Kinetic Proofreading can Enhance Single Nucleotide Discrimination in a Non-enzymatic DNA Strand Displacement Network

Rakesh Mukherjee[1], Aditya Sengar[1], Javier Cabello-Garcia[2], Thomas E. Ouldridge[1]*

[1]Department of Bioengineering, Imperial College London, UK. [2]Department of Chemistry, University College London, UK.

Email: t.ouldridge@imperial.ac.uk

SUMMARY

Kinetic proofreading is used throughout natural systems to enhance the specificity of molecular recognition. At its most basic level, kinetic proofreading uses a supply of chemical fuel to drive a recognition interaction out of equilibrium, allowing a single free-energy difference between correct and incorrect targets to be exploited two or more times. Despite its importance in biology, there has been little effort to incorporate kinetic proofreading into synthetic systems in which molecular recognition is important, such as nucleic acid nanotechnology. In this letter, we introduce a DNA strand displacement-based kinetic proofreading motif, showing that the consumption of a DNA-based fuel can be used to enhance molecular recognition during a templated dimerization reaction. We then show that kinetic proofreading can enhance the specificity with which a probe discriminates single nucleotide mutations, both in terms of the initial rate with which the probe reacts and the long-time behaviour.

INTRODUCTION

Specificity of molecular interactions is at the heart of biology. In particular, selective nucleotide base pairing drives information propagation in DNA replication, RNA transcription and protein translation. Most simply, molecular recognition can be driven by differences in binding free energy between two candidate molecules and a substrate, resulting in an equilibrium with a biased toward one candidate. However, in many biosynthetic processes, the specificity of correct product formation is orders of magnitude higher than that suggested by free-energy differences, despite additional complications that make discrimination more challenging than implied by the equilibrium picture[1]. Translation of protein from mRNA operates with an error rate[3] of 10^{-4}, and DNA replication with an astonishingly low error rate of 10^{-9}. To describe such unusually high accuracy, Hopfield[5] and Ninio[6] independently proposed “kinetic proofreading” (KP) (Fig. 1a), in which small free-energy differences are utilised repeatedly over multiple steps in a reaction cycle to achieve a significant overall discrimination. The idea has been widely adopted, and the same basic principle has been identified in antigen recognition by T-cells[7], aminoacylation of tRNA[8], disentanglement of DNA by topoisomerases[9], specific protein degradation via ubiquitination[10], and chromatin remodelling [11] .

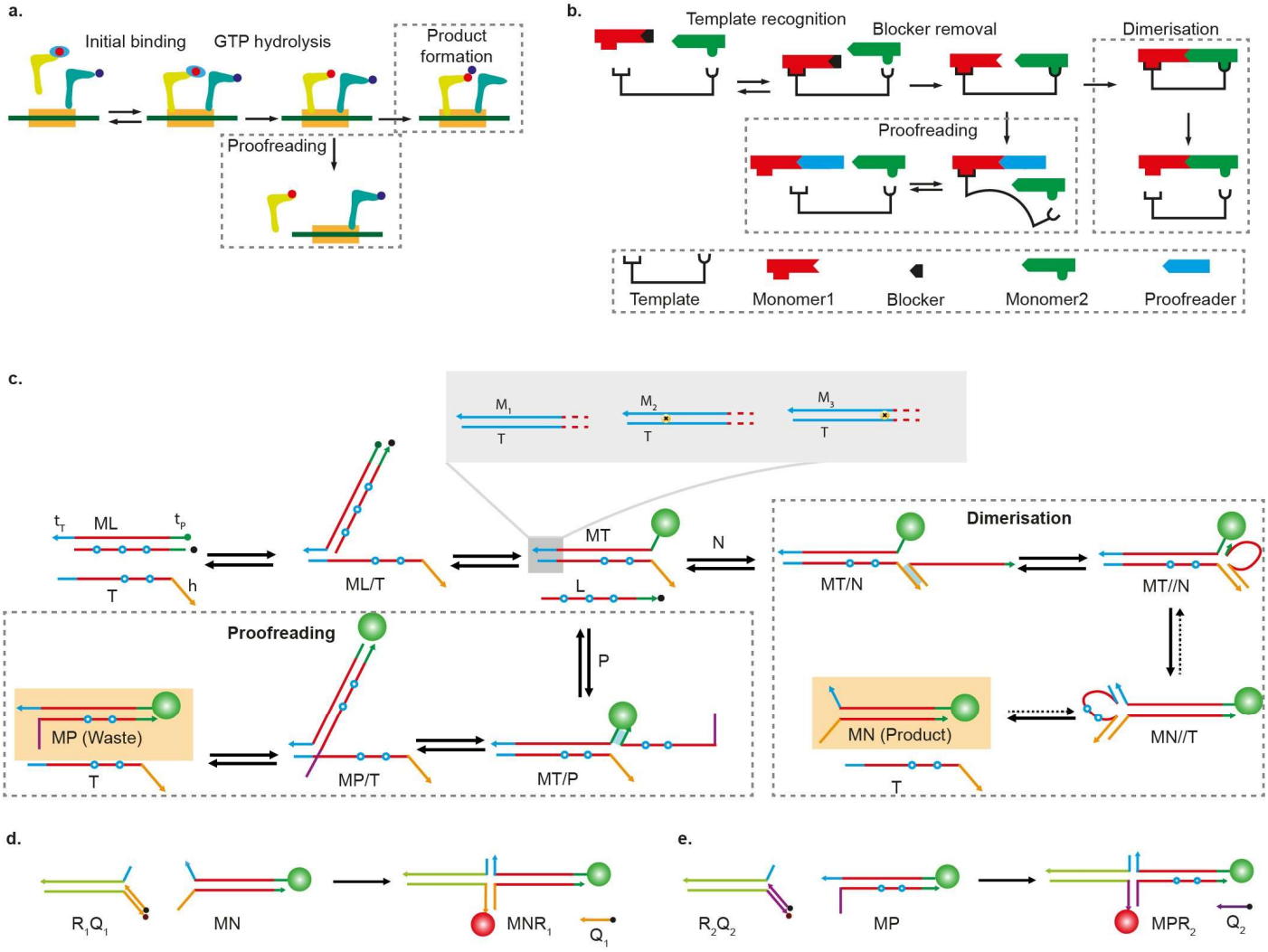

Fig. 1: Natural and synthetic kinetic proofreading networks. In all cases, reactions that are strongly thermodynamically disfavoured are shown with dotted arrows. (a) A simplified diagram of tRNA selection by the ribosome during translation, following Hopfield's kinetic proofreading model. The existence of two bound states prior to product formation allows for two stages of discrimination via unbinding rates. (b) A schematic representation of the proposed synthetic kinetic proofreading mechanism, showing its fundamental similarity with (a). The proposed system performs catalytic templated dimerization, with kinetic proofreading applied to the first monomer in the dimer. The catalytic requirement means that the second monomer does not bind the template strongly, but instead undergoes rapid binding and unbinding. (c) Domain-level design of a DNA-based kinetic proofreading network. Blocked monomer *ML* binds to the template *T* in two steps via TMSD to form the intermediate *MT*; initial discrimination is performed based on the complementarity of the blue toehold. We have three strands as test candidates (blow-up in grey): M_1 is the correct one with perfect toehold complementarity with the template. M_2 and M_3 are the incorrect ones having single nucleotide mismatch between them and the template toehold. The intermediate can be converted into waste complex *MP* by the proofreader *P or* dimerised via an HMSD reaction with the second monomer *N* to form the product *MN*, allowing a second stage of discrimination. After either process, *T* is released to act as a substrate in another reaction cycle. Blue circles with white core indicate strategic mismatches with the monomers and are common for all the monomers. The invasion of *ML* by T is reported by the increase in fluorescence of the *M* molecule. Product and waste formation are reported by two external reporters R_1Q_1 (d) and R_2Q_2 (e). Further design details are in Supplemental Fig. S1-2.

