The design space of strand displacement cascades with toehold-size clamps

Boya Wang[1⋆], Chris Thachuk[2⋆⋆],
Andrew D. Ellington[1⋆⋆⋆], and David Soloveichik[1†]

[1] University of Texas at Austin, Austin, USA
[2] California Institute of Technology, Pasadena, USA

Abstract. DNA strand displacement cascades have proven to be a uniquely flexible and programmable primitive for constructing molecular logic circuits, smart structures and devices, and for systems with complex autonomously generated dynamics. Limiting their utility, however, strand displacement systems are susceptible to the spurious release of output even in the absence of the proper combination of inputs—so-called leak. A common mechanism for reducing leak involves clamping the ends of helices to prevent fraying, and thereby kinetically blocking the initiation of undesired displacement. Since a clamp must act as the incumbent toehold for toehold exchange, clamps cannot be stronger than a toehold. In this paper we systematize the properties of the simplest of strand displacement cascades (a translator) with toehold-size clamps. Surprisingly, depending on a few basic parameters, we find a rich and diverse landscape for desired and undesired properties and trade-offs between them. Initial experiments demonstrate a significant reduction of leak.

1 Introduction

DNA strand displacement is a powerful mechanism for molecular information processing and dynamics [11]. A strand displacement reaction is the process where two strands hybridize with each other and displace a pre-hybridized strand. The displaced strand could then serve, in turn, as the displacing strand for downstream strand displacement events. Through concatenation of strand displacement reactions, a variety of programmable behaviors have been experimentally achieved, such as performing logical computation [5], engineering molecular mechanical devices [4], and implementing chemical reaction networks [1]. In strand displacement cascades, single-stranded DNA typically fulfills the role of *signals* that carry information, while pre-hybridized DNA complexes drive their interaction (and are consequently called the *fuels*).

⋆ bywang@utexas.edu. Supported by NSF grants CCF-1618895 and CCF-1652824.
⋆⋆ thachuk@caltech.edu. Supported by NSF grant CCF-1317694.
⋆⋆⋆ andy.ellington@mail.utexas.edu. Supported by NSF grant DBI-0939454, international funding agency ERASynBio 1541244, and by the Welch Foundation F-1654.
† david.soloveichik@utexas.edu. Supported by NSF grants CCF-1618895 and CCF-1652824.

Although the DNA strand displacement mechanism has shown great promise for programming molecular systems, the current scale of these systems remains limited. The main obstacle is *leak*, which occurs when undesired reactions get spontaneously triggered in the absence of triggering strand. Since leak results from a spurious interaction of the fuel complexes, fuels are necessarily kept at low concentration to reduce the leak reaction, which limits the general speed of the cascade. Instead of seconds, complex cascades often take hours [5].

To combat leak, a number of approaches have been tested such as introducing Watson-Crick mismatches [2, 3], or adding a threshold species that can consume the leaked signal at a faster rate than it propagates to downstream components [5, 8]. It is understood that leak occurs as a result of fraying at the end of a double helix which exposes a nucleation point for spurious displacement. Adding 1 to 3 nucleotides as the *clamp* domains [5, 8, 9] has proven useful in reducing undesired leak, since leak can only occur after the entire clamp and one or more additional nucleotides that are adjacent to the clamp fray. Clamps, in some form, are now commonly used in the majority of strand displacement systems.

Since longer clamps should better prevent fraying, and the probability of spontaneously opening the end of double-stranded helix decreases exponentially with the length of the clamp, we want to make the clamps as long as possible. However, with clamp domains, the intended strand displacement reaction must be a toehold exchange reaction, which limits the size of the clamp to that of a toehold [12]. In this paper, we extend the size of clamp to its maximum—toehold size—and generalize a design principle for strand displacement systems: Every fuel has a toehold-size clamp, and every reaction is a toehold exchange reaction. We consider the simplest kind of strand displacement cascade—a cascade of translators (which are logically equivalent to repeater gates). In a single translator, a signal strand serves as the input and through a series of strand displacement reactions, an output signal strand is produced whose sequence is independent of the input strand. Chaining single translators allows us to build systems of translators. Such translator chains have been used in a molecular automata system that can selectively target cellular surface [6]. Translators can also be composed to perform logic *OR* computation through having two translators convert two different input signals to the same output signal. A catalytic system can be constructed if a translator chain's output is the same as its input. Although translators are the simplest kind of strand displacement module, they can already exhibit complex and useful behavior.

Our simple formulation allows rigorous formal arguments about leak reduction and other desired properties that affect the intended reaction pathway. We prove that although every reaction is reversible, the system completion level does not decrease arbitrarily with the depth of the cascade, which allows long cascades to be constructed. We show that by adjusting two parameters which define the length of the double-stranded region (N) in a fuel and the minimal distance between two fuels (*shift*), a variety of schemes can be achieved, each with unique properties. We prove a tradeoff theorem which says that no scheme

satisfies all the properties we could want. Thus understanding the taxonomy of schemes is necessary to make proper design choices.

To analyze properties of interest for the various schemes, we use a thermodynamic argument which assumes that "enthalpy" and "entropy" are the dominating factors deciding whether a configuration is favorable or not. More specifically, we assume the main contribution of enthalpy is the number of bound toehold-size domains and that of entropy is the number of separate complexes.[1] Thus when comparing two configurations, if the number of bound toehold-size domains is the same while one configuration has n fewer separate complexes than the other, then this configuration is considered unfavorable as it incurs n units of entropic penalty relative to the other. Similarly, all else being equal, a configuration with n fewer bound toehold-size domains than another configuration is unfavorable and has a relative enthalpic penalty of n units.[2]

Recently, a leak reduction method relying on increasing a redundancy parameter N_r has been proposed where leak requires binding of N_r separate fuels ("*NLD* scheme" in [10]).[3] However, in the *NLD* scheme, a leaked upstream signal can start a cascade which gains one unit of enthalpy for every downstream strand displacement step. Thus to ensure the "leakless" property, the *NLD* system needs to have enough entropic penalty to compensate for this enthalpic driving force. In contrast, since every reaction is a toehold exchange reaction in our toehold-size clamp design, a leaked upstream signal cannot be driven by the enthalpy of forming new bonds downstream. Unlike the leak reduction method of [10], which solely relies on the entropic penalty, our design has an additional enthalpic penalty for leak, which scales as N/shift. Leak reduction based on the additional enthalpic penalty could be preferable especially at high concentrations, where the entropic penalty to leak is smaller. Note that high concentration regimes are of particular interest because they result in faster kinetics.

We also show that if the clamps in the *NLD* scheme are extended to toehold size, the *NLD* scheme can be categorized into one class of the toehold-size clamp design with the parameter N representing the length of double-stranded region and *shift* representing the length of one long domain. According to the taxonomy

[1] Although our use of the words enthalpy and entropy are meant to evoke the respective physical chemistry concepts, the mapping is not 1-1. We note especially that the contribution of forming additional base pairs to the free energy has both substantial enthalpic and entropic parts (which can be physically distinguished based on their temperature dependence).

