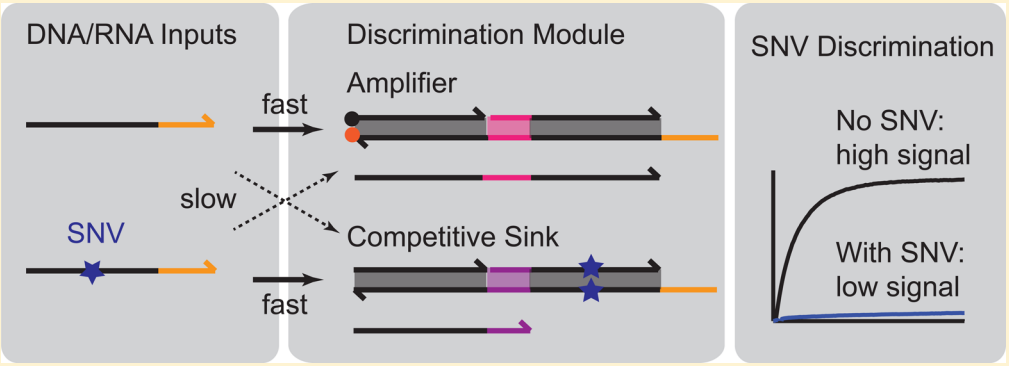
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

An Engineered Kinetic Amplification Mechanism for Single Nucleotide Variant Discrimination by DNA Hybridization Probes

Sherry Xi Chen[†] and Georg Seelig*[,†,‡]

[†]Department of Electrical Engineering, University of Washington, Seattle, Washington 98195, United States
[‡]Department of Computer Science & Engineering, University of Washington, Seattle, Washington 98195, United States

Supporting Information



ABSTRACT: Even a single-nucleotide difference between the sequences of two otherwise identical biological nucleic acids can have dramatic functional consequences. Here, we use model-guided reaction pathway engineering to quantitatively improve the performance of selective hybridization probes in recognizing single nucleotide variants (SNVs). Specifically, we build a detection system that combines discrimination by competition with DNA strand displacement-based catalytic amplification. We show, both mathematically and experimentally, that the single nucleotide selectivity of such a system in binding to single-stranded DNA and RNA is quadratically better than discrimination due to competitive hybridization alone. As an additional benefit the integrated circuit inherits the property of amplification and provides at least 10-fold better sensitivity than standard hybridization probes. Moreover, we demonstrate how the detection mechanism can be tuned such that the detection reaction is agnostic to the position of the SNV within the target sequence. in contrast, prior strand displacement-based probes designed for kinetic discrimination are highly sensitive to position effects. We apply our system to reliably discriminate between different members of the let-7 microRNA family that differ in only a single base position. Our results demonstrate the power of systematic reaction network design to quantitatively improve biotechnology.

INTRODUCTION

Single base differences in DNA or RNA sequences can have important biological consequences, and developing tools for detecting such single nucleotide variants (SNVs) has long been a focus of nucleic acid biotechnology. Applications ranging from DNA amplification by the polymerase chain reaction[1] to single nucleotide polymorphism genotyping[2] to modern sequencing-by-synthesis[3] all rely on the specificity of Watson−Crick base pairing. Most commonly, hybridization specificity is ensured by performing binding experiments near the probe melting temperature where binding of a perfect complement is energetically favorable but binding of a mismatched strand is unfavorable. Because of the need to perform experiments at elevated and precisely controlled temperatures such approaches can be challenging to implement, especially in low resource settings or in assays where many probes are used in parallel. To mitigate these limitations, there have been attempts to engineer synthetic nucleotide analogues with improved molecular specificity,[4−8] as well as hybridization probes with greater specificity due to supramolecular architecture.[9−13] Rationally designed hybridization probes that rely on DNA strand displacement or strand exchange mechanisms are particularly intriguing because of the high level of discrimination they provide and because they work robustly for a wide range of temperatures and buffer conditions.[10,12,13]

Toehold-mediated DNA strand displacement is a competitive hybridization reaction whereby an incoming DNA strand outcompetes an incumbent strand for binding to a complementary binding partner (Figure 1A). In spite of its simplicity, DNA strand displacement has become a key enabling technology not only for the design of specific hybridization probes but for dynamic DNA nanotechnology more broadly.[14] Strand displacement was introduced to the DNA nanotechnology field by Yurke and co-workers as a means to drive

