagger} + 3 \times \left(RT \ln \frac{V}{V_0} \right)$

(E) $\Delta G_E = \Delta G^\circ_{m-product} + \Delta G^\circ_{n-product} + 3 \times \left(RT \ln \frac{V}{V_0} \right)$

(F) $\Delta G_F = \Delta G^\circ_{m-product} + \Delta G^\circ_{n-product} + 2 \times \left(RT \ln \frac{V}{V_0} \right)$

$$\begin{aligned} \Delta G^\circ_{k_1} &= \Delta G_C - \Delta G_B \\ &= \Delta G^\circ_{reporter-complex} - \Delta G^\circ_{complex} - \Delta G^\circ_{reporter} \end{aligned}$$

$$\begin{aligned} \Delta G^\circ_{k_2} &= \Delta G_E - \Delta G_C \\ &= \Delta G^\circ_{m-product} + \Delta G^\circ_{n-product} - \Delta G^\circ_{reporter-complex} \end{aligned}$$

$$\begin{aligned} \Delta G^\circ &= \Delta G_F - \Delta G_A \\ &= \Delta G_E - \Delta G_B \\ &= \Delta G^\circ_{k_1} + \Delta G^\circ_{k_2} \end{aligned}$$

Figure 5.4: Energy landscapes for toehold-initiated four-way branch migration. (A) and (C) show example diagrammatic energy landscapes for the open case where $(m = 6, n = 0)$ and for the closed case where $(m = 6, n = 6)$. The free energy differences $\Delta G^\circ_{k_1}$, $\Delta G^\circ_{k_2}$ and the overall ΔG° are indicated. (B) and (D) show the corresponding elementary step energy landscapes, in which each point indicates the making or breaking of a single base pair. In all energy landscapes, we define the 0 kcal energy reference point as the state in which all four strands of DNA that make up the Reporter and Complex are separated with no base pairs in the system. The energies of each system state are calculated using NUPACK energies adjusted by the entropic penalty $(RT \ln V/V_0)$ as described in the text. (The reader will notice stretches of "flat" steps in the elementary step landscapes, when one might expect an uninterrupted sawtooth diagram; the data plotted reflects the treatment of dangles in the NUPACK model.) Labels correspond to the indicated state along the reaction coordinate; energy equations are shown below.

of branch migration, **(E)** branch migration completed and the two products still co-localized, and **(F)** the two products separate within the volume. This reaction pathway allows us to examine two hypothesized energy barriers: co-localization and initiation of branch migration. We use the secondary structure nearest neighbor model energies of the aforementioned key states to estimate the barriers present in this primary reaction pathway.

Free energies of system states are calculated as follows. A state consists of the secondary structure description for a hypothetical box of volume V containing one copy of each molecule (the four strands of the Reporter-Complex, either together or separate) at an effective concentration of 100 nM. We define the 0 kcal mol^{-1} energy reference point for the system as the state in which all four strands are fully disassociated. We estimate the system energy by summing the nearest-neighbor model energies for the complexes as calculated by NUPACK [Zadeh et al., 2010] and then, as in Multistrand [Schaeffer, 2012], including the entropic penalty $RT \ln(V/V_0) = +9.54$ kcal mol^{-1} for each additional co-localized molecule, where V_0 is the volume occupied by one molecule in a 1 M solution. For example, we add two of these entropic penalty terms to the NUPACK predicted energies for the complexes in states **A** (Reporter and Complex) and **F** (m-product and n-product) in Figure 5.4 to account for the cost of bringing two strands together, twice. States **B** and **E** show two separate molecules that are co-localized, and thus pay an entropic penalty despite not being bound together in one molecule. We add one more energetic term (for a total of three) to the NUPACK predicted energies for the complexes in states **B**, **C**, **D** and **E** to account for the entropic cost of bringing the Reporter and the Complex together. These extra terms cancel in the overall reaction shown in Figure 5.1A (the energy difference between states **A** and **F** in Figure 5.4).

We now explore a path from **A** to **F** using the full secondary structure state space. While there are many possible trajectories that can take us from **A** to **F** in this state space, we will consider a particular "minimum energy barrier pathway" which can be used to define a reaction coordinate [Moulton et al., 2000]. The pathway from **A** to **F** is as follows: the **m** toehold binds first, followed by the **n** toehold (if present), and branch migration proceeds forward by the two bond-breaking steps followed by two bond-formation steps that comprise a single step of the branch migration (Figure 5.4B and D). An additional trajectory is provided in Figure C.11.

These elementary step energy landscapes provide a few key insights into the mechanism. In contrast to the diagrammatic energy landscapes, the actual trajectories are nuanced: the two main energy barriers in the open case are co-localization (state **B**) and closing the multi-loop (state **D**), but in the elementary step landscape we also observe the extended branch migration "plateau"; any effect this plateau may have on kinetics can only be examined in the more detailed model (Figure 5.4A and B). The major energy barrier in the closed case is co-localization (state **B**), while the barrier to the first step of branch migration (state **D**) is now comparable with those of other branch migration steps (Figure 5.4C and D). During toehold binding (the states between **B** and **C**), we can clearly distinguish the first toehold binding, the cost of closing the multiloop, and the second toehold binding. The analysis below shows that the dominant factor controlling overall reaction kinetics is the probability that, from the toehold bound state **C**, the molecules fall apart without reacting (first return to state **B**) or initiate and complete branch migration (first reach state **E**, going through **D**).