[2] Roughly speaking, "one unit of enthalpic penalty" corresponds to an average of $l \cdot 1.5$ kcal/mol, where l is the length of the domain (typically 5-10 nucleotides for a toehold). "One unit of entropic penalty" at concentration C M corresponds to $\Delta G^{\circ}_{\text{assoc}} + RT\ln(1/C) \approx 1.96 + 0.6\ln(1/C)$ kcal/mol [7]. With these numbers, at roughly 650 nM concentration, binding an additional $l = 7$ domain is equal to one unit of entropy. At low concentrations the entropic penalty becomes dominant, while the enthalpic penalty prevails at high concentrations.

[3] Our length parameter N is related to the redundancy parameter of [10], but whereas we count the number of short domains, they count the long domains.

in this paper, the extended *NLD* scheme for any redundancy N_r has the property that toehold occlusion [5, 9] and spurious strand displacement—(partial) displacement of a strand on a fuel by a spurious invader—cannot be avoided. In this sense, the toehold-size clamp design principle has a broader design space, allowing for more flexibility in balancing desired and undesired properties.

We conclude with an experimental demonstration of the toehold-size clamp design with one set of parameters. Leak reduction compares positively both in terms of the kinetic leak rate and the maximum amount of leak ever generated with the previously proposed *NLD* schemes. (Although the absolute leak rate is smaller for our scheme, the lower completion level due to toehold exchange reaction reversibility results in an overall smaller "signal to noise ratio" compared with the *NLD* schemes.)

2 Design space of toehold-size clamp translators

2.1 System description

We first introduce the conventions used in this paper. We use the domain level abstraction for DNA strands. A domain represents a concatenation of DNA bases treated as a unit, which can hybridize to or dissociate from a complementary domain. Unlike the traditional representation that divides domains into two classes where long domain indicates irreversible binding and short (toehold) domain indicates reversible binding, here all the domains have equal length of a toehold. As a result, if two strands are only held by a single domain, they can dissociate (see Fig. 1a). We assume that all domains are orthogonal (no cross-talk). Note that this is a strong assumption because the size of a domain is restricted to that of a toehold and there are a limited number of distinct toeholds that could be designed (see Discussion). The desired pathway consists entirely of toehold exchange strand displacement reactions. Additionally, to capture leak, we consider blunt-end strand displacement, which is not preceded by toehold binding but rather is mediated by fraying at the ends of helices. We assume that fraying cannot open a whole domain. The unique domains can be aligned in a row, and their identity is represented as their horizontal position (i.e., numbering on top of Fig. 1b). Domains aligned vertically have the same or complementary sequence.

By a *domain instance*, we mean a particular domain on a particular complex. In contrast, when we refer to a (domain) *position*, we mean all domain instances that have the same or complementary sequences, and are drawn vertically aligned in our figures.

Domain instances can be either *single-stranded* or *double-stranded*. Single-stranded domain instances are subdivided into *toehold* and *overhang* types. Note that toehold domain instances initiate the toehold mediated strand displacement. Double-stranded domain instances are subdivided into *left flank*, *clamp*, and *unused*. Specifically, *clamp*s are the domain instances located at the right end of a double-stranded helix, *left flank*s are the domain instances located at

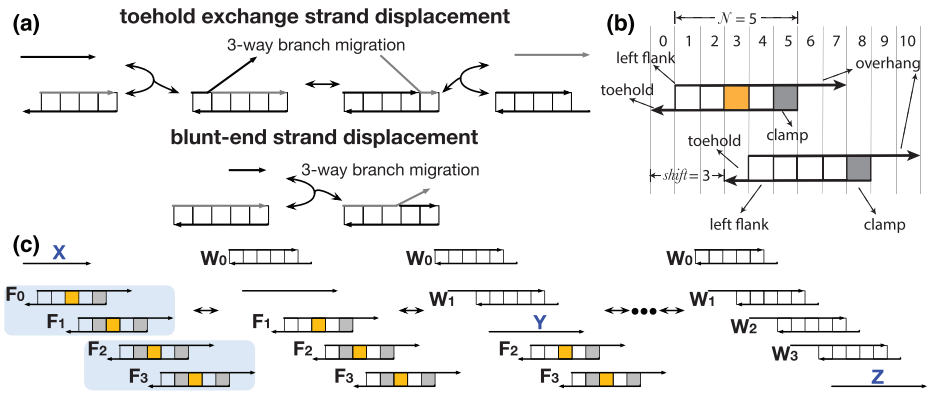

Fig. 1: (a) Fundamental reaction steps we consider. (b) The conventions of the toehold-size clamp design. (c) Desired reaction pathway of a 2-translator cascade ($X \to Y \to Z$). In the presence of the input signal strand, after 4 elementary translation steps, the signal strand X is translated to signal strand Z.

the left end of a double-stranded helix (see Fig. 1b). The remaining domain instances are *unused*. The name "clamp" comes from historical use, as structurally similar domains were added to previous schemes to "clamp-down" the ends of helices to reduce leak. Note that without the clamp domain on the second fuel in Fig. 1b, the overhang of the first fuel can initiate blunt-end strand displacement.

The coloring of double-stranded domain instances refers to whether or not there are toehold (orange) or clamp (gray) domain instances at the same position. In particular, the color orange indicates that there is a toehold domain instances at the same position If a domain position overlaps with both toehold and clamp domain instances, it is colored in both orange and gray (e.g., see Fig. 2). If a domain position does not have toehold or clamp domain instances, it is colored in white.

2.2 Translator design

A translator, composed of different fuels where each fuel is responsible for an elementary translation step, can translate an input signal strand to an independent output signal strand. When the input signal strand is present, it reacts with the first fuel displacing the top strand, which then serves as the input to trigger the downstream fuel (Fig. 1c). Note that the figure shows two translators, of two fuels each (since Y is sequence independent from X, and Z is sequence independent from Y).

