Remote toehold: Supporting Information

Anthony J. Genot, David Yu Zhang, Jonathan Bath, and Andrew J. Turberfield

S1. Domain structures used and domain notation.

The domain structures of DNA strands used in experiments described in the main paper and in Supporting Information are shown in Figures S1, S2 and S3 respectively. Domains are continuous stretches of DNA that act as a unit in hybridization or dissociation. A barred number denotes a domain complementary to that denoted by the corresponding unbarred number (e.g. $\bar{2}$ is complementary to 2).

Dom.	Sequence	Length (nt)
0	5' - TT - 3'	2
1	5' - ATAGATCCTGATAGC - 3'	15
2	5' - GAGACCTAGCAACCTGAAACCA - 3'	22
2a	5' - GAGAC - 3'	5
2b	5' - CTAGCAA - 3'	7
3_{14}	5' - CATATCCAACTCAC - 3'	14
3_{11}	5' - CATATCCAACT - 3'	11
3_{10}	5' - CATATCCAAC - 3'	10
3_9	5' - CATATCCAA - 3'	9
3_8	5' - CATATCCA - 3'	8
3_6	5' - CATATC - 3'	6
3_4	5' - CATA - 3'	4
3_2	5' - CA - 3'	2
4_{23}	5' - CACATCTCCTATCTACACAAC - 3'	21
4_{20}	5' - CACATCTCCTATCTACAC - 3'	18
4_{17}	5' - CACATCTCCTATCTA - 3'	15
4_7	5' - ATCTA - 3'	5
5_{23}	5' - ACCTCATACAATACCTTACTC - 3'	21
5_{20}	5' - ACCTCATACAATACCTTA - 3'	18
5_{17}	5' - ACCTCATACAATACC - 3'	15
5_8	5' - TACATA - 3'	6
6	5' - TTCTTC - 3'	6
7	5' - TTAGTGA - 3'	7
8	5' - G - 3'	1