In a classic KP motif, as illustrated in Fig. 1a, a candidate molecule first binds to a substrate. Then the molecule can either unbind or undergo a chemical transition to another bound state. This moment is the first discrimination point: candidate molecules with less affinity for the substrate will unbind faster and are thus less likely to proceed to the second bound state. From the second bound state, there is another opportunity to detach prior to incorporation into the final product, giving a second opportunity to discriminate. The two stages of discrimination doubly enrich the correct product over the incorrect product.

A crucial component of KP is that it necessarily functions out of equilibrium[12,13]. Just as a population inversion in a laser can only be maintained by pumping atoms into excited states, chemical free energy must be used to maintain a non-equilibrium ratio of correct to incorrect molecules in the second bound state. Typically, this free energy is supplied by the consumption of chemical fuel molecules, such as nucleotide triphosphates, allowing the substrate to be systematically driven around a cycle of states and avoid equilibrium.

Discriminating similar molecules is also a ubiquitous motif in synthetic nanotechnology, whether in computational strand displacement cascades[14], tile assembly systems[15] or diagnostics-based applications[16]. The ability to discriminate

between perfectly matching and almost matching sequences is particularly relevant to single nucleotide polymorphism (SNP) detection. SNPs are single nucleotide mutations in specific gene sequences and are markers for various diseases[17]. Current strategies for accurate genotyping involve enzymatic reactions, are time-consuming, and costly[18]. Alternate non-enzymatic strategies utilising DNA binding or strand displacement[19-25] often face challenges with sensitivity or specificity.

Synthetic nanotechnologists have designed elegant systems in which discrimination between intended and unintended complexes is enhanced relative to a baseline[22,25-28]. Typically, the approach is to modify a design to increase the number of mismatched base pairs that must be formed to make the unintended product. Although such an approach implements a form of proofreading, it does not constitute non-equilibrium KP, where a single free-energy difference is exploited multiple times. Indeed, to the best of our knowledge, *de novo* KP has not previously been demonstrated in synthetic systems, despite its importance in nature.

We introduce non-enzymatic DNA-based KP that discriminates between very similar recognition domains through a fuel-consuming cycle. We apply KP to enhance the specificity with which a molecule is incorporated into a two-stranded DNA dimer by a catalytic template, a synthetic analogue of tRNA charging. We first characterise the individual discrimination steps, then demonstrate that the proofreading motif enhances the overall specificity of dimerization. Finally, we modify the network to demonstrate that KP can enhance the detection of generic SNPs in ssDNA.

RESULTS

Network Design for Kinetic Proofreading. We construct an enzyme-free, DNA-based synthetic KP system. DNA is used purely as an engineerable molecule that predictably assembles into well-defined structures driven by Watson-Crick-Franklin base pairing. Our KP system exploits the widely-used motif of toehold mediated strand displacement[29-32] (TMSD). In TMSD, a single invader strand displaces an incumbent from an incumbent-target duplex (see Fig. 1c). The process is accelerated and pushed thermodynamically downhill by a toehold of available bases in the target to which only the invader can bind. Alongside TMSD, we also exploit the handhold mediated strand displacement (HMSD) reaction[33,34] (also shown in Fig. 1c). HMSD is analogous to TMSD, but the binding to the target is accelerated by a handhold in the incumbent rather than the target.

The network is illustrated at a domain level in Fig. 1c. Domains are sections of single-stranded (ss)DNA that are designed to bind as a collective unit. In this network, the monomers *M* and *N* that form *MN* are each single strands of DNA (rather than individual nucleotides or amino acids). Dimerization in solution is inhibited by the presence of a blocker strand *L* bound to *M,* but can occur rapidly via a template T*,* which acts as the substrate in our system. First, using a toehold (t_T), *M* can bind to *T* via TMSD and remove *L,* revealing a second short toehold domain (t_P). The newly available toehold t_P can be used by the proofreader strand *P* to initiate a second TMSD reaction to form a waste duplex *W*, or a second monomer *N* can bind to the handhold h and then complete HMSD to form a dimeric product *MN.*

We consider three versions of the first monomer: M_1, M_2 and M_3. These monomers differ by a single base in the toehold recognition domain t_T; M_1 has a perfectly matching toehold for template recognition, whereas M_2 and M_3 have one mismatch at different positions in the toehold. The scheme is intended to provide two stages at which the free energy of toehold binding can be used to distinguish between M_1 and M_2 or M_3. First, after the initial binding of *M* to T*,* the strength of the toehold t_T determines the probability that the monomer will remain bound for long enough for blocker displacement[35]

$$ML + T \rightarrow MT + L. \quad (1)$$

Second, a stronger toehold interaction can also reduce the probability of successful displacement by the proofreader[35],

$$MT + P \rightarrow MP + T, \quad (2)$$

allowing for the concentration of M_1T to be further enriched relative to M_2T *or* M_3T and thereby enhancing the relative rate at which M_1N is formed.

Complete design details, including sequences, are provided in Supplemental Fig. S1 and Table S13. Long binding domains (red domain, Fig. 1c) contain mismatched base pairs (blue circles) in the initial *ML* complexes; these mismatches are gradually eliminated during subsequent reactions. The mismatches are common to all monomers and provide a "hidden thermodynamic drive"[30,36] towards formation of the final product *MN*.

We initially consider a fixed set of monomer, template, and blocker strands. We explore the performance of the individual substeps and overall network for different designs of the proofreader strand.

Monomers binding to the template. We consider the rates of invasion of blocked monomer complexes *ML* by the template *T*. The monomers were labelled with Cy3 fluorophore, and the blocker strand *L* was labelled with quencher IowaBlack-FQ. The progress of the reaction was monitored by the increase in Cy3 fluorescence when *L* was displaced by *T* (Fig. 2a). We performed the reaction with a range of *T* concentrations for 8nM of M_1, M_2 and M_3; the results for [*T*]=10nM are plotted in Fig. 2b, showing the correct monomer M_1 forming a duplex with *T* faster and with a higher equilibrium yield than the mismatched monomers (Supplemental Fig. S32). Solid lines in Fig. 2b show fits to an ODE

model for each system (Supplemental Note 11); these fits used a single set of parameters for each system to fit all concentrations of *T*.