Received: January 9, 2016
Published: March 24, 2016

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

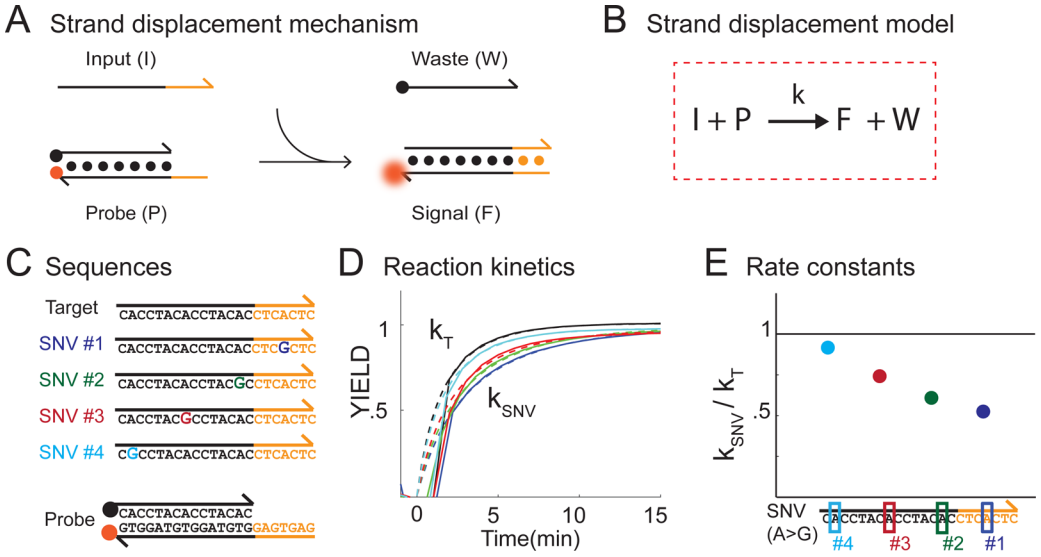


Figure 1. Toehold-mediated DNA strand displacement is sensitive to SNVs. (A) Strand displacement reaction. The single-stranded input (I) binds to the partially double-stranded probe (P) producing a fluorescent, double-stranded product (F) and a single-stranded waste product (W). The reaction is initiated at complementary single-stranded toehold domains (orange). The probe is labeled with a fluorphore (red dot) and quencher (black dot) such that fluorescence is initially low. Displacement of the quencher-labeled strand results in increased fluorescence at the reaction end point. (B) Strand displacement kinetics are well-approximated by a bimolecular rate law. Here, k is the reaction rate constant. (C) Input and probe sequences. The bottom strand in the probe is designed to be the exact complement of the target input sequence. All SNV inputs contain the same CAC > CGC mutation, but the SNV position varies. (D) Kinetics traces for different inputs. Experiments were performed in 1× TAE with 12.5 mM Mg^{2+} at room temperature. Initial concentrations of probe and input were 10 nM and 5 nM, respectively. Fluorescence data for the first 2 min after input addition could not be recorded resulting in the apparent discontinuity in the data. We fitted each trace to a bimolecular rate law to obtain rate constants (dashed lines). The best fit reaction rate constant for the target input k_T is 1.2×10^6 M^{-1} s^{-1} while the rate constants for inputs SNV1 through SNV4 are 6.3×10^5, 7.3×10^5, 8.9×10^5 and 1.1×10^6 M^{-1} s^{-1}. (E) Ratio of reaction rates of each SNV input to the target input. The reaction with SNV1 (mutation in the toehold) is the slowest. Reaction rates increase as the mutation position moves toward the end of branch migration region and reaction with SNV4 is almost as fast as the reaction with target input.

a DNA-based molecular motor.[15] Subsequent work showed that multiple strand displacement reactions could be modularly connected into reaction cascades in which a strand released in one reaction could serve as an input in a downstream reaction.[16,22] Such multistage cascades have enabled the construction of a very wide range of functional biochemical reaction networks including computational circuits[16−20] and catalytic signal amplification mechanisms.[21−24]

Here, we show that rational hybridization probe design and biochemical reaction network design can be integrated to solve the SNV detection problem. We assume that a given sample either contains a nucleic acid target without a mutation or a mutated nucleic acid but not a mixture of the two. An ideal discrimination probe would bind the intended target and produce signal but would not do so if the SNV target is present in the sample of interest. To build such a probe we use an engineered, catalytic signal amplification mechanism in parallel with a competitive inhibitor or "sink." Our architecture is reminiscent of the seesaw logic gates recently introduced by Qian and Winfree,[18] but both the aims of our work and the details of the molecular implementation are distinct from that work.

