J|A|C|S ARTICLES

Published on Web 11/06/2009

Control of DNA Strand Displacement Kinetics Using Toehold Exchange

David Yu Zhang and Erik Winfree*

California Institute of Technology, MC 136-93, 1200 E. California Boulevard, Pasadena, California 91125

Received August 18, 2009; E-mail: winfree@caltech.edu; dzhang@dna.caltech.edu

Abstract: DNA is increasingly being used as the engineering material of choice for the construction of nanoscale circuits, structures, and motors. Many of these enzyme-free constructions function by DNA strand displacement reactions. The kinetics of strand displacement can be modulated by toeholds, short single-stranded segments of DNA that colocalize reactant DNA molecules. Recently, the toehold exchange process was introduced as a method for designing fast and reversible strand displacement reactions. Here, we characterize the kinetics of DNA toehold exchange and model it as a three-step process. This model is simple and quantitatively predicts the kinetics of 85 different strand displacement reactions from the DNA sequences. Furthermore, we use toehold exchange to construct a simple catalytic reaction. This work improves the understanding of the kinetics of nucleic acid reactions and will be useful in the rational design of dynamic DNA and RNA circuits and nanodevices.

Introduction

DNA nanotechnology has emerged as a method of constructing structures,[1−4] motors,[5−9] and circuits[13−16] at the nanometer scale. Major reasons for the success of nucleic acids as nanoscale engineering materials are the predictability of their double-helical structure and of their Watson−Crick binding thermodynamics.[17−32] This predictability has allowed the rational design of remarkably complex static structures that self-assemble from synthetic DNA oligonucleotides of defined sequence.[1−4] Furthermore, DNA's programmable interactions can be used to mediate the molecular interactions of other materials, such as carbon nanotubes[33,34] and gold nanoparticles,[35,36] and other chemistries.[37−40]

(1) Kallenbach, N. R.; Ma, R. I.; Seeman, N. C. *Nature* **1983**, *305*, 829.
(2) Rothemund, P. *Nature* **2006**, *440*, 297.
(3) Winfree, E.; Liu, F.; Wenzler, L. A.; Seeman, N. C. *Nature* **1998**, *394*, 539.
(4) He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. *Nature* **2008**, *452*, 198.
(5) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605.
(6) Dirks, R. M.; Pierce, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15275.
(7) Pei, R.; Taylor, S. K.; Stefanovic, D.; Rudchenko, S.; Mitchell, T. E.; Stojanovic, M. N. *J. Am. Chem. Soc.* **2006**, *128*, 12693.
(8) Green, S. J.; Bath, J.; Turberfield, A. J. *Phys. Rev. Lett.* **2008**, *101*, 238101.
(9) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318.
(10) Shin, J. S.; Pierce, N. A. *J. Am. Chem. Soc.* **2004**, *126*, 10834.
(11) Sherman, W. B.; Seeman, N. C. *Nano Lett.* **2004**, *4*, 1203.
(12) Omabegho, T.; Sha, R.; Seeman, N. C. *Science* **2009**, *324*, 67.
(13) Stojanovic, M. N.; Semova, S.; Kolpashchikov, D.; Macdonald, J.; Morgan, C.; Stefanovic, D. *J. Am. Chem. Soc.* **2005**, *127*, 6914.
(14) Penchovsky, R.; Breaker, R. R. *Nat. Biotechnol.* **2005**, *23*, 1424.
(15) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585.
(16) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121.
(17) Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *SIAM Rev.* **2007**, *49*, 65.
(18) Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737.
(19) Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406.
(20) SantaLucia, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460.
(21) SantaLucia, J.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415.
(22) Bommarito, S.; Peyret, N.; SantaLucia, J. *Nucleic Acids Res.* **2000**, *28*, 1929.
(23) Peyret, N. Prediction of Nucleic Acid Hybridization: Parameters and Algorithms, Doctoral thesis, Wayne State University, 2000.
(24) Tan, Z. J.; Chen, S. J. *Biophys. J.* **2006**, *90*, 1175.
(25) Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A. *Biochemistry* **2008**, *47*, 5336.
(26) Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D. *J. Mol. Biol.* **2004**, *342*, 775.
(27) Pyshnyi, D. V.; Ivanova, E. M. *Russ. Chem. Bull.* **2002**, *51*, 1145.
(28) Vasiliskov, V. A.; Prokopenko, D. V.; Mirzabekov, A. D. *Nucleic Acids Res.* **2001**, *29*, 2303.
(29) Pyshnyi, D. V.; Ivanova, E. M. *Nucleosides, Nucleotides, Nucleic Acids* **2004**, *23*, 1057.
(30) Petruska, J.; Goodman, M. F. *J. Biol. Chem.* **1995**, *270*, 746.
(31) Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nollmann, M. *Biophys. Chem.* **2001**, *93*, 215.
(32) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.
(33) Zheng, M.; Jagota, A.; Semke, E. D.; Diner, B. A.; Mclean, R. S.; Lustig, S. R.; Richardson, R. E.; Tassi, N. G. *Nat. Mater.* **2003**, *2*, 338.
(34) Maune, H. T.; Han S.; Barish R. D.; Rothemund, P. W. K.; Goddard, W. A.; Bockrath, M.; Winfree, E. Nat. Nanotechnol. 2009, in press.
(35) Elghanian, R.; Storhoff, J. J.; Mucic, R. C.; Letsinger, R. L.; Mirkin, C. A. *Science* **1997**, *277*, 1078.
(36) Park, S. Y.; Lytton-Jean, A. K. R.; Lee, B.; Weigand, S.; Schatz, G. C.; Mirkin, C. A. *Nature* **2008**, *451*, 553.
(37) Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F. *Science* **2008**, *321*, 1795.
(38) Gianneschi, N. C.; Ghadiri, M. R. *Angew. Chem., Int. Ed.* **2007**, *46*, 3955.
(39) Gartner, Z. J.; Kana, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124*, 10304.
(40) Zhu, L.; Lukeman, P. S.; Canary, J. W.; Seeman, N. C. *J. Am. Chem. Soc.* **2003**, *125*, 10178.

 © 2009 American Chemical Society

Nonequilibrium DNA devices with dynamic function have also been constructed. These devices utilize DNA hybridization, branch migration, and dissociation for active state reconfiguration in which the kinetic pathways were designed.[5,8,9,16,41] Although the kinetics of DNA hybridization, branch migration, and dissociation processes have each been studied individually,[32,42−47] currently no method exists that accurately predicts the kinetics of DNA strand rearrangement processes from the nucleotide sequence. Establishment of a well-understood kinetic model of DNA interactions would further facilitate the widespread use of DNA devices as a technology for mediating nanoscale interactions with active behaviors.

Here, we specifically study one class of DNA reactions, that of strand displacement, in which one strand of DNA displaces another in binding to a third strand with partial complementarity to both. It is highly desireable, from an engineering perspective, to exactly predict the kinetics of strand displacement reactions from only the strand and domain sequences, as this class of reactions has been used to construct a variety of DNA devices, including logic gates,[15,48,49] catalysts,[16,50−52] and motors.[5,8−12] In these systems, a short single-stranded overhang region (known as a *toehold*) initiates the strand displacement reaction. Previous characterization of the kinetics of strand displacement reactions observed an exponential dependence of kinetics on the length of the toehold[53] but did not establish a general method to quantitatively predict kinetics from toehold sequence. The general solution to this problem is complicated by kinetics of hybridization nucleation and the unfolding of unintended secondary structures.

Our current work takes a step toward this goal by providing a three-step model of strand displacement that, under certain conditions, allows quantitative estimations of strand displacement rate constants from the thermodynamics of olignucleotide hybridization[17,20,21] for secondary structure-free molecules. The rate constants predicted by our model were within 1 order of magnitude of the experimental best fit rate constants for all 85 different strand displacement reactions we ran for this paper, and within a factor of 2 for 71 of the 85 reactions that we tested. To further evaluate the predictive ability of the model for networks of reactions, we examine the kinetic behavior of a simple noncovalent DNA catalytic reaction network, again obtaining reasonable agreement with experiments. Remarkably, our model needed to fit values for only two parameters: rate constants for DNA hybridization and for branch migration. The former was fitted to three different values for three different classes of sequences.

The kinetics of nucleic acid strand displacement and hybridization are often confounded by the existence of unwanted secondary structure in strands or regions intended to be single-stranded.[32,54,67] For the purposes of this paper, we wished to study the kinetics of strand displacement in isolation of such complications; thus, we carefully designed the experimental DNA sequences to possess minimal unwanted secondary structure. The model of strand displacement presented in this paper is expected to yield accurate rate constant predictions only when the underlying DNA sequence possess relatively little unwanted secondary structure.

Domain Notation. In theory any guanine (G) can bind to any cytosine (C), but in practice the reverse reaction is so fast that the expected lifetime of an isolated one base-pair binding is very short. A stretch of several consecutive nucleotides must be complementary in order for the bound state to exist as an intermediate for further reaction at longer time scales. One useful abstraction for understanding hybridization-based constructions that exploit such intermediates is the *domain*, a consecutive stretch of nucleotides designed to act as a unit in binding. In this paper, domains are represented by Greek letters (Figure 1A). Barred Greek letters denote domains complementary to the domains represented by unbarred Greek letters (e.g., $\bar{\beta}$ is complementary to β).

Because we perform experiments with a series of DNA molecules differing from each other only by a few bases, in this paper we further subdivide domains using superscript and subscript to denote the 5′ and 3′ portions of the domain, respectively. A superscript value of m indicates the 5′-most m bases of the full domain, while a subscript value of m indicates all but the 5′-most m bases of the domain. Thus the concatenation of β^m and β_m form the full β domain for all values of m.

Single-stranded molecules of DNA (*strands*) consist of one or more concatenated domains and are represented by capital English letters. DNA *complexes* are composed of several strands bound noncovalently to each other and are also represented by capital English letters.

The domain abstraction simplifies the sequence design process for hybridization-based DNA devices. The interactions among DNA strands are determined by domain complementarities; the exact sequences of the domains have relatively little impact on the binding properties of the strands, except insofar as poor sequence design could lead to spurious binding between noncomplementary domains. This sequence flexibility allows many instances of hybridization-based DNA systems to be constructed and run simultaneously.

