Catalytic DNA Polymerization Can Be Expedited by Active Product Release

Pepijn G. Moerman[1], Momcilo Gavrilov[2], Taekjip Ha[2,3,4], and Rebecca Schulman[1,5,6]

[1] *Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA*
[2] *Biophysics & Biophysical Chemistry, Johns Hopkins University, Baltimore, MD 21205, USA*
[3] *Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA*
[4] *Howard Hughes Medical Institute, Baltimore, MD 21205, USA*
[5] *Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA*
[6] *Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA*

Abstract

The sequence-specific hybridization of DNA facilitates its use as a building block for designer nanoscale structures and reaction networks that perform computations. However, the strong binding energy of Watson-Crick base pairing that underlies this specificity also causes the DNA dehybridization rate to depend sensitively on sequence length and temperature. This strong dependency imposes stringent constraints on the design of multi-step DNA reactions, because a small deviation from the optimal conditions slows down the process dramatically. Here we show how an ATP-dependent helicase, Rep-X, can drive certain dehybridization reactions in designed DNA reaction networks at rates independent of sequence length, thereby decoupling the rates of hybridization and dehybridization. To illustrate this principle, we show that Rep-X extends the range of conditions where the primer exchange reaction, which catalytically adds a domain provided by a hairpin template to a DNA substrate, proceeds rapidly: in the strong substrate-hairpin binding regime, Rep-X expedites the reaction almost one hundred-fold. Our results provide an example of how ATP consumption can drive specific dehybridization reactions in designed DNA reaction networks and how this consumption can be harnessed to expedite reactions beyond their equilibrium rates.

1 Introduction

Sequence complementarity is the central design rule for building nanostructures and reaction networks out of DNA[1, 2]. It enables DNA computers to recognize and report disease-related RNAs among a slew of native oligonucleotides[3, 4]. It guides thousands of short strands simultaneously to their intended positions in two- and three-dimensional structures[5, 6], some of which can be reconfigured[7, 8]. And it makes possible complex computations that take DNA strands as inputs and produce different DNA strands as outputs[9, 10, 15]. The binding specificity of oligonucleotides that makes these applications possible comes from the strong Watson-Crick base pairing: under standard conditions each pair contributes $1 - 4\ k_BT$[12], so that a strand strongly favors binding to its full complement over a spurious target with as little as one mismatch[13].

Yet this strong dependence of the binding energy on oligonucleotide length can also be an Achilles' heel in designing multi-step reactions or reaction cascades. In such processes an individual sequence domain can participate in multiple reaction steps in which it has different functions (e.g. Fig. 1a). These different steps may require conflicting binding and unbinding rates. For example, a long domain may provide the binding energy required to speed up the formation of one complex by stabilizing it, but then slow down a reaction elsewhere in the network that requires a high off-rate. This conflict creates an upper limit on the effective rate of a multi-step reaction, which can only be achieved at a optimal domain length and temperature. Consequently, many DNA reaction networks operate on the timescale of hours [14, 15, 18]. The constraint that on- and off-rates are coupled is a consequence of thermodynamic equilibrium: The upper limit on reaction rates is generic to any multi-step, reversible chemical processes. In heterogeneous catalysis, it is known as Sabatier's principle [19],which states that reactions only proceed if substrate-catalyst binding is not too weak, but product-catalyst binding not so strong that it poisons the catalyst (Fig. 1b). Addressing this fundamental limit on the composite rate of multi-step reactions requires energy input to subvert equilibrium.

Here we ask how an exergonic reaction can be used to decouple the off-rates of DNA hybridization reactions that involve the same binding domain and thus re-

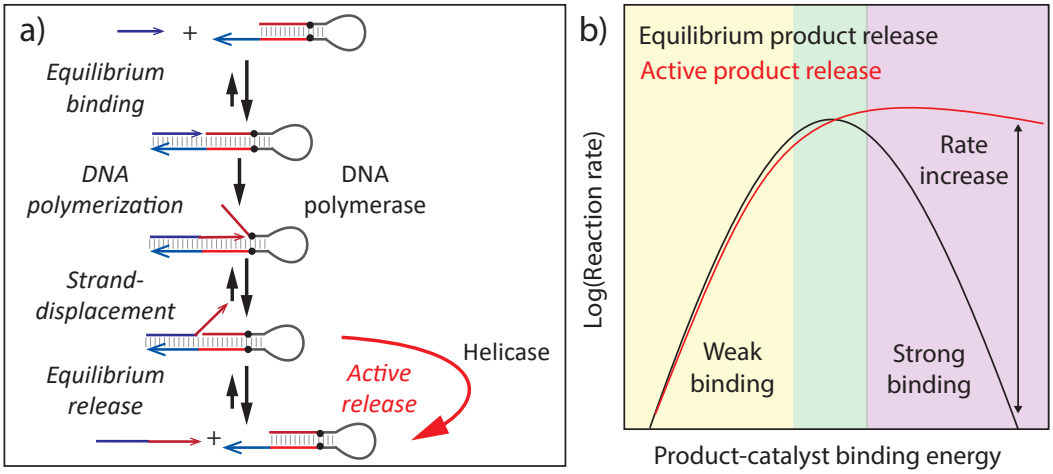

Figure 1: **Active product release expedites a catalytic DNA polymerization reaction** a) The primer exchange reaction involves two competing equilibrium hybridization steps: the reactant binding to the catalyst and the product binding to the catalyst. A helicase can dissipatively and selectively remove product from the catalyst, freeing the catalyst to bind to new reactant, and thereby expedite the reaction. b) A "volcano plot" captures the peaked reaction rate as a function of binding energy that is typical for catalytic reactions. Both weakly and strongly binding catalysts are ineffective: in the weak-binding regime because no reactant is adsorbed and in the strong-binding regime because no product is released. Only at intermediate binding energy does the reaction proceed rapidly. Active removal of the product could prevent catalyst poisoning and expedites the reaction in the strong-binding regime.

duce the dependency of the process's rate on the binding strengths. We aim to use DNA helicases—a class of ATP-dependent proteins that separate double-stranded DNA into its single-stranded components—to couple ATP hydrolysis to DNA unwinding. *In vivo*, helicases unwind parts of long double-stranded DNA whose rates of dehybridization would otherwise be negligible to prepare genomic DNA for replication by exposing a template strand. We ask how helicases could be used to fulfil a similar role in DNA nanotechnology and selectively increase the off-rates of DNA hybridization reactions.