5.6 Mechanistic Model

The phenomenological model that guided our experiments in Section 3 (the two-step reaction mechanism shown in Figure 5.1A) was only concerned with the two reactants, one intermediate, and the two products of the toehold-mediated four-way branch migration reaction (Figure 5.5A). While this model guided our experiments, we cannot use it to explain how our measured rate constants depend upon m and n. In contrast, the general-purpose elementary-step model using a nearest-neighbor secondary structure energy landscape provides a unified means to predict rate constants for all (m, n) pairs – but does not provide simple analytic understanding. Our goal in this section is to capture the accuracy of the Multistrand simulations with (nearly) the simplicity of the coarse-grained approach that guided our experiments.

We derive a mechanistic model that explains all of the data with only four new parameters (k_2^{open}, k_2^{closed}, $\Delta G^{fit}_{k_2^{open}}$ and $\Delta G^{fit}_{k_2^{closed}}$), a hybridization rate constant (k_f) and the established NUPACK parameters. One difference between the mechanistic and phenomenological models is the former's consideration of the barrier to forming a stable Holliday junction from which the uni-

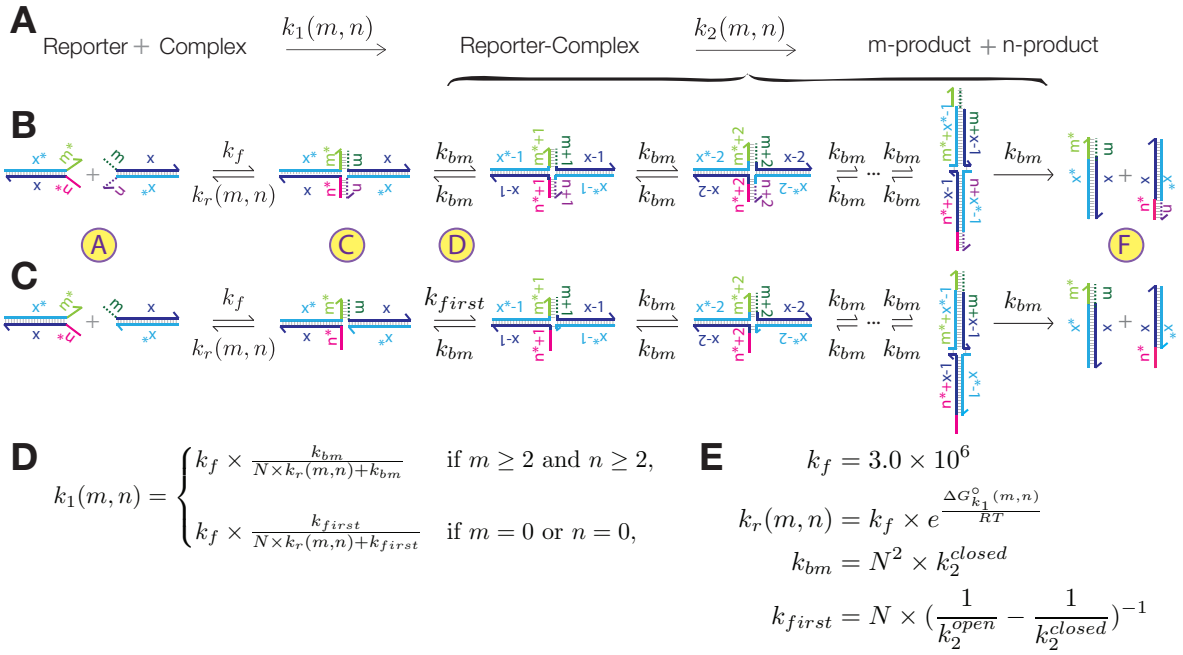


Figure 5.5: A summary of our phenomenological and mechanistic models. (A) The phenomenological reaction with rates $k_1(m,n)$ and $k_2(m,n)$. (B) A mechanism for the closed reaction ($m \geq 2$ and $n \geq 2$) shows branch migration steps, each corresponding to four individual base pair breaking and formation steps in Figure 4. (C) A mechanism for the open reaction ($m = 0$ or $n = 0$) also shows branch migration steps, but now distinguishes the first step, where initiating branch migration requires closing the multiloop. Yellow circles label states that correspond to those in Figure 5.4, State D corresponds to an intermediate transition state. (D) Formulas for calculating $k_1(m,n)$ as derived in Section 5. (E) Formulas for k_f, $k_r(m,n)$, k_{bm}, and k_{first} as derived in Section 5.

molecular branch migration can proceed. As a result, the mechanistic model uses a small parameter set to fit all of the data, whereas the phenomenological model requires 32 independent parameters to separately fit 16 toehold pairs.

The mechanistic model explicitly considers individual branch migration steps, but with less detail than the NUPACK one-dimensional landscape. The model is simple: we ignore sequences and only consider the key free energies associated with the unimolecular and bimolecular processes of the reaction. We did not test how differing sequences might affect the kinetics of the reaction, rather we assume that different sequences will behave similarly (however, sequence effects are taken into account by NUPACK in the calculation of complex free energies). In addition, we chose to model the unimolecular branch migration process as an unbiased random walk (ignoring branch

sequence dependence as well).