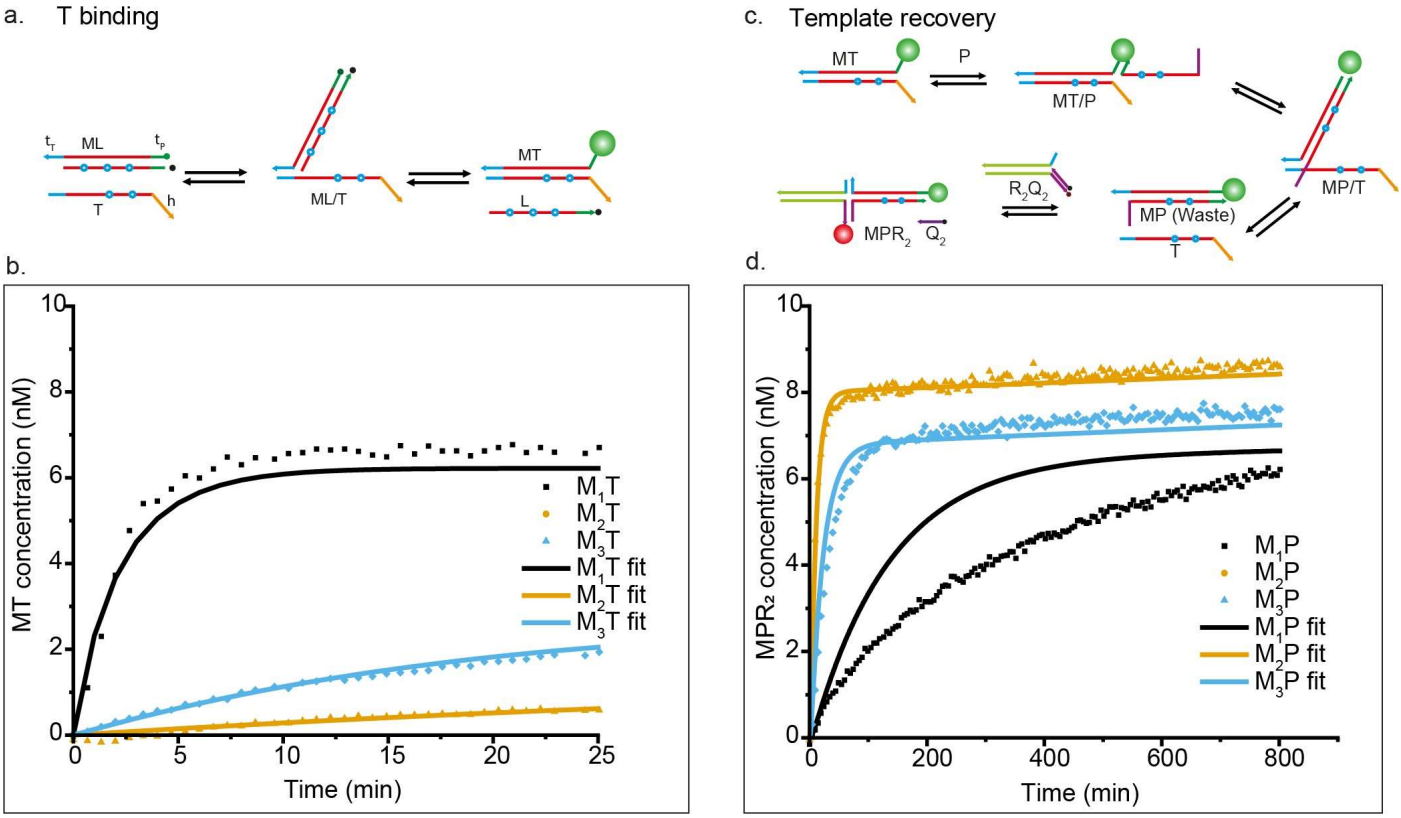

Fig. 2: Two stages of discrimination in the KP experiment. (a) Schematic of template binding showing that template *T* invades blocked monomer *ML* by TMSD to form *MT*. (b) Kinetics of the template binding process obtained from the Cy3 signal of the M strand. 8 nM M_1L, M_2L, and M_3L were added in separate wells and the reactions were initiated by injecting 10 nM *T* in each well. Obtained fluorescence signals are converted into concentrations using calibration curves of known concentrations (Supplemental Note 11). Solid lines are fits to an ODE model fitted to several experimental conditions simultaneously. (c) Template recovery is the removal of *M* from *T* by the proofreader molecule *P*. Formed waste MP is reported by the external reporter R_2Q_2, which is labelled with AlexaFluor-647. (d) Kinetics of the template recovery step were obtained from the AlexaFluor-647 signals. 8 nM M_1T, M_2T, and M_3T and 20 nM of their respective reporter complex were added in separate wells. 50 nM P_8 was added in each well to trigger the reaction. Solid lines are fits to an ODE model (Supplemental Note 12) fitted to several experimental conditions simultaneously (Supplemental Fig. S37-39).

Sequence-specific proofreading. As a second step, we looked at the efficacy and sequence-specificity with which proofreader strands *P* remove monomers from the template, allowing "template recovery" to participate in another reaction. We explored three different designs for *P*; each had the same 5-base toehold t_P, but we modified the thermodynamic strength of the *MP* waste product by truncating the displacement domain of *P* (shown in red in Fig. 2c). We label the designs P_6, P_7 and P_8, with the 6, 7, and 8 being the number of bases in the toehold and displacement domain of *M* that are left unpaired after the proofreading reaction. P_6 thus represents the most thermodynamically favoured proofreader, and P_8 the least. In all cases, reactions were driven by adding large excess of *P*, which constitute a fuel reservoir.

For all combinations of proofreaders and monomers, a range of concentrations of *MT* complexes were added to a large excess of proofreaders; the formation of M_1P waste was reported by an external reporter R_2Q_2, which was labelled with AlexaFluor-647-IowaBlack RQ fluorophore-quencher pair and was already present in solution (Fig. 2c, Supplemental Note 4, Supplemental Fig. S34-36). The results for *MT*= 8nM and P_8 are shown in Fig. 2d; full results are given in Supplemental Fig. S37-39. Again, we also report fits to an ODE model for each system; these fits used a single set of parameters for each proofreader, fitted simultaneously to all template recovery experiments as well as subsequent experiments on the full proofreading cycle.

As expected, P_6, which had the strongest binding to *M*, was the fastest in forming the waste duplex *PM*. Other proofreaders were slower, but still demonstrated template recovery. Crucially, P_7 and P_8 showed clear discrimination between M_1, M_2 and M_3, with M_1 being removed from *T* much more slowly, as evident for P_8 in Fig. 2d and borne out by the fitted rate constants (Supplemental Table S11). Alongside the discrimination present in the initial template binding, the existence of this second discrimination step demonstrates the potential for kinetic proofreading. For P_6, removal by proofreader appears slightly slower for M_1 than M_2 and M_3, but the distinction is smaller, and the relatively slow reporter rate complicates the analysis.

The action of the proofreaders is fundamentally different from the presence of a large reservoir of blocker strands. Adding a large concentration of blocker monomers to *MT* complexes does result in the removal of *M* and the recovery of T, as shown outlined in Supplemental Note 2 and Fig. S33. However, removal of M_2 and M_3 is not substantially faster than M_1. The relative rates for template displacing blocker and blocker displacing template are fundamentally constrained by the free-energy change of the process. The overall discrimination is equilibrium in nature, and if the full equilibrium discrimination is manifested in the rate of template binding, no discrimination will be observed in the reverse reaction[37]. In contrast, using a distinct proofreader strand allows, in principle, the initial template binding and subsequent template recovery steps to be thermodynamically decoupled by fuel consumption. In practice, we rely on a single additional mismatch between blocker and template to manifest this difference; previous work[29] has shown that a well-placed single mismatch can have large kinetic and thermodynamic effects.

Complete discard mechanism. We now analyse the entire catalytic discard mechanism in which the blocked monomers first bind to the template to form *MT* duplexes, and are subsequently converted into waste complexes by *P* (Fig.3a). We added *T* to solutions of M_1, M_2 or M_3, P_6, P_7 or P_8, and the appropriate waste reporter, and tracked both the amount of monomers that had been liberated from blocker strands and the amount of waste in solution. Results for P_8 and T= 4 nM and are shown in Fig. 3b, 3c, and full results in Supplemental Fig. S40-42. Fits of the data to an ODE model were performed simultaneously with the data on template recovery, using a single set of parameters for each proofreader (Supplemental Note 13).

The speed of *MP* complex formation depends on both the rate of initial template binding and proofreader-driven template recovery. The first step is faster for M_1, and the second step is faster for M_2 and M_3. More importantly, the system exhibits clear catalytic turnover as multiple monomers can be converted into waste for each template (most evident for lower template concentrations, Supplemental Fig. S40-42), showing that proofreading returns a functional template as required.