Toeholds. When two strands share a common domain, they compete for binding to strands with complementary single-stranded domains. In Figure 1B, strand X (green) and strand Y (red) both possess the β domain and thus compete for binding to the $\bar{\beta}$ domain on the base strand (blue). Strand X, however, possesses an extra *toehold* domain γ, the complement of which ($\bar{\gamma}$) also exists on the base strand. This toehold domain allows X to be colocalized to substrate complex S, even though strand Y is already bound. Branch migration of the β domain then allows strand X to displace strand Y. Strand Y possesses its own unique domain α and can further react with other DNA complexes once released. This process is an example of *toehold-mediated strand displacement*, in which a toehold facilitates a strand displacement reaction.

The reverse reaction, where strand Y displaces strand X from complex L, occurs at a rate up to 6 orders of magnitude slower

(41) Zhang, D. Y.; Winfree, E. *J. Am. Chem. Soc.* **2008**, *130*, 13921.
(42) Panyutin, I. G.; Hsieh, P. *J. Mol. Biol.* **1993**, *230*, 413.
(43) Radding, C. M.; Beattie, K. L.; Holloman, W. K.; Wiegand, R. C. *J. Mol. Biol.* **1977**, *116*, 825.
(44) Panyutin, I. G.; Hsieh, P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2021.
(45) Morrison, L. E.; Stols, L. M. *Biochemistry* **1993**, *32*, 3095.
(46) Green, C.; Tibbetts, C. *Nucleic Acids Res.* **1981**, *9*, 1905.
(47) Weinstock, P.; Wetmur, J. *Nucleic Acids Res.* **1990**, *18*, 4207.
(48) Hagiya, M.; Yaegashi, S.; Takahashi, K. *Nanotechnol.: Sci. Comput.* **2006**, 293.
(49) Frezza, B. M.; Cockroft, S. L.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2007**, *129*, 14875.
(50) Bois, J. S.; Venkataraman, S.; Choi, H. M. T.; Spakowitz, A. J.; Wang, Z. G.; Pierce, N. A. *Nucleic Acids Res.* **2005**, *33*, 4090.
(51) Seelig, G.; Yurke, B.; Winfree, E. *J. Am. Chem. Soc.* **2006**, *128*, 12211.
(52) Turberfield, A. J.; Mitchell, J. C.; Yurke, B.; Mills, A. P.; Blakey, M. I.; Simmel, F. C. *Phys. Rev. Lett.* **2003**, *90*, 118102.
(53) Yurke, B.; Mills, A. P. *Genet. Prog. Evol. Mach.* **2003**, *4*, 111.
(54) Gao, Y.; Wolf, L. K.; Georgiadis, R. M. *Nucleic Acids Res.* **2006**, *34*, 3370.

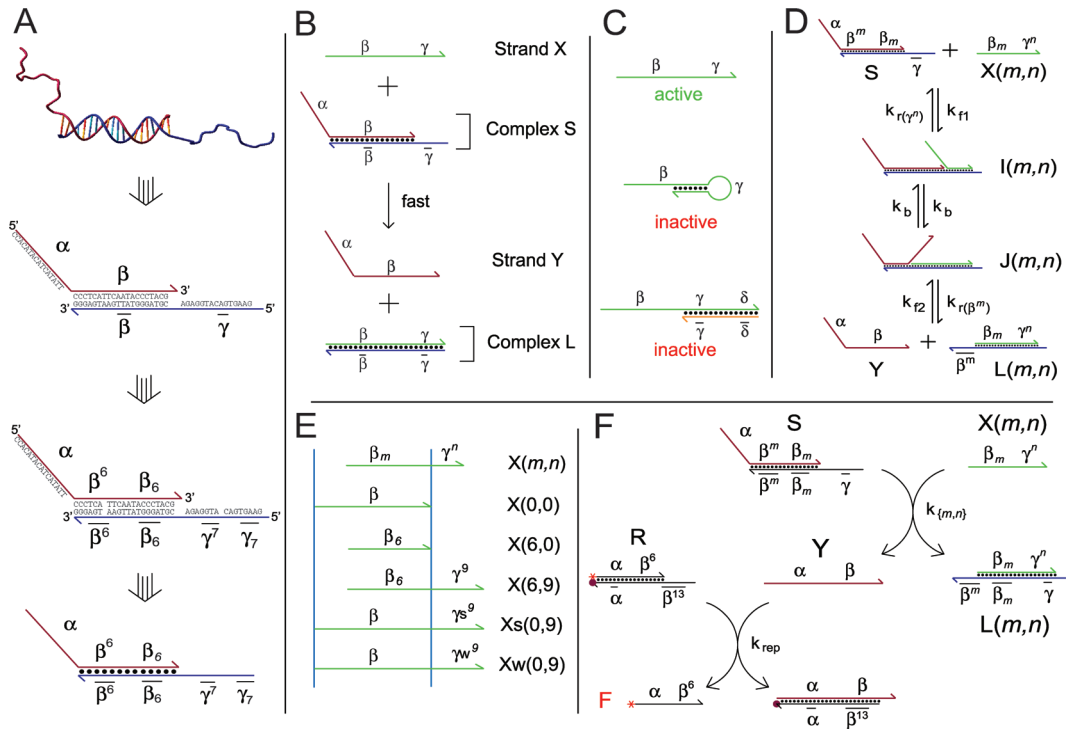

Figure 1. (A) DNA abstraction. A DNA complex (top) is typically abstracted as several directional lines, one for each strand, with bases identities shown. Here, we abstract DNA strands and complexes one step further by grouping contiguous nucleotides into domains, functional regions of DNA that act as units in binding. Because the principles and mechanisms studied in this paper are expected to be generalizable to most DNA sequences, we typically do not show the sequences of DNA strands in figures. For sequences, refer to Table 1. (B) A toehold-mediated strand displacement reaction. The displacement of strand Y by strand X is facilitated by strand X's toehold domain γ. (C) Two examples of toehold sequestration. A strand of DNA can be rendered unreactive by inactivating its toehold domains. In the figure, toehold γ is sequestered through isolation in a hairpin (middle) and through hybridization to a complementary domain (bottom). (D) A toehold exchange reaction and its mechanism. Invading strand X(m, n) binds to substrate complex S by toehold $\overline{\gamma^n}$ (known as the invading toehold) to form intermediate I(m, n). Intermediate I(m, n) represents all branch migration states in which Y is bound to more bases of $\overline{\beta_m}$ than X(m, n). Intermediate I(m, n) rearranges to form intermediate J(m, n), which analogously represents all states in which X(m, n) binds more bases of $\overline{\beta_m}$ than Y. Domain β^m (known as the incumbent toehold) spontaneously dissociates, releasing products Y and L(m, n). The toehold exchange reaction is reversible, because strand Y can subsequently bind to complex L(m, n) via strand Y's toehold β^m. (E) Comparison of various invading strands X(m, n). Strand X(m, n) is the concatenation of domains β_m and γ^n and consequently has length $(b + n - m)$ nt, where b is the length of the full β domain. In a toehold exchange reaction using X(m, n), the invading toehold has length n and the incumbent toehold has length m. For our experiments, we used three sets of invading toeholds, γ^n, γs^n, and γw^n. The sequence composition of the latter two are purely A/T's and purely G/C's, respectively, to characterize the kinetics of toehold exchange given weak and strong toeholds, respectively. Substrates using γs are labeled Ss, and inputs using γs are labeled Xs(m, n), and similarly for γ and γw. (F) Schematic of the experimental system used for rate measurements. Reporter complex R reacts stoichiometrically with product Y to yield increased fluorescence.

Table 1. Domain Sequences

domain	sequence	length (nt)
α	5′- CCACATACATCATATT -3′	16
β	5′- CCCTCATTCAATACCCTACG -3′	$b \equiv 20$
γs	5′- CCCGCCGCCG -3′	10
γ	5′- TCTCCATGTCACTTC -3′	15
γw	5′- ATTTATTATA -3′	10

and therefore can be considered effectively unreactive. This is because the complex L does not possess a single-stranded complement to strand Y's α domain, so Y cannot be easily colocalized to complex L. Thus, the presence and properties of the toehold domain are instrumental to the kinetic control of DNA strand displacement reactions.[5,53]

Because of the toehold's role in initiating strand displacement reactions, strands can be rendered effectively unreactive if the toehold domain is made inaccessible by *toehold sequestering*. Toehold sequestering can be achieved in a number of ways, the two most common of which are hybridization of the toehold to a complementary domain[9,15,16] and isolation of the toehold in a short hairpin structure where helix formation is difficult[6,52,55] (see Figure 1C). Programmed sequestering and subsequent exposure of toehold domains allows precise control of order and timing over the reactions and has been used in conjunction with toehold-mediated strand displacement to construct molecular motors,[5,8,9] polymerization reactions,[6,9] catalytic reactions,[9,51,52] and logic gates.[15,48,49]

Recently, the *toehold exchange* mechanism was introduced as a method for improved control of strand displacement kinetics.[16] Toehold exchange is similar to toehold-mediated strand displacement in that an invading strand (X) binds by a toehold to initiate branch migration but differs from the latter in that the incumbent strand (Y) possesses a unique toehold that must spontaneously dissociate for the reaction to complete. Expanding on the example strand displacement reaction in Figure 1B, the toehold exchange reaction and mechanism that we study experimentally is illustrated in Figure 1DE for an invading toehold of length n and an incumbent toehold of length m ($m, n > 0$): Strand X(m, n) binds to complex S via invading toehold γ^n and displaces strand Y's β_m domain by branch migration. Strand Y's incumbent toehold β^m then spontaneously dissociates, yielding free strand Y and complex L(m, n). The end result of the toehold exchange reaction is that the originally active toehold γ^n is sequestered while the formerly sequestered

(55) Green, S. J.; Lubrich, D.; Turberfield, A. J. *Biophys. J.* **2006**, *91*, 2966.

toehold β^m is activated. Thus, the active toehold is "exchanged" from γ^n to β^m. Note that the α domain and the overhang on the γ domain in Figure 1D are not pertinent to the toehold exchange reaction; they are present in the figure to accurately reflect the experimental reaction (α is used to used to trigger a downstream fluorescent reporter, see Experimental System).