To design a translator system with toehold-size clamps, two parameters are necessary and sufficient. We use the parameter N to represent the number of the double-stranded domains in a fuel and the parameter *shift* to represent the minimal distance between two single-stranded toeholds (see Fig. 1b). Since a

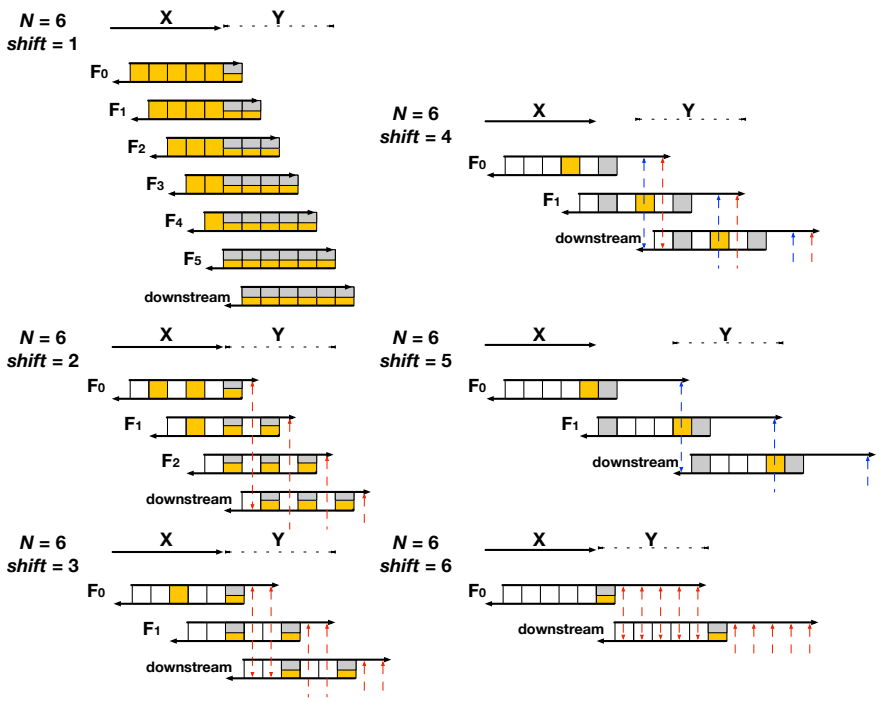

Fig. 2: The example schemes of translators with $N = 6$ and all values of *shift*. Every domain instance is colored according to the domain types of the domain instances at that position. The dashed red lines represent possible spurious strand displacement events between different fuels. The dashed blue lines represent possible toehold occlusion events when a toehold domain instance and an overhang domain instance exist in the same position.

toehold-size clamp is in every fuel, *shift* should be between 1 and N. We illustrate the diversity of schemes for $N = 6$ for all values of *shift* in Fig. 2.

Once the parameters N and *shift* are determined, the translator design and the types of each domain instance can be assigned. Using 0 indexing, we start from the 0th fuel which is responsible for the first elementary translation step (e.g. Fig. 1c). The toehold domain instance of the 0th fuel is located at position 0. Since the length of the double-stranded domains in a fuel is N, the clamp instance in this fuel lies at position N. To ensure that every desired reaction is a toehold exchange strand displacement reaction (i.e., there is a clamp), the number of overhang domain instances in a fuel should be $\mathit{shift} - 1$ (we write $\mathit{overhang} = \mathit{shift} - 1$).

Generalizing these rules, the domain types of each domain instance in every fuel can be determined. The position of the ith fuel (which is responsible for the $i + 1$ elementary translation step) is shifted to the right by $\mathit{shift} \cdot i$. Thus the toehold domain instance in the ith fuel lies at the position $\mathit{shift} \cdot i$; the left flank

domain instance lies at the position $shift \cdot i + 1$; and the clamp domain instance lies at the position $shift \cdot i + N$. How many fuels do we need to compose a single translator? Recall that single translator consists of as many fuels as are necessary to generate an output signal that is sequence independent of the input. Thus the number of fuels per translator is $\lfloor \frac{N+overhang}{shift} \rfloor$, which can be equivalently written as $\lfloor \frac{N+shift-1}{shift} \rfloor = \lfloor \frac{N-1}{shift} \rfloor + 1$.

2.3 Useful lemmas

In subsequent sections, we prove properties of schemes based on their parametrization of N and *shift*. Many of our arguments rely on showing whether regularly spaced positions with certain domain instances (such as toeholds) can intersect with other regularly spaced positions or intervals with different domain instances (such as overhangs). To simplify those arguments we first establish the following claims.

Lemma 1. *Let p, q, and r be natural numbers with $p > 0$ and $r - q > 0$. $\forall i \in \mathbb{N}, \exists j \in \mathbb{N}$ such that $j \cdot p + q = i \cdot p + r$ if and only if $r - q$ is a multiple of p.*

Proof. Fix any $i \in \mathbb{N}$. Suppose $j \cdot p + q = i \cdot p + r$, for some $j \in \mathbb{N}$. Then $j = i + \frac{r-q}{p}$ and $r - q$ must be a multiple of p. □

Lemma 2. *Let p, q, and r be natural numbers with $p > 1$ and $r - q > 0$. $\forall i \in \mathbb{N}, \exists j \in \mathbb{N}$ such that $j \cdot p + q$ is contained in the interval $[i \cdot p + r, (i+1) \cdot p + r - 2]$ if and only if $r - q - 1 > 0$ and $r - q - 1$ is not a multiple of p.*

Proof. Call $k \cdot p + q$ a *valid position*, for any $k \in \mathbb{N}$. Fix any $i \in \mathbb{N}$. First we consider the case when $r - q - 1 = 0$ (and thus $r = q + 1$). Consider any interval $[i \cdot p + r, (i+1) \cdot p + r - 2] = [i \cdot p + q + 1, (i+1) \cdot p + q - 1]$. Suppose by contradiction that a valid position $j \cdot p + q$ intersects the interval. Then we have the following: (i) $i \cdot p + q + 1 \leq j \cdot p + q$ which simplifies to $j \geq i + \frac{1}{p}$ and thus $j \geq i + 1$ (since $j \in \mathbb{N}$). (ii) $j \cdot p + q \leq (i+1) \cdot p + q - 1$ which simplifies to $j \leq i + 1 - \frac{1}{p} < i + 1$. Contradiction.

Finally, consider two cases when $(r - q - 1) > 0$: (1) $(r - q - 1)$ is a multiple of p. Then $i \cdot p + q + (r - q - 1) = i \cdot p + r - 1$ is a valid position and the next valid position occurs at $(i+1) \cdot p + r - 1$. Thus there is no valid position in the interval $[i \cdot p + r, (i+1) \cdot p + r - 2]$. (2) $(r - q - 1)$ is not a multiple of p. Let δ be the remainder; $1 \leq \delta \leq p - 1$. Thus the smallest valid position larger than $i \cdot p + q + (r - q - 1)$ occurs at $i \cdot p + r - 1 + \delta$. Since $1 \leq \delta \leq p - 1$, we know that it falls in the interval $[i \cdot p + r, (i+1) \cdot p + r - 2]$. □

3 Thermodynamic properties

An effective translator cascade ideally has a high signal-to-leak ratio, even at thermodynamic equilibrium. We next show that by varying scheme parameters N and *shift*, we vary the thermodynamic barrier to leak. We go on to show that while every translator cascade scheme is reversible, there is a lower bound on fraction of output signal when input is present, regardless of its depth.