Table S0: Domain sequences

Figure 2a

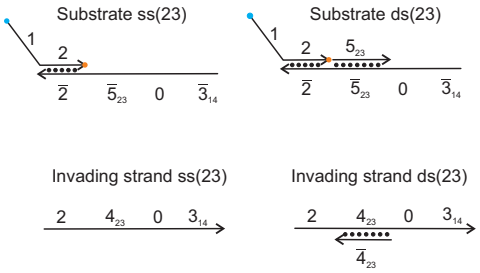

Figure 2b

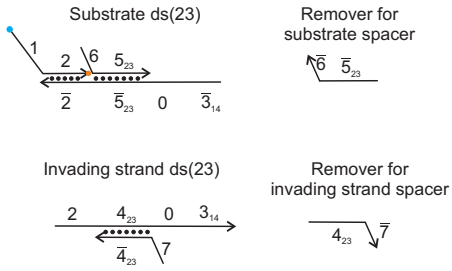

Figure 3

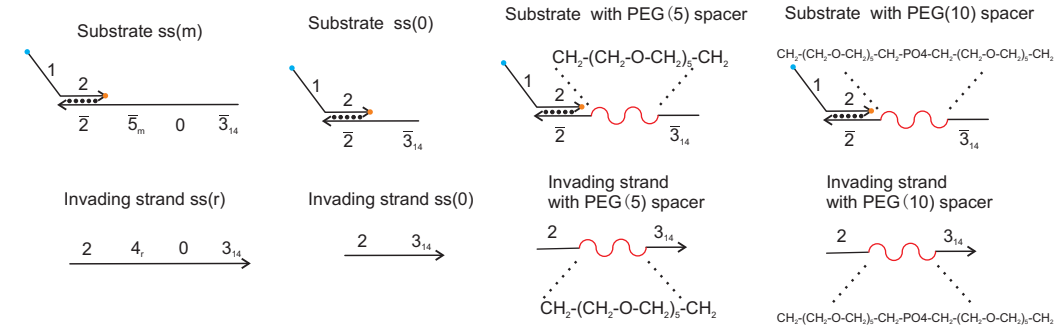

Figure 4a

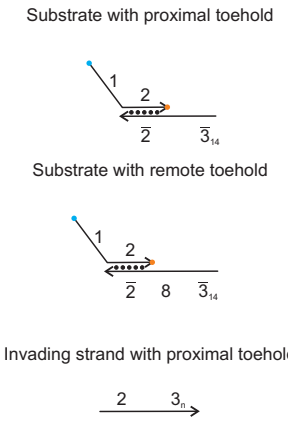

Figure 4b

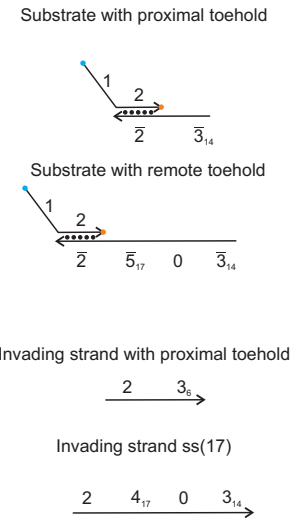

FIG. S1: Domain structures of strands used for the experiments in the main paper.

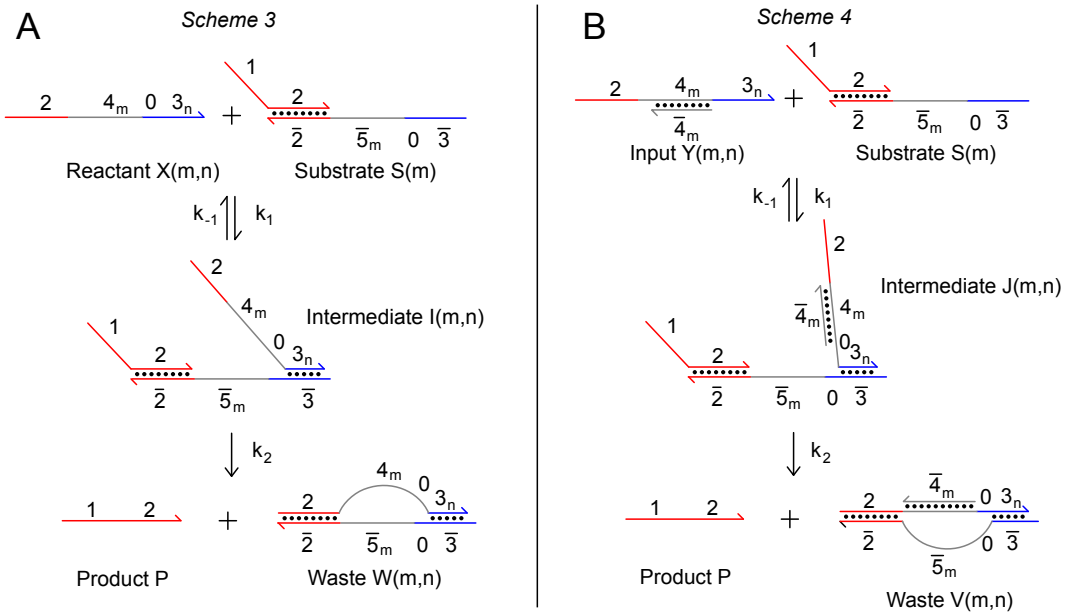

FIG. S2: Detailed schematics of reactions for ss-ss and ss-ds experiments. **(A)** Scheme 3: Remote toehold strand displacement with single-stranded spacer on invading strand and single-stranded spacer on substrate complex. **(B)** Scheme 4: Remote toehold strand displacement with double-stranded spacer on invading strand and single-stranded spacer on substrate complex.F.