As a case study, we investigate how helicase-driven dehybridization could increase the rate of the primer exchange reaction (PER). PER is a DNA nanotechnology tool that appends new domains with user-defined sequences onto single-stranded input strands (primers) (Fig. 1a) [17], and is part of a family of template extension reactions that can recognize inputs of a specific sequence and amplify them. These reactions, which include Polymerase/Exonuclease/Nickase (PEN)[16] circuits, are of interest for molecular and medical applications such as RNA and protein imaging[21, 22] and for directing active self-organization[23, 24].

PER appends new domains with user-defined sequences onto single-stranded input strands (primers) in a four-step process (Fig. 2a)[17]. First, a hairpin with a single-stranded 3' overhang reversibly binds to the primer (*equilibrium binding*). Then a DNA polymerase extends the primer by copying the template domain on the hairpin (*DNA polymerization*). During this polymerization step, the nascent strand displaces the top strand in the hairpin. Next, the displaced hairpin domain competes for binding to the template domain on the hairpin with the nascent strand in a reversible strand-displacement reaction (*strand-displacement*). Finally, the product is reversibly released from the hairpin (*equilibrium release*). PER is done at high polymerase concentrations so that either the reactant-catalyst binding or the product-catalyst unbinding, but not the polymerization step is rate-limiting. PER can extend primers of 10-12 nucleotides in just minutes at 37 ^{o}C [17], but extension of longer or shorter primers is much slower (Supp. Fig. 2), consistent with the notion that the binding strength between reactant and catalyst can be neither too weak nor too strong (Fig. 1b).

We show how to expedite PER in the strong-binding regime by coupling the reaction dissipative dehybridization of DNA by ATP-dependent helicases. Specifically, we use Rep-X, which is an engineered "super" helicase that has a higher unwinding activity than its wild type counterpart Rep and selectively targets DNA duplexes with an 3' single-stranded overhang[26]. This selectivity facilitates the design of catalytic reactions where the product-catalyst duplex is separated by Rep-X, but the reactant-catalyst complex does not have a 3' overhang—as is the case for PER—and is protected. As a result, a helicase can more rapidly remove a reaction product, speeding up the rate of the last step of a catalytic process without slowing down the initial substrate-catalyst binding step. We will show that Rep-X selectively un-

winds product-catalyst but not primer-catalyst duplexes, increases the product's off-rate, and thus speeds up the reaction in the strong binding regime.

To develop this expedited Primer Exchange Reaction, we first establish an analytical model that predicts the dependence of the PER rate on reaction temperature and primer length, and captures why PER occurs quickly only in a narrow range of primer lengths for a given temperature. We then use this model to predict the effect of helicase activity on the PER rate. Next, we test helicase activity on DNA complexes with and without a 3' overhang. Finally, we measure the PER rate in the presence of helicase and show that it agrees well with our prediction, demonstrating how helicases can be used as a predictable tool in DNA nanotechnology.

2 Results and Discussion

2.1 Analytical model of PER rate

To understand how the PER rate depends on the binding energy between the reactant/product and catalyst strands, we develop a simple analytical model that combines features of the three-step model for toehold-mediated strand displacement reactions[27] and from Michaelis-Menten kinetics[28]. In our model, the primer binds the hairpin during *equilibrium binding* with forward and reverse (or on- and off-) rate constants k_{1f} and k_{1r} respectively (Fig. 2a). We model the *DNA polymerization* and *strand-displacement* steps as a single, irreversible reaction with an effective rate constant, k_2 (Fig. 2a). Finally, the product is released from the hairpin during *equilibrium release* with forward and reverse rate constants k_{3f} and k_{3r} respectively (Fig. 2a).

When PER proceeds at steady-state and there is much more reactant than catalyst, the reaction can be modeled as a process in which only the reactant concentration $[R]$ and the product concentration $[P]$ change over time, *i.e.* the concentrations of the unoccupied catalyst $[C]$, the catalyst-reactant complex $[RC]$, and the catalyst-product complex $[PC]$ remain constant. This model is analogous to Michaelis-Menten kinetics for enzymatic reactions where the catalyst strand takes the role of the enzyme[28], except that we consider the conversion of reactant to product and the release of the product from the catalyst to be two separate steps, analogous to the three-step-model for DNA strand-displacement reactions[14]. Under these assumptions the differential equations governing the reaction are:

$$\frac{d[RC]}{dt} = k_{1f}[R][C] - k_{1r}[RC] - k_2[RC] = 0, \quad (1)$$

$$\frac{d[PC]}{dt} = k_{3f}[P][C] - k_{3r}[PC] + k_2[RC] = 0. \quad (2)$$

Assuming that the reactant and product bind equally strongly to the catalyst, with equilibrium constant K ($k_{1f} = k_{3f} \equiv k_f$ and $k_{1r} = k_{3r}$), and that the reactant concentration is much larger than the catalyst concentration, we can solve Equations 1 and 2 at steady-state to find that the reactant is consumed as $[R](t) = R_0 \exp(-t/\tau)$, where the reaction half-time τ is given by (see Supplementary Discussion 1 for derivation)

$$\tau = \left(\frac{1}{k_2} + \frac{K}{k_f}\right)\left(\frac{R_0}{C_0} + \frac{1}{KC_0}\right). \quad (3)$$

We assume kf is a standard hybridization rate between two short DNA strands, reflecting previous findings that this rate of hybridization is not strongly dependent on sequence length or base composition for 10-100 nucleotide reactants Equation (3) shows that in the limit of very strong binding between catalyst and reactant (large K: $\tau \approx \frac{K}{k_f}\frac{R_0}{C_0}$), τ is proportional to K, whereas in the limit of weak binding (small K: $\tau = \frac{1}{k_2}\frac{1}{KC_0}$), τ is proportional to $1/K$. In both cases τ is large and the reaction is slow. Only at intermediate binding energy—$K = \sqrt{\frac{1}{R_0}\frac{k_f}{k_2}}$—does τ have a minimum value that corresponds to a peak in reaction rate.

2.2 Comparison to experiments

To check that our model captures the essential features of PER, we next measure the reaction rate as a function of the binding energy between reactant and catalyst. In our experiments we vary temperature and the length of the binding domain on the catalytic hairpin as control parameters to tune this binding energy. We relate τ, the typical reaction half-time, to the domain length noting that the equilibrium constant depends on the free energy of hybridization between the primer and catalyst, $K = \exp[-\Delta G^o/k_B T]$. The ΔG^o of hybridization is proportional to strand length and can be calculated using the nearest-neighbour model as the sum of the free energies of each of the pairs of hybridized bases[12]. We use that $k_f \approx 3 \times 10^6 M^{-1}s^{-1}$ [14]. In our experiments C_0 is either 10 nM or 100 nM and R_0 is either 100 nM or 200 nM. The only unknown parameter in the model is k_2, the polymerization rate of Bst Large Fragment Polymerase, which Deng *et al.* measured to be around $10^{-3}s^{-1}$[29]. Using these input parameters, Equation (3) predicts that the reaction rate is maximal for 10-nucleotide primers at 25°C and for 12-nucleotide primers at 37°C and that shorter or longer primers lead to slower reactions.