Functionally, toehold exchange offers two main advantages over toehold-mediated strand displacement. First, the partially double-stranded product resulting from a toehold exchange reaction L(m, n) possesses a single-stranded domain $\overline{\beta^m}$, allowing it to undergo further toehold-mediated reactions. Thus in toehold exchange, two reactive reactants (X(m, n) and S) yield two reactive products (Y and L(m, n)), rather than yielding just one as in toehold-mediated strand displacement. This preservation of the number of active molecules, in combination with the reaction's reversibility, can be useful in larger reaction networks.[56−58]

Second, toehold exchange weakens the coupling between the kinetics of strand displacement and the thermodynamics of the reaction. In toehold-mediated strand displacement reactions operating below their maximum rate, there is a strong coupling between the kinetics and the thermodynamics: to speed up the strand displacement reaction, the invading toehold must be made stronger thermodynamically, for example by increasing the toehold length.[53] Consequently, faster strand displacement reactions are more thermodynamically favorable in the net reaction. (The reaction kinetics of toehold-mediated strand displacement saturates for sufficiently long toeholds, breaking the coupling between thermodynamics and kinetics.) In contrast, in toehold exchange, when both the invading and the incumbent toeholds (γ^n and β^m) are made stronger, the kinetics of both the forward and the reverse reactions are sped up, within the limits of sufficiently low concentration. Thus, a strand displacement reaction based on toehold exchange can be fast despite being only weakly thermodynamically favorable or even thermodynamically unfavorable.

Toehold-mediated strand displacement and toehold exchange encompass much larger classes of reactions than the examples presented in Figure 1B and Figure 1D. Both the substrate S and the input X(m, n) may be part of larger DNA complexes (as in refs 16, 56, and 58) or may be functionalized to other materials. As one example, inputs may be expressed on the surface of gold nanoparticles,[35] while substrates are regularly arrayed on a DNA origami,[2,59] to create an autonomous moving device analogous to those constructed from deoxyribozymes.[13] Here, we experimentally characterize only reactions of the type presented in Figure 1D, but it is expected that the results on the kinetics of toehold exchange can be generalized to a wide range of molecular designs.

Three-Step Model of Toehold Exchange. We simplify the biophysics of the toehold exchange process to the three-step model shown in Figure 1D:

$$\mathrm{X}(m,n) + \mathrm{S} \underset{k_{\mathrm{r}(\gamma^n)}}{\overset{k_{\mathrm{f1}}}{\rightleftharpoons}} \mathrm{I}(m,n) \underset{k_{\mathrm{b}}}{\overset{k_{\mathrm{b}}}{\rightleftharpoons}} \mathrm{J}(m,n) \underset{k_{\mathrm{f2}}}{\overset{k_{\mathrm{r}(\beta^m)}}{\rightleftharpoons}} \mathrm{Y} + \mathrm{L}(m,n)$$

The rate constants k_{f1} and k_{f2} denote the hybridization rates of γ^n and β^m to their complements, respectively. For simplicity, here we assume that these two rate constants are equal in value, $k_{\mathrm{f1}} = k_{\mathrm{f2}} = k_{\mathrm{f}}$. Later, we see that the base compositions of the γ^n and β^m domains can cause the values of k_{f1} and k_{f2} to vary significantly.[71]

The value of $k_{\mathrm{r}(\gamma^n)}$ denotes the first-order rate at which the toehold γ^n dissociates and is calculated so as to satisfy the equilibrium between X(m, n) + S and I(m, n):

$$k_{\mathrm{r}(\gamma)} = k_{\mathrm{f1}} \times \frac{2}{b-m} \times e^{\Delta G^\circ(\gamma^n)/RT} \qquad (1)$$

where $\Delta G^\circ(\gamma^n) < 0$ is the binding energy between γ^n and its complement, b is the length of the full β domain, and $b - m$ is the length of the β_m domain (since the state I(m, n) contains $(b - m)/2$ isoenergetic branch migration intermediate states). For the systems we examine, $b - m$ varies from 13 to 20 nt. The $2/(b - m)$ term is necessary to ensure the correct relative concentrations of the states at equilibrium.[72] The value of $k_{\mathrm{r}(\beta^m)}$ is calculated similarly. Finally, k_{b} represents the effective rate at which the branch migration junction crosses the middle of β_m, and depends on the length of the branch migration domain β_m.

To quantitatively test the rate constants predicted by our model, we collected kinetic data on a series of 85 different strand displacement reactions. To generate the 85 different predicted rate constants, our model required values for k_{f1}, k_{f2}, and k_{b}, in addition to the generally accepted nucleic acid hybridization thermodynamics parameters. Two of these parameters, k_{f1} and k_{b}, were fitted to our experimental data, while the last (k_{f2}) was assumed to be the same as k_{f1}. Note that three different values of k_{f1} were fitted, corresponding to the three different γ domains experimentally tested. We expect these parameter values to be roughly the same for systems with similar salt concentrations, temperatures, and domain lengths and compositions; consequently, we expect our model to be able to roughly predict the kinetics of DNA hybridization-based systems and circuits that utilize similar strands and complexes. Our predictions may not be accurate for strands that possess significant unwanted secondary structure.

Bimolecular Reaction Model of Toehold Exchange. For ease of designing, modeling, simulating, and data fitting, it is convenient to model toehold exchange and toehold-mediated strand displacement as simple bimolecular reactions. In fact, the kinetics of toehold exchange and toehold-mediated strand displacement have been previously observed to be well-approximated by bimolecular reactions with second-order rate constants.[16,53] In terms of the example system shown in Figure 1D,

(56) Soloveichik, D.; Seelig, G.; Winfree, E. In *DNA Computing: 14th International Meeting on DNA Computing*, Proceedings of the 14th Annual Conference on DNA Computing, Prague, Czech Republic, June 2−9, 2008; Goel, A.; Simmel, F.; Sosik, P., Eds.; Springer Berlin, Heidelberg; 2009, pg. 57−69.
(57) Qian, L.; Winfree, E. In *DNA Computing: 14th International Meeting on DNA Computing*, Proceedings of the 14th Annual Conference on DNA Computing, Prague, Czech Republic, June 2−9, 2008; Goel, A.; Simmel, F.; Sosik, P., Eds.; Springer Berlin, Heidelberg; 2009, pg. 70−89.
(58) Cardelli, L. Proceedings of the 15th Annual Conference on DNA Computing, June 8−11, 2009; Fayetteville, AR,2009.
(59) Rinker, S.; Ke, Y.; Liu, Y.; Chhabra, R.; Yan, H. *Nat. Nanotechnol.* **2008**, *3*, 418.
(60) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry: Part III: The Behavior of Biological Macromolecules*; W. H. Freeman: San Francisco, CA, 2000.
(61) Dimitrov, R. A.; Zuker, M. *Biophys. J.* **2004**, *87*, 215.
(62) Segel, L. A.; Slemrod, M. *SIAM Rev.* **1989**, *31*, 446.
(63) Dirks, R. M.; Lin, M.; Winfree, E.; Pierce, N. A. *Nucleic Acids Res.* **2004**, *32*, 1392.
(64) Reynaldo, L. P.; Vologodskii, A. V.; Neri, B. P.; Lyamichev, V. I. *J. Mol. Biol.* **2000**, *297*, 511.

$$\mathrm{X}(m,n) + \mathrm{S} \underset{k_{(\gamma^n,\beta_m,\beta^m)}}{\overset{k_{(\beta^m,\beta_m,\gamma^n)}}{\rightleftharpoons}} \mathrm{Y} + \mathrm{L}(m,n)$$

The associated forward and reverse second-order rate constants $k_{(\beta^m,\beta_m,\gamma^n)}$ and $k_{(\gamma^n,\beta_m,\beta^m)}$ for this bimolecular reaction model (BM) of strand displacement are dependent on the sequences of the β^m, β_m, and γ^n domains, as well as the ambient temperature and salt concentrations.

Using our three-step model of toehold exchange, we can analytically derive the BM rate constant under quasi-steady-state (QSS) conditions[62] (see Text S1, Supporting Information):

$$k_{(\beta^m,\beta_m,\gamma^n)} \equiv \frac{k_{\mathrm{r}(\beta^m)} k_{\mathrm{f}}\, k_{\mathrm{b}}}{k_{\mathrm{r}(\gamma^n)} k_{\mathrm{r}(\beta^m)} + k_{\mathrm{r}(\gamma^n)} k_{\mathrm{b}} + k_{\mathrm{r}(\beta^m)} k_{\mathrm{b}}} \qquad (2)$$

The above equation yields the expression for the BM rate constant, but there is a subtle inaccuracy to it. The bimolecular reaction model concentrations $[\mathrm{X}(m,n)]_{\mathrm{BM}}$ and $[\mathrm{S}]_{\mathrm{BM}}$ differ from their three-step model counterparts $[\mathrm{X}(m,n)]$ and [S] because the BM does not account for intermediate concentrations $[\mathrm{I}(m,n)]$ and $[\mathrm{J}(m,n)]$. In order for the BM model and rate constant to accurately describe the kinetics of strand displacement processes, the concentrations of $\mathrm{X}(m,n)$ and S must be sufficiently low. For example, the kinetics of Y production are accurate to within 50% when the initial concentrations of $[\mathrm{X}(m,n)]$ and [S] are below

$$c_{\mathrm{crit}} = \frac{0.1}{k_{\mathrm{f}}} \times \frac{k_{\mathrm{r}(\gamma^n)} k_{\mathrm{r}(\beta^m)} + k_{\mathrm{r}(\gamma^n)} k_{\mathrm{b}} + k_{\mathrm{r}(\beta^m)} k_{\mathrm{b}}}{k_{\mathrm{b}} + k_{\mathrm{r}(\beta^m)}} \qquad (3)$$

The 0.1 in the numerator of the first term varies nonlinearly with the required accuracy of the BM model (with larger numbers leading to higher inaccuracy). The derivation of this critical concentration is shown in Text S2, Supporting Information. Within these concentration limits, the BM rate constants for a particular system can be derived from the parameters of the three-step model, which can in turn be predicted based on the sequences of the strands.

The Experimental System. A typical fluorescence kinetics experiment contains substrate complex S, input strand $\mathrm{X}(m,n)$, and fluorescent reporter complex R mixed in solution (Figure 1F), with the input $\mathrm{X}(m,n)$ added last to trigger the start of the reaction. $\mathrm{X}(m,n)$ reacts with S to form Y and $\mathrm{L}(m,n)$, the former of which undergoes further reaction with reporter R to yield an increased fluorescence signal. In addition to allowing high-resolution kinetic data, the reporter complex R reduces the back-reaction of Y with product complex $\mathrm{L}(m,n)$ by quickly removing free Y from solution. Another reason for using a separate reporter complex rather than directly labeling X or S with fluorophores and quenchers is the extent to which fluorophore−quencher binding alters the thermodynamics.[66] The indirect reporter complex R isolates this effect from toehold exchange thermodynamics and kinetics. When bound in S, strand Y does not react with R because the relevant toehold is sequestered. Strand $\mathrm{X}(m,n)$ does not react significantly with R because it lacks the α domain.