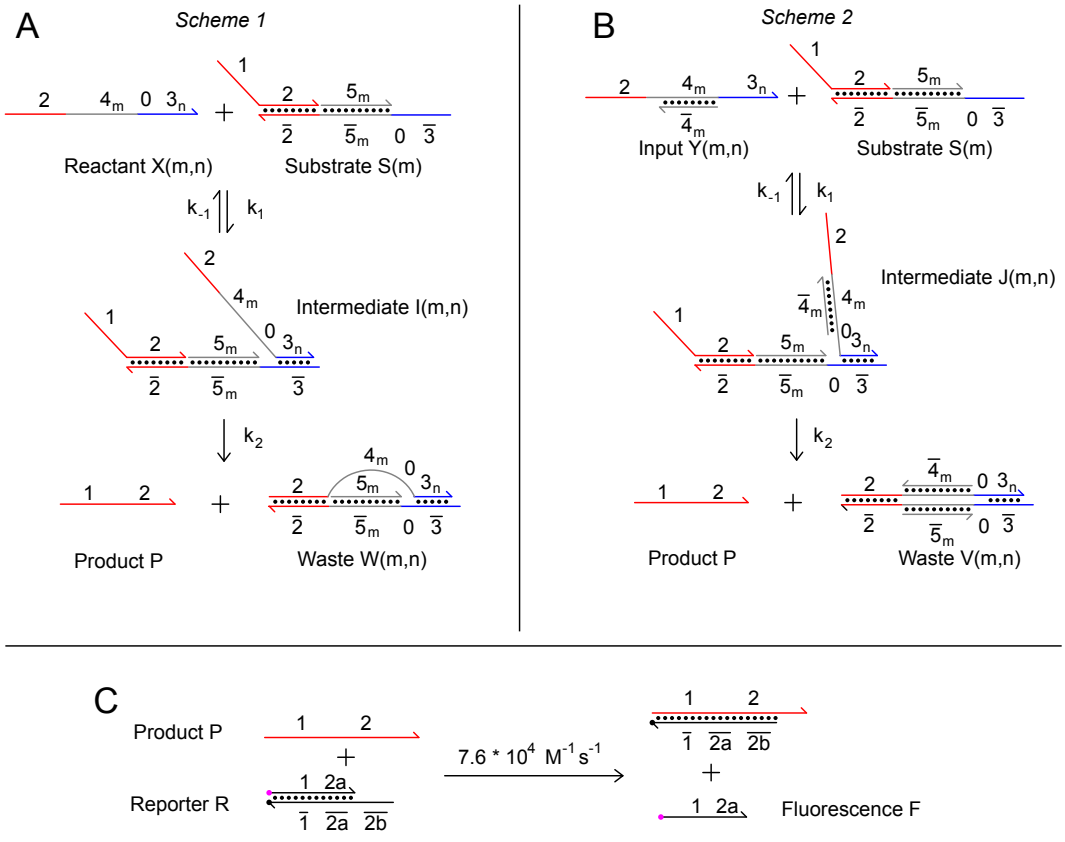

FIG. S3: Schematics and domain structures of strands for the experiments described in the supplementary material. **(A)** Scheme 1: Remote toehold strand displacement with single-stranded spacer on invading strand and double-stranded spacer on substrate complex. The subscript variables m and n denote the number of bases in the corresponding domains; for example, X(15,10) denotes that the 4_m domain is 15 nt long and the 3_n domain is 10 nt long. **(B)** Scheme 2: Remote toehold strand displacement with double-stranded spacers. **(C)** Fluorescence reporter. Reporter complex R reacts stoichiometrically with P to yield an unquenched ROX fluorophore-labeled strand F.

S2. Further characterization of ds-ss and ds-ds systems.

Figure S4 shows the effect of varying the spacer lengths for ds-ss and ds-ss systems, Figure S5 shows the effect of varying the toehold length for ds-ss systems and Figure S6 shows a PAGE analysis of ds-ss and ds-ds sytems. In these experiments, an indirect reporter complex was employed that reacts stoichiometrically with the product of the reaction (Fig. S3C). The rate constant of the reaction between the product and the reporter complex was measured to be 7.6×10^4 M^{-1} s^{-1} (data not shown). Because the reporter concentration was always at least 20 nM in excess of the maximum product concentration, the reporting delay is no more than 12 minutes, which is short compared to the timescale of the remote toehold reactions characterized here.