The purpose of KP is to enrich the population of template-bound M_1 monomers relative to template-bound M_2 and M_3. The concentrations of each template-bound monomer can be estimated by subtracting the waste concentration form the unblocked monomer concentration (subtracting Fig. 3c from Fig. 3b for P_8 and T= 4 nM). Although this approach is crude, it provides clear evidence for a spike in in M_1T at short times, with M_2T and M_3T highly suppressed at all time points (Fig. 3d-f and Supplemental Fig. S43). This behaviour is qualitatively consistent with predictions of an ODE model, parameterized by fits to earlier experiments (Supplemental Fig. S43). We suspect that the apparent ~1 nM yield of M_3T at long times is likely an artefact of the crude methodology. This interpretation is supported by the fact that the apparent yield of M_3T is almost as high for P_6 as P_8, despite P_6 having much higher affinity for M_3.

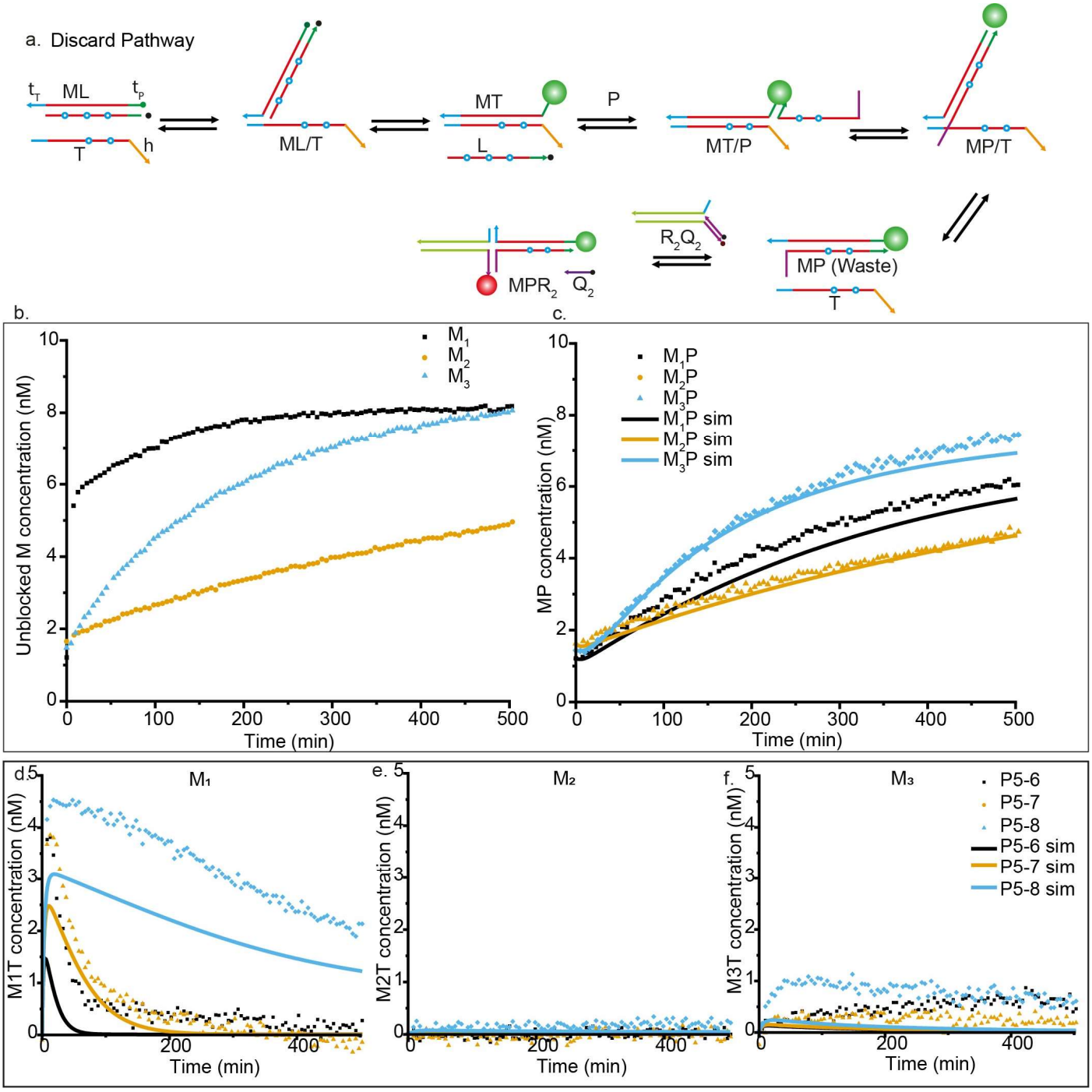

Fig. 3: Kinetics of the full discard pathway. (a) The full discard pathway starts with blocked monomer, proofreader and reporter, and is triggered by the injection of *T*, which displaces the blocker from *M* and is subsequently displaced itself by the proofreader strand. (b) Concentration of *M* liberated from its blocker after injection of 4 nM *T*, as inferred from Cy3 fluorescence, for 8 nM M_1L, M_2L and M_3L in different wells, each with 50 nM P_8 and 20 nM of their corresponding reporters. (c) Formation of MP waste in the same reactions, as inferred from reporter fluorescence. Solid lines are fits to an ODE model fitted to several experimental conditions simultaneously (Supplemental Note 13). (d) Concentration of intermediate M_1T over time for three different proofreaders, P_6, P_7, and P_8. Concentration of M_1T was estimated by subtracting AlexaFluor647 signals from Cy3 signals. (e,f) Concentration of M_2T and M_3T over time in equivalent experiments. Solid lines in (d,e,f) are estimates of MT from the fitted ODE model.

Templated dimer formation with kinetic proofreading. Having discriminated between the correct and incorrect monomers at both the template binding and template recovery steps, we redesigned the DNA strands to allow for dimerization through HMSD, using a handhold domain in a new template, *T'*. We also redesigned the monomers, now called M'_1, M'_2, and M'_3, to accommodate three mismatches between the blocker and monomers (Fig.1c), rather than two. One of those mismatches is eliminated in template binding and remaining two during HMSD. In this redesign, we retained the domain lengths from the previous characterisation process but changed the sequences (Supplemental Table S13-15).

The handhold recognises the second monomer *N*, allowing dimerization, via HMSD (Fig. 1c). The intermediate *M'T'* can also be invaded by the redesigned proofreader *P'* to form waste *M'P'*, allowing kinetic proofreading of *M'*. We explored this process by adding various template concentrations to mixtures of monomers and proofreaders, with M_1, M_2 and M_3

probed separately. Dimer concentration was reported by an external reporter similar to that used for waste characterisation (Fig.1d). We observed that the reporter had a small but significant tendency to react directly with the monomers (Supplemental Fig. S44). We therefore monitored lock strand removal by the increase in Cy3 signal intensity, and added external reporter only after the Cy3 signal reached a plateau for all template-containing systems.

We tested the three variants *P'6, P'7* and *P'8*, with the results for *P'6* and 2nM *T* shown in Fig. 4a-b, and the others in Supplemental Fig. S44-52. All systems showed discrimination at the first step, which is not proofreader-dependent; formation of M_1T is substantially faster than M_2T or M_3 T, as evidenced by the Cy3 signal. In the absence of proofreader, however, systems with *M'1, M'2* or *M'3* all reach high concentrations of *M'N* eventually, as evidenced by the dimer reporter signal. By contrast, in the presence of proofreaders *P'6* or *P'7*, the eventual production of M'_2T and M'_3T is strongly supressed; the level of incorrect product in these cases is comparable to the inherent leak of the reporters (Supplemental Fig. S44-52).