Figure 5.5 summarizes the mechanistic model. In Figure 5.5B and 5.5C we distinguish between the two cases under which four-way migration may proceed: the "closed" case in which both toeholds are present on both reactants and the "Reporter-Complex" intermediate forms a Holliday junction, and the "open" case in which only one toehold is able to form in the Reporter-Complex intermediate resulting in a branched DNA complex with no initial Holliday junction. What in the phenomenological model was considered simply the bimolecular rate is in the mechanistic model broken down into a forward and a reverse rate. The forward rate, k_f, is a fixed constant independent of toehold length in both the closed and open cases of the reaction. The reverse rate, $k_r(m,n)$, of intermediate formation is the rate at which the intermediate disassociates into the original complexes before branch migration completes, and is dependent on the length and sequence (or "strength") of toeholds **m** and **n**. This allows us to introduce the assumption that non-successful interactions are due to the release of toeholds before branch migration is successful.

Given the detailed balance condition, $k_r(m,n)$ can be expressed in terms of k_f and $\Delta G^\circ_{k_1}$, the free energy of intermediate formation (i.e. toehold binding), as follows:

$$k_r(m,n) = k_f \times e^{\Delta G^\circ_{k_1}/RT}, \tag{5.1}$$

where $\Delta G^\circ_{k_1} = \Delta G^\circ_{k_1^{open}}$ or $\Delta G^\circ_{k_1^{closed}}$ as appropriate. The overall unimolecular rate, either k_2^{closed} or k_2^{open}, describes the transition from the intermediate structure to the formation of the two products. This rate includes the time to initiate and complete branch migration (we assume that initiation time is negligible in the closed case), and thus will depend on the rates of the individual unimolecular steps, k_{bm} and k_{first}.

Below, we show how both the closed and open cases of the mechanistic model can be derived from an elementary step consideration of the branch migration mechanism. A branch migration can be thought of as a random walk process in one dimension: at any given position in the branch migration process, the next step can be a bond breaking and bond formation step that moves the Holliday junction either one step to the left or one step to the right (Figure 5.5B and Figure 5.5C). Since branch migration is a random walk along an N-step path, the probability,

$P(m,n)$, of successfully completing branch migration is dependent on the number of steps N (in this paper $N = 21$). We ignore the sequence dependent differences observed in the NUPACK energy landscape for simplicity, and we assume that once the Reporter-Complex intermediate is formed, the molecule has already begun the process of branch migration, with an equal chance of moving the Holliday junction to the left (toward disassociation of the reactants) or to the right (toward completion of the branch migration in the direction of the products).

The value of $k_1(m,n)$ can be defined as $k_f \times P(m,n)$, where $P(m,n)$ is the probability that the Reporter and Complex will complete branch migration once the two complexes are joined in the intermediate state. This probability is derived differently in the closed and open cases.

$$k_1(m,n) = k_f \times P(m,n). \tag{5.2}$$

5.6.1 Closed Model

In order to model $k_1(m,n)$ for $m \geq 2$ and $n \geq 2$ we define it in terms of k_2^{closed}, the unimolecular rate that we extracted from the long-toehold experiments, k_f, the hybridization rate of single stranded DNA, and $k_r(m,n)$, the reverse rate of intermediate formation. In the closed case, we assume that all of the elementary steps along the branch migration pathway are equivalent, including the first step, hence the rate of each individual step is k_{bm}.

The probability of initiating four-way branch migration (or moving one step to the right or toward the products) from the Reporter-Complex intermediate state is the probability of going to the right divided by the sum of the probabilities of moving left or right. Since the rate of branch migration is k_{bm} and the rate of disassociation is $k_r(m,n)$, the probability of making the first step toward the products in this random walk is $\frac{k_{bm}}{k_r(m,n)+k_{bm}}$. If the branch migration begins, it will successfully complete with probability $\frac{1}{N}$, before the complexes disassociate [Feller, 1968].

Since we assume that the probability of branch migrating to the left or to the right is 0.5, it follows that the probability of successfully completing branch migration from the intermediate starting point before the two complexes disassociate is $\frac{1}{N}$, and the probability of returning to the initial starting point is $\frac{N-1}{N}$. If the branch migration is started, and returns to the intermediate

state before it completes, then it will once again have probability $P(m,n)$ of successfully completing branch migration. Thus, $P(m,n)$ is the probability of beginning branch migration from the intermediate step multiplied by the probability of reaching the end before the two complexes disassociate:

$$P(m,n) = \frac{k_{bm}}{k_r(m,n)+k_{bm}} \times \left(\frac{1}{N} + \frac{N-1}{N} \times P(m,n)\right) \tag{5.3}$$