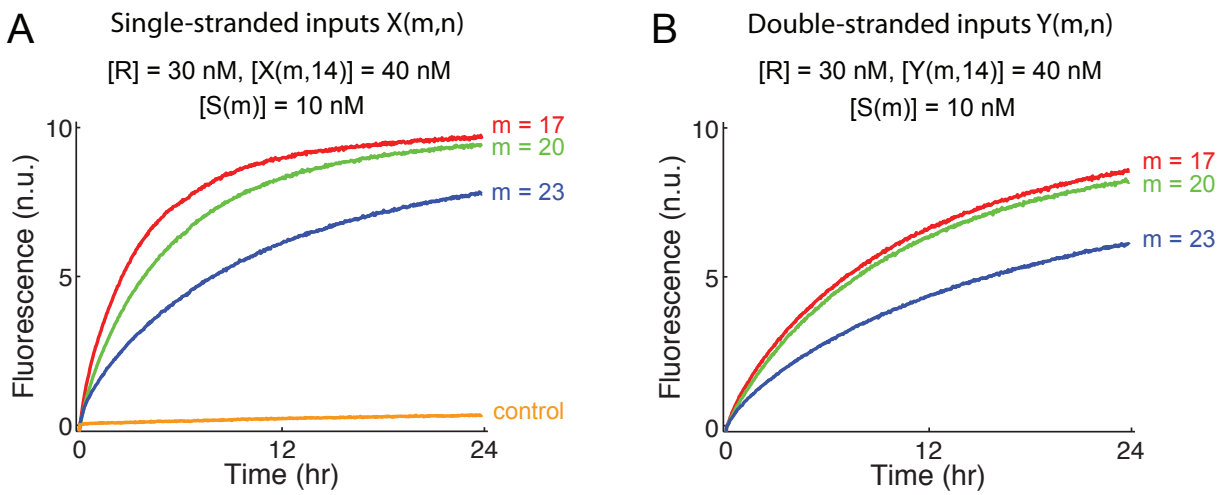

FIG. S4: Effects of covarying spacer lengths on remote toehold strand displacement kinetics. **(A)** Lengthening the spacers of the single-stranded input and the double-stranded substrate causes a significant decrease in strand displacement kinetics, as expected. The orange "control" trace shows the kinetics of strand displacement when no remote toehold exists. The rate constants k_2 were fitted to be 8.9×10^{-5} s^{-1}, 5.9×10^{-5} s^{-1}, and 3.0×10^{-5} s^{-1} for m = 17, 20, and 23, respectively. **(B)** When both spacers are double-stranded, the kinetics of strand displacement also decreases for longer spacers, but other factors such as helicity also contribute to determining kinetics. The rate constants k_2 were fitted to be 3.1×10^{-5} s^{-1}, 2.8×10^{-5} s^{-1}, and 1.4×10^{-5} s^{-1} for m = 17, 20, and 23, respectively.

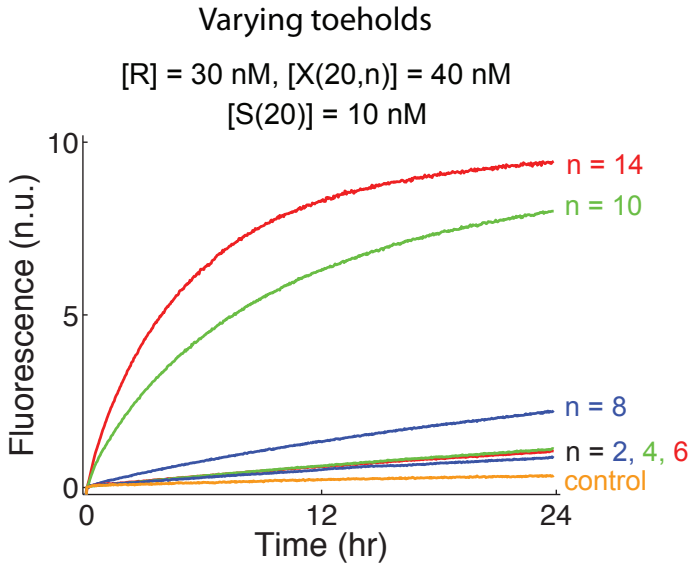

FIG. S5: Effects of varying toehold lengths on remote toehold strand displacement kinetics. Increasing the toehold length n decreases the value of k_{-1}, the rate constant of toehold dissociation, leading to a higher probability of completing strand displacement. For short toeholds 6 nucleotides or less, the kinetics of of strand displacement appear to be very slow; this could be because the toehold binding is insufficient to thermodynamically drive the reaction to completion, because of entropic penalties to the final structures V and W.