This performance indicates successful proofreading, in which the alternative proofreading discard pathway competes effectively with incorporation into the dimerised state – although we note that the apparent mechanism for *P'6* was a little unexpected (Supplemental Note 9). By contrast, *P'8* does not show effective proofreading. We believe that, in this case, the proofreading pathway is too slow, allowing M'_2T and M'_3T to form even when *P'8* is present. Model predictions based on the experiments in Fig. 2 (Supplemental Fig. S53) show high yields of M'_2T for *P'7* and *P'8*; in the case of *P'8*, slower proofreading contributes to this response. As with template recovery, the reservoir of proofreading molecules acts in a fundamentally different manner to an excess of blocker strands. In Supplemental Fig. S54, we show that an excess of blockers does not prevent the eventual formation of M'_2T and M'_3 T, and only has a weak effect on kinetics.

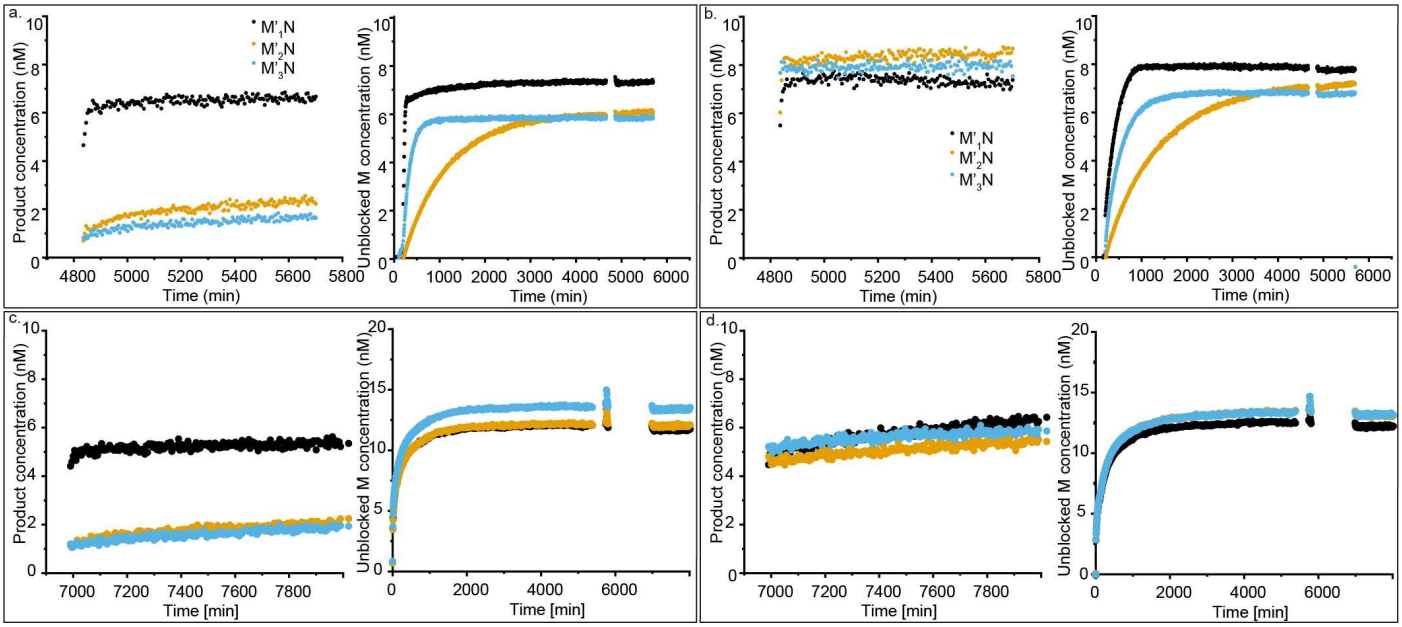

Fig. 4: Kinetic proofreading in a dimerization process. (a) Left: Dimeric product formation by HMSD in presence of proofreaders. 8nM blocked monomersM'1, *M'2* and *M'3* in separate wells were mixed with 10 nM *N*, 50 nM *P'6* and the reactions were triggered by injecting 2 nM *T'*. After three days, the extent of product formation was assayed by adding 20nM of the corresponding reporters; product concentration is inferred from reporter fluorescence, as outlined in Supplemental *Note* 15 and Fig. S44-49. We observe a much higher yield of M'_1N than for the alternatives. Right: Blocker strand removal in the same reaction, with unblocked monomer inferred from Cy3 fluorescence. (b) Left: Dimeric product formation from identical experiments in absence of the proofreader show almost equal product formation from all monomers. Right: Blocker strand removal in the same reactions. (c) Left: product formation in a competitive system initiated with 5nM each of M_1L, M_2L, and M_3L, 15nM *N* and 50nM *P'6*. Three replicas were analysed using the same approach as in (a), one to track each of M_1, M_2, and M_3 separately. As in (a), we observe a much higher yield of M'_1N than for the alternatives. Right: total concentration of monomers liberated from their blocker strands in these experiments. (d) Left: identical experiments to (c) but performed without *P'6*. All products M'_1N, M'_2N or M'_3N are observed with a high yield.

To demonstrate that our KP network enhances discrimination from a mixed pool of distinct monomers, we prepared an equimolar mixture of M'_1L, M'_2L and M'_3L with an equal concentration of each duplex and an excess of *N*. As in the individual experiments, *P'6* and P'5-7 again significantly suppressed the formation of incorrect dimers M'_2N and M'_3N while allowing M'_1N to be formed at a high level (Fig. 4c and d, Supplemental Fig. S55-60).

SNP identification via kinetic proofreading. We next modify the KP network to detect SNPs in a DNA strand where the mutations are anywhere in the displacement domains (Fig. 5a). In this context, it is natural for the strand being tested – either the wild type (*WT*) or SNP-bearing strand (*SNP*) – to be single stranded. A *Probe* strand is fully complementary to *WT* whereas the *SNP*s have a mismatch with the *Probe* at the point of mutation. The candidates (*WT* or *SNP*s) bind to the initially blocked *Probe* via TMSD, providing initial discrimination. This TMSD reveals a second toehold that facilitates the slow formation of a stable complex with fluorescent reporters. Alternatively, proofreader molecules P_{snp}

can react with intermediates via the second toehold to form a *Probe-P*$_{snp}$ complex that is sequestered by a sink *S*. This proofreading discard pathway provides a second opportunity to discriminate between *WT* and *SNP*.

We generated a random *WT* conforming to our design and introduced a single mismatch at different positions of the displacement domain to give six *SNP* strands (Supplemental Table S16). We used different nucleotide mismatches to vary the free energy of the mismatched pair. By design, the same proofreader and probe strands can be used regardless of the specific competing *SNP*. We triggered the system by adding a blocked *Probe* to a solution containing the candidate strand, reporter complex and sink complex (Supplemental Notes 8, 16). Fluorescent traces with and without 20nM proofreader are shown in Fig. 5; results for other concentrations of proofreader are given in Supplemental Fig. S61.

The proofreader improves accuracy in two ways. Firstly, by selectively reacting with the intermediate, the initial rate of product formation for *SNP*s is reduced by an additional factor of 2-3 relative to *WT* when proofreaders are used (Supplemental Fig. S62, Table S12). Secondly, the existence of a fuel-consuming discard pathway allows *SNP*-triggered probes to be consumed (Fig. 5c), preventing them from eventually reacting with the reporter, as is typical in the absence of a proofreader strand (Fig. 5b).