To measure the concentration of product over time, we use the reporting scheme outlined in Figure 2b. The reporter was designed to have a 6-base overhang so that the rate constant for the reaction between product and reporter $k_{rep} \approx 10^{-3}\text{nM}^{-1}\text{s}^{-1} \gg k_{cat}$[27]. We verified this by measuring the product concentration over time both based on fluorescence measurements using the reporter in Fig. 2b and directly using gel electrophoresis, and we found good agreement (Supp. Fig. 3 & 4).

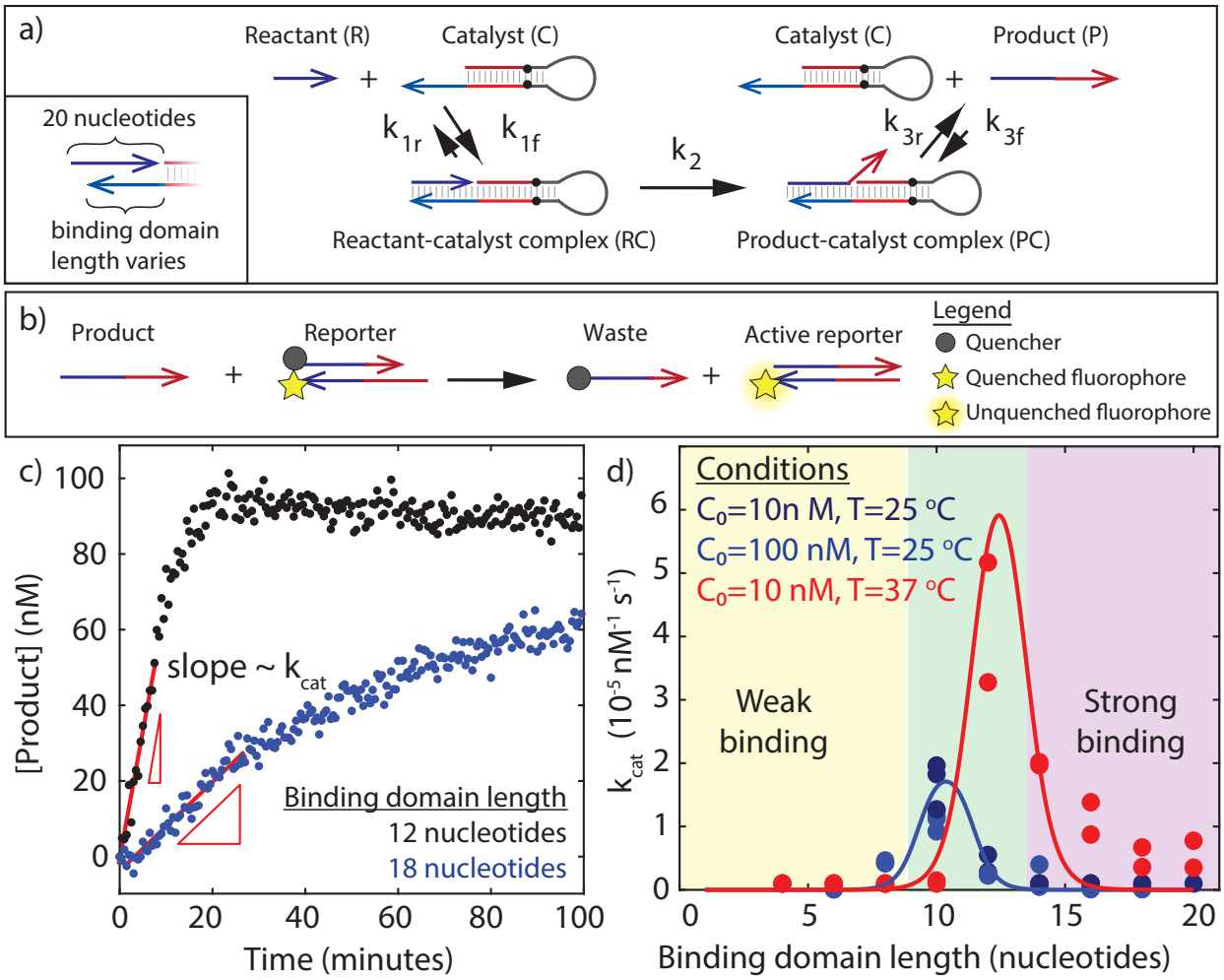

Figure 2: **PER rate is peaked for primer lengths of 10 to 12 nucleotides.** a) Overview of the PER reaction. The reactant or primer (blue) is a 20 nucleotide single-stranded DNA. It binds to the blue single-stranded binding domain on the catalytic hairpin. This binding domain can vary in length from 4 to 20 nucleotides. Black dots represent a stop sequence for DNA polymerase. Dark and light red and dark and light blue strands each have complementary sequences. b) Reporting scheme for measuring the output of PER. The PER product reacts with the reporter via a 6 nucleotide toehold strand-displacement reaction. This reaction separates the quencher-labeled strand from the fluorophore-labeled strand in the reporter complex and produces a fluorescent signal proportional to the product strand concentration. c) Measured PER product concentration as a function of time. The turnover frequency k_{cat} was determined by dividing the initial slope by R_0 and C_0. The experiment was conducted at 37 ^{o}C using catalyst hairpins with binding domains 12 (black) and 18 (blue) nucleotides in length. Both binding domains only contained A's and T's. $C_0 = 10$ nM and $R_0 = 100$ nM. d) The turnover frequency k_{cat} as a function of binding domain length for a range of experimental conditions. Dark and light blue dots represent experiments conducted at 25 ^{o}C and red dots represent experiments conducted at 37 ^{o}C. $C_0 = 10$ nM in all experiments except the light blue ones, where $C_0 = 100$ nM. R_0 is either 100 nM or 200 nM (see Supplementary Methods for details). The curves represent fits of Equation (3) to the data with k_2 as the only adjustable parameter. We find that $k_2 \approx 0.002$ at 25 ^{o}C and $k_2 \approx 0.008$ at 37 ^{o}C. In the model we use $R_0 = 100$ nM.

Figure 2c shows product formation over time for a typical PER experiment (See Supp. Fig. 5 for conversion from fluorescence to concentration). We measured the initial rate at which product strand is formed (Fig. 2c) and divided it by C_0 to obtain k_{cat}, a measure for reaction rate that is independent of catalyst concentration and is related to τ as $1/\tau \approx C_0 k_{cat}$ for $t \ll \tau$.