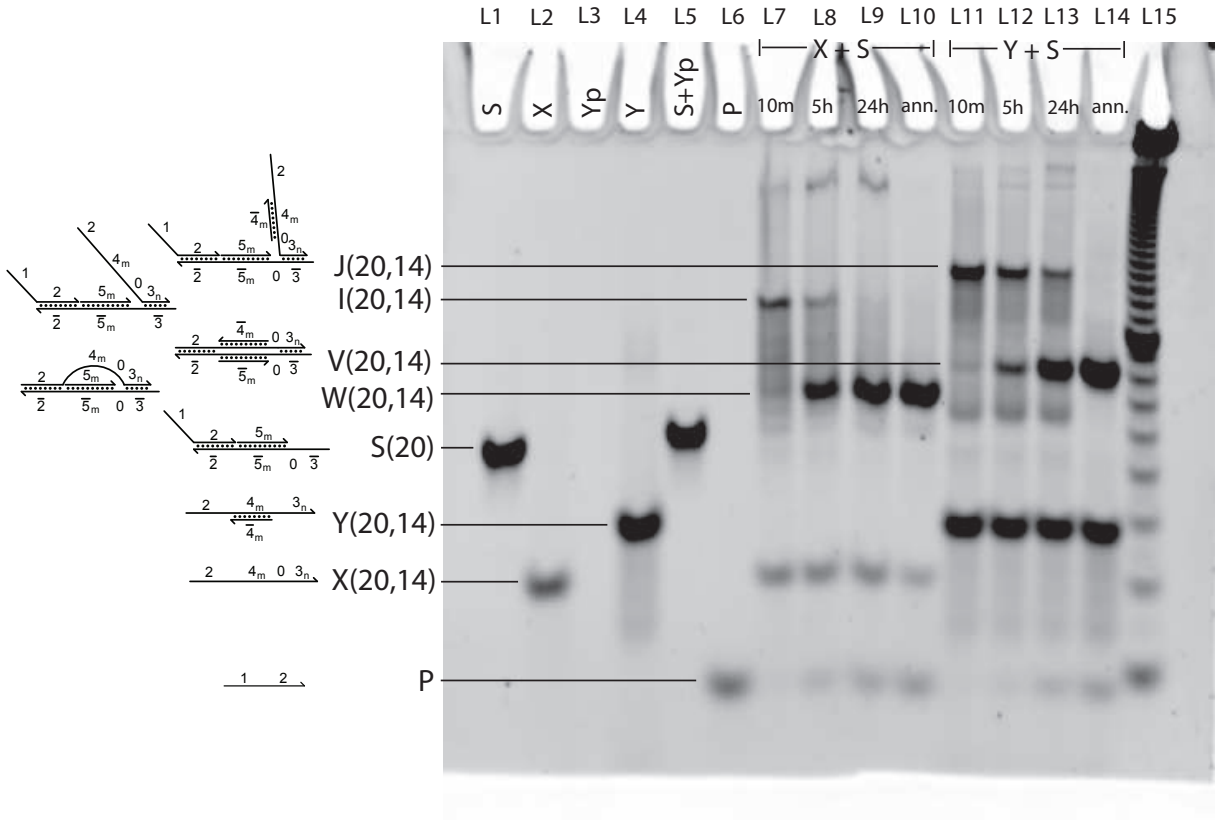

FIG. S6: An analysis of ds-ss and ds-ds strand displacement. Native polyacrylamide gel electrophoresis (PAGE) gel ($m = 20$ and $n = 14$). Lanes 1-6: controls. Lanes 7-10 show a time series and the annealed endpoint for the reaction between input $X(20, 14)$ (with a single-stranded spacer) and complex $S(20)$ (with a double-stranded spacer). The heavier intermediate I can clearly be identified, and product P accumulates over time. Lanes 11 through 14 similarly show the time series and annealed endpoint for the reaction between $Y(20, 14)$ and $S(20)$. Lane 15: 10 bp duplex ladder.