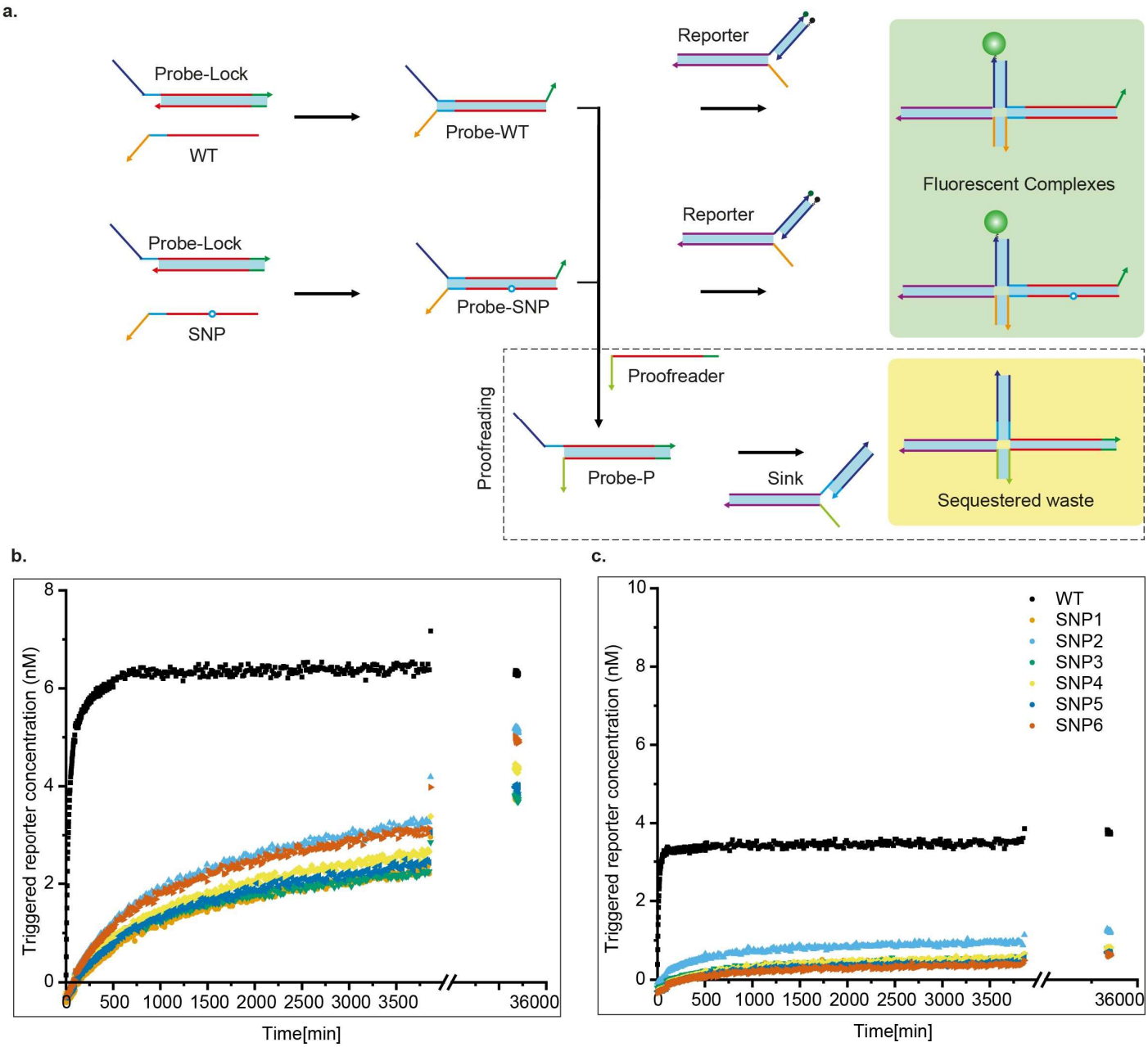

Fig. 5: SNP identification scheme utilising KP to suppress output of mismatched target. (a) Reaction scheme: *WT* and *SNP* strands bind to the blocked probe to form the intermediate *WT-Probe* or *SNP-Probe,* which trigger the fluorescence by displacing the quencher strands from the external reporters. Proofreader molecules can also invade these intermediates to convert the *Probe* into *Waste*, which is in turn sequestered into an inert complex by the *Sink*. (b) Fluorescence signal triggered by the *WT* and 6 different *SNP*s when 10 nM of *Probe-blocker* is added to a mixture containing 15 nM of candidate strands, 20 nM of corresponding reporters, and 20 nM of *Sink* complex. (c) The same reactions are performed in presence of 20 nM Proofreader *P*$_{snp}$. While the *WT* still shows rapid growth up to about 4 nM in the reported strand concentration, the SNPs show limited increase the signal even after several days. At least some

of the residual activation apparent in (Fig. 5c) is due to a leak reaction between *Probe-blocker* and reporter (Supplemental Fig. S62).

Discussion

We have developed a non-enzymatic DNA strand displacement network that performs catalytic dimerization, with kinetic proofreading applied to one of the monomers. To the best of our knowledge, this work represents the first time a kinetic proofreading system has been engineered *de novo*. This successful demonstration of KP in a synthetic context sets the groundwork for applying the motif more broadly and opens pathways to explore the fundamental principles of KP in engineerable systems.

We have shown how KP can enhance the discrimination between wild type DNA strands and single nucleotide mutants. In this context, the relatively small difference between targets is unavoidable, making it challenging to redesign a system to increase the discrimination at equilibrium between *WT* and many variants simultaneously. Re-exploiting a single free-energy difference through KP is therefore especially valuable. We note that even if discrimination exists without KP, enhancing that discrimination enables reliable detection of low concentrations of a target from within a pool of competitors.

KP relies on pushing the template through fuel-driven non-equilibrium cycles. In our design, the large excess of proofreader tends to drive the template through the cycle, although backwards steps that reverse the proofreading reaction are not impossible. The presence of downstream reactions, such as the sequestration of waste by reporter molecules, can provide an additional contribution to the thermodynamic drive. Although such a design feature may enhance KP, it is not required for our system – as evidenced by the successful KP in the dimeric production experiments.

When describing natural systems, it is common to think of KP as increasing the rate at which correct molecules are incorporated into a product relative to incorrect molecules. That behaviour is observed here. But we also observe a second effect, arising from the non-biological context of a finite pool of input molecules. The discard pathway provided by KP allows the incorrect species to be rendered inert, so that incorrect products are never made at all, rather than just delayed. This fact, as much as the rate advantage provided by KP, will likely aid in the design of more precise synthetic molecular networks. This behaviour is distinct from thresholding[38], in which a rapid reaction with a thresholding molecule is used to consume inputs so that output signals are only produced when the input concentrations exceed the concentration of the threshold molecule. Mechanistically, KP does not rely on depletion of the proofreader to propagate a signal, and so doesn't require a minimal input concentration to generate an output. Moreover, thresholding molecules will, if anything, tend to sequester correct species faster than their mutant counterparts.

Experimental methods

Sequence design

All sequences were generated using the NUPACK web server to minimise undesired secondary structures. Some strand designs were manually altered to incorporate strategic features such as mismatches. All the strands were purchased from Integrated DNA Technologies as HPLC purified at 100 μM concentrations in IDTE buffer, pH 8. All the sequences used in the paper are listed in Tables S13-14. Detailed design considerations are shown in Supplemental Fig. S1-S2.