Three different versions of S and $\mathrm{X}(m,n)$ were used, corresponding to three different sequences for domain γ. The default sequence, which we refer to as γ with no further annotations, is representative of toeholds with typical binding energies for their lengths; it consists of a mixture of G/C/A/T bases with all subdomains γ^n containing roughly equal numbers of G/C and A/T base pairs (e.g., γ^8 contains four A/T bases and four G/C bases). For weak toeholds and strong toeholds, we used domains γw and γs, which consist exclusively of A/T and G/C base pairs, respectively (see Table 2). For these three sequence choices, the corresponding strands and complexes are referred to as $\mathrm{X}(m,n)$, S, and $\mathrm{L}(m,n)$ for the typical toeholds, $\mathrm{Xw}(m,n)$, Sw, and $\mathrm{Lw}(m,n)$ for the weak toeholds, and $\mathrm{Xs}(m,n)$, Ss, and $\mathrm{Ls}(m,n)$ for the strong toeholds.

Table 2. Calculated Toehold Binding Energies (in kcal/mol)

toehold	binding energy	toehold	binding energy
γ^0	+1.9	$\gamma\mathrm{s}^0$	+1.9
γ^1	+0.2	$\gamma\mathrm{s}^1$	−1.1
γ^2	−1.7	$\gamma\mathrm{s}^2$	−3.2
γ^3	−3.0	$\gamma\mathrm{s}^3$	−5.0
γ^4	−4.7	$\gamma\mathrm{s}^4$	−8.0
γ^5	−6.9	$\gamma\mathrm{s}^5$	−10.3
γ^6	−8.3	$\gamma\mathrm{s}^6$	−12.1
γ^7	−9.2	$\gamma\mathrm{s}^7$	−15.1
γ^8	−11.9	$\gamma\mathrm{s}^8$	−17.3
γ^9	−12.9	$\gamma\mathrm{s}^9$	−19.2
γ^{10}	−14.8	$\gamma\mathrm{s}^{10}$	−21.2
γ^{15}	−21.8		
β^0	+1.2	$\gamma\mathrm{w}^0$	+1.9
β^1	−0.6	$\gamma\mathrm{w}^1$	+0.2
β^2	−2.7	$\gamma\mathrm{w}^2$	−0.8
β^3	−4.5	$\gamma\mathrm{w}^3$	−2.1
β^4	−5.6	$\gamma\mathrm{w}^4$	−3.8
β^5	−6.7	$\gamma\mathrm{w}^5$	−4.3
β^6	−9.5	$\gamma\mathrm{w}^6$	−5.3
β^7	−10.2	$\gamma\mathrm{w}^7$	−7.0
		$\gamma\mathrm{w}^8$	−7.5
		$\gamma\mathrm{w}^9$	−9.0
		$\gamma\mathrm{w}^{10}$	−8.9

The domain sequences were carefully designed to avoid secondary structure in single-stranded species so that the kinetics of toehold exchange could be separated from the first-order kinetics of unfolding unintended secondary structures. To this end, one useful sequence design heuristic was to minimize the number of G's in single-stranded species (e.g., $\mathrm{X}(m,n)$ and Y). Not only are unintentional G-C bases much more stable than A-T ones, G's also further contribute to undesired secondary structures through G-T wobbles and G-quartets.[32] For this reason, there are no G's in the α domain and one G each in the β and γ domains.

To systematically characterize the kinetics of toehold-mediated strand displacement and toehold exchange, we measured the BM rate constants of the reaction between substrates Ss, S, and Sw and a series of different inputs $\mathrm{Xs}(m,n)$, $\mathrm{X}(m,n)$, and $\mathrm{Xw}(m,n)$ for various values of m and n. Double-stranded waste products $\mathrm{Ls}(m,n)$, $\mathrm{L}(m,n)$, and $\mathrm{Lw}(m,n)$ possess 5′ overhangs on the bottom strand whenever n is smaller than the length of the corresponding γ domain. For convenience, the BM rate constants for forward reactions ($k_{(\beta^m,\beta_m,\gamma^n)}$ by the previous convention) are relabeled $k\mathrm{s}_{\{m,n\}}$, $k_{\{m,n\}}$, and $k\mathrm{w}_{\{m,n\}}$ for the rate constants using the three sets of sequences presented in this paper. For discussions not specific to sequence, we use the notation for typical toehold strengths, e.g., $k_{\{m,n\}}$.

Materials and Methods

DNA Sequences and Design. The sequences presented here are based on those of Zhang et al.[16] and were designed by hand to possess minimal secondary structure and crosstalk (binding between

(65) Puglisi, J. D.; Tinoco, I. *Methods Enzymol.* **1989**, *180*, 304.
(66) Marras, S. A.; Kramer, F. R.; Tyagi, S. *Nucleic Acids Res.* **2002**, *30*, 122.

unrelated domains). NUPACK[17] calculates there to be no more than four paired bases between any pair of domains at 25 °C, even at 1 μ M concentration. Thus, the domains we use can be approximated as structure-free. Similar sequences could have been designed using computer-aided methods.[63] Substantial secondary structure is known to slow down branch migration and interfere with hybridization.[54]

Toehold Binding Energy Calculations. The value of the predicted BM rate constant depends sensitively on the toehold binding energies $\Delta G^\circ(\gamma^n)$ and $\Delta G^\circ(\beta^m)$. We calculated each these as difference in standard free energies of two relevant complexes:

$$\Delta G^\circ(\gamma^n) = \Delta G^\circ(\mathrm{I}(0,n)) - \Delta G^\circ(\mathrm{S})$$
$$\Delta G^\circ(\beta^m) = \Delta G^\circ(\mathrm{J}(m,0)) - \Delta G^\circ(\mathrm{L}(m,0))$$

The calculation of the standard free energies of each complex used the initiation free energy and base stacking energy parameters reported by Santa Lucia et al.[21] and dangling end parameters reported by Bommarito et al.[22] Additionally, we needed three energy parameters that have not been conclusively established in literature: the correction for DNA in a solution containing magnesium ions, the free energy contribution of the coaxial stack at a nick, and the contribution of "coaxial stack dangles", dangles adjacent to two coaxial helices (e.g., the β domain in state I(m, n) in Figure 1D).

Several different models for the energetics of DNA hybridization in magnesium solutions have been proposed and investigated,[23−25] but it was unclear which model would best apply to our DNA complexes. However, all of these models suggested that the energetics of DNA hybridization were similar in a 1 M Na^+ solution as in a 10 mM Mg^{2+}. Consequently, for the calculation of the default toehold binding energies presented in the main paper, we simply used the energetics parameters for DNA hybridization in a 1 M Na^+ solution. Application of Owczarzy et al.'s magnesium correction formula[25] yielded standard free energies that were slightly less energetic than reported in the main paper (see Text S3, Supporting Information). The BM rate constants predicted using these modified parameters showed similar quality of fit (see Figure S2D, Supporting Information).

Similarly, several different sets of coaxial stacking energetics parameters have been reported.[23,26−28] In some cases, these values differed from each other by up to 2.5 kcal/mol. Using the values reported by Protozanova et al.,[26] with minor adjustments for temperature[30,31] (see Text S3), yielded the best agreement between model-predicted BM rate constants and experimental best-fit rate constants. Using values reported by Pyshnyi et al.[27] yielded similar energetics for S and Ss but significantly stronger binding energetics for Sw (by about 0.8 kcal/mol) and led to slightly worse quality of fit for BM rate constants (see Figure S2A). Furthermore, Pyshnyi et al.[29] later reported that the energetics of coaxial stacking near a nick depends significantly on the nearest neighbors (bases one away from the nick). Note that several publicly available and commonly used DNA folding programs, such as mFold[19,61] and NUPACK,[17] by default do not include any coaxial stacking term for their energetics calculations.

To the best of our knowledge, the energetics of coaxial stack dangles has not been comprehensively characterized. DNA folding programs such as mFold and NUPACK by default use the energetics parameters of "terminal dangles" (those adjacent to a single helix, reported by Bommarito et al.[22]) as a substitute, but there is no a priori reason to believe that the energetics of coaxial stack dangles would be correlated with the energetics of terminal dangles. For our default energetics calculations, we assumed that the ΔG° contribution from these coaxial stack dangles is 0. Using the energetics parameters of terminal dangles instead of 0 for coaxial stack dangles led to slightly worse quality of fit for BM rate constants (see Figure S2B,C).

The default toehold binding energies (as described above using works by Santa Lucia et al.,[21] Bommarito et al.,[22] and Protozanova et al.[26]) are reported in Table 2. These values were the ones used to generate the model-predicted BM rate constants in Figures 3, 4, and 6. A detailed step-by-step procedure of how to derive the toehold binding energies is shown in Text S3.

The uncertainty of the absolute values of the toehold binding energies may be 1 kcal/mol or more, but uncertainty in the relative binding energies of toeholds of different lengths is smaller, because the latter depend primarily on the well-established energy parameters for stacks and terminal dangles. A consequence of the former, however, is that the value of parameter k_b is known to only 1 order of magnitude. Using some of the alternative energetics parameters discussed above, the fitted value of k_b ranges from 0.5 s^{-1} to 7 s^{-1}.

Buffer Conditions. DNA oligonucleotides were stored in TE buffer (10 mM Tris·HCl pH balanced to 8.0, with 1 mM EDTA·Na_2, purchased as 100× stock from Sigma-Aldrich) at 4 °C. Directly preceding experiments, TE buffer with 62.5 mM $MgCl_2$ was added at 1:4 ratio to the sample, achieving a final $MgCl_2$ concentration of 12.5 mM (of which 1 mM is bound to EDTA). This buffer is henceforth known as "TE/Mg^{2+}" buffer. All experiments and purifications were performed at 25 ± 0.5 °C, with temperature controlled using an external temperature bath.

Substrate Purification. DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT), with HPLC purification. Where applicable, fluorophores were attached by IDT as well. Concentrations were determined from the measured absorbance at 260 nM using an Eppendorf Biophotometer and the calculated extinction coefficients.[65]

Substrate complexes S, Ss, and Sw, and the reporter complex R were further purified by nondenaturing (ND) polyacrylamide gel electrophoresis (PAGE) as follows: Strands for each sample were prepared with nominally correct stochiometry at 20 μM and annealed. The samples were then run on 12% ND PAGE at 180 V for 6 h.