This reduces to:

$$P(m,n) = \frac{k_{bm}}{N \times k_r(m,n)+k_{bm}}. \tag{5.4}$$

By substituting equation (5.4) for the value of $P(m,n)$ in equation (5.2) we get:

$$k_1^{calc} = k_f \times \frac{k_{bm}}{N \times k_r(m,n)+k_{bm}}. \tag{5.5}$$

All that remains to be shown is how we arrive at a value for k_{bm}. k_2^{closed} is the overall rate of the unimolecular branch migration reaction, or the time it takes to complete branch migration once it is initiated. From the probability theory of one dimensional random walks, we know that the expected time to reach a position x from the origin is x^2 [Feller, 1968]. Thus, we can derive the step rate, k_{bm}, from k_2^{closed} as follows: the expected time for a single branch migration step is $\frac{1}{k_{bm}}$, the expected time to complete the whole branch migration is $\frac{1}{k_2^{closed}}$, and therefore

$$\frac{1}{k_2^{closed}} = \frac{N^2}{k_{bm}}, \tag{5.6}$$

so that k_{bm} is equal to $k_2^{closed} \times N^2$. When we substitute this value for k_{bm} into equation (5.5) we arrive at:

$$k_1^{calc} = k_f \times \frac{k_2^{closed}}{\frac{k_r(m,n)}{N} + k_2^{closed}}. \tag{5.7}$$

5.6.2 Open Model

The open model has an additional consideration: the barrier to forming a Holliday junction from the Reporter-Complex intermediate. In this case we now have a large uphill first step to initiate branch migration because the complexes must overcome the entropic cost of forming a loop (as seen in Figure 5.4).

If branch migration is fast with respect to the rate of the first step, k_{first} (which consists of closing the loop and initiating branch migration), then k_2^{open} would be dominated by the rate of this first step, and

$$k_2^{open} \approx \frac{1}{N} \times k_{first}. \tag{5.8}$$

As above, $\frac{1}{N}$, is the probability that once branch migration is initiated it will complete before the complex returns to the initial starting point (the Reporter-Complex intermediate). We can use the value of k_2^{closed} to calculate the expected time of the first step:

$$\mathbb{E}[Time_{C\to E}] = N \times \mathbb{E}[Time_{C\to D}] + \mathbb{E}[Time_{D\to E}] \tag{5.9}$$

where $\mathbb{E}[Time_{C\to E}]$ is the expected time to complete the unimolecular reaction (moving from state C to state E in Figure 5.4), $\mathbb{E}[Time_{C\to D}]$ is the expected time to complete the first step (moving from state C to state D in Figure 5.4) and $\mathbb{E}[Time_{D\to E}]$ is the expected time to complete branch migration after the first step (moving from state D to state E in Figure 5.4). N is the number of attempts required to initiate branch migration before successful completion. $\mathbb{E}[Time_{D\to E}]$ is equivalent to the expected time to complete the unimolecular reaction in the closed case, which is $\frac{1}{k_2^{closed}} = \frac{N^2}{k_{bm}}$. Thus,

$$\frac{1}{k_2^{open}} = \frac{N}{k_{first}} + \frac{N^2}{k_{bm}}, \tag{5.10}$$

which yields:

$$k_2^{open} = \frac{1}{\frac{N}{k_{first}} + \frac{1}{k_2^{closed}}}. \quad (5.11)$$

Now we can solve for k_{first} from measurable quantities.:

$$k_{first} = \frac{N}{\frac{1}{k_2^{open}} - \frac{1}{k_2^{closed}}}. \quad (5.12)$$

The probability of successfully closing the loop and initiating branch migration is

$$P(m,n) = \frac{k_{first}}{N \times k_r(m,n) + k_{first}}. \quad (5.13)$$

Thus,

$$k_1^{calc} = k_f \times \frac{k_{first}}{N \times k_r(m,n) + k_{first}}. \quad (5.14)$$

5.6.3 Results

The mechanistic model presented here is built on parameters from the literature (k_f and ΔG°), parameters that were directly measured (k_2^{open} and k_2^{closed}), and two parameters that were fit to our data to adjust predicted ΔG° values (described below). A summary of all parameters and equations used in the model can be found in Figure 5.5. We now return to the energy landscape discussion that informed the mechanistic model in order to explain additional fit parameters and to compare the model directly to both the experiments and simulations.

The diagrammatic energy landscapes (Figure 5.4A and C) show the two main energy barriers to the completion of a toehold-mediated four-way branch migration reaction: $\Delta G^\circ_{k_1}$ corresponds to the standard free energy of the bimolecular reaction in which the Reporter and Complex bind by their toeholds to form the Reporter-Complex intermediate, and $\Delta G^\circ_{k_2}$ corresponds to the standard free energy of the unimolecular reaction in which the Reporter-Complex intermediate completes the branch migration to yield the m-product and n-product. Both are negative by convention, and