General annealing protocol

The *ML* duplexes were prepared by annealing 100 nM *M* with 20% excess *L* to ensure that all *M* strands were bound to *L* strands. For *MT* and *MP* complexes, 100 nM *M* strands were annealed with 20% excess *T* or *P* strands. Reporter complexes *RQ* were comprised of three strands, *Rcomp*, $r_{i\text{-}j}$ (collectively called *R*) and *Q* (details in Supplemental Fig. S2, Supplemental Tables S13-16). For this, 200 nM *Rcomp* strands were annealed with 20% excess of $r_{i\text{-}j}$ and *Q* strands. The volumes of the annealing mixtures were 100 or 200 μL depending on the experimental requirements. The required strands were mixed to have the final concentrations in 1x Tris-Acetate-EDTA buffer containing 1M NaCl. This mixture was first heated to 95°C and held there for 5 minutes, then gradually cooled down to 20C at 1°C per minute and stored at 4°C until used.

Fluorescence Kinetics measurements and calibration

All fluorescence measurements were performed in a BMG Labtech ClarioSTAR fluorescence plate reader in Greiner μClear flat-bottomed 96-well plates. Reactions were initiated, unless otherwise stated, by injecting 50 μL of trigger (usually a combination of single stranded or duplexed DNA oligonucleotides with buffer) into 150 μL of the other reactants using the instrument's in-built injectors (pump speed 430 μL/ sec). The final mixtures were then shaken for 10-30 seconds at 400-500 rpm before measuring the fluorescence.

Monomers M_1, M_2, M_3, M'_1, M'_2, M'_3, and *Rsnp* were labelled with fluorophore Cy3. *L*, *L'*, and *Qsnp* were labelled with IowaBlack FQ quencher. *Rcomp-AF* and *RcompP* were labelled with AlexaFluor-647. Q_1 and Q_2 were labelled with IowaBlack RQ quencher.

For the Cy3 and AlexaFluor-647 measurements, presets from the instruments were used for excitation and emission wavelengths. Specifically, for Cy3, excitation: 530-20 nm, emission: 580-30 nm, gain: 2100; and for AlexaFluor-647, excitation: 625-30 nm, emission: 680-30 nm, gain 2700). For each well, each measurement was taken as 20 flashes in a spiral scan with a maximum diameter of 4 mm. Step-by-step experimental methods used in each experiment are given in Supplemental Notes 1-8.

Fluorescence calibrations were performed for fluorophore-labelled complexes to quantify the fluorescence signals obtained from unit concentration of such complexes. Cy3 (for *ML*, *MP*, *MT* complexes) and AlexaFluor-647 (for *RQ* and *MPR* complexes) signals were monitored in a volume of 200 μL with different candidates' concentrations ranging from 0-20 nM in 1x TAE buffer with 1 M NaCl. Only M_1 and P_6 were used rather than testing all possible permutations of monomers and proofreaders because the local environments of the fluorophores in all complexes were same.

Data Processing and analysis

This section describes the methods used for the estimation of the reaction rate constants for the various intermediary reactions. All rate constants are obtained from the experimental reaction dataset as obtained from the Cy3 and AlexaFluor-647 measurements shown in Supplemental Fig. S16, S19, S22, S26. The AlexaFluor-647 channel gives a fluorescent signal in the presence of *RQ*, *MPR*, and *PR* whereas Cy3 gives signal for *ML*, *MT*, *MP* and *MPR*.

Trajectory analysis and rate constant estimation

We used the ParametricNDSolve built-in function in Mathematica 12.3 to fit mass action models of reaction kinetics to the processed data. The ODE models used to describe each experiment are given in Supplemental Notes 10 – 13. In general, the procedure involved fitting a single set of rate constants to a number of kinetic traces, at a range of distinct initial conditions. In each case, an array K_n of known rates (determined by earlier fits to different experiments) and an array K_{un} of unknown rates were defined. The system of ODEs was constructed inside the ParametricNDSolve function, and all the starting conditions (initial concentration values and K_n) were provided. ParametricNDSolve was then used to minimize the mean squared error (MSE)

$$MSE(K_{un})_i = \sum_{0\le t<t_1} \sum_j w_1([M]_{i,j}(K_{un},t) - [E]_{i,j}(t))^2 + \sum_{t_1\le t\le T} \sum_j ([M]_{i,j}(K_{un},t) - [E]_{i,j}(t))^2 \quad (3)$$

over possible values of K_{un}. Here, $[M]_{i,j}(K_{un},t)$ is the predicted concentration of the chosen species *i* in kinetic trace *j* at time T, given K_{un} as the values of the unknown rate constants. $[E]_{i,j}(t)$ is the equivalent data observed in experiment. To improve the accuracy with which fits captured the crucial behaviour at early time points, weights w_1=1 and w_2=0.2 were assigned to data points with $t < t_1$ and $t_1 \le t$, respectively. t_1 is approximately equal to the time instance when the species concentration has plateaued and was manually defined for each reaction. Each element in set k_{un} is sampled from a variable range of values that can be found in Supplemental Notes 10-13.

The data for binding of blocked monomers to templates was used to identify the rate constants with which the template displaced blockers from monomers, and vice versa. The procedure is outlined in Supplemental Note 10. The fitted parameters can be found in Table S9 and the comparison between experimental observations and simulation predictions is shown in Supplemental Fig. S32. Subsequently, data for the triggering of reporter complexes by pre-prepared *MP* duplexes was used to fit rate constants for the reporter. As outlined in Supplemental Note 11, the total concentration of monomers could not be unambiguously identified in these reactions. We therefore generated eight different sets of rate constants for each reporter based on estimates of the total monomer concentration from within a plausible range. Each of these eight sets of reporter rate constants was then used, alongside the rate constants for template binding, as inputs to a simultaneous fit of the template recovery and full discard pathway experiments. The set of rate constants that gave the lowest total error for template recovery and discard pathway was then selected. The procedure is outlined in more detail in Supplemental Note 11. The resulting fits for reporter characterization are shown in Supplemental Fig. S34-36; the fits for template recovery are shown in Supplemental Fig. S37-39; and the fits for discard pathway are shown in Supplemental Fig. S40-S42. The fitted rate constants are given in Tables S10-11.

Model-based predictions for intermediates and dimerization

The values of rate constants obtained from fits were used to predict the time-varying concentration of *MT* complexes during the full discard pathway experiments, Supplemental Fig. S43. The predictions were made by integrating Eqs. S35-S39 using the rate constants in Tables S9-S11 and the experimental initial conditions as parameters.

Although the dimerization system used re-designed monomers and templates, we nonetheless used the rate constants obtained from initial fits to predict the yield of dimers over time. The predictions were made by integrating Eqs. S51-S53 using the rate constants in Tables S9-S11 and the experimental initial conditions as parameters. A plausible dimerization rate constant of 3.50E+05 M^{-1} s^{-1} was also used. The results are plotted in Supplemental Fig. S53. At variance with the experimental results, the model predicts a high dimerization yield for all monomers and proofreaders in the long-time limit. We suspect that this prediction is due to an excessive reverse proofreading (rebinding of proofread monomers to the template) rate constant for mismatched monomers; these constants are not well-constrained by the data.

DATA AVAILABILITY

All the fluorescence data as obtained from the plate reader, the Mathematica scripts used to process those data, convert to concentration, and fit them to the ODE models, and the Fig. are available https://doi.org/10.5281/zenodo.8132461.

ACKNOWLEDGEMENT

The research is supported by ERC research grant agreement No. 851910. TEO, RM, and JCG designed the project. RM performed the experiments. AS and RM analysed the data. AS constructed the mathematical models and did the fittings. All the authors contributed to writing the manuscripts. The authors are grateful to Dr Wooli Bae and Francesca Smith for their helpful discussions. RM and AS are supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 851910).