3.1 Thermodynamic barrier to leak in translator cascades

Suppose there is a hypothetical experiment implementing a single copy of every fuel in a translator cascade of arbitrary depth, coupled with a single copy of a downstream reporter. For any scheme, a downstream reporter is identical to a fuel of that scheme but contains no overhang. Leak occurs when the top and bottom strands of the reporter dissociate into separate complexes — in typical experimental settings, this would increase a fluorescence signal used as a proxy to measure the produced output. To give a barrier to leak based on N and *shift*, we first develop some useful lemmas.

Lemma 3. *If a position contains an overhang domain instance, then it does not contain a clamp domain instance.*

Proof. If $shift = 1$ there are no overhang domain instances and we are done. Suppose $shift > 1$. The ith clamp domain instance lies at position $i \cdot shift + N$. The jth overhang domain instance lies at positions between $j \cdot shift + N + 1$ and $(j+1) \cdot shift + N - 1$. Let $r = N + 1$, $p = shift$ and $q = N$. By Lemma 2, the position that has a clamp domain instance cannot intersect the position that has an overhang domain instance since $r - q - 1 = 0$. □

Lemma 4. *Let c_i and c_{i+1} be the positions of the clamp domain instances for neighboring fuels i and $i+1$, respectively. Then positions in the (possibly empty) interval $[c_i + 1, c_{i+1} - 1]$ contain overhang domain instances.*

Proof. $c_i = i \cdot shift + N$ and $c_{i+1} = (i+1) \cdot shift + N$. The overhang domain instances for fuel i lie at positions between $i \cdot shift + N + 1$ and $(i+1) \cdot shift + N - 1$, establishing the claim. (Note that if the interval is empty, then fuels do not have overhangs and the claim in trivially true.) □

For the purposes of the next argument, it is convenient to refer to domain instances as either a *top* domain instance (if it occurs on a top strand), or a *bottom* domain instance (if it occurs on a bottom strand). We will refer to a double-stranded domain instance as a top domain instance that has a *bond* to a bottom domain instance. A *configuration* is a matching between top and bottom domain instances, where each match is one bond. Our arguments are based solely on counting the maximum number of possible bonds, given certain constraints. The barrier to leak implied by our result, even in the presence of pseudoknots, is entirely enthalpic in nature since it assumes no entropic penalty for joining two complexes into one.

Theorem 1. *Given a translator cascade of arbitrary depth, a downstream reporter, and no input signal, if M is its maximum possible number of bonds then any configuration having the two strands of the reporter in distinct complexes can have at most $M - \lceil \frac{N}{\text{shift}} \rceil$ bonds.*

Proof. Let M be the maximum possible number of bonds, of any configuration, for the translator cascade with downstream reporter in the absence of input.

Suppose the reporter complex is in the ith layer; then its toehold domain instance is at position $i \cdot \mathit{shift}$ and the remainder of its domain instances lie in positions between $p = i \cdot \mathit{shift} + 1$ and $q = i \cdot \mathit{shift} + N$.

Let L be any maximally bound configuration of top and bottom domain instances in $[0, p-1]$, and let M_L equal the number bonds in L. Let R be the intended configuration (in the absence of input) of top and bottom domain instances in $[p, q]$: all bottom domain instances of fuel j have a bond to a top domain instance, also from fuel j, if their position is in $[p, q]$ (and similarly for the reporter). Let M_R equal the number of bonds in R. Since there are no toehold domain instances in $[p, q]$, then that interval contains at least as many top as bottom domain instances. Thus R is a maximally bound configuration of top and bottom domain instances in $[p, q]$, and $M_L + M_R = M$.

Let R' be a maximally bound configuration of top and bottom domain instances in $[p, q]$ subject to no top domain instance in the reporter complex being bound to a bottom domain instance in the reporter complex. Let $M_{R'}$ be equal to the number of bonds in R'. For each position $j \in [p, q]$, there are two possibilities: (i) There is an excess of top domain instances, and thus one of those can bind the reporter bottom domain instance, keeping the total number of bonds in position j unchanged. (ii) There is an equal number of top and bottom domain instances, and thus position j now has one fewer bond. Therefore, the difference $M_R - M'_R$ can be determined by counting the number of positions in $[p, q]$ with an equal number of top and bottom domain instances. Let p' be the maximal position containing a clamp domain instance where $p' < p$. Note that p' must exist since in any translator design there is a clamp domain instance at position N, and since $p > N$ as otherwise the reporter would have domain instances in common with the input signal. By Lemma 3 and Lemma 4 every position in $[p', q]$ has either overhang domain instances or clamp domain instances, but not both. The same is true for positions in $[p, q]$ and none of those positions have instances of a toehold domain. Thus, the number of positions in $[p, q]$ with an equal number of top and bottom domains is exactly the number of clamp domain instances in that interval. There is a clamp domain instance at position q, and every position $q - k \cdot \mathit{shift} \geq p$, for $k \in \mathbb{N}$. Since $q - p + 1 = N$, then there are $\lceil \frac{N}{\mathit{shift}} \rceil$ positions that have clamp domain instances in $[p, q]$. Therefore $M_R - M_{R'} = \lceil \frac{N}{\mathit{shift}} \rceil$. By (i) and (ii) above, every position in $[p, q]$ has at least one unbound top domain instance. Let R'' be a reconfiguration of R', in the obvious way of swapping bonds, such that the reporter top strand forms its own complex. Since $M_{R''} = M_{R'}$, with $M_{R''}$ being equal to the number of bonds in R'', then $M - M_{R''} = \lceil \frac{N}{\mathit{shift}} \rceil$ establishing the claim. □

This theorem implies that in the absence of input there is an enthalpic barrier of $\lceil \frac{N}{\mathit{shift}} \rceil$ bonds to separate the reporter strands from a maximally bound state. In contrast, when the input is present, the signal can be propagated all the way until the reporter, where separating the reporter strands incurs the loss of only 1 bond (breaking the bonding of the top and bottom clamp domains on the reporter, which have no other binding partners). Thus by increasing N we can

enlarge the enthalpic barrier to leak without increasing the enthalpic barrier to correct output.

3.2 Asymptotic completion level of translator cascades

With a cascade of effectively irreversible strand displacement reactions (not relying on toehold exchange), it is safe to assume that most of the input signal should propagate through to the end. However, with a cascade of reversible reactions such as those we necessarily obtain with toehold size clamps, it might seem that the signal will decrease with the length of the cascade if the signal "spreads out" across the layers. Does this mean that translators with toehold size clamps cannot be composed into long cascades? In this section we prove a lower-bound on the amount of final signal output by a chain of translators that is *independent* of the length of the chain, which shows that long cascades are indeed feasible.

To analyze a system with a cascade of translators, we simplify each translator reaction to be a bimolecular reversible toehold exchange reaction $X+F \rightleftharpoons Y+W$, where X is the input signal, F is the fuel, Y is the output signal and W is the waste species. Assuming that the two toeholds in a toehold exchange reaction (i.e., toehold and clamp domain instances in our nomenclature) have the same thermodynamic binding strength, the net reaction of a translator has $\Delta G^o \approx 0$ and the equilibrium constant of each reaction can be treated as 1. Thus for a single translator, if the initial concentration for the reactants are $[X]_0 = \alpha$, $[F]_0 = 1$, at chemical equilibrium, the concentration of output strand Y will be $\frac{\alpha}{\alpha+1}$.