The amplification system is designed to preferentially amplify a signal in the presence of the correct target strand while the sink is designed to preferentially bind and inactivate a target input containing a specific mutation of interest thus reducing the amount of signal that can be produced from that mutated target. We mathematically show that this system can achieve quadratically better discrimination than a system based on competition alone. Subsequently, we perform experimental studies to validate our theoretical findings using existing, off-the-shelf components. We find that this naive implementation performs as predicted but only for a subset of mutations that occur near the "toehold" of the single-stranded analyte. Finally, we use rational reaction engineering to modify the detection system such that a high degree of specificity is achieved for mutations anywhere in the input strand.

Strand Displacement Kinetics Are Sensitive to SNVs. Strand displacement is initiated at single-stranded complementary overhangs ("toeholds," orange in Figure 1A), proceeds through a branch migration process and, finally, results in the spontaneous release of the outgoing strand when the branch point approaches the end of the double-stranded domain. To experimentally follow a strand displacement reaction, the probe (P) is labeled with a fluorophore on the target-complementary strand, and a quencher on the other strand. The probe is initially dark but binding of the input (I) displaces the quencher-labeled strand resulting in an observable fluorescence signal (F) and a byproduct (waste, W).

Although strand displacement reactions involve breaking and (re)forming many individual base pairs, the overall kinetics is well approximated by a bimolecular rate law for a wide range of experimental conditions (Figure 1B).[10,25−27] Importantly, the reaction rate is highly sensitive to the length and sequence composition of the toehold domain. Li et al. recognized that mismatches within or near the toehold domain could reduce the binding rate of a mismatched SNV input compared to a fully matched target input and exploited these kinetic differences to create a specific SNV detection system.[10]

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

Machinek et al. used mismatches in toeholds to predictably control the rates of strand displacement reactions.[28]

Here, we compare the rates for the reaction between a probe, a matching target input and four different SNV inputs (Figure 1C). The four SNV inputs are numbered from right (SNV1 in toehold) to left (SNV4 at end of branch migration region) and are marked by different colors (SNV1: blue, SNV2: green, SNV3: red, SNV4: cyan). in order to have a consistent basis for comparison, all SNVs were designed to have the same mutation and nearest neighbors (from CAC to CGC) even though they occur in different positions within the input. Probe sequences were designed to be self-similar to minimize the effects of nonlocal structure. Input and probe sequences are given in Supplementary Tables S1 and S2, respectively. The CAC > CGC mutation is the most difficult SNV to discriminate against, because CGC will form a G−T wobble with the GTG complement subsequence of the CAC target. At 25 °C in 1 M Na+, the $\Delta\Delta G^{\circ}$ of the CGC/GTG wobble pair compared to CAC/GTG is only +2.11 kcal/mol[29] (compared to +6.61 kcal/mol for CGC > CCC, for example). Because of the small energy penalty, the CAC > CGC mutation provides a good test for the selectivity of our integrated biochemical circuit.

Experimental data for the reactions of the different inputs with the probes are shown as solid traces in Figure 1D. Reaction rate constants were found by fitting the data to a bimolecular rate law (dashed lines). The ratio of reaction rate constants for each SNV input to the rate constant for the target input is shown Figure 1E. The rate for the SNV4 input with a mismatch by the end of the branch-migration region is comparable to the rate measured with the target input. In contrast, the rate for the reaction with the SNV1 input carrying a mismatch in the toehold is significantly slower than the rate of the reaction with the target input. Importantly, since the energy penalties for all SNV inputs are the same, any differences in reaction kinetics results only from varying the mismatch positions.

It is not surprising that the rate of a strand displacement reaction strongly depends on the exact position of the mutation within the SNV input. Strand displacement rates are not expected to be sensitive to mutations that occur close to the end of the branch migration region away from the toehold. Once the strand that is being displaced is only attached to its original binding partner by a few base pairs, these bonds will dissociate spontaneously even if the invading strand cannot replace all the lost base pairs because it carries an SNV.