REFERENCES

1 Poulton, J. M., Ten Wolde, P. R. & Ouldridge, T. E. Nonequilibrium correlations in minimal dynamical models of polymer copying. *Proceedings of the National Academy of Sciences* **116**, 1946-1951 (2019). https://doi.org:10.1073/pnas.1808775116

2 Bouadloun, F., Donner, D. & Kurland, C. G. Codon-specific missense errors in vivo. *The EMBO Journal* **2**, 1351-1356 (1983). https://doi.org:10.1002/j.1460-2075.1983.tb01591.x

3 Kurland, C. G., Rigler, R., Ehrenberg, M. & Blomberg, C. Allosteric mechanism for codon-dependent tRNA selection on ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 4248 (1975). https://doi.org:10.1073/pnas.72.11.4248

4 Kunkel, T. A. & Bebenek, K. DNA Replication Fidelity. *Annual Review of Biochemistry* **69**, 497-529 (2000). https://doi.org:10.1146/annurev.biochem.69.1.497

5 Hopfield, J. J. Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proceedings of the National Academy of Sciences* **71**, 4135-4139 (1974). https://doi.org:10.1073/pnas.71.10.4135

6 Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **57**, 587-595 (1975). https://doi.org:10.1016/s0300-9084(75)80139-8

7 Mckeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the National Academy of Sciences* **92**, 5042-5046 (1995). https://doi.org:10.1073/pnas.92.11.5042

8 Hopfield, J. J., Yamane, T., Yue, V. & Coutts, S. M. Direct experimental evidence for kinetic proofreading in amino acylation of tRNAIle. *Proc Natl Acad Sci U S A* **73**, 1164-1168 (1976). https://doi.org:10.1073/pnas.73.4.1164

9 Yan, J., Magnasco, M. O. & Marko, J. F. A kinetic proofreading mechanism for disentanglement of DNA by topoisomerases. *Nature (London)* **401**, 932-935 (1999). https://doi.org:10.1038/44872

10 Bartlett, D. W. & Gilbert, A. M. A kinetic proofreading model for bispecific protein degraders. *Journal of Pharmacokinetics and Pharmacodynamics* **48**, 149-163 (2021). https://doi.org:10.1007/s10928-020-09722-z

11 Schiessel, H. & Blossey, R. Pioneer transcription factors in chromatin remodeling: The kinetic proofreading view. *Physical Review E* **101** (2020). https://doi.org:10.1103/physreve.101.040401

12 Murugan, A., Huse, D. A. & Leibler, S. Discriminatory Proofreading Regimes in Nonequilibrium Systems. *Physical Review X* **4** (2014). https://doi.org:10.1103/physrevx.4.021016

13 Qian, H. Phosphorylation Energy Hypothesis: Open Chemical Systems and Their Biological Functions. *Annual Review of Physical Chemistry* **58**, 113-142 (2007). https://doi.org:10.1146/annurev.physchem.58.032806.104550

14 Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368-372 (2011). https://doi.org:10.1038/nature10262

15 Woods, D. *et al.* Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature* **567**, 366-372 (2019). https://doi.org:10.1038/s41586-019-1014-9

16 Zhang, D. Y., Chen, S. X. & Yin, P. Optimizing the specificity of nucleic acid hybridization. *Nature Chemistry* **4**, 208-214 (2012). https://doi.org:10.1038/nchem.1246

17 Shastry, B. S. SNP alleles in human disease and evolution. *Journal of Human Genetics* **47**, 0561-0566 (2002). https://doi.org:10.1007/s100380200086

18 Kwok, P.-Y. & Chen, X. Detection of Single Nucleotide Polymorphisms. *Current Issues in Molecular Biology* (2003). https://doi.org:10.21775/cimb.005.043

19 Xiao, Y. *et al.* Fluorescence Detection of Single-Nucleotide Polymorphisms with a Single, Self-Complementary, Triple-Stem DNA Probe. *Angewandte Chemie International Edition* **48**, 4354-4358 (2009). https://doi.org:10.1002/anie.200900369

20 Guo, Z., Liu, Q. & Smith, L. M. Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nature Biotechnology* **15**, 331-335 (1997). https://doi.org:10.1038/nbt0497-331

21 Chen, S. X., Zhang, D. Y. & Seelig, G. Conditionally fluorescent molecular probes for detecting single base changes in double-stranded DNA. *Nature Chemistry* **5**, 782-789 (2013). https://doi.org:10.1038/nchem.1713

22 Chen, X. *et al.* Thermodynamics and kinetics guided probe design for uniformly sensitive and specific DNA hybridization without optimization. *Nature Communications* **10** (2019). https://doi.org:10.1038/s41467-019-12593-9

23 Tyagi, S. & Kramer, F. R. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nature Biotechnology* **14**, 303-308 (1996). https://doi.org:10.1038/nbt0396-303

24 Xiao, Y. *et al.* An Electrochemical Sensor for Single Nucleotide Polymorphism Detection in Serum Based on a Triple-Stem DNA Probe. *Journal of the American Chemical Society* **131**, 15311-15316 (2009). https://doi.org:10.1021/ja905068s

25 Hwang, M. T. *et al.* Highly specific SNP detection using 2D graphene electronics and DNA strand displacement. *Proceedings of the National Academy of Sciences* **113**, 7088-7093 (2016). https://doi.org:10.1073/pnas.1603753113

26 Schulman, R., Wright, C. & Winfree, E. Increasing Redundancy Exponentially Reduces Error Rates during Algorithmic Self-Assembly. *ACS Nano* **9**, 5760-5771 (2015). https://doi.org:10.1021/nn507493s

27 Soloveichik, D., Cook, M. & Winfree, E. Combining self-healing and proofreading in self-assembly. *Natural Computing* **7**, 203-218 (2008). https://doi.org:10.1007/s11047-007-9036-x

28 Winfree, E. & Bekbolatov, R. in *DNA Computing.* (eds Junghuei Chen & John Reif) 126-144 (Springer Berlin Heidelberg).

29 Irmisch, P., Ouldridge, T. E. & Seidel, R. Modeling DNA-Strand Displacement Reactions in the Presence of Base-Pair Mismatches. *Journal of the American Chemical Society* **142**, 11451-11463 (2020). https://doi.org:10.1021/jacs.0c03105

30 Haley, N. E. C. *et al.* Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement. *Nature Communications* **11** (2020). https://doi.org:10.1038/s41467-020-16353-y

31 Machinek, R. R. F., Ouldridge, T. E., Haley, N. E. C., Bath, J. & Turberfield, A. J. Programmable energy landscapes for kinetic control of DNA strand displacement. *Nature Communications* **5**, 5324 (2014). https://doi.org:10.1038/ncomms6324

32 Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res* **41**, 10641-10658 (2013). https://doi.org:10.1093/nar/gkt801

33 Cabello-Garcia, J., Bae, W., Stan, G.-B. V. & Ouldridge, T. E. Handhold-Mediated Strand Displacement: A Nucleic Acid Based Mechanism for Generating Far-from-Equilibrium Assemblies through Templated Reactions. *ACS Nano* **15**, 3272-3283 (2021). https://doi.org:10.1021/acsnano.0c10068

34 Cabello-Garcia, J. *Handhold-mediated strand displacement: a mechanism for non-equilibrium and catalytic templating* PhD thesis, Imperial College London, (2022).

35 Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *Journal of the American Chemical Society* **131**, 17303-17314 (2009). https://doi.org:10.1021/ja906987s

36 Broadwater, D. W. B., Jr. & Kim, H. D. The Effect of Basepair Mismatch on DNA Strand Displacement. *Biophys J* **110**, 1476-1484 (2016). https://doi.org:10.1016/j.bpj.2016.02.027

37 Ouldridge, T. E. The importance of thermodynamics for molecular systems, and the importance of molecular systems for thermodynamics. *Natural Computing* **17**, 3-29 (2018). https://doi.org:10.1007/s11047-017-9646-x

38 Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332**, 1196-1201 (2011). https://doi.org:10.1126/science.1200520