We then ask how much output signal will be produced if multiple translators are cascaded together. Suppose we can have the n-layer reaction system, where each reaction represents a translator reaction:

$$\begin{gathered} X_1 + F_1 \rightleftharpoons X_2 + W_1 \\ \cdots \\ X_i + F_i \rightleftharpoons X_{i+1} + W_i \\ \cdots \\ X_n + F_n \rightleftharpoons X_{n+1} + W_n \end{gathered}$$

The system starts with all the fuels (F_i, $i = 1, 2, \ldots, n$) at concentration 1 and X_1 at α. If there is relatively little signal compared with fuel (we set $\alpha < \frac{1}{2}$), the reactions are driven forward by the imbalance between F_i and W_i. By conservation of mass, $F_i + W_i = 1$. Since $W_i \leqslant \alpha$ (we can't produce more waste than there was input), we get $F_i = 1 - W_i \geqslant 1 - \alpha$. Thus, at chemical equilibrium, for each reaction we have:

$$\frac{X_{i+1}}{X_i} = \frac{F_i}{W_i} \geqslant \frac{1-\alpha}{\alpha}.$$

Letting $\beta = \frac{1-\alpha}{\alpha}$, we obtain a lower bound for X_{i+1}: $X_{i+1} \geqslant \beta X_i$.

Since the total concentration of all signal strands is conserved, we have:

$$X_1 + X_2 + \cdots + X_i + \cdots + X_{n+1} = \alpha$$

$$(\beta^{-n} + \beta^{-(n-1)} + \cdots + \beta^{-1} + 1) \cdot X_{n+1} \geqslant \alpha$$

Since $\sum_{i=-n}^{0} \beta^i = \beta^{-n} \frac{1-\beta^{n+1}}{1-\beta}$, the above equation can be simplified as

$$X_{n+1} \cdot \beta^{-n} \frac{1-\beta^{n+1}}{1-\beta} \geqslant \alpha$$

Thus the concentration of X_{n+1} is

$$X_{n+1} \geqslant \frac{\alpha\beta^n(1-\beta)}{1-\beta^{n+1}} = \frac{\alpha(\frac{1}{\alpha}-2)}{\frac{1}{\alpha}-1-\frac{1}{(\frac{1}{\alpha}-1)^n}} \geqslant \frac{\alpha(\frac{1}{\alpha}-2)}{\frac{1}{\alpha}-1} = \alpha\frac{1-2\alpha}{1-\alpha}.$$

This result indicates that increasing the number of reaction layers does not affect the lower bound of the equilibrium concentration of the output signal. Therefore, concatenating the translators composed of toehold exchange reactions can always generate at least a constant fraction of signal independent of the number of layers.

4 Kinetic properties

Beyond thermodynamic properties, the kinetic properties of translator schemes can vary depending on the choice of N and *shift*. In this section, we show that schemes are susceptible to undesirable properties such as toehold-occlusion, spurious strand displacement, or reconfiguration of fuels, to varying degrees. As we will see, certain schemes preclude some of these phenomena entirely.

4.1 Toehold occlusion

In strand displacement systems, reaction kinetics can be controlled by the strength of a toehold [12]. Stronger toeholds can enable faster reaction kinetics; however, if the toehold strength is too strong, toehold dissociation can become a rate limiting step. This is problematic when overhangs of fuel can bind to toeholds of other fuel, since fuel is typically present in high concentration. This creates so-called *toehold occlusion* [5, 9], which can significantly slow down the intended reaction kinetics in the presence of input signal.

Theorem 2. *Toehold occlusion is not possible in a translator scheme if and only if N is a multiple of* shift*.*

Proof. Toeholds are occluded when overhangs can bind to them. The toehold domain instance of fuel i lies at position $i \cdot \mathit{shift}$. Overhang domain instances of fuel i lie at positions between $i \cdot \mathit{shift} + N + 1$ and $i \cdot \mathit{shift} + N + \mathit{shift} - 1$. The claim follows by Lemma 2 ($p = \mathit{shift}$, $q = 0$ and $r = N + 1$). □

4.2 Spurious strand displacement

Spurious strand displacement events, even if they do not lead to leak of output signal or dissociation of strands of any kind, are unproductive reactions that can slow down the intended kinetics of the system. Spurious displacement occurs when any proper prefix or suffix of a fuel's double-stranded helix is displaced by a spurious invader. A spurious invader of a fuel is any complex not equal to its intended input strand. We partition our analysis into two categories: (i) spurious displacement in the absence of input, and (ii) spurious displacement in its presence.

Spurious displacement in the absence of input. Spurious displacement between fuels can become increasingly problematic with respect to the intended kinetics of a cascade—since fuels involved in spurious displacement can be unavailable for their intended reaction—as the concentration of the system is increased.

We find it convenient to refer to specific *top* or *bottom* domain instances, as in Section 3.1. We begin by looking at spurious displacement of bottom domain instances which can only occur in positions containing toehold domain instances as these are the only positions with bottom domains in excess.

Lemma 5. *In the absence of input, spurious displacement of bottom domain instances in a translator scheme is possible (i) in left flank domain instances if and only if* shift $= 1$ *(i.e. the toehold domain instance can invade the left flank domain instance, see Fig. 2), and (ii) in clamp domain instances if and only if N is a multiple of* shift*.*

Proof. (i) Left flank domain instances are offset of toehold domain instances by 1. By Lemma 1, setting $p = \mathit{shift}$, $q = 0$ and $r = 1$, it follows that $r - q = 1$ is a multiple of p, and thus $\mathit{shift} = 1$, if and only if a domain position overlaps with both a toehold domain instance and a left flank domain instance. (ii) Clamp domain instances are offset of toehold domain instances by N. By Lemma 1, setting $p = \mathit{shift}$, $q = 0$ and $r = N$, it follows that $N = r - q$ is a multiple of $\mathit{shift} = p$ if and only if a domain position overlaps with both a toehold domain instance and a clamp domain instance. □

Now consider spurious displacement of top domain instances which can only occur in positions containing overhang domain instances as these are the only positions with top domain instances in excess.

Lemma 6. *In the absence of input, spurious displacement of top domain instances in a translator scheme is possible if and only if $N - 1$ is not a multiple of* shift*.*

Proof. By construction, when $\mathit{shift} = 1$ there are no overhang domain instances and therefore spurious displacement of top domain instances is not possible. Assume $\mathit{shift} > 1$.

We first establish that it is not possible to spuriously displace top domain instances of clamp instances, because the positions of clamp domain instances cannot intersect that of overhang domain instances, by Lemma 3.