The acrylamide (19:1 acrylamide:bis) was diluted from 40% acrylamide stock (Ambion). ND loading dye containing xylene cyanol FF in 50% glycerol was added to all samples, achieving final gycerol concentration of 10% by volume. Gels were run at 25 °C using a Novex chamber with external temperature bath.

The proper bands were cut out and eluted in 2 mL of TE/Mg^{2+} buffer for 2 days. Purified complexes were quantitated by measurement of absorbance at 260 nm using an Eppendorf Biophotometer and calculated extinction coefficients.[65] Typical yields ranged from 40% to 60%.

Annealing. All annealing processes were performed with an Eppendorf Mastercycler Gradient thermocycler. The samples were brought down from 95 °C to 20 °C at a constant rate over the course of 90 min.

Spectrofluorimetry Studies. Spectrofluorimetry studies were done using a SPEX Fluorolog-3 (Horiba) with 1.6 mL synthetic quartz cells (Hellma 119-004F). The excitation was at 588 nm, while emission was at 602 nm (optimal signal for ROX fluorophore). In all spectrofluorimetry experiments, the total reaction volume was 1.5 mL. For net reaction studies in which the concentration of the reporter was in excess of 1 nM (see Supporting Information Tables S2, S3, and S4), 2 nm band-pass slits were used for both excitation and emission monochrometers; for experiments where reporter concentration was less than 1 nM, 4 nm slits were used. Experiments shown in Figures 2D, 3A, 4A, 5B, and 7B, as well as experiments not shown, were done with integration time of 10 s for every 60 s time-point. The experiment shown in Figure 7A was done with integration time of 10 s for every 15 s time-point.

Prior to each experiment, all cuvettes were cleaned thoroughly: 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. For the slit size, concentrations, and times chosen, no measurable photobleaching was observed. Fluorescence measurements are linear in the concentration of the free fluorescent strand F. All experimental results were within the linear regime of

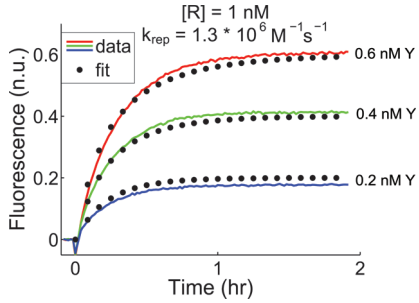

Figure 2. Kinetic characterization of the reporter complex R. "n.u." denotes normalized units; all fluorescent results were normalized so that 1 n.u. corresponds to 1 nM of free F. Reporter R was initially in solution at the indicated concentration, and Y was added to solution at $t \approx 0$ to achieve the final concentration displayed. The dotted line shows simulation traces of a second-order displacement reaction with rate constant $k_{rep} = 1.3 \times 10^6\ M^{-1}\ s^{-1}$.

the spectrofluorimeter detector, according to specification sheets provided by the manufacturer.

Fluorescence Normalization. Fluorescence is normalized so that 1 normalized unit (n.u.) of fluorescence corresponds to 1 nM of unquenched fluorophore-labeled strand F. This normalization is based on the fluorescence levels of annealed samples: a negative control with [R] = 30 nM, [X(0, 10)] = 20 nM and [S] = 0 nM, and a positive control with [R] = 30 nM, [X(0, 10)] = 20 nM and [S] = 10 nM. Day-to-day and sample-to-sample variations are estimated to be less than 5%.[16]

Carrier Strands. In our procedures, DNA sticks nonspecifically to pipet tips, so that serial dilutions led to stocks more dilute than expected.[16] Unfortunately, this loss is not consistent, so we could not compensate for tip loss with additional reagent. Instead, we introduced into all dilute stocks (1 μM and below) a nonreactive 20 nt poly-T "carrier" strand, at a concentration of 1 μM. Because pipet tip loss is nonspecific, the majority of DNA loss would be of the carrier strand, and serially diluted stocks are only slightly more dilute than expected. Poly-T strands have minimal influence on the reactions of other DNA molecules in this system.[16]

Parameter Fitting. The best-fit rate constants to experimental data were fitted using the 'fminunc' function in Matlab to minimize the error between experimental data and the reaction model. Sample code is shown in Text S4. The error is calculated as follows:

$$\text{error} = \sum_{t,\text{traces}} \frac{(F_d(t) - F_m(t))^2}{F_m(t)}$$

where $F_d(t)$ is the fluorescence value of the data at time t, and $F_m(t)$ is the fluorescence value predicted by the ODE model at time t. The sum of square error is divided by the expected fluorescence value $F_m(t)$ because the fluorescence readings yielded by the photomultiplier tube follow a Poisson distribution.

Rate constants were fitted one at a time; in every case, all but one rate constant were previously fitted. For example, the rate constant $k_{rep} = 1.3 \times 10^6\ M^{-1}\ s^{-1}$, established by experimentation and fitting (Figure 2), was used to fit the value of $k_{\{0,5\}}$. Matlab yielded $k_{\{0,5\}} = 1.0 \times 10^6\ M^{-1}\ s^{-1}$ as the best-fit value of $k_{\{0,5\}}$ for the following reaction network to fit the data in Figure 3A:

$$S + X(0,5) \xrightarrow{k_{\{0,5\}}} Y + L(0,5)$$

$$Y + R \xrightarrow{k_{rep}} F$$

Errors in the quantitation of DNA oligonucleotide concentrations can be reflected as errors in the inferred BM rate constants. For example, DNA strands and complexes being quantitated to be 5% more concentrated than their actual concentration would lead to an inferred rate constant 5% lower than the actual value.

Error bars shown in Figure 3B,C were generated using "leave one out" error: best-fit rate constants were generated for each pair of data traces (e.g., the 0.2 nM X(0, 5) and the 0.4 nM X(0, 5) traces of Figure 3A), and the standard deviation of these three pairs is calculated. The error bars show two standard deviations above and below the rate constant fitted using all three data traces. For subsequent figures, no error bars are plotted because each data point represents the best-fit BM rate constant to a single experimental trace.

For short invading toehold lengths ($n < 3$), the initial data points with fluorescence less than 1 nM were ignored in evaluating the error score. This is because in these experiments, there was an initial fast reaction that very quickly yielded about 0.5 nM of F; this fast reaction is assumed to be due to synthesis errors in the oligonucleotides and thus is not representative of the reaction of interest.

Results and Discussion

Reporter Complex Characterization. Figure 2 shows the kinetics of the reaction between product Y and reporter complex R. A second order reaction with rate constant $k_{rep} = 1.3 \times 10^6\ M^{-1}\ s^{-1}$ fits the data well at the concentrations tested; this parameter value was used for all subsequent modeling.

Toehold-Mediated Strand Displacement, $m = 0$. The rate constant $k_{\{0,5\}} = 1.0 \times 10^6\ M^{-1}\ s^{-1}$ yielded the best fit to the experimental data in Figure 3A. Similar experiments and simulations were done for other inputs X(0, n), Xs(0, n), and Xw(0, n) (data not shown), and all best-fit rate constants $k_{\{0,n\}}$ showed similar or better quality of fit to experimental data. The concentrations of reporter, substrate, and input used for each set of experiments are listed in Table S2, S3, and S4. The concentrations were chosen so that the experiments finished within a reasonable amount of time (between 10 min and 3 days). Subsequent analysis showed that the concentrations were below the critical concentration c_{crit} derived earlier.

The best-fit values of $ks_{\{0,n\}}$, $k_{\{0,n\}}$, and $kw_{\{0,n\}}$ are plotted semilogarithmically against the length of the toehold n in Figure 3B (small dots with error bars) and against the predicted toehold binding energies in Figure 3C. In both, the experimental best-fit rate constants follow a "kinked line" distribution, with an initial sloped region and an eventual flat region.

This kinked line behavior is predicted by our three-step model (shown as solid lines in Figure 3B,C). Recall eq 2, the expression for the BM rate constant of toehold exchange:

$$k_{\{m,n\}} \equiv k_{(\beta^m,\beta_m,\gamma^n)} = \frac{k_f\, k_{r(\beta^m)} k_b}{k_{r(\gamma^n)} k_{r(\beta^m)} + k_{r(\gamma^n)} k_b + k_{r(\beta^m)} k_b}$$

For the toehold-mediated strand displacement experiments in this section, $m = 0$, so $k_{r(\beta^0)} \gg k_b$, and the $k_{r(\gamma^n)} k_b$ term can be eliminated from the denominator because it is dominated by $k_{r(\gamma^n)} k_{r(\beta^0)}$ term.

In the flat region, the binding energy of the invading toehold is strong enough that $k_b > k_{r(\gamma^n)}$, and the expression can be approximated as:

$$k_{\{0,n\}} \equiv k_{(\beta^0,\beta_0,\gamma^n)} \approx \frac{k_f\, k_{r(\beta^0)} k_b}{k_{r(\beta^0)} k_b} = k_f \qquad (4)$$

Different values of k_f were fitted to the three different sets of traces (with S, Ss, or Sw). In general, higher G-C content appears to increase the hybridization rate of complementary strands (k_f). For the substrate and inputs using the γs^n toehold,

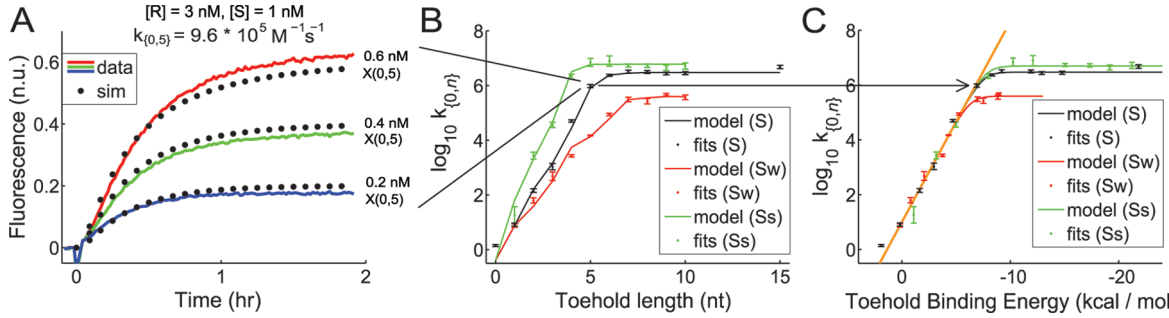

Figure 3. Toehold-mediated strand displacement experiments ($m = 0$). Results presented in this figure also use domains γs and γw in place of domain γ where specified. (A) Sample trajectories for $n = 5$. S and R were initially in solution at the displayed concentrations, and X(0, 5) was added to solution at $t \approx 0$ to achieve the final concentration displayed. The black dotted lines labeled "fit" denote simulations of a bimolecular reaction with with the experimental best-fit rate constant $k_{\{0,5\}} = 1.0 \times 10^6\ \mathrm{M^{-1}\ s^{-1}}$, and the reporter reaction with rate constant $k_{\mathrm{rep}} = 1.3 \times 10^6\ \mathrm{M^{-1}\ s^{-1}}$. The black lines spanning A, B, and C indicate that the three traces shown in A are represented by a single data point in B and C. (B) Summary of strand displacement rate constants plotted against the invading toehold length n. The y-axis shows the base-10 logarithm of the experimental best-fit values ("fits") and three-step model predicted values ("model") of $ks_{\{0,n\}}$, $k_{\{0,n\}}$, and $kw_{\{0,n\}}$, expressed in $\mathrm{M^{-1}\ s^{-1}}$. (C) Summary of strand displacement rate constants plotted against the calculated binding energy of the toehold. The orange line shows the asymptotic behavior predicted by the three-step model as k_f approaches ∞.