overall,

$$\Delta G^\circ = \Delta G^\circ_{k_1} + \Delta G^\circ_{k_2}. \tag{5.15}$$

While we have confidence in the NUPACK-predicted free energy ΔG°, we found that the NUPACK-predicted values for $\Delta G^\circ_{k_1}$ did not result in good fits of the mechanistic model equations (Figure 5.5D and E) to the experimental data. (NUPACK's values predict on average a 408-fold slow-down of closed reactions relative to experimental results, and a speed up of open reactions by as much as a factor of 65.8.) This could be due to inaccuracies in the predicted energy of the intermediate state **C** (e.g. because of unmodeled coaxial stacking effects), due to oversimplifications of the mechanistic model relative to the NUPACK energy landscape, due to inaccuracies in the experimentally measured k_2 values, or mostly likely a combination of the above. Therefore, we chose to incorporate an empirical adjustment to $\Delta G^\circ_{k_1}$ for use within the mechanistic model. To do so we can use either $\Delta G^\circ_{k_1}$ or $\Delta G^\circ_{k_2}$ to adjust NUPACK's free energy predictions, since estimating one of these values gives us the other and we treat NUPACK's prediction of ΔG° as reliable. While $\Delta G^\circ_{k_1}$ depends on each combination of m and n, NUPACK predicts $\Delta G^\circ_{k_2}$ to be independent of toeholds m and n beyond whether they belong to the open or closed case. Thus we chose to calculate $\Delta G^\circ_{k_1}$ using NUPACK's prediction for ΔG° and two fit values for $\Delta G^\circ_{k_2}$, one for the open case and one for the closed case. $\Delta G^{fit}_{k_2^{open}}$ and $\Delta G^{fit}_{k_2^{closed}}$ are each a least squares fit of experimental data points in Figure 5.6B to values of k_1 calculated as in Figure 5.5, where all other parameters are fixed. We arrived at the following values: $\Delta G^{fit}_{k_2^{open}}$ is -2.41 kcal mol^{-1} and $\Delta G^{fit}_{k_2^{closed}}$ is -5.39 kcal mol^{-1}. For reference, if we had calculated these values from NUPACK, we would have arrived at $\Delta G^{NUPACK}_{k_2^{open}} = -1.36$ kcal mol^{-1} and $\Delta G^{NUPACK}_{k_2^{closed}} = -8.95$ kcal mol^{-1}.

In total the mechanistic model has four fit parameters (k_2^{open}, k_2^{closed}, $\Delta G^{fit}_{k_2^{open}}$ and $\Delta G^{fit}_{k_2^{closed}}$) in addition to k_f and ΔG° values from the literature. Figure 5.6 compares the mechanistic models for open and closed-loop toehold-mediated four-way branch migration to the experimentally fit mean k_1 rates and to rates predicted from the elementary step simulations. We now see excellent agreement. An exception is the three slowest open-loop cases, where the mechanistic model predicts lower rates than are observed. It is unclear whether this discrepancy reflects an inaccuracy of

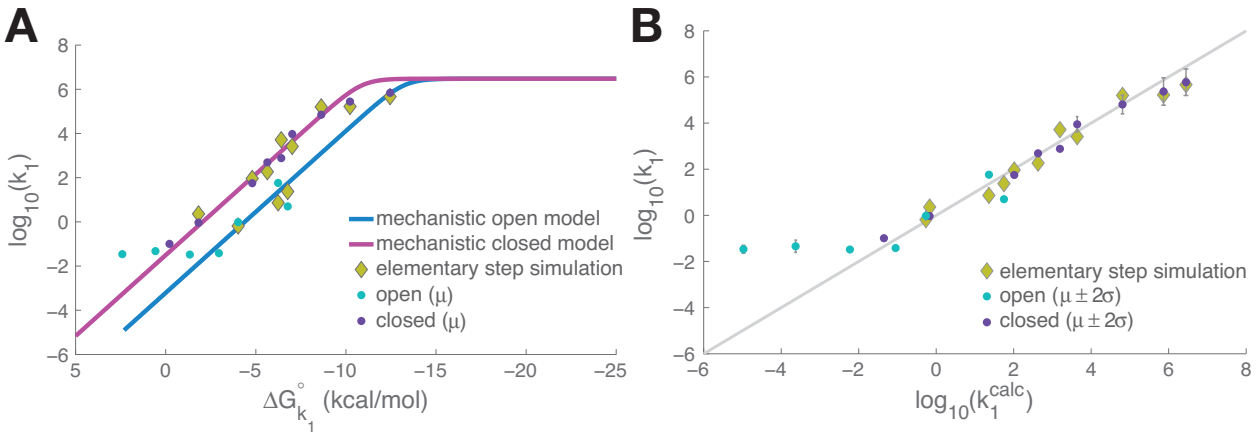


Figure 5.6: (A) Plot of adjusted $\Delta G^{\circ}_{k_1}$ versus $\log_{10}(k_1)$ comparing the mechanistic models for open (blue) and closed (magenta) loop toehold-mediated four-way branch migration to the experimentally fit mean k_1 rates. Dots correspond to open (blue) and closed (purple) loop reactions, and an elementary step simulation using predicted NUPACK free energies (yellow diamonds). Note that neither the $\Delta G^{fit}_{k_2}$ parameter nor the k_2^{fit} parameter was used in the elementary step model, but the x-axis $\Delta G^{\circ}_{k_1}$ is plotted with the adjusted value. (B) Plot comparing the $\log_{10}(k_1)$ rate calculated by the mechanistic model to the experimentally fit mean rates. Error bars show two standard deviations of error in experimental measurements. The gray line indicates the points at which $x = y$.

the mechanistic model (e.g. for small or positive values of $\Delta G^{\circ}_{k_1}$ where the nature of the energy landscape changes, c.f. Figure S11), or whether it simply reflects a limitation in our experimental technique's ability to accurately measure very slow reactions (e.g. below the rate of 0.01 M^{-1} sec^{-1}). Our ability to accurately measure the kinetics of the fastest reactions is also limited, as reflected by greater variance in the data. The measured values are not unreasonable; even the fastest reaction rates are lower than the k_f value from prior work that we use in the mechanistic model.