Thus any displacement of top domain instances of a fuel must be a proper prefix of its helix and therefore must include a left flank domain instance. Left flank domain instances are offset by 1, relative to toehold domain instances. Setting $p = \mathit{shift}$, $q = 1$ and $r = N + 1$ it follows from Lemma 2 that the positions of left flank domain instances can intersect that of overhang domain instances if and only if $r - q - 1 = N - 1$ is not a multiple of $p = \mathit{shift}$. □

Spurious displacement in the presence of input. A second type of spurious displacement is when a free signal strand (including the input), can act as a spurious invader of a fuel other than its designed target. In this case, particularly when the input concentration is significantly lower than fuel as is typical, signal strands can become involved in numerous unproductive reactions thus slowing (possibly significantly) signal propagation through every layer of the cascade.

Lemma 7. *Spurious displacement between signal strands and fuels is not possible in a translator scheme if and only if* $\text{shift} \geq N - 1$.

Proof. Domain instances of signal strand i lie at positions between $i \cdot \mathit{shift}$ and $i \cdot \mathit{shift} + N - 1$. Signal strand i is a spurious invader if it can displace any domain instances on some fuel $j > i$. Signal strand i cannot displace the clamp domain instance on fuel i, at position $i \cdot \mathit{shift} + N$, and therefore cannot displace the clamp domain instance of fuel j, at position $j \cdot \mathit{shift} + N > i \cdot \mathit{shift} + N$. Suppose signal strand i is a spurious invader of fuel j; it must invade a prefix of fuel j's double-stranded domain instances (its helix) which necessarily includes its left flank domain instance at position $j \cdot \mathit{shift} + 1$. It follows that $j \cdot \mathit{shift} + 1 \leq i \cdot \mathit{shift} + N - 1$, and thus $\mathit{shift} \leq \frac{N-2}{j-i} \leq N - 2$ since $j > i$. Finally, suppose signal strand i is not a spurious invader of any fuel $j > i$; then it cannot displace the left flank domain instance of fuel j, so $j \cdot \mathit{shift} + 1 > i \cdot \mathit{shift} + N - 1$ which implies $\mathit{shift} > \frac{N-2}{j-i} = N - 2$ when $j = i + 1$. □

Spurious displacement with or without input. By Lemma 5, Lemma 6, and Lemma 7 we have the following.

Theorem 3. *Spurious displacement is not possible in a translator scheme if and only if* $\text{shift} = N - 1$.

4.3 Reconfiguration

Spurious strand displacements can be more complicated when multiple species are involved. Some may result in the formation of complex multi-stranded structures. In this section, we show contrasting examples of possible reconfiguration after spurious strand displacement. The first example requires a bimolecular reaction to undo, while unimolecular reconfiguration is sufficient for the second.

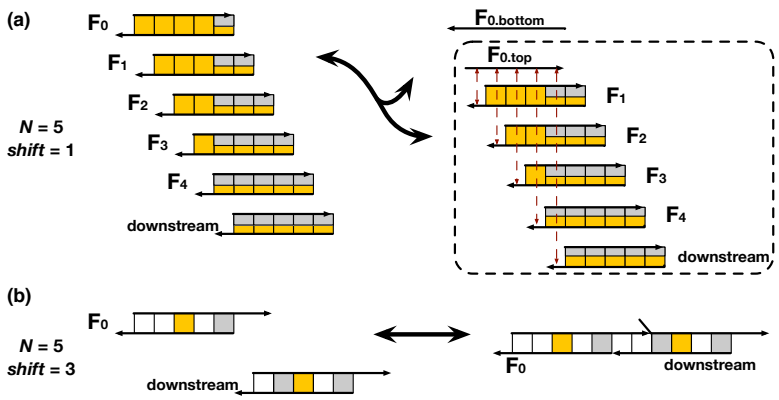

Fig. 3: Examples of configurations reachable with spurious strand displacement.

(To the first approximation, even at the "high" concentrations used here, bimolecular reactions are relatively slower than unimolecular reactions.)

Consider the extreme case when $shift = 1$ in the absence of input signal. By Lemma 5, we have shown that the left flank bottom domain instance can be displaced by a toehold domain instance on the next fuel. Since $shift = 1$, the ith left flank domain instance appears at position $shift \cdot i + 1 = i + 1$, for all $i \in \mathbb{N}$. Therefore, multiple spurious displacement events could result in all of the bottom domain instances of one fuel being displaced by toehold domain instances of other fuels. This results in a free (unbound) bottom strand. To restore the original configuration, a bimolecular reaction pathway is needed. Although the multiple blunt-end displacement events to cause this reconfiguration are unlikely, once formed, it requires a slow bimolecular reaction to undo. Fig. 3a demonstrates this pathway.

In other cases, spurious strand displacement seems unable to cause any major reconfiguration problem. For example, when $N = 5$ and $shift = 3$ any reconfiguration can be undone via fast unimolecular steps. An example of a structure that can form via spurious displacement in this scheme is shown in Fig. 3b.

These two examples suggest that there is likely a separation between schemes with harmful spurious displacement and those in which spurious displacement occurs but can be quickly undone. Formally differentiating the two cases is an area for further research.

5 Trade-offs between the properties

We have shown that different choices of N and $shift$ yield translator schemes with varying thermodynamic and kinetic properties. By Theorem 2 and Theorem 3 we have the following fundamental kinetic trade-off.

Corollary 1. *There is no translator scheme that avoids both toehold occlusion and spurious displacement.*

Thus, every translator scheme has some undesirable kinetic property. There is a quantitative trade-off for the thermodynamic property of enthalpic barrier to leak, given by Theorem 1, based on the ratio of N and *shift*. Schemes can also have a trade-off between unfavorable kinetic and thermodynamic properties. By Theorem 3, only schemes with $shift = N - 1$ can avoid spurious strand displacement. However, by Theorem 1, these schemes only have a constant enthalpic barrier to leak.

Corollary 2. *A large enthalpic barrier to leak is incompatible with avoiding spurious displacement.*

In fact, schemes with the largest enthalpic barrier to leak also have the most potential spurious interactions between fuels and signal strands. As an example of this, compare the spurious interactions caused by the X input signal with fuels, besides the initial fuel, as *shift* increases in Fig. 2.

In summary, there is no best translator scheme with respect to all thermodynamic and kinetic properties studied here. Instead, the entire taxonomy we develop informs the choice of translator scheme, and one should be chosen based on the expected conditions of its planned use. For example, high concentration conditions may be best served by a scheme with no toehold occlusion and a balance between its enthalpic barrier to leak and its potential number of spurious displacement reactions.

6 Preliminary experimental verification

To experimentally test the kinetic leak reduction design strategy with toehold-size clamps, we chose one of the parameter pairs $N = 5$ and $shift = 3$. This combination has the desired property that (1) in the absence of the input signal strand, leak requires at least two units of enthalpic penalty (breaking two bonds) compared with the maximally bound state, and (2) even if spurious strand displacement can occur, there exists a unimolecular reaction pathway that can reverse the spurious interactions and restore the original configuration of the system (see Section 4.3).