Our derivation of τ (Supp. Disc. 2) assumes that the reaction rate is either limited by reactant binding or by product release, and that the polymerization itself is not rate-limiting in PER. Consistent with that assumption we found that decreasing the concentration of DNA polymerase 10-fold does not decrease the reaction rate (Supp. Fig. 6).

Figure 2d shows both the predicted and measured reaction rates as function of the binding domain length at 25^oC and 37^oC. At 25^oC, the experimentally observed peak in reaction rate lies at around 10 nucleotides and at 37^oC the peak is around 12 nucleotides, in agreement with our predictions. Values of $k_2 = 2 \times 10^{-3} s^{-1}$ at 25^oC and $k_2 = 8 \times 10^{-3} s^{-1}$ at 37^oC produce a close correspondence between the model and the experiment (See Supplementary Discussion 2 for a list of the used parameters). These polymerization rates are consistent with the ones measured by Deng *et al.* who also found that the rate increases with temperature[29].

Despite an overall good agreement, the measured rates for long binding domains are higher than our predicted values, which can be vanishingly small. One reason for this may be that low DNA reaction rates can be difficult to measure precisely in bulk because some DNA strand may have sequence errors that allow them to react faster [30] and our strands are unpurified after solid-state synthesis so a fraction of strands is expected to contain deletions.

The agreement between the prediction of Equation (3) and our experimental findings shown in Figure 2 supports the idea that the PER is fast only when the occupancy time of the product is within a particular range. The occupancy time must be short enough that the product detaches, allowing the reaction to complete, but not so short that the reactant, which has the same occupancy time, cannot bind long enough for the polymerase to extend it while it is bound. At a given temperature, these occupancy times depend exponentially on the hybridization energy, meaning that PER is only efficient for sequences in a very narrow range of energies. Next, we ask whether the PER rate can be sped up by using an enzyme that separates DNA duplex regions at a rate independent of the hybridization energy. In this case, when the hybridization is fast and binding strong, product-catalyst separation would occur primarily because of enzymatically-driven separation, decoupling the PER rate from the hybridization energy. To test this idea, we next explore how the addition of an ATP-dependent helicase separates DNA complexes at a sequence length-independent rate and thereby enables a wider range of lengths for PER.

2.3 Predicted effect of helicase on PER

Helicases, a class of ATP-dependent enzymes that unwind double stranded DNA, can help expedite PER by increasing the product off-rate beyond the equilibrium rate. We use the engineered helicase Rep-X, which selectively targets complexes with a single-stranded 3' overhang. This selectivity is a desirable feature in PER because it causes Rep-X to remove product from the catalyst without affecting the residence time of the reactant on the catalyst. While in the ideal case Rep-X only unwinds complexes with 3' overhangs, Rep-X also unwinds double-stranded DNA without 3' overhangs, albeit at a lower rate[26]. We will measure this selectivity in Section 2.4.

To quantify how Rep-X affects the PER rate, we include terms in Equations (1) and (2) to account for the unwinding of the product-catalyst complex at rate k_h (see Fig. 3a) and the unintended removal of the reactant from the catalyst with a leak rate $k_l = L \times k_h$ (see Fig. 3b):

$$\frac{d[PC]}{dt} = k_{3f}[P][C] - (k_{3r} + k_h)[PC] + k_2[RC] = 0. \quad (4)$$

$$\frac{d[RC]}{dt} = k_{1f}[R][C] - (k_{1r} + k_2 + k_l)[RC] = 0. \quad (5)$$

Here, k_h is a rate constant with units s^{-1} and the leak parameter L is a dimensionless constant between 1 and 0 that captures the relative rate at which Rep-X unwinds complexes without 3' overhangs compared to complexes with 3' overhangs. L is 0 for a leak-free reaction and is 1 if the 3' overhang makes no difference. We follow the same derivation as outlined in section 2.1, but have to make an additional simplification (details in Supplementary Discussion 1) to arrive at an analytical expression for the reaction timescale in the presence of helicase:

$$\tau = \left(\frac{1}{K_2} + \frac{K}{k_f}\left(1 + \frac{k_l}{k_2}\right)\right)\left(\frac{\frac{R_0}{C_0}}{1 + K\frac{k_h}{k_f}} + \frac{1}{KC_0}\right) \quad (6)$$

Equation (6) shows that the addition of helicase introduced a second off-rate, k_h (and k_l which is proportional to k_h), which is similar to k_r, but not related to the on-rate via the equilibrium constant. Note that if $k_h = 0$, Equation (6) equals the expression in Equation (3) in which we did not consider a helicase, as it should.

Figure 3c depicts the predicted turnover frequency $k_{cat} = \frac{1}{C_0 \tau}$ as a function of binding domain length for varying helicase rates, considering a perfectly selective helicase ($L = 0$). It shows that the reaction rate is affected by helicase only in the strong binding regime. Before the peak, the reaction rate is limited by the on-rate of reactant and unaffected by the addition of helicase.