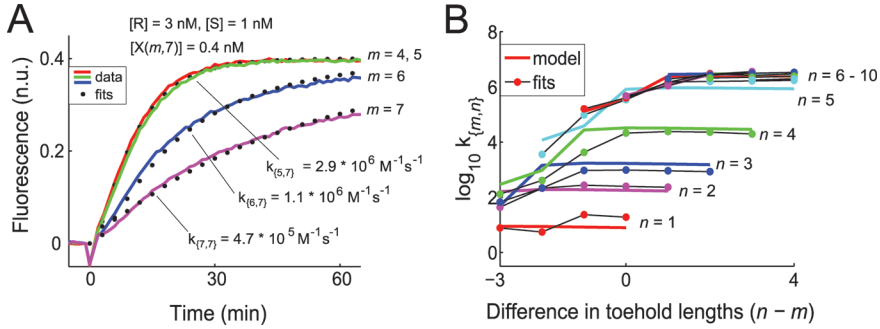

Figure 4. Toehold exchange. All results presented in this figure use only domain γ (and not γs or γw). (A) Sample trajectories for $n = 7$ for various m. R and S were present in solution initially, and X(m, 7) was added at $t \approx 0$. The black dotted lines labeled "fit" denote simulations of a bimolecular reaction with the experimental best-fit rate constants $k_{\{m,n\}}$ shown in the figure. (B) Summary of toehold exchange rates. The solid lines labeled "model" show the BM rate constants $k_{\{m,n\}}$, with $k_f = 3 \times 10^6\ \mathrm{M^{-1}\ s^{-1}}$, and $k_b = 1.0 \times (b/(b-m))^2\ \mathrm{s^{-1}}$. These predictions were not fitted to the data in this figure; they depended only on the values of k_f and k_b fitted from the earlier experiments on toehold-mediated strand displacement. The dots labeled "fits" show the experimental best-fit rate constants. Unlike in Figure 3, only one concentration of each X(m, n) was tested here.

the best-fit value of k_f was the highest at $6 \times 10^6\ \mathrm{M^{-1}\ s^{-1}}$. The values $k_f = 3 \times 10^6\ \mathrm{M^{-1}\ s^{-1}}$ and $k_f = 4 \times 10^5\ \mathrm{M^{-1}\ s^{-1}}$ fit best for substrates and inputs with the γ and γw toeholds, respectively.

In the sloped region, $k_{r(\gamma^n)} > k_b$, and we can approximate the BM rate constant expression as:

$$k_{\{0,n\}} \approx \frac{k_f k_{r(\beta 0)} k_b}{k_{r(\gamma^n)} k_{r(\beta 0)}} = \frac{k_f k_b}{k_{r(\gamma^n)}} \tag{5}$$

$$= k_b \frac{b-m}{2} e^{-\Delta G^\circ(\gamma^n)/RT} [M^{-1}] \tag{6}$$

Taking the logarithm (base 10) of the BM rate constant,

$$\log_{10}(k_{\{0,n\}}) \approx \log_{10}\left(\frac{k_b(b-m)}{2}\right) - \frac{\log_{10} e}{RT} \Delta G^\circ(\gamma^n) \tag{7}$$

and it can be seen that the slope and y-intercept of Figure 3B are predicted by the three-step model to be $\log_{10} e/RT \approx 0.732$ and $\log_{10}(k_b(b-m)/2)$, respectively. Taking the 0.732 slope as a given, the best-fit value of the y-intercept is calculated by Matlab to be 1.0. For toehold-mediated strand displacement, the length of the branch migration is $b - m = 20 - 0 = 20$ nt, so the corresponding best-fit value of k_b is $1.0\ \mathrm{s^{-1}}$. This value of k_b fit well for all three sets of experiments as expected, because k_b should be dependent only on the branch migration domain β_m, which is the same for all three sets of experiments.

We do not currently have a good method of estimating k_f from first principles: The γs toehold possesses a k_f about twice as large as the γ toehold, and about twenty times as large as the γw toehold. Thus, strong base-pairs seem to have a disproportionate effect on determining k_f. It is likely that k_f reflects the biophysics of hybridization nucleation.[32,60,69]

Toehold Exchange, $m > 0$. Experiments to characterize the kinetics of toehold exchange are similar to those of toehold-mediated strand displacement. Figure 4A shows the data and simulations using the best-fit $k_{\{m,n\}}$[73] for the experiments using the set of inputs X(m, 7), where m ranges between 4 and 7. The best-fit $k_{\{m,n\}}$ are plotted as dots in Figure 4B and compared against the values of $k_{\{m,n\}}$ predicted by our three-step model, using the value $k_f = 3 \times 10^6\ \mathrm{M^{-1}\ s^{-1}}$. The value of k_b is calculated as $1.0 \times (b/(b-m))^2\ \mathrm{s^{-1}}$ to account for changes in the length of the branch migration region, in accordance with previous unbiased random walk models of branch migration kinetics.[42,43] Notably, no new model parameters were fitted to generate the model-predicted rates in Figure 4B.

(67) Flamm, C.; Hofacker, I. L. *Monatsh. Chem.* **2008**, *139*, 447.
(68) Sun, W.; Mao, C.; Liu, F.; Seeman, N. C. *J. Mol. Biol.* **1998**, *282*, 59.
(69) Yuan, B.; Zhuang, X.; Hao, Y.; Tan, Z. *Chem. Commun.* **2008**, *48*, 6600.
(70) Feller, W. *An Introduction to Probability Theory and Its Applications*; John Wiley and Sons: Singapore, Singapore, 1958; chapter 5.

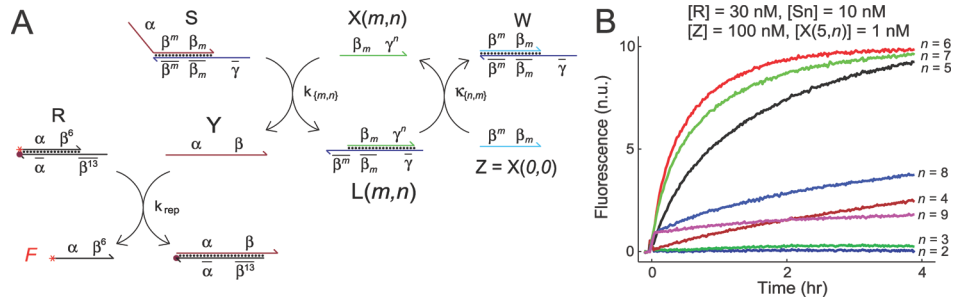

Figure 5. Using toehold exchange to construct a simple catalytic system. (A) Schematic of the catalytic system. Strand Z (the same as X(0, 0)) is present in solution initially with S and R. In the absence of X(m, n), reaction between Z and S to form W and Y is slow (≈ 1 M^{-1} s^{-1}). Both of reverse reactions (Y + L(m, n) → S + X(m, n) and X(m, n) + W → Z + L(m, n)) are present and modeled, with rates constants $k_{\{n,m\}}$ and $\kappa_{\{m,n\}}$, respectively. (B) Sample catalysis results. Catalyst X(5, n) was added $t \approx 0$.

The toehold exchange experimental results showed that the predicted BM rate constants agreed reasonably well to the experimental best-fit rate constants (with most values correct to within a factor of 2, and all values correct to within a factor of 10). Observed differences between the model-predicted and best-fit rate constants are at least partially due to errors in the binding energies of the toeholds. Unintended secondary structures, particularly in the cases of Xs, Ss, Xw, and Sw, may have complicated the kinetics of strand displacement. Finally, strand synthesis impurities and inaccuracies in DNA extinction coefficients may have caused error in DNA quantitation, that would have been reflected as an error in observed rate constants.

The BM rate constants of toehold exchange ($k_{\{m,n\}}$) were similar in value to those of toehold-mediated strand displacement $k_{\{0,n\}}$ when $m \leq n$. When $m > n$, the rate constant of toehold exchange decreases sharply with increasing m, both experimentally and as predicted by the model. This can be understood intuitively: Below the critical concentration, the initial association of X(m, n) and S by the toehold is the rate-limiting step. When an input molecule X(m, n) binds to the substrate S, the resulting complex can dissociate either into Y + L, or into X(m, n) + S. The probability of the former case is determined by the relative binding energies of the incumbent and invading toeholds and is evaluated (using eq 1) as:

$$\Pr(\mathrm{I} \rightarrow \mathrm{Y} + \mathrm{L}) \approx \frac{k_{\mathrm{r}(\beta^m)}}{k_{\mathrm{r}(\beta^m)} + k_{\mathrm{r}(\gamma^n)}} \tag{8}$$

$$= \frac{e^{\Delta G^\circ(\beta^m)}}{e^{\Delta G^\circ(\beta^m)/RT} + e^{\Delta G^\circ(\gamma^n)/RT}} \tag{9}$$

As $n - m$ increases, the probability of dissociating into Y + L approaches 1, and the BM rate constant approaches the hybridization rate of the toehold k_f. As $n - m$ decreases, the probability of dissociating into Y + L and the BM rate constant decays roughly exponentially.