6.1 Leak reduction with a single translator

We compare with the previously proposed leak reduction method (*NLD* scheme) based exclusively on an entropic penalty [10]. More specifically, we choose two *NLD* redundancy parameter values $N_r = 1$ and $N_r = 2$. For $N_r = 1$, the scheme is the typical leaky translator (named *SLD* scheme). For $N_r = 2$, the scheme is the previously described "leakless" translators (named *DLD* scheme).

In our experiments, every fuel is kept at 5 μM, which is 50–100 times larger than the typical concentration used in strand displacement systems. Fig. 4a compares the kinetic behaviors of *SLD*, *DLD*, and the toehold-size clamp design, all in the absence of input signal. As expected, the leak rate in the *SLD* scheme (measured at the first 20 minutes) is 30 times higher than that of the *DLD*

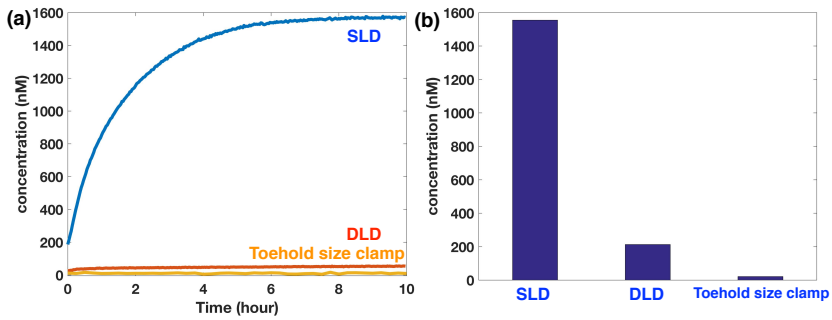

Fig. 4: Comparison of leak between the *SLD*, *DLD* and toehold-size clamp design ($N = 5$, $shift = 3$) from (a) a kinetics and (b) a thermodynamic equilibrium perspective. Since leak is measured, the systems do not contain the input signal strands. To measure leak at thermodynamic equilibrium, fuels and reporter are slowly annealed. The concentrations for fuels are 5 μM. Reporter in the toehold-size clamp design is at 5 μM. Reporters in the *SLD* and *DLD* designs are at 6 μM. The reaction temperature is 37°C. See the sequences and methods in the Appendix.

scheme. However, the *DLD* leak rate is still roughly 0.03 nM/min throughout the 10 hours. In contrast, after quickly generating 10 nM initial leak (which is hypothesized to be caused by misfolded fuel structure or synthesis error of DNA strands), the toehold-size clamp design does not show gradual leak at our experimental setting.

In addition to kinetic measurements, we tested how much leak each design has at thermodynamic equilibrium, which sets an upper bound of the total leak for an isolated translator. To achieve thermodynamic equilibrium, fuels and reporter are slowly annealed together. Fig. 4b compares the total leak amount of these designs. The toehold-size clamp design shows the least amount of leak even at thermodynamic equilibrium.

6.2 Leak reduction with translator cascade

Beyond a single translator, we wanted to know (1) how much leak (without input signal) and (2) desired output signal (with input signal) a translator cascade can generate with increasing number of translators.

Fig. 5a shows that in the time period of the experiment, in the absence of input signal strand, the translator cascades of 1 to 6 fuels (1 to 3 translators) all show no apparent leak. In the presence of input signal, the completion level decreases with the number of layers. However, as more layers are added, the completion level does not decrease linearly, and indeed seems to approach an asymptote, a behavior consistent with the theoretical prediction of Section 3.2.

Fig. 5b studies the leak of translator cascades of varying depth at thermodynamic equilibrium. The leak at equilibrium increases as the number of fuels increases. However, even if there are 6 fuels (3 translators), the total leak concentration is still less than 3% of the fuel concentration.

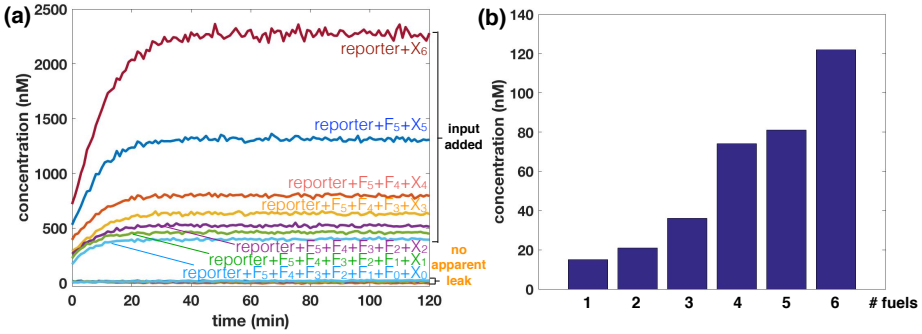

Fig. 5: Kinetics and thermodynamic equilibrium of translator cascades with the toehold-size clamp design ($N = 5$, $shift = 3$). (a) Kinetic behaviour in the presence and absence of input signal, for cascades of different length. (b) The total amount of leak in the absence of input signal at thermodynamic equilibrium, for cascades of different length. The concentrations of the reporter and the fuels are around 5 μM. The concentration of each input is 2.5 μM. The reaction temperature is 37°C. See the sequences and methods in the Appendix.

These results suggest that the absolute leak concentration of the toehold-size clamp design is significantly less than the previously proposed *DLD* design. Nonetheless, the relative positive signal to background noise ratio of the toehold-size clamp design is smaller than of the *DLD* design because of the significant smaller completion level due to the reversibility of all displacement steps. Our results suggest that the toehold-size clamp design could be preferable, especially when a system requires absolutely smaller leak, such as when concatenating the translators with downstream catalytic or auto-catalytic systems.

7 Discussion

In this work, we study schemes for constructing strand displacement systems which utilize toehold size clamps to decrease leak. The full diversity of such schemes for translators is accessible by varying two parameters N and *shift*. We provide rigorous guarantees on the enthalpic barrier to leak as a function of these parameters. We further prove that certain parameter values result in other desirable properties like no spurious displacement, and no toehold occlusion. We prove a tradeoff theorem which says that no scheme satisfies all desired properties; consequently, understanding the properties of the full assortment of schemes helps to make the proper design choices. Since no single scheme can be judged to be "best", and tradeoffs are inherent, future work will also experimentally compare different parameter sets in different experimental regimes.

In contrast to previously reported methods for arbitrarily decreasing leak which rely on entropic barrier arguments, we describe how the enthalpic barrier to leak can be raised arbitrarily. The enthalpic barrier argument is particularly

germane for the high concentration regime where the entropy penalty for joining complexes is smaller.