S2.1 Materials and Methods.

Buffer conditions. The buffer for all experiments was TE (10 mM Tris · HCl pH 8.0, 1 mM EDTA), with 12.5 mM $MgCl_2$ added. Because EDTA chelates magnesium ions, the effective concentration of Mg^{2+} is 11.5 mM. All experiments and purifications were performed at 25 °C.

Native polyacrylamide gel electrophoresis (PAGE). A 12% acrylamide (19:1 acrylamide:bis) native polyacrylamide gel was run to prepare Fig. S6. Acrylamide solution was diluted from 40% stock acrylamide (Ambion). Native loading dye containing XCFF in 50% glycerol was added to all samples in 1:4 ratio, achieving final glycerol concentration of 10% by volume. Gels were run at 120 V for 90 minutes at 25 °C, with temperature controlled using a Novex chamber external temperature bath. Gels were stained with Sybr-Gold stain (Invitrogen), and scanned with a Bio-Rad Molecular Imager.

Spectrofluorimetry studies. Spectrofluorimetry studies were performed using a SPEX Fluorolog-3 (Horiba) with 1.6 mL synthetic quartz cells (Hellma 119-004F). Sample solutions were excited at 588 nm, and emission at 602 nm was observed (optimal signal for ROX fluorophore in our buffer). Slit size used were 2 nm for both excitation and emission monochromators. Integration time: 10 seconds.

Before each experiment, all cuvettes were cleaned thoroughly: each cuvette was washed 15 times in distilled water, once in 70% ethanol, another 5 times in distilled water, and finally once more in 70% ethanol. Appropriate volumes of DNA stock solutions, typically 1 μM in concentration, were added to TE/Mg^{2+} buffer to achieve the correct final concentrations with a total volume of 1.5 mL in the cuvettes. Solutions were mixed by pipetting 250 μL of solution in and out of the cuvette 32 times. Stir bars were not used because they have been observed to significantly contribute to fluorophore decay. Unpublished data indicate that this may be due to bleach (used for cleaning) slowly desorbing from the surface of the stir bars, that in turn cause nonspecific fluorophore quenching [1].

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 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-labelled strand F. This normalization is based on the fluorescence levels of annealed samples: a negative control with only $[R] = 30$ nM (set to to 0 n.u.), and a positive control with $[R] = 30$ nM and $[P] = 10$ nM (normalized to 10 n.u.). Day-to-day and sample-to-sample variations are measured to be less than 5%.

Parameter Fitting. Rate constants for our model of remote toehold strand displacement were fitted using the 'fminunc' function in Matlab to minimize the error between experimental data and the reaction model. The error is calculated as follows:

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

where $F_d(t)$ is the fluorescence value of the data, and $F_m(t)$ is the fluorescence value predicted by the ODE model at time t.

S3. Biophysical modelling of the internal displacement rate.

In the following section, we introduce a worm-like chain model to account for the effect of the spacer length on the displacement speed in the case of a ds-ss system.

We are interested in computing the effective concentration c^* of the end of the blue strand in the red box of side a (Figure S7). The red box represents a reactive volume where the displacement reaction can be initiated. Note that in principle the reaction can be initiated at both ends of the duplex, so the rate we compute will be a lower limit. L is the contour length of the ssDNA spacer, s its persistence length. Let $p(R)dR$ be the probability that the blue chain is extended to an end-to-end distance R within a range dR. The probability distribution $p(R)$ depends on the model used to account for the biophysics of DNA.