The success of both the mechanistic model (using NUPACK-predicted energies and empirical adjustments) and the elementary step simulations (using the full NUPACK energy landscape without adjustments) suggests, first, that the NUPACK energy landscape is not too far off (i.e. inaccuracies due to coaxial stacking and divalent salt interactions at the junction are not fatal), and therefore, second, that the empirical adjustments k_2^{fit} and $\Delta G^{fit}_{k_2}$ serve rather to accommodate coarse-graining effects found in the mechanistic model.

5.7 Discussion and Conclusions

In our model we experimentally derived unimolecular reaction rates using complexes with long-toeholds. We assume that k_2^{closed} is the rate for an overall N-step branch migration to complete with no initiation barrier. We found that the time to half completion for the closed case is 7.7 minutes. We derive the step time by dividing by N^2 where N is the length of the double stranded region that is exchanged during four-way branch migration. This yields a step time of 1.05 seconds ($\approx \frac{1}{k_{bm}}$). This is comparable with other experimentally measured rates for four-way branch migration stepping times. McKinney et al. found that the time between steps in DNA branch migration is sequence dependent and ranges between 1 and 2 seconds per step (experiments were conducted at 25 °C with comparable salt concentrations to those that we used) [McKinney et al., 2005]. Panyutin and Hsieh measured the step time of branch migration with and without 10 mM Mg^{++}, but at a higher temperatures than that used here. We estimate a step time of approximately 2.7 seconds at 25 °C by assuming that the exponential slowing of branch migration kinetics that they observed from 50 °C to 37 °C continues to hold at lower temperatures [Panyutin and Hsieh, 1993]. Our experimental data is reasonably consistent with prior studies of four-way branch migration stepping time.

We also examined how our experimentally-fit unimolecular rates match those that would be predicted by the energy landscape in Figure 5.4. Consider the energy barrier between State C and the branch migration plateau in our energy landscape, which is State D. The height of this barrier, ΔG_{D-C}, should determine k_{bm} and k_{first}, much as the energy barrier to toehold dissociation, $-\Delta G^{\circ}_{k_1}$ determines $k_r = k_f \times e^{\Delta G^{\circ}_{k_1}/RT}$ in Equation 5.1. We should be able to predict k_{bm} and k_{first} in a fashion similar to the way we define k_r (i.e. $k_r = k_f \times e^{\Delta G^{\circ}_{k_1}/RT}$ from Equation 5.1). This suggests that k_{bm} should be proportional to $e^{-\Delta G^{closed}_{D-C}/RT}$ and that k_{first} should be proportional to $e^{-\Delta G^{open}_{D-C}/RT}$ with the same kinetic pre-factor that reflects the base rate for unimolecular transitions in our model. Thus we should be able to extract the ratio

$$\frac{k_{bm}}{k_{first}} = \frac{e^{-\Delta G^{closed}_{D-C}/RT}}{e^{-\Delta G^{open}_{D-C}/RT}} \tag{5.16}$$

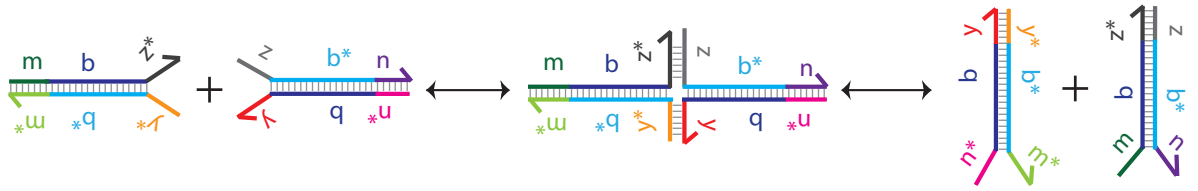


Figure 5.7: A toehold-mediated four-way strand exchange mechanism. Note that the mechanism is identical to our toehold-mediated four-way branch migration scheme except for the addition of two more domains on each of the complexes. The reaction is reversible.

from the energy landscape. In the NUPACK energy landscape, $\Delta G^{closed}_{D-C} = 3.01$ and $\Delta G^{open}_{D-C} = 11.14$. This gives us a ratio of $\approx 760,000$. When we compare this value to the ratio of ≈ 73 that we get from our experimental data where $k_{bm} = 0.95$ sec^{-1} and $k_{first} = 0.013$ sec^{-1}, it is off by four orders of magnitude. However if we correct the energies of State C using our $\Delta G^{fit}_{k_2}$ values, we get $\Delta G^{closed}_{D-C} = 6.57$ and $\Delta G^{open}_{D-C} = 10.09$ and a predicted k_{bm}/k_{first} ratio of 353, which is within one order of magnitude of the ratio derived from our experimental results. These results fit reasonably well and justify our intuition.