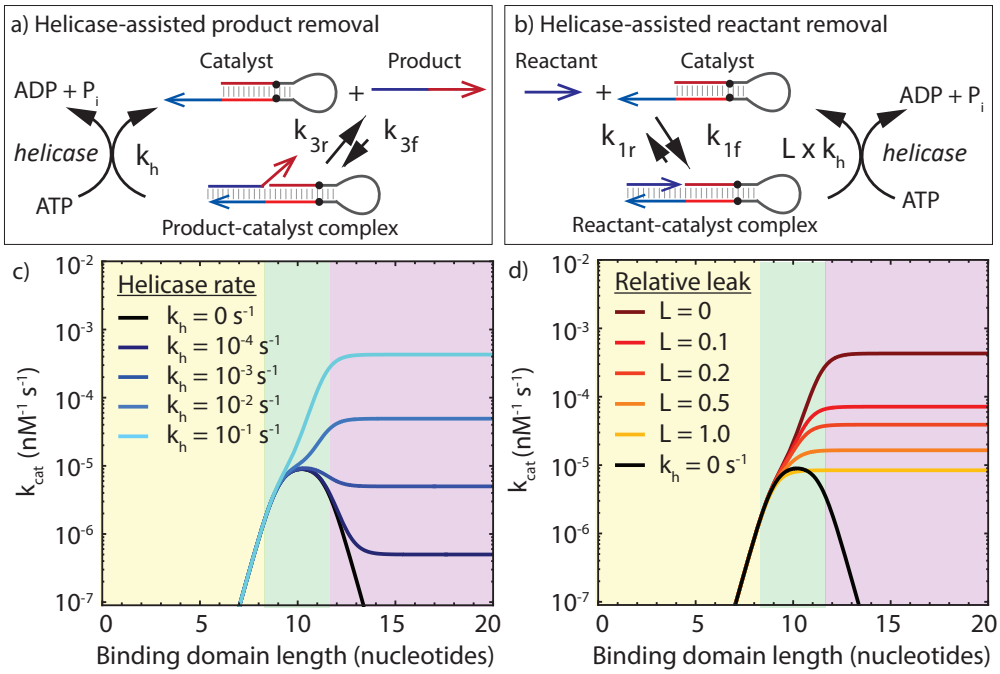

Figure 3: **PER rate depends on helicase-assisted and leak dehybridization rates.** a) Schematics of the intended reaction in which helicase removes product from the catalyst strand (top). b) Schematics of the leak reaction in which helicase removes unreacted primer from the catalyst strand. c) Turnover frequency versus binding domain length for a range of helicase-assisted product removal rates, as predicted by Equation (6). A higher helicase rate results in a faster reaction for long binding domain lengths. The rate at short binding domain lengths is unaffected. d) Turnover frequency versus binding domain length for a range of leak rates using $k_h = 0.1$ s^{-1}. Even a nonselective helicase ($L = 1$) expedites PER for large binding lengths, but not as effectively as selective helicases. The peak rate (at optimal binding domain length) can only be increased by a selective helicase. In all the calculations $C_0 = 100$ nM, $R_0 = 200$ nM, $T = 25$ ^{o}C, and $k_2 = 2 \times 10^{-3}$ s^{-1}.

After the peak, the reaction rate is limited by the product off-rate and increases due to the addition of helicase. The increase only manifests in the regime where $k_h > k_r$.

Figure 3d shows the influence of the unintended helicase-assisted removal of the reactant from that catalyst (with rate $L \times k_h$) on the PER rate. Notably, it shows that a selective helicase is not required to expedite PER, but higher selectivity results in a larger rate increase. Taken together, these findings show that a helicase could dramatically reduce the PER rate's sensitivity to domain length in the strong binding regime, even if it is not entirely selective.

2.4 Helicase unwinding rate and leak

To predict the effect of Rep-X on the PER rate, we measure k_h and the leak rate of Rep-X using the two reporter complexes shown in Figure 4a. These complexes have identical sequences except that one of the two reporters, $R_1 : R_1'$ (depicted in purple), has a 3' overhang whereas the other, $R_2 : R_2'$ (depicted in green), has a 5' overhang. When the reporter complexes are hybridized, the fluorophore on one reporter's strand is in close proximity to a quencher on the other, dampening the fluorescent signal. In equilibrium, the spontaneous off-rate of the $R_1 : R_1'$ complex is negligible and all R_1 is hybridized to R_1'. The fluorescent signal thus indicates the concentration of unhybridized R_1, from which we can calculate k_h.

Figure 4b shows the concentration of R_1 over several hours, beginning directly after the addition of Rep-X helicase and ATP to a solution of $R_1 : R_1'$ complex. Initially, most of the 100 nM reporter complex was unhybridized, indicating high Rep-X activity. Over time, $[R_1]$ decreased, suggesting that the Rep-X unwinding rate decreased over time. We found that this decrease is due to ATP depletion, as adding additional ATP causes the fluorescence signal to increase and subsequently decay again (Supp. Fig. 7) and higher ATP concentrations result in slower decays (Supp. Fig. 8).

We used the measurements in Figure 4b to obtain an order of magnitude estimate of the helicase rate by noting that the non-zero concentration $[R_1]$ is due to a competition between the helicase-mediated off-rate k_h and the on-rate k_f. The binding domain of R_1 to R_1' is 15 nucleotides, so the equilibrium off-rate k_r is negligible and in equilibrium $[R_1]$ should be near zero. We thus calculate the k_h values at the three Rep-X concentrations tested at times $t = 0$ and $t = 30$ minutes from $[R_1]$ at those times using $k_h = k_f \frac{[R_1][R_1']}{[R_1:R_1']}$.

The measured values of k_h are shown in the inset of Figure 4b, which show that k_h increases with Rep-X concentration. After 30 minutes k_h is smaller for all tested Rep-X concentrations than it was at time 0. The difference in the rates at these two times also increases as Rep-X concentration does.

Next, we estimate the leak of Rep-X helicase—that is the relative rate of unwinding of complexes without a 3' overhang—by comparing the amount of unbound reporter strand in the experiment containing the purple complex with a 3' overhang to the experiment containing the green complex with a 5' overhang, shown in Figure 4c. Interestingly, the leak reaction rate appears to depend only weakly on the Rep-X concentration. As a consequence, the leak is approximately 1% for 1 µM Rep-X but close to 10% for 100 nM Rep-X. The leak reaction is likely due to fraying at the blunt end of the R_2 complex, resulting in temporary single stranded 3' overhangs that are substrates for Rep-X.

Based on these measurements of Rep-X's DNA unwinding performance, we can refine our prediction of whether Rep-X will speed up PER and by how much. We found that the 100 nM Rep-X resulted in $10^{-1}\text{s}^{-1} < k_h < 10^{-3}\text{s}^{-1}$ in the 10 to 30 minute window, in which we expect most of the reaction to complete. At those experimental conditions we find that the leak rate is on the order of 10% which should reduce the efficacy of helicase on expediting PER slightly, as shown in Fig. 3d. Using those values we expect the PER rate to be unaffected by Rep-X in the weak binding regime (0-10 nucleotides), but sped up by at least an order of magnitude in the strong binding regime (10-20 nucleotides) as shown in Fig. 3c.

This predicted speed-up in reaction rate does not come freely and requires the consumption of ATP. As an aside, we quantify the rate of fuel consumption based on the rate decay due to ATP depletion shown in Fig. 4b. An exponential fit to the data for 100 nM Rep-X shows that at those conditions the ATPase rate is on the order of $6 \times 10^{-4}\text{s}^{-1}$ (Supp. Fig. 9). That means that at the start of the reaction, where [ATP]= 1 mM, each Rep-X molecule consumes 6 ATP molecules per second.

2.5 Helicase increases PER rate

Equipped with estimates for the helicase-directed off-rate and relative leak of Rep-X helicase, we moved on to test the prediction that Rep-X can increase the PER rate in the strong binding regime by expediting the off-rate of the product without affecting the reactant on-rate.

Figure 5a shows the product concentration as a function of time for a PER reaction with a 16 nucleotide binding domain. In absence of helicase, the strong product-catalyst bond prevents rapid conversion even with the increased catalyst concentration. The addition of 100 nM Rep-X and 1 mM of ATP increases the initial rate 30-fold.