Our argument relies on the strong assumption that all domains are orthogonal. In reality, given the limited size of a domain (that of a toehold), as the number of distinct domains increases, it is not possible to make all the domains orthogonal. Nonetheless we note that in certain cases having the same sequence in multiple domain positions seems to pose no problem. (For example, in schemes where N is not a multiple of *shift*, all the clamp domain instances could have the same sequences.) Future work could further explore how to assign the same domains without undesired interactions and how the number of orthogonal domains needed scales with the length of double-stranded region N.

References

1. Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, 8(10):755–762, 2013.
2. Yu Sherry Jiang, Sanchita Bhadra, Bingling Li, and Andrew D Ellington. Mismatches improve the performance of strand-displacement nucleic acid circuits. *Angewandte Chemie*, 126(7):1876–1879, 2014.
3. Xiaoping Olson, Shohei Kotani, Jennifer E Padilla, Natalya Hallstrom, Sara Goltry, Jeunghoon Lee, Bernard Yurke, William L Hughes, and Elton Graugnard. Availability: A metric for nucleic acid strand displacement systems. *ACS Synthetic Biology*, 2017.
4. Tosan Omabegho, Ruojie Sha, and Nadrian C Seeman. A bipedal DNA brownian motor with coordinated legs. *Science*, 324(5923):67–71, 2009.
5. Lulu Qian and Erik Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196–1201, 2011.
6. Maria Rudchenko, Steven Taylor, Payal Pallavi, Alesia Dechkovskaia, Safana Khan, Vincent P Butler Jr, Sergei Rudchenko, and Milan N Stojanovic. Autonomous molecular cascades for evaluation of cell surfaces. *Nature Nanotechnology*, 8(8):580–586, 2013.
7. John SantaLucia Jr and Donald Hicks. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.
8. Georg Seelig, David Soloveichik, David Yu Zhang, and Erik Winfree. Enzyme-free nucleic acid logic circuits. *Science*, 314(5805):1585–1588, 2006.
9. Niranjan Srinivas. *Programming chemical kinetics: engineering dynamic reaction networks with DNA strand displacement*. PhD thesis, California Institute of Technology, 2015.
10. Chris Thachuk, Erik Winfree, and David Soloveichik. Leakless DNA strand displacement systems. In *International Workshop on DNA-Based Computers*, pages 133–153. Springer, 2015.
11. David Yu Zhang and Georg Seelig. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry*, 3(2):103–113, 2011.
12. David Yu Zhang and Erik Winfree. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131(47):17303–17314, 2009.

Appendix

DNA sequences. All the unlabelled oligonucleotides were purchased PAGE purified from Integrated DNA Technologies. The oligonucleotides with fluorophore or quencher modifications were ordered HPLC purified.

Sample preparation. All the DNA oligonucleotides were suspended in water and then quantified by NanoDrop. Fuels and reporters were annealed in TAE/Mg^{2+} buffer from 90°C to 25°C. PAGE purification was then conducted [5] to remove malformed structures or single-stranded DNA. Purified duplexes were quantified by NanoDrop again and also concentrated by centrifugal filters (Amicon Ultra-0.5 mL, 10K device) to achieve a high stock concentration.

Fluorescence measurement. Fluorescence signal is measured by BioTech Cytation 5. The NBS (non-binding surface) 384 well plates with clear flat bottom (# 3544) were used. Fluorescence was measured from bottom.

SLD.$F_{1.top}$	TTCCA TCACA TAACA TCCTT CTAAT CAATC
SLD.$F_{1.bottom}$	GTGTT GATTG ATTAG AAGGA TG
SLD.$F_{2.top}$	TCTCA AACCT ATACA TTCCA TCACA TAACA
SLD.$F_{2.bottom}$	AAGGA TGTTA TGTGA TGGAA TG
DLD.$F_{1.top}$	AACCT ATACA TTCCA TCACA TAACA TCCTT CTAAT CAATC
DLD.$F_{1.bottom}$	GTGTT GATTG ATTAG AAGGA TGTTA TGTGA TG
DLD.$F_{2.top}$	TCCAC TACTT TCTCA AACCT ATACA TTCCA TCACA TAACA
DLD.$F_{2.bottom}$	AAGGA TGTTA TGTGA TGGAA TGTAT AGGTT TG
SLD.$Rep_{.top}$	/5IAbRQ/TCTCA AACCT ATACA
SLD.$Rep_{.bottom}$	TGGAA TGTAT AGGTT TGAGA/3Rox_N/
DLD.$Rep_{.top}$	/5IAbRQ/TCCAC TACTT TCTCA AACCT ATACA
DLD.$Rep_{.bottom}$	TGGAA TGTAT AGGTT TGAGA AAGTA GTGGA/3Rox_N/
SLD.X_0	CA TCCTT CTAAT CAATC AACAC
DLD.X_0	CA TCACA TAACA TCCTT CTAAT CAATC AACAC
Toehold.X	AACATC CTTCTA ATCAAT CAACAC ACATAT
Toehold.$F_{1.top}$	CTATAC ATTCCA TCACAT AACATC CTTCTA ATCAAT CAACAC
Toehold.$F_{1.bottom}$	ATATGT GTGTTG ATTGAT TAGAAG GATGTT ATGTGA
Toehold.$F_{2.top}$	ACTTTC TCAAAC CTATAC ATTCCA TCACAT AACATC
Toehold.$F_{2.bottom}$	TAGAAG GATGTT ATGTGA TGGAAT GTATAG GTTTGA
Toehold.$rep_{.top}$	/5IAbRQ/TCCACT ACTTTC TCAAAC CTATAC
Toehold.$rep_{.bottom}$	TGGAAT GTATAG GTTTGA GAAAGT AGTGGA/3Rox_N/
Toe.Line.X_0	TCAAAC AATCCA CACTAC ATCTCA TCAAAT
Toe.Line.$F_{1.top}$	TCCCAT TTCATT CACTTT TCAAAC AATCCA CACTAC ATCTCA
Toe.Line.$F_{1.bottom}$	ATTTGA TGAGAT GTAGTG TGGATT GTTTGA AAAGTG
Toe.Line.$F_{2.top}$	TCAATA CATATA CTTCCT TCCCAT TTCATT CACTTT TCAAAC
Toe.Line.$F_{2.bottom}$	TGGATT GTTTGA AAAGTG AATGAA ATGGGA AGGAAG
Toe.Line.$F_{3.top}$	CAACAC ACATAT CCTCAC TCAATA CATATA CTTCCT TCCCAT
Toe.Line.$F_{3.bottom}$	AATGAA ATGGGA AGGAAG TATATG TATTGA GTGAGG
Toe.Line.$F_{4.top}$	AACATC CTTCTA ATCAAT CAACAC ACATAT CCTCAC TCAATA
Toe.Line.$F_{4.bottom}$	TATATG TATTGA GTGAGG ATATGT GTGTTG ATTGAT