We investigated a limited set of toehold lengths due to cost and time constraints; however the Multistrand simulator can be used to investigate the entire range of toehold lengths at the cost of only additional simulation time. There are many remaining unknowns in controlling the kinetics of four-way branch migration reactions. For example, we have not explored how our model will generalize to other sequences in either the toehold or branch regions. We classified the "open" and "closed" cases of four-way branch migration, but we don't know how a one base toehold would affect the unimolecular step (e.g. when $m = 16$ and $n = 1$), nor how mismatches in the toehold sequence would behave. In principle, Multistrand can be used to simulate all of these conditions and make predictions about kinetics that can guide experimental design. Multistrand can be used to identify sequence designs that have unintended features such as alternate reaction pathways or kinetic traps, which may be difficult to diagnose in an experimental setting. Another feature of using a simulator in this context lies in its ability to make predictions about the kinetics of reactions that are beyond our experimental resolution, such as the very slow short toehold reactions that were discussed in this paper (see Figure 5.6).

While this characterization of toehold-mediated four-way branch migration explores irre-

versible processes, we recognize that more interesting dynamic behavior may be generated by reversible toehold-mediated four-way branch migration reactions. An analog to this system is the reversible toehold-mediated three-way branch migration system that was studied by [Zhang and Winfree, 2009]. Reversible three-way systems have been used to construct complex chemical logic circuits [Zhang et al., 2007, Qian and Winfree, 2011, Qian et al., 2011] and appear essential to more complex theoretical constructions [Soloveichik et al., 2010,Cardelli, 2011]. To our knowledge, four-way strand exchange systems have not been constructed yet, but may prove useful. This manuscript presents a first-step toward understanding reversible four-way branch migration reactions as a reversible system demonstrates the same kinetics as an irreversible system when one toehold is stronger than the other. Just as we can derive toehold-mediated three-way strand exchange kinetics from toehold-mediated three-way strand displacement reactions, we expect that we can do the same using toehold-mediated four-way branch migration reactions.

Our work presents a model for controlling the rate of a four-way branch migration using toeholds, based on the length and strength of these toeholds. We have shown that we can control the rate of a four-way branch migration reaction over at least seven orders of magnitude, and that these rates are exponential in the free energy of the toehold binding step. We experimentally characterized how the length and strength of toeholds affects the kinetics of the bimolecular rate. We have used both thermodynamic (NUPACK) and kinetic (Multistrand) modeling to elucidate this mechanism and have derived a mechanistic model with four fit parameters and a rate constant that is consistent with an elementary step model simulating the full nearest neighbors secondary structure energy landscape. We hope the model presented here will be used by DNA nanotechnologists to design more complicated and robust dynamical systems.

5.8 Materials and Methods

Experimental System A typical fluorescence kinetics experiment contains a Complex with toeholds (m, n), and a Reporter (labeled with a fluorophore and quencher pair on opposite strands) mixed together in solution (Figure 5.1A), where the Complex is added last to trigger the reaction. As the Complex and and Reporter exchange strands, the fluorophore and quencher pair on the Re-

porter are separated from each other, yielding an increased fluorescence signal in the solution. At the end of the experiment another strand of DNA, typically n-displace, was added into the solution in order to fully displace all unreacted quencher strands on the Reporter (Figure 5.1B). This "displacement" strand was added in at least $50\times$ excess to the concentration of Reporter in the solution to ensure that this reaction quickly goes to completion. We use the final completion level to normalize our fluorescence signals.

DNA Sequences and Design The sequences presented here (in Figure 5.1 and in Tables C.1 and C.6) are based on those used in a previous insertional polymerization motor [Venkataraman et al., 2007]. These sequences were designed using the NUPACK web application [Zadeh et al., 2010,Zadeh et al., 2011] and our in-house DNA Design software package to minimize the presence of any unanticipated secondary structures that might interfere with the kinetics under investigation.

Toehold Binding Energy Calculations The free energy ΔG° of all complexes was found using the NUPACK web application [Dirks et al., 2007, Zadeh et al., 2010]. The structure and sequence information was entered into the utility function on Nupack.org with ion concentrations set to 0.05 M Na+ and 0.0125 M Mg++, and dangles were set to "some". The free energy of a reaction was determined by subtracting the partition function-based free energies of the reactants from the free energies of the products. In calculating intermediate structures a mole fraction correction was used [Dirks et al., 2007]. The $\Delta G^\circ_{k_1}$ values presented in Table 5.2 are the NUPACK values corrected by a $\Delta G^{fit}_{k_2}$ parameter to correct for possible coaxial stacking effects unaccounted for in current thermodynamic models of DNA.

Buffer Conditions DNA oligonucleotides were stored in TAE 12.5 mM Mg++ buffer (purchased as $50\times$ stock TAE from Invitrogen, and solid Magnesium Acetate from Sigma) at 4°C directly preceding experiments. All experiments and purifications were performed at 25°C, with temperature controlled using an external temperature bath.

Annealing All annealing processes were performed with an Eppendorf Mastercycler Gradient thermocycler. The samples were brought down from 95°C to 16°C at a constant rate over the course of 90 min.

Substrate Purification DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT), with standard desalting purification, except for strands with a

quencher, fluorophore or a $5'$ toehold involved in the four-way branch migration. These strands were ordered with HPLC purification. Concentrations of individual strand stocks were determined from the measured absorbance at 260 nM using an Eppendorf Biophotometer and the calculated extinction coefficients provided by IDT.