Catalysis Based on Toehold Exchange. From the previous section, the kinetics of toehold exchange are seen to approach that of toehold-mediated strand displacement ($m = 0$) when $n \geq m$ and the initial concentrations of S and X(m, n) are sufficiently low. Our model similarly suggests that the kinetics of the reverse reaction approaches that of toehold-mediated strand displacement when $m \geq n$. It follows then that when $m \approx n$, and the concentrations are below the critical concentrations c_{crit}, both the forward and the reverse reactions are reasonably fast, and the system is in dynamic equilibrium.

We utilize this property to construct a simple reaction network in which X(m, n) acts as catalyst, sustainably speeding up the net reaction S + Z → Y + W (Figure 5A). Here X(0, 0) is renamed Z because it serves a different role from X(m, n). In the absence of catalyst X(m, n), sequestration of toehold β^m in S should prevent S from reacting with Z; strand Z was shown previously to react with S with a rate constant of about 1 M^{-1} s^{-1}, which is consistent with previous studies on blunt end strand exchange kinetics,[64] and is slow enough to be neglected for the purposes of this section. When the two toeholds are of similar intermediate lengths, toehold exchange will ensue rapidly, with X(m, n) reacting with S to yield Y and L. Partially double-stranded L, due to its newly activated toehold β^m, can react quickly with Z to form W and regenerate X(m, n). When excess Z and S are present, each X(m, n) molecule should be able to turn over multiple reactions. Z can therefore be considered a fuel for the catalytic conversion of substrate S to product Y, and W can be considered the waste product.

The net reactions of the system are summarized as:

$$\mathrm{X}(m,n) + \mathrm{S} \underset{k_{\{n,m\}}}{\overset{k_{\{m,n\}}}{\rightleftharpoons}} \mathrm{Y} + \mathrm{L}(m,n)$$

$$\mathrm{Z} + \mathrm{L}(m,n) \underset{k_{\{m,n\}}}{\overset{k_{\{n,m\}}}{\rightleftharpoons}} \mathrm{X}(m,n) + \mathrm{W}$$

$$\mathrm{Y} + \mathrm{R} \xrightarrow{k_{\mathrm{rep}}} \mathrm{F}$$

Previously, we used $k_{\{m,n\}}$ as shorthand notation for $k_{(\beta^m, \beta_m, \gamma^n)}$. Here, we also need to consider the BM rate constant $k_{(\gamma^n, \beta_m, \beta^m)}$ of the reaction between Z and L(m, n), which we denote as $\kappa_{\{n,m\}}$. In general, $k_{\{m,n\}} \neq \kappa_{\{m,n\}}$, because different sequences are used for the initiation and dissociation toeholds. Unlike the values of $k_{\{m,n\}}$, we do not have any experimental verification on the accuracy of the predicted values of $\kappa_{\{m,n\}}$. The values of $\kappa_{\{m,n\}}$ were evaluated using the calculated binding energies and assuming $k_{f2} = 3 \times 10^6$ M^{-1} s^{-1}.

The efficacy of X(m, n) as a catalyst depends on both the relative and the absolute values of m and n. Figure 5B shows the catalytic activities of the set of inputs X(5, n), with n ranging between 2 and 9. When the value n is 2, 3, or 4, the X(m, n) + S reaction is the rate-limiting step of the catalytic cycle. The production rate of F is consistently slow, with $n = 2$ being slowest due to having the slowest forward reaction. When $n = 8$ or 9, the Z + L(m, n) reaction is the rate-limiting step; catalyst X(m, n) is not released rapidly from L by reaction with Z. Consequently, the first turnover is fast, while all subsequent ones are slow, with $n = 9$

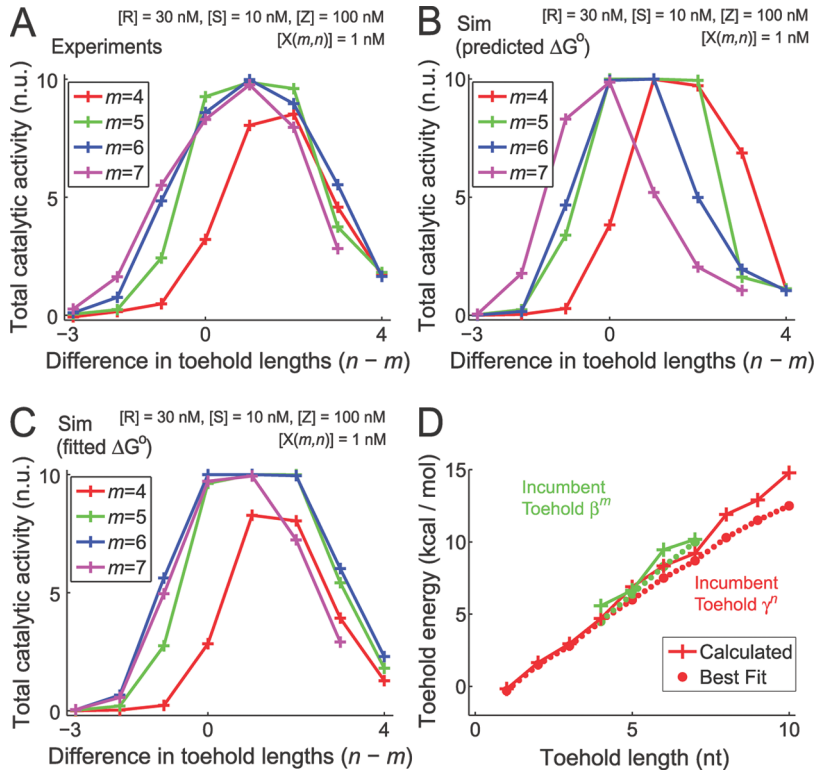

Figure 6. Characterization of catalysis based on toehold exchange. (A) Experimental results on the total catalytic turnover over 4 h. Catalytic activity is greatest when the length of the invading toehold (n) is slightly larger than that of the incumbent toehold (m). (B) Simulation results using default energy parameters. In all simulations, $k_f = 3 \times 10^6$ M^{-1} s^{-1}, and $k_b = 1.0 \times (b/(b - m))^2$ M^{-1} s^{-1}. Results are qualitatively similar to experimental data, but show more variation for $n > m$ than was experimentally observed. (C) Simulation results using fitted energy parameters. Energy parameters were fitted by hand to yield good quantitative fitting to the experimental data. (D) Differences between energy parameters used in parts B and C.

yielding the faster initial turnover and slower subsequent speed. When $n = 5$, 6, or 7, both the binding and the release of X(m, n) are fast, reflected in the data by the fast catalytic production of Y: In Figure 5B, the reactions with X(5, 5), X(5, 6), and X(5, 7) approach completion despite substrate S being at a 10-fold higher concentration than the catalyst.

In these catalytic experiments, the initial concentration of Z was 10-fold higher than that of S; this was done to minimize the slowdown as [Z] decreases through the course of the reactions. One side effect of this concentration disparity is to slightly favor the catalytic activity of strands with shorter dissociation toeholds ($m < n$): The reaction between Z and L has higher rate (as opposed to rate constant) and thus less likely to be the rate-limiting step.

The final normalized fluorescence values at $t = 4$ h, which we call the total catalytic activities, are plotted in Figure 6A for experiments with $m = 4$, 5, 6, and 7. Because [X] = 1 nM for these experiments, the total catalytic activities (in nM) corresponds to the turnover (number of reactions catalyzed on average by each catalyst molecule during the course of the 4 h experiments). The total catalytic activities predicted by the three-step model using the default energy parameters are shown in Figure 6B. For the modeling in this section, the full three-step models of both the S + X(m, n) ⇌ L(m, n) + Y and the L(m, n) + Z ⇌ X(m, n) + W reactions were simulated. The reporter reaction Y + R → F was modeled as a bimolecular reaction with the previously measured rate constant because the k_f and k_b for that reaction have not been characterized.

The simulation results agreed only qualitatively with experimental data; the discrepancy between the model predictions and results can likely be attributed to inaccuracies in the calculated binding energy of the toeholds, due to the uncertainties discussed in Materials and Methods. The three-step model is sensitive to small changes in the relative binding energies of the toeholds because of the exponential role that energy plays in determining rate constants.

To test this hypothesis, binding energies for toeholds were individually fitted to the catalytic data; the best-fit results are shown in Figure 6C and show significantly better agreement with experimental data. Although there were many free parameters in this fit (10 for the invading toehold energies and 4 for the incumbent toehold energies), the fitted values were all relatively close to predicted (Figure 6D). This provides evidence in support of both the three-step model of toehold exchange, as well as the binding energy models.

Limitations of the BM Rate Constant. The BM rate constant is a reasonably accurate predictor of the kinetics of toehold exchange only when [X(m, n)] and [S] are always below a critical concentration (see Text S2), shown earlier as eq 3:

$$c_{crit} = \frac{0.1}{k_f} \times \frac{k_{r(\gamma^n)} k_{r(\beta^m)} + k_{r(\gamma^n)} k_b + k_{r(\beta^m)} k_b}{k_b + k_{r(\beta^m)}}$$

The critical concentration decreases as the binding energies of the toeholds β^m and γ^n increase. For the systems that we investigated in this paper, the lowest critical concentration occurs in the case of (m, n) = (7, 10), at the value 3.2 nM using the value of $\Delta G°(\beta^7) = -10.2$ kcal/mol fitted in the previous section).

We experimentally test the accuracy of our analysis by comparing toehold exchange kinetics using inputs X(m, 10) at high and low concentrations (Figure 7). At low concentrations of [S] = 1 nM and [X(m, 10)] = 0.4 nM, the kinetics of all

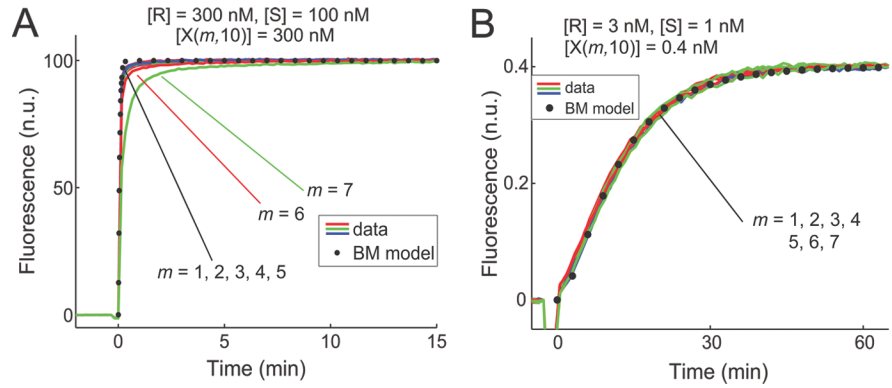

Figure 7. Limitations of models. (A) Toehold exchange reactions run at high concentration. The black dotted line shows the predicted behavior for all m under the BM and predicts reaction kinetics significantly faster than experimentally measured for $m = 6$ and 7. The data and three-step-model results for $m = 6$ differ significantly from those of $m = 7$. The energies used to calculate the reaction rates are those fitted in Figure 6D. (B) Toehold exchange reactions run at low concentration. The full three-step model and the BM rate constants produce nearly identical results for all of the strands X(m, 10) (shown by the black dotted line).