We measured the PER rates for binding domain lengths varying from 6 to 18 nucleotides with and without helicase in triplicate and the results are shown in Figure 5b. The black curve is a fit of Equation (3) to the data without helicase using the value for $k_2 = 2 \times 10^{-3}$ we found in Figure 2. The red curve is a fit of Equation (6) to the data with helicase, using the same value

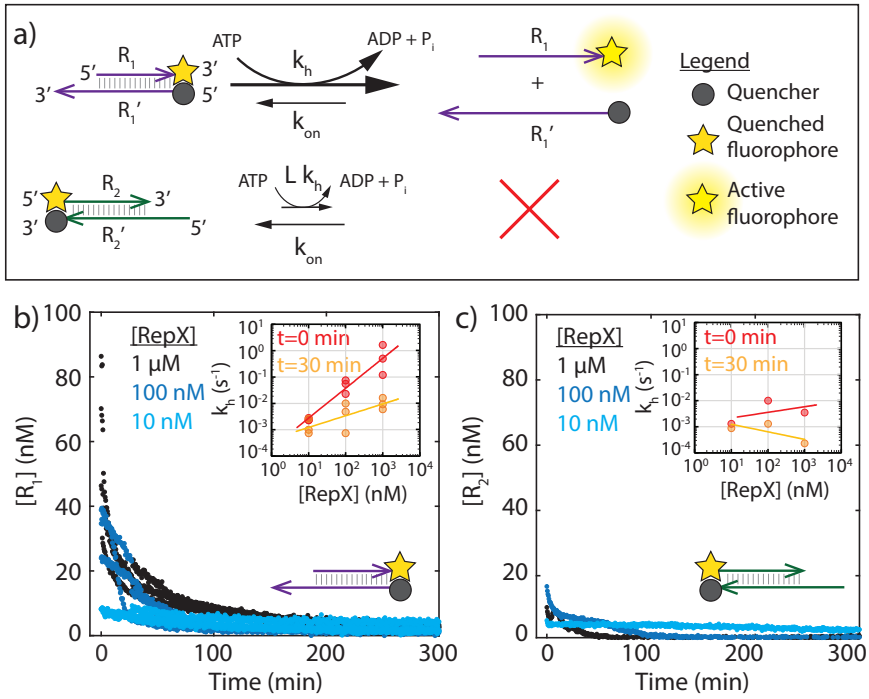

Figure 4: **Rep-X preferentially unwinds DNA complexes with a single-stranded 3' overhang.** a) Overview of the experiment to test Rep-X helicase performance. Strands of the same color have complementary sequences and hybridize. The dark yellow stars indicate a quenched FAM fluorophore, the gray sphere indicates the Iowa Black quencher. Measured fluorescence increases with increasing concentration of R_1 or R_2. Rep-X helicase preferentially unwinds complexes with a 3' (rather than 5') single-stranded overhangs. b) Concentration of free reporter strand R_1 in a sample with the reporter complex R_1:R_1' directly after the addition of Rep-X helicase and 1 mM ATP. We calculated $[R_1]$ by comparing the fluorescent signal during the experiment with the fluorescent signal of a sample with separate fluorophore and quencher strands. The inset shows the inferred k_h for a range of Rep-X concentrations directly after mixing and after half an hour. The decrease in helicase activity over time is due to ATP depletion (Supp. Fig. 7 and 8). c) Same data as in b) but for a reporter with a 5' overhang. The helicase-mediated unwinding rate is substantially lower for $R_2 : R_2'$ than for $R_1 : R_1'$.

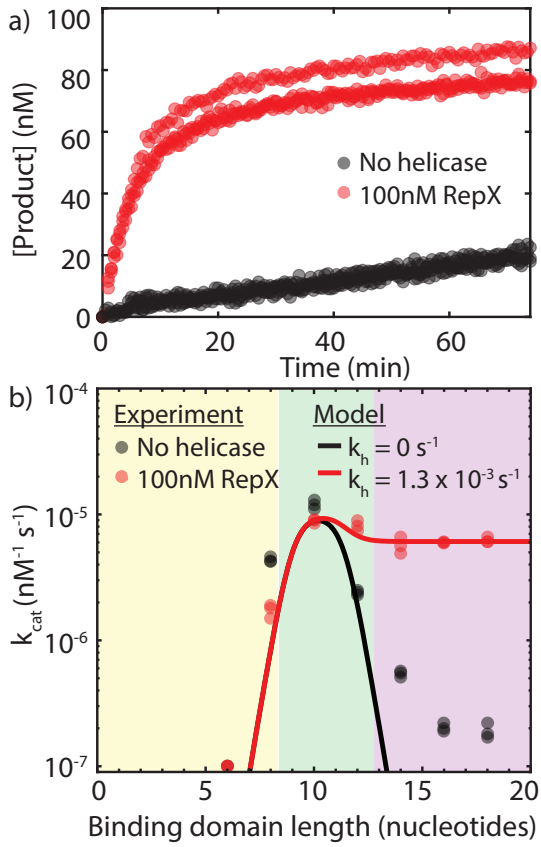

Figure 5: **Rep-X expedites PER.** a) Product concentration increase over time in PER reactions with a binding domain length of 16 nucleotides. The black dots depict experiments in the absence of helicase. The red dots are samples containing 100 nM Rep-X helicase and 1 mM ATP. In both experiments $C_0 = 100$ nM and $R_0 = 200$ nM. Each experiment is done in triplicate. b) The turnover frequency k_{cat} is plotted versus the binding domain length on a semilog plot, resulting in the classical volcano plot. Curves represent the predictions from Equation (3) and (6). The black curve is a fit to the data in absence of helicase with the only adjustable parameter $k_2 = 2 \times 10^{-3}$. The red curve is a fit to the data in presence of helicase using $k_2 = 2 \times 10^{-3}$ with $k_h = 1.3 \times 10^{-3}$ and $L = 0.1$ as the only adjustable parameters. Data points represent individual experiments. Each experiment is done in triplicate.

for k_2 and with the helicase and leak rates as the only adjustable parameters.

The model for the PER rate in presence of helicase matches the experimentally obtained rates well using $k_h = 1.3 \times 10^{-3}$ and $L = 0.1$. This helicase rate is on the low end of the range measured in Figure 4 which is possibly due to a higher overall DNA concentration in the PER experiments (400 nM compared to 100 nM). Notably the hairpins also have a single-stranded 3' overhang so a substantial portion of the helicase action is likely wasted on opening hairpins instead of removing product from hairpins.

The presence of helicase limited the yield of the PER reaction(Supp. Fig. 10,11), so we studied the reaction at high catalyst concentrations. Under these conditions the quasi-steady state assumption is longer valid. Surprisingly, our model nonetheless captured the experimentally observed reaction rates as functions of binding energy. This is likely because the main purpose of the model is to capture a transition from reactant binding being the rate-limiting step to product release being rate limiting. This transition does not rely on the quasi-steady-state assumption.