Assuming that $p(R)$ is almost constant in the reactive volume, that is, a is much smaller than the persistence length s, then the probability P for the chain end to be in the reactive volume is:

$$P = p(r_0)a\frac{V_a}{V_s}$$

where $V_a = a^3$ is the volume of the red box and $V_s = 4\pi r_0^2 a$ is the volume of the shell of radius r_0 and thickness a considering that $a \ll r_0$. This simplifies to:

$$P = p(r_0)\frac{a^3}{4\pi r_0^2}$$

Finally we get the effective concentration c^* in $particles/m^3$ by dividing P by the reactive volume a^3. To convert to a result in mole/L, we divide by $1000N_a$, where N_a is Avogadro's number.

$$c^* = \frac{p(r_0)}{4\pi r_0^2} * \frac{1}{1000N_a}$$

We then assume that the reaction rate is as if the duplex containing the target strand saw an invading strand free in solution at a concentration c^* and containing only a displacement domain with no toehold . The displacement rate for a displacement domain bearing no toehold has been measured (data not shown). The rate constant for this blunt-end strand displacement is $k_b = 0.8M^{-1}s^{-1}$.

We now only need the distribution $p(R)$. We model short single-stranded DNA as a Worm-Like Chain. The probability $p(R, s, L)dR$ for a worm-like chain of length L and persistence length s to be extended to a length R within a range dR is [2]:

$$p(R, s, L) = \frac{1}{L}\frac{4\pi A r^2}{(1-r^2)^{9/2}} exp(-\frac{3t}{4(1-r^2)})$$

with

$$A = \frac{4(3t/4)^{3/2}exp(3t/4)}{\pi^{3/2}(4 + \frac{12}{3t/4} + \frac{15}{(3t/4)^2})}, t = L/s, r = R/L$$

We can now compute the internal displacement rate k_s. It is simply the rate for a blunt-end reaction multiplied by the effective concentration of displacement domains inside the reactive volume.

$$k_s = k_b c^* = k_b \frac{p(r_0)}{4\pi r_0^2} * \frac{1}{1000N_a}$$

Comparison between measured (Figure S4A) and predicted rates are shown in table S1. For ssDNA structural parameters we use a persistence length of 2 nm and an inter-nucleotide distance of 68 nm [3]. The model yields rates which are within one order of magnitude of those measured. The model also properly accounts for the effect of increasing the spacer length on the displacement kinetics.

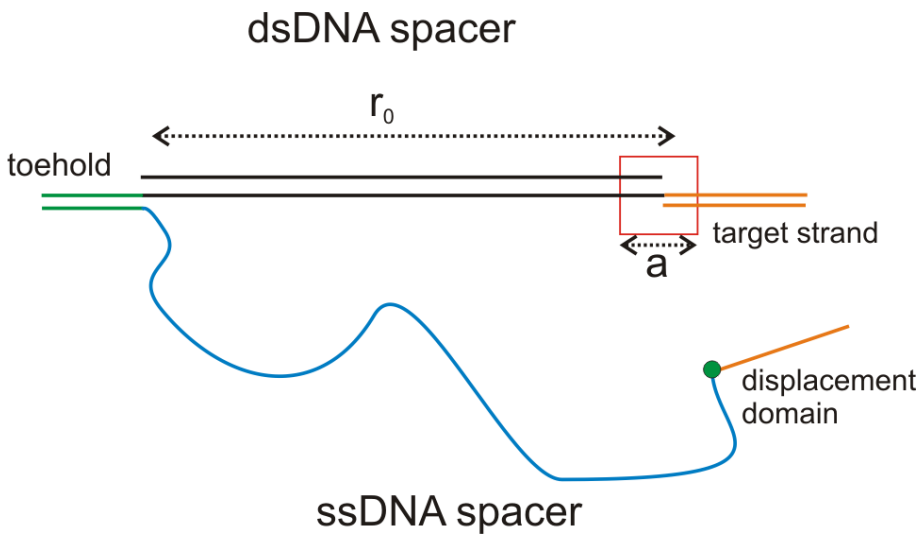

FIG. S7: Notation used for the WLC model. The blue strand is a ssDNA spacer, the orange domain is the displacement domain and the black domains is a dsDNA spacer. The red box is a reactive volume where branch-migration and strand displacement can be initiated.

Spacer length in nt	Fitted experimental rate	Predicted rate
23	$3 * 10^{-5} s^{-1}$	$2.7 * 10^{-4} s^{-1}$
20	$5.9 * 10^{-5} s^{-1}$	$4 * 10^{-4} s^{-1}$
17	$8.9 * 10^{-5} s^{-1}$	$6 * 10^{-4} s^{-1}$

TABLE S1: Comparison between fitted displacement rates and predicted rates for a ds(n)-ss(n) system with n=17,20,23.