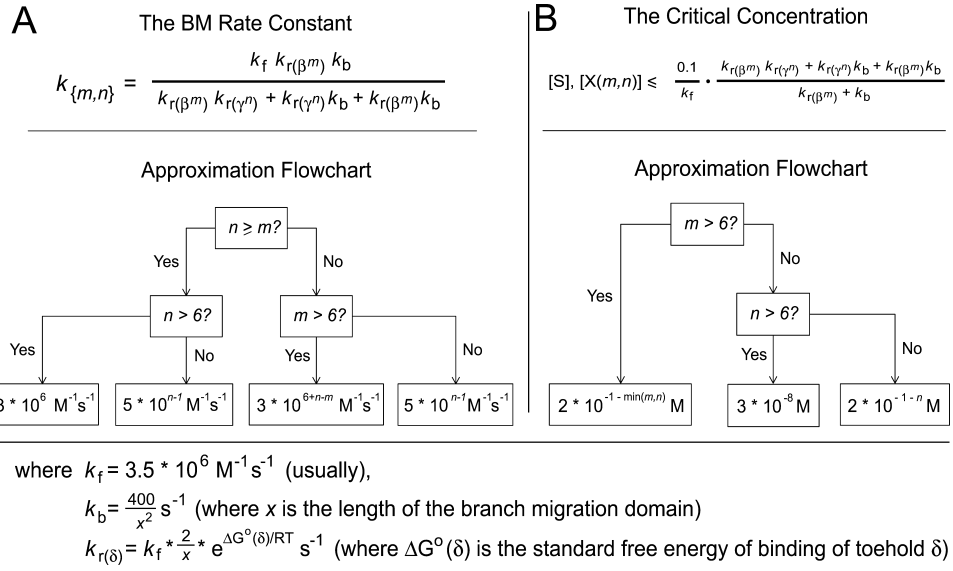

Figure 8. (A) Flowchart for order of magnitude estimate of $k_{\{m,n\}}$. This flowchart assumes domains sequences with average binding strength and no secondary structure and approximates the rate constant at 25 °C and 11.5 mM Mg^{2+} or 1 M Na^{+}. (B) Flowchart for estimating the critical concentration, below which the BM rate constant ($k_{\{m,n\}}$) accurately predicts kinetics.

trajectories are well characterized by the BM rate constants (which are nearly identical for all $m \leq 7$). In constrast, at high concentrations of [S] = 100 nM and [X(m, 10)] = 300 nM, kinetics differ significantly from those predicted by the BM rate constants (black dotted line) for $m = 7$.

Discussion

By characterizing and quantitatively modeling toehold exchange, this work serves as a "user's manual" for using toehold exchange as a modular design component for hybridization-based dynamical DNA systems. The expressions for the BM rate constant (2) and the critical concentration (3) can be evaluated given the sequences of the nucleic acid strands involved. Figure 8 shows an even simpler method of estimating of these two values based on only the lengths of toeholds involved. These estimates assume average strength toeholds and are derived in text S5. A Matlab script for computing these values are shown in text S6.

Despite the success of the three-step model, some aspects are still imperfectly understood or obviously inaccurate: First, the reasons for dependence of the DNA strand association rate constant (k_f) on the sequence of the toehold domain are not well-understood, differing by a factor of 20 from the fastest sequence to the slowest. The hybridization rate constant k_f is thought to depend on hybridization nucleation (i.e., the rate of formation of the first few base pairs).[32,60,69] However, the sequence dependence of k_f is not yet quantitatively predictable, even for secondary structure-free strands.

Second, branch migration is modeled as a single phenomenological rate (k_b) with no intermediate states. This is not only physically inaccurate, but also difficult to predict from first principles. Previous studies have modeled three-stranded branch migration as a random walk process, with time scale $N^2 t_s$, where N is the length of the branch migration domain and t_s is the time scale of an elementary step.[42,43] For our toehold-mediated strand displacement experiments, the length of the branch migration domain β_0 is 20 nt. Naively, our fitted value of $k_b = 1.0\ s^{-1}$ implies an elementary step time scale of $1/1.0\ s^{-1} \times 1/20^2$ steps = 2.5 ms/step. This value is significantly higher than previously reported results on the rate branch migration of genomic-length DNA (between 10 and 100 μs).[42−44]

However, it is likely that our phenomenological k_b incorporates not only the time need for branch migration but also the probability of displacement failure: After X(m, n) and S bind to each other by toehold γ^n, branch migration by the β_m domain ensues. However, branch migration is inherently a random walk process, and sometimes the junction moves backward. A significant fraction that start the branch migration process will return to the initial position where X(m, n) is bound to S by only γ^n. In this state, X(m, n) may dissociate before restarting branch migration. The probability of this occurrence can be estimated mathematically: When n is small, $k_{r(\gamma^n)} \gg k_b$, and the probability of dissociation before restarting branch migration is large and can be approximated as 1. The probability that an unbiased random walk starting at +1 will reach +20 before reaching 0 is simply 1/20.[70] Consequently, the actual rate constant of branch migration would then be 20 s^{-1}, corresponding to an elementary step time scale of 125 μs/step, which is much more consistent with previous reported values.

Furthermore, similarly to k_f, the value of k_b may be sequence-dependent; varying the sequence of the branch migration region to discover the range of k_b will be helpful in further understanding the conditions under which our model is accurate. Four-stranded branch migration has been observed to be sequence dependent;[68] three-stranded branch migration can be expected to be more so because of the predominant role that secondary structure can play in single-stranded DNA.[68,54]

Finally, it is hoped that this model is applicable across a range of temperatures and salt concentrations. For example, temperature and salt corrections to the binding thermodynamics of DNA are well-understood,[21,23,25] and the values of $k_{r(\beta^m)}$ and $k_{r(\gamma^n)}$ can be calculated for different conditions. When the effects of these conditions on k_b are known, it should be possible to predict accurate BM rate constants $k_{(\beta^m, \beta_m, \gamma^n)}$ across a wide range of sequences and conditions without performing any additional experiments (for molecules without significant unwanted secondary structure).

We expect that the same biophysical principles used here to estimate the rate constants of DNA strand displacement are also applicable for understanding toehold exchange within RNA and synthetic nucleic acid analogues (such as LNA and PNA). However, because these other nucleic acid systems generally have stronger Watson−Crick binding interactions and less specificity (for example, G-U wobbles in RNA are signficantly more thermodynamically favorable than G-T wobbles in DNA), unwanted secondary structure may be more difficult to avoid. It is unclear whether higher temperatures and/or more stringent sequence design would be sufficient for obtaining the degree of predictability observed here for DNA toehold exchange kinetics.

Further characterization of the toehold exchange process may lead to improved understanding of the biophysics of nucleic acid hybridization, branch migration, and displacement. Rational design is based on reliable and modular components; detailed characterization of underlying mechanisms and principles will ease the construction of superior nucleic acid logic gates, thresholds, and amplification elements. By interfacing these hybridization-based nucleic acid constructions with other nanoscale engineering accomplishments, DNA, RNA, and synthetic nucleic acids can be used for regulation and timing of more general chemical and biological processes. Previous works in which DNA controlled other nanoscale chemistries[33−40] can be expanded to allow finer control based on the programmed kinetics of DNA. Such advances will enable the construction of robust chemical reaction networks that allow embedded control of biology and chemistry, such as monitoring and modulating the concentrations of biologically relevant molecules.

Acknowledgment. We thank Xi Chen for a very careful reading of this paper and useful suggestions regarding the binding energy calculations. We thank Karthik Sarma for useful suggestions regarding rate constant fitting. We thank Niles Pierce, Justin Bois, and Joe Zadeh for discussion on the energetics parameters used by NUPACK. We thank Niles Pierce, Anne Condon, and Victor Beck for many helpful suggestions in the revision of the manuscript. We thank Bernard Yurke, Georg Seelig, and Joseph Schaeffer for insightful discussions. D.Y.Z. and E.W. were supported by NSF grants 0506468, 0622254, 0533064, 0728703, and 0832824. D.Y.Z. is supported by the Fannie and John Hertz Foundation.

Supporting Information Available: Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

JA906987S

(71) On the basis of the dependence of k_{f1} on the sequence composition of the toehold, we believe that k_{f2} is likely closest in value to $k_{f1} = 3 \times 10^6$ M^{-1} s^{-1} for inputs X(m, n) with toeholds containing both strong G-C pairs and weak A-T pairs. However, we have no way of knowing the actual value of k_{f2} without performing a new series of experiments on the β^m toehold, similar to those shown in Figure 3.

(72) To take a simple example, consider when the invading and incumbent toeholds are both 4 nt but the branch migration domain is 1000 nt long. Molecules will spend a significant amount of time in a three-stranded complex, with the branch migration junction vacillating near the middle of the branch migration domain. Our model, with the $2/(b - m)$ correction, will account for the high multiplicity of branch migration microstates.

(73) ODEs used for fitting $k_{\{m, n\}}$ simulated the two reactions X(m, n) + S → Y + L(m, n) and Y + R → F. The reverse reaction Y + L(m, n) → S + X(m, n) is not included because (1) it is desireable to fit one parameter at a time and (2) because the value of the reverse rate constant $\kappa_{\{n, m\}}$ is constrained poorly by data from experiments of the type shown in Figure 3A. Furthermore, the concentration of the reporter R is in 3× excess over that of the substrate S, and the rate constant of the reaction between Y and R is high; consequently, it is expected that the reaction between Y and R will dominate that between Y and L(m, n). Fitting $k_{\{m, n\}}$ to a full model including the reverse reaction with its rate constant $\kappa_{\{m, n\}}$ set to the value predicted by our model yielded values very similar to those presented in the paper (difference of no more than 10%).