Reaction complexes were further purified by nondenaturing (ND) polyacrylamide gel electrophoresis (PAGE) as follows: Strands for each sample were prepared with nominally correct stoichiometry at 10 nM and annealed. The acrylamide (19:1 acrylamide:bis) was diluted from 40% acrylamide stock (Ambion). ND loading dye (containing Bromphenol Blue in 50% glycerol) was added to all samples, achieving a final gycerol concentration of 10% by volume. The samples were then run on 12% ND PAGE at 120 V for 6 h.

Gels were run at room temperature ($\approx 25°$C) using a Hoefer Vertical Slab Gel unit. The proper bands were cut out and eluted in 1 mL of 12.5 mM TAE Mg++ buffer for two days. Purified complexes were quantitated by measurement of absorbance at 260 nm using an Eppendorf Biophotometer and extinction coefficients calculated by summing the IDT provided extinction coefficients for individual strands.

Spectrofluorimetry Studies Spectrofluorimetry studies were done using a SPEX Fluorolog-3 (Horiba) with external water bath and 1.6 mL synthetic quartz cells (Hellma 119-004F). The excitation was at 584 nm, while emission was at 604 nm. In all spectrofluorimetry experiments, the total reaction volume was 1.5 mL and the temperature was $25°$C. For net reaction studies in which the concentration of the reporter was in excess of 1 nM, 2 nm band-pass slits were used for both excitation and emission monochrometers; for experiments in which the reporter concentration was less than 1 nM, 4 nm slits were used. Experiments were conducted with an integration time of 10 s for every 60 s time-point. The fifth data set in the long toehold experiments was conducted with an integration time of 10 s for every 15 s time-point. Prior to each experiment, all cuvettes were cleaned as follows: each cuvette was washed 15 times in Milli-Q water, 5 times in 70% ethanol, another 15 times in Milli-Q water, and finally once more in 70% ethanol and then Milli-Q water. For the slit size, concentrations, and times chosen, no measurable photobleaching was observed.

Carrier Strands It is well-known that DNA sticks nonspecifically to pipet tips. Since this loss

is inconsistent, we introduced 20-nucleotide-long poly-T "carrier" strands into our experiments to coat our pipet tips. We used the carrier strands only in experiments in which one or more of our complexes occurred at a final concentration of less than 10 nM. In these experiments, the pipette tip used to add a complex kept at a stock of less than 1 μM was first dipped into a stock of carrier strand at 100 μM concentration; this stock was pipetted up and down into the tip 15 times, before being released. After this the same pipette tip was used to add the low concentration complex to the cuvette. Poly-T strands have minimal influence on the reactions of other DNA molecules in this system [Zhang et al., 2007, Zhang and Winfree, 2009].

Fluorescence Normalization Fluorescence is normalized so that one normalized unit of fluorescence corresponds to 1 nM of unquenched fluorophore-labeled strand Reporter-2. This normalization is based on the fluorescence levels of annealed samples with a minimal fluorescence measurement taken of the diluted Reporter complex before the experiment was initiated, and a maximal fluorescence value taken at the end of the experiment, after the m-displace (or n-displace) strand is added to displace all unreacted fluorophore-quencher pairs.

Parameter Fitting At least three traces for each toehold combination were analyzed and each toehold combination was investigated with at least two different concentration sets (see Appendix C for concentrations used, and all traces). The best-fit rate constants to experimental data were fitted using the "fminunc" or the "polyfit" function in Matlab to minimize the mean squared error between experimental data and our model (see sample code in Supplementary Information). All traces that did not reach a completion level of 20% within 24 hours were fit linearly using the polyfit function in Matlab. Each experimental trace was fit separately.

In order to determine k_2 we designed reaction complexes with a 16-basepair toehold. We assumed $k_1 = k_f = 3 \times 10^6$, a value taken from prior work studying toehold-mediated strand displacement kinetics [Zhang and Winfree, 2009]. After fitting the k_2^{open} and k_2^{closed} values using these "long-toehold" experiments, we used these values in fitting the k_1 rates on the smaller toehold length experiments.

We noted a lack of completion in our faster reactions with very low concentrations. In the $(m = 4, n = 6)$, $(m = 6, n = 2)$, $(m = 6, n = 4)$, and $(m = 6, n = 6)$ experiments, we adjusted the simulation to set the maximal concentration to the maximal fluorescence signal before

the displacement strand was added.

Finally, after fitting the reaction rates, we fit the data to our model by adding a $\Delta G_{k_2}^{fit}$ parameter to the ΔG° values collected from NUPACK. We found this parameter by using a minimum least squares error fit between the data and our model (see explanation in Section 5.6.3 and sample code in Appendix C).

As each trace was fit individually, the $k_1^{fit}(m,n)$ values and standard deviations reported in Table 2 is the mean value and standard deviation across all experiments for each (m,n) pair. Error bars used in Figure 5.6 show two standard deviations above and below the rate constant fitted using all three (or more) data traces.

Trajectory Simulations Simulated reaction rates (k_1^{sim}) were calculated using Multistrand [Schaeffer, 2012], an analysis tool that simulates the kinetics of multistranded DNA systems with single-basepair resolution utilizing the NUPACK energetics model. Sample size indicates the number of trajectories simulated. Final values were normalized by computing a scaling factor by minimizing the mean multiplicative factor that best fit the raw multistrand results to the experimental results, which in this case was a factor of 20. Values that are upper-bounded indicate reactions that did not have a single forward result.