In summary, we showed that Rep-X-assisted product removal can expedite PER in the strong binding regime. This finding suggests that Rep-X could also be used to expedite other multi-step DNA reactions where the dehybridization step is rate-limiting. However, for Rep-X to become a useful tool in DNA nanotechnology, there needs to be a way to protect DNA complexes that must not be dehybridized from from unwinding by Rep-X. We already showed that DNA complexes without 3' overhangs are protected from Rep-X-mediated unwinding. Here we asked if specific complexes with 3' overhang can also be protected. We tested the replacement of the 3' overhang from DNA to methylated RNA, because methylated RNA has similar binding properties to DNA and can form Watson-Crick base pairs with DNA strands, but it is not recognized as a substrate by most enzymes. Indeed we found that unwinding rate of Rep-X is dramatically reduced for complexes with methylated RNA toeholds compared to DNA toeholds (Supp. Fig. 12). This suggests a design strategy for protecting DNA complexes from unwanted unwinding. It can however not be applied to PER because the 3' methylated RNA binding domain is also not recognized as a template by the DNA polymerase (Supp. Fig. 13).

Taken together these data show that a helicase can be used to expedite DNA reactions where the off-rate is the rate limiting step and that methylated RNA can be used to protect DNA duplexes from unwinding by the helicase.

3 Conclusions

Here we asked whether helicases, enzymes that catalyze the processive dehybridization of DNA strands, can be incorporated into designed DNA reaction networks to selectively increase off-rates and thus increase reaction flux. In this case study, we have shown, both theoretically and experimentally, that the PER rate can be increased more than 30-fold compared to the equilibrium rate in the strong binding regime at the cost of ATP-hydrolysis, thus circumventing Sabatier's principle. These findings suggest could also expedite many other DNA reactions where the off-rates are limiting[14, 15, 18].

A key advantage of Rep-X is its propensity to unwind only some duplexes (those with 3' overhangs) which will allow its use as a sequence-specific agent within programmed reaction cascades. To direct helicase activity, complexes that should be actively dehybridized in a reaction could present 3' overhangs, while duplexes whose separation could lead to unwanted interactions could be protected from helicase action by either removing their 3' overhang or replacing the bases on these overhangs with RNA or methylated RNA.

The concept of active removal of products from catalysts is used broadly—albeit less explicitly—in the polymerase chain reaction (PCR) reaction, where the temperature is oscillated to alternate between strong primer binding and quick product release. Also during the loop-mediated isothermal amplification of DNA, LAMP, dissipation by a polymerase drives product removal [31]. Milligan and Ellington showed that RecA, an ATP-dependent DNA-binding protein, could also speed up DNA reaction cycles[20]. Non-enzymatic catalytic DNA reactions remove product strands via toehold-mediated strand displacement, dissipating energy by forming low energy, fully hybridized waste-products[32]. In this work, we developed a mechanistic understanding of how dissipation can be harnessed that, by its relation to general ideas in chemistry, can be used to drive the design of a wider range of dissipative reaction processes to circumvent kinetic limitations. This framework could conceivably also serve as a foundation for a wider range of incorporation of active agents in DNA networks.

The finding that a dissipative process can be used to expedite a reaction beyond its equilibrium limit imposed by Sabatier's principle raises the question of how much energy needs to be minimally be dissipated to expedite a reaction by a certain amount[33]. We are certainly far from the efficiency limit, because in our experiments with 100 nM Rep-X, one enzyme hydrolyzed on average 6 ATP per second while only separating on average one base pair.

Theoretical work by Hopfield from 1974 shows the driven release of molecules from a template is required for kinetic proofreading, a process that increases reaction specificity at the cost of energy consumption[34]. The dissipative strand-separating function of helicases could potentially also be used to increase specificity in DNA reactions *via* this kinetic-proofreading method.

4 Acknowledgements

P.G.M. and R.S. acknowledge funding from the ARL award W911NF2020281, DOE grant number 90085886 and the ARO grant 90089252. M.G. acknowledges NSERC Canada PDF. Work in the Ha lab is supported by the NIH grant (R35 GM 122569). We thank Moshe Rubanov and Sam Schaffter for critically reading the manuscript. T.H. is an investigator with the Howard Hughes Medical Institute.

5 Author contributions

P.G.M. and R.S. conceived the experiments. P.G.M. performed the experiments and analysis. M.G. produced and purified the helicase. All authors contributed to writing of the paper.

References

[1] Neville R. Kallenbach, Rong-Ine Ma, Nadrian C. Seeman, *Nature* 1983, **305**, 829-831. Experimental introduction of designed stable DNA junctions that underlie much of structural DNA nanotechnology.

[2] Bernard Yurke, Andrew J. Turberfield, Allen P. Mills Jr., Friedrich C. Simmel & Jennifer L. Neumann, *Nature* 2000, **406**, 605-608. Introduction of designed DNA strand-displacement reactions that underlie many dynamical DNA nanotechnology.

[3] Yaakov Benenson, Binjamin Gil, Uri Ben-Dor, Rivka Ador & Ehud Shapiro, *Nature* 2004, **429**, 423-429. Demonstrates the power of DNA nanotechnology by building a DNA computer that recognizes disease-related mRNA in vitro and produces an output DNA.

[4] Randolph Lopez, Ruofan Wang & Georg Seelig, *Nature Chemistry* 2018, **10**, 746-754. Experimentally built a DNA computer that recognizes patterns in mRNA concentration that signal disease.

[5] Paul W. K. Rothemund, *Nature* 2006, **440**, 297-302. Demonstration of DNA origami: the folding of DNA into complex well-defined 2D shapes.

[6] Swarup Dey, Chunhai Fan, Kurt V. Gothelf, Jiang Li, Chenxiang Lin, Longfei Liu, Na Liu, Minke A. D. Nijenhuis, Barbara Saccà, Friedrich C. Simmel, Hao Yan & Pengfei Zhan, *Nature Reviews Methods Primers* 2021, **1**, 13. Reviews methods, applications, and understanding of DNA origami.

[7] Guido Grossi, Mette Dalgaard Ebbesen Jepsen, Jørgen Kjems & Ebbe Sloth Andersen, *Nature Communications* 2017, **8**, 992. Built a reconfigurable DNA box that contains an enzyme to control its activity.