SNV Detection with Strand Displacement. In principle, at least some SNVs can be detected simply by following reaction kinetics as shown above. However, there are at least three concerns that make this approach impractical. First, in most diagnostic applications we do not have control over analyte concentrations. A sample containing a high concentration of the SNV input could result in a faster initial reaction than a sample with the target input at low concentrations. Second, following the full reaction time course is cumbersome and requires advanced instrumentation. An assay that provides end-point discrimination would be preferable. Third, kinetic discrimination cannot work when there is no difference in the reaction rates of the SNV input and intended input (e.g., SNV4 above). An assay operating at equilibrium could overcome this issue.

Li et al.[10] addressed the first two concerns and showed that such kinetic differences could be converted into distinct signals at the reaction end point in a setup where two probes are competing for binding to an input. Specifically, they designed one probe to exactly match the target input and a "sink" complementary to the SNV input. Because any given input strongly prefers binding to its fully complementary partner, the probe is triggered more strongly by the target than by the SNV target. Zhang and collaborators later addressed the third issue and showed that probes based on toehold exchange, a reversible form of strand displacement, could result in even higher specificity and, moreover, could detect mutations anywhere in the incoming strand and not just near toehold domains.[12,13]

Here, we build on these results in two ways. First, we demonstrate that any system that combines competition with amplification will not only have better sensitivity but also better specificity than a competitive system without amplification. Then, to make our detection system insensitive to the relative position of the mutation, we design an improved competitive probe that uses a two-step detection reaction including a reversible first step.

RESULTS

Reaction Model Analysis. We consider three different discrimination systems, one based on competition between two hybridization probes, a second consisting of a single probe capable of signal amplification through a catalytic reaction pathway, and a third that combines competition with amplification. In this Section, we present a mathematical analysis to establish the maximal discrimination and sensitivity achievable in each case. For our analysis we assume that either a target input I_T or an input with an single nucleotide variant I_{SNV} is added to the discrimination system but not both. All systems, at the minimum, contain a probe P that is designed to bind the target input.

Discrimination Factor and Yield. As a quantitative metric of probe selectivity we will use the discrimination factor

$$\mathrm{DF} = [F_T]/[F_{SNV}] \tag{1}$$

Here, F_T refers to the fluorescent signal species produced when I_T reacts with the probe and F_{SNV} refers to the product of the reaction between I_{SNV} and probe. (Note that in practice these are the same molecular species but produced in different experiments.) A highly specific assay will result in a large discrimination factor while a failure to discriminate between the two types of inputs results in DF = 1. Below, we will mostly be interested in the value of DF_∞, the discrimination factor at the reaction end point.

An ideal detection system should not only be specific but also sensitive. Biological analytes are often present at low concentrations and a sensitive probe would result in a strong signal even for small amounts of a target. As a quantitative measure for the sensitivity we will use the yield

$$\chi = [F_T]/[I_T]_0 \tag{2}$$

that is the ratio of amount of signal produced to the amount of target added initially. We note that in a system without amplification the yield cannot exceed $\chi = 1$ since a single target molecule can at best result in a single copy of the F_T molecule.

Discrimination by Competition. For discrimination by competition both a probe and a sink are used. The probe P is specific to the target input I_T while the sink S is designed to exactly match the SNV input I_{SNV}. The rate for binding of an input to its cognate target is assumed to be faster (rate constant