S4. Kinetic modelling of toehold-mediated strand displacement.

In this section we propose an analytical function to capture the interplay between the invading strand concentration, the toehold binding strength and the internal displacement rate. The model was used as a guide to design and analyze experiments. We start from the qualitative model of reactions:

$$Subtrate + Invading\ Strand \underset{k_r}{\overset{k_f}{\rightleftharpoons}} Intermediate \overset{k_s}{\rightharpoonup} Released\ Target\ Strand \tag{1}$$

$$k_r = k_f e^{\frac{-\triangle G}{RT}}$$

where k_f is the rate constant for hybridization between the two ssDNA toeholds and $\triangle G$ is the toehold binding strength. Note that the model describes both proximal and remote toeholds.

When the invading strands are in large excess, their concentration can be supposed constant (c) and the model can be solved analytically. Let $x(t)$ be the number of released target strands as a proportion of the total number of target strands:

$$x(t) = 1 + \frac{k_f c k_s e^{-\frac{1}{2}k_+ t}}{k_f c k_s - (k_f c + k_r + k_s)k_+ + 3/4k_+^2} + \frac{k_f c k_s e^{-\frac{1}{2}k_- t}}{k_f c k_s - (k_f c + k_r + k_s)k_- + 3/4k_-^2} \tag{2}$$

$$k_+ = k_f c + k_r + k_s + \sqrt{(k_f c + k_r + k_s)^2 - 4k_f c k_s}$$

$$k_- = k_f c + k_r + k_s - \sqrt{(k_f c + k_r + k_s)^2 - 4k_f c k_s}$$

The kinetic regime of strand displacement is determined by the dominant terms in the sum $k_f c + k_f e^{\frac{-\triangle G}{RT}} + k_s$. Kinetic regimes presented in the main text are limiting cases.

Toehold binding energy threshold.

When the toehold binding strength $\triangle G$ is strong enough that $k_f e^{\frac{-\triangle G}{RT}} \ll k_f c + k_s$, i.e. $k_r \ll k_f c + k_s$, then

$$k_+ \simeq k_f c + k_s + \mid k_f c - k_s \mid$$

$$k_- \simeq k_f c + k_s - \mid k_f c - k_s \mid$$

One of the rate equals $2k_f c$ and the other rate equals $2k_s$. The proportion of released strands $x(t)$ from equation 2 can be written as:

$$x(t) = 1 + \frac{k_s e^{-k_f ct}}{k_f c - k_s} + \frac{k_f c e^{-k_s t}}{k_s - k_f c} \tag{3}$$

The equation 3 shows that the strand displacement kinetics are insensitive to toehold binding strength for $k_r \equiv k_f e^{\frac{-\triangle G}{RT}} \ll k_f c + k_s$. Consequently the binding energy threshold $\triangle G_{0,threshold}$ can be approximated by $\triangle G_{0,threshold} = -RT\ ln(c + \frac{k_s}{k_f})$, with c in $mol/liter$. This expression depends on the concentration of invading strands. For typical proximal toeholds $k_s \sim 1s^{-1}$ and $\frac{k_s}{k_f} \sim 10^{-6}M$ [4], so the effect of concentration on the energy threshold is not seen for submicromolar concentrations.

Concentration robust regime.

The 'concentration robust' regime corresponds to a strong toehold and a slow internal displacement rate, $k_r \ll k_f c + k_s$ and $k_s \ll k_f c$ and in that case the equation 3 further simplifies to:

$$x(t) = 1 - e^{-k_s t}$$

The displacement kinetics are effectively first order and only depend on the internal displacement rate.

[1] Personal communication from Bernard Yurke, Boise State University, September 2009.
[2] Ha, B. Y.; Thirumalai, D., Academic Press: San Diego, CA. 135., 1998.
[3] Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T. J., Biophysical Journal 2004, 86 (4), 2530-2537.
[4] Zhang, D. Y.; Winfree, E., Journal of the American Chemical Society 2009, 131, 17303-17314.