[8] Ebbe S. Andersen, Mingdong Dong, Morten M. Nielsen, Kasper Jahn, Ramesh Subramani, Wael Mamdouh, Monika M. Golas, Bjoern Sander, Holger Stark, Cristiano L. P. Oliveira, Jan Skov Pedersen, Victoria Birkedal, Flemming Besenbacher, Kurt V. Gothelf & Jørgen Kjems, *Nature* 2009, **459**, 73-76. Demonstrated a reconfigurable DNA origami in the form of a box with a lid that opens on demand.

[9] Leonard M. Adleman, *Science* 1994, **266** 1021-1024. Experimental demonstration of a DNA-based computation.

[10] Georg Seelig, David Soloveichik, David Yu Zhang & Erik Winfree, *Science* 2006, **314** 1585-1588. Experimental demonstration of DNA logic gates.

[11] Kevin M. Cherry & Lulu Qian, *Nature* 2018, **559**, 370-376. Demonstration of a DNA-based neural network that performs pattern recognition tasks.

[12] John SantaLucia Jr, *Proc. Natl. Acac. Sci. USA* 1998, **95**, 1460-1465. Summarized measurements of nucleic acid binding energies into a single set. Value at 37^{o}C and 1 M NaCl from table 1.

[13] David Yu Zhang, Sherry Xi Chen & Peng Yin, *Nature Chemistry* 2012, **4**, 208-214. Lowest theoretical upper bound value taken from the mismatch case in Figure 1c.

[14] David Yu Zhang & Georg Seelig, *Nature Chemistry* 2011, **3**, 103-113. Review showing various DNA-strand displacement circuits. Most occur over many hours.

[15] Kevin M. Cherry & Lulu Qian, *Nature* 2018, **559**, 370-376. Developed a neural network that operates on the timescale of 3 to 20 hours based on DNA hybridization reactions involving over 100 strands.

[16] Kevin Montagne, Raphael Plasson, Yasuyuki Sakai, Teruo Fujii & Yannick Rondelez, *Mol. Sys. Bio.* 2011, **7**, 466. First introduction of the PEN toolbox.

[17] Jocelyn Y. Kishi, Thomas E. Schaus, Nikhil Gopalkrishnan, Feng Xuan & Peng Yin, *Nature Chemistry* 2018, **10**, 155-164. First introduction of the primer exchange reaction. Fig. 2d shows the optimal rate.

[18] Lulu Qian & Erik Winfree, *Science* 2011, **332**, 1196-1201. Showed computations that include seesaw-gates which rely on the reversible binding of 5-nucleotide toeholds and operate on the timescale of hours.

[19] Gadi Rothenberg. Catalysis: Concepts and Green Applications. John Wiley &Sons, Ltd. Publising, 2008. Textbook on the basics of catalysis. Sabatier's principle is discussed in section 2.3.

[20] J. N. Milligan & A. D. Ellington, *Chemical Communications* 2015, **51**, 9503-9506. Used RecA to modulate binding energy in catalytic DNA reactions and thereby speed them up.

[21] Guillaume Gines, Roberta Menezes, Kaori Nara, Anne-Sophie Kirstetter, Valerie Taly & Yannick Rondelez, *Science Advances* 2020, **6**, 6. Used PEN-toolbox for sensitive micro RNA detection.

[22] Sinem K. Saka, Yu Wang, Jocelyn Y. Kishi, Allen Zhu, Yitian Zeng, Wenxin Xie, Koray Kirli, Clarence Yapp, Marcelo Cicconet, Brian J. Beliveau, Sylvain W. Lapan, Siyuan Yin, Millicent Lin, Edward S. Boyden, Pascal S. Kaeser, German Pihan, George M. Church & Peng Yin, *Nature Biotechnology* 2019, **37**, 1080-1090. Used PER to amplify signal in an in-vivo protein imaging tool.

[23] G. Gines, A. S. Zadorin, J.-C. Galas, T. Fujii, A. Estevez-Torres & Y. Rondelez, *Nature Nanotechnology* 2017, **12**, 351-359. Built reaction networks using PEN toolbox that controls positioning of microscopic particles.

[24] Anton S. Zadorin, Yannick Rondelez, Guillaume Gines, Vadim Dilhas, Georg Urtel, Adrian Zambrano, Jean-Christophe Galas & Andé Estevez-Torres, *Nature Chemistry* 2017, **2017**, 990-996. Used PEN toolbox to create spatial pattern that controlled particle aggregation.

[25] Thomas E. Schaus, Sungwook Woo, Feng Xuan, Xi Chen & Peng Yin, *Nature Communications* 2017, **8**, 696. Used PER to built a molecular-event recorder that tracks the sequence in which molecules have been in close proximity.

[26] Sinan Arslan, Rustem Khafizov, Christopher D. Thomas, Yann R. Chemla & Taekjip Ha, *Science* 2015, **348**, 344-347. Engineering and characterization of Rep-X helicase.

[27] David Yu Zhang, Erik Winfree, *J. Am. Chem. Soc.* 2009, **131**, 17303-17314. Introduction of theoretical model to predict toehold-mediated DNA strand-displacement rates and its experimental validation. $k_f = 3 \times 10^6$ M/s from Fig. 3 was used.

[28] Kenneth A. Johnson & Roger S. Goody, *Biochemistry* 2011, **50** 8264-8269. Translation of 1913 paper by Michaelis and Menten on enzyme kinetics. Our derivation is based on this paper.

[29] Yingnan Deng, Liang Ma, Qianqian Han, Changyuan Yu, Alexander Johnson-Buck & Xin Su, *Nano Lett.* 2020, **20**, 2688-2694. Measurements of polymerization rate at different temperatures are estimated based on Figure S7.

[30] Jamal Temsamani, Michael Kubert & Sudhir Agrawal, *Nucleic Acid Research* 1995, **23**, 1841-1844. Sequence analysis of incomplete side products formed during DNA synthesis shows that deletions occur anywhere in the strand with increased frequency near the 3' end.

[31] Tsugunori Notomi, Hiroto Okayama, Harumi Masubuchi, Toshihiro Yonekawa, Keiko Watanabe, Nobuyuki Amino & Tetsu Hase, *Nucleic Acids Research* 2000, **28**, 63. First introduction of LAMP, an isothermal DNA amplification tool that uses Bst polymerase to displace strands.

[32] David Yu Zhang, Andrew J. Turberfield, Bernard Yurke & Erik Winfree, *Science* 2007, **16** 1121-112. First experimental work to show an isothermal autocatalytic DNA reaction.

[33] Henry Hess, 2018, arXiv preprint arXiv:1803.07415. Theoretical work asking the minimum rate of energy dissipation required for a molecular machine to speed up a chemical reaction.

[34] J. J. Hopfield, *Proc. Natl. Acad. Sci.* **71**, 1974, 4135-4139. Theoretical work showing that a dissipative process can decrease error rates in templated reactions.