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

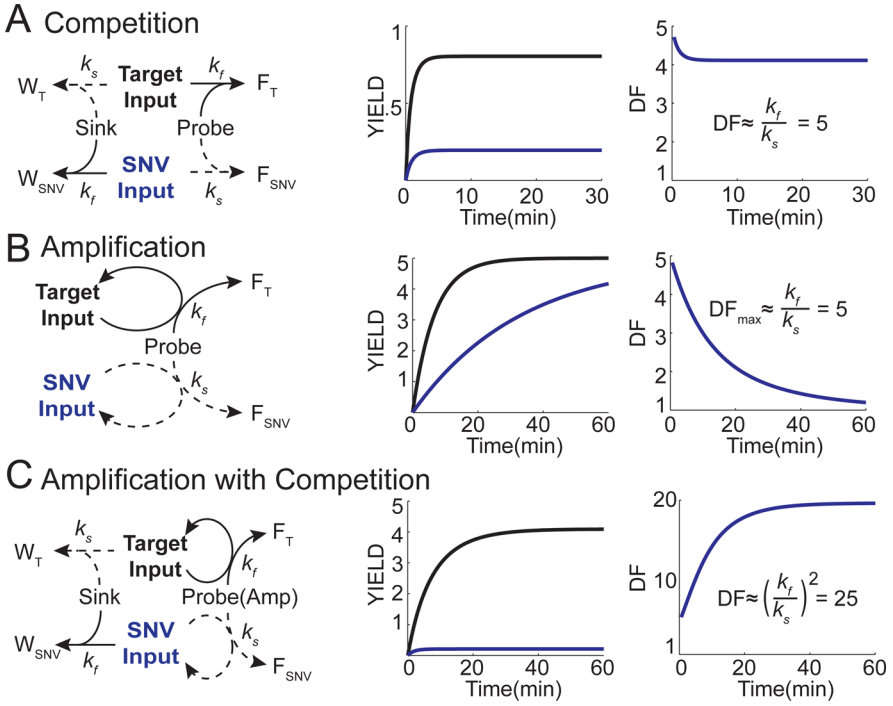


Figure 2. Different approaches to single nucleotide discrimination. (A) SNV discrimination by competition. (B) SNV discrimination with a catalytic amplifier. (C) Competitive hybridization combined with amplification. The left panel in each subfigure shows a graph of all reactions for both the target input and SNV input. The rate constant for all intended reactions (full lines) is k_f while the rate constant for all nonintended reactions (e.g., target input binding to a sink, dashed lines) is k_s. The middle graph in each row shows the simulated time evolution of the detection system for the target input (black trace) and the SNV input (blue trace). The right-hand panel shows the time evolution of the discrimination factor. The analytical result for the end point discrimination factor is shown in the inset.

k_f) than the rate for a reaction that results in a mismatch (rate constant k_s).

To obtain the discrimination factor and yield for a competitive strand displacement system, we first calculate how much signal $[F_T]$ is produced by a given initial amount of target $[I_T]_0$. With I_T as the input, the corresponding chemical reaction network is

$$I_T + P \xrightarrow{k_f} F_T \tag{3}$$

$$I_T + S \xrightarrow{k_s} W_T \tag{4}$$

where W_T is a "waste" product not observed in an experiment. Additional byproducts may be produced in both reactions (see Figure 1A) but are not included here for convenience.

The coupled differential equations that can be derived from this CRN do not have a closed-form solution. However, if we make the (realistic) assumption that $[P]_0, [S]_0 \gg [I_T]_0$, so that $[P]$ and $[S]$ are effectively constant through the course of the reaction (see Supplementary Text S1 for details) we find that the signal strength at equilibrium $[F_T]_\infty$ is given by

$$[F_T]_\infty = \frac{k_f[P]_0}{k_f[P]_0 + k_s[S]_0}[I_T]_0 \tag{5}$$

This expression has an intuitive interpretation since it is effectively the ratio of the rates for the productive reaction to the total reaction rate for the input. An expression for $[F_{SNV}]_\infty$ can be derived using an analogous procedure and has exactly the same form as the expression for $[F_T]_\infty$ but with the roles of k_f and k_s interchanged. Substituting our results into the definition for the (end point) discrimination factor, eq 1, we find that

$$DF_\infty = k_f/k_s \tag{6}$$

for an idealized competitive system. The yield can be similarly calculated as $\chi_T = \frac{k_f}{k_f + k_s}$.

Numerical time course simulations for $[F_T]$, $[F_{SNV}]$ and DF are shown in Figure 2A. For our simulations, we arbitrarily set $k_s = k_f/5$ which, according to our analysis, should result in an end point DF = 5 and yield $\chi = 5/6$. In practice, the numerical value for the discrimination stabilizes at a steady state value slightly below the analytically derived value of DF = 5. The slight deviation from the theoretically predicted value is not surprising since for the simulations the concentrations of probe and sink are not constant over the course of the reaction as we assumed in our analytical derivation.

Discrimination by Amplification. Next, we consider an amplification probe P_{amp}, which is catalytically converted to a fluorescent signal species by I_T or I_{SNV} (Figure 2B). In practice, we will realize this system using the entropy driven catalyst[24] (Supplementary Figure S1) but first we consider an idealized catalytic reaction mechanism. The amplification probe is specific to I_T, so the rate constant for the reaction with I_T is k_f while the rate constant for the reaction with I_{SNV} is k_s. In the case where I_T is added to the amplification probe the corresponding chemical reaction is

$$I_T + P_{amp} \xrightarrow{k_f} I_T + F_T \tag{7}$$

The corresponding differential equations can be integrated to yield $[F_T] = [P_{amp}]_0(1 - e^{-k_f[I_T]_0 t})$. The equation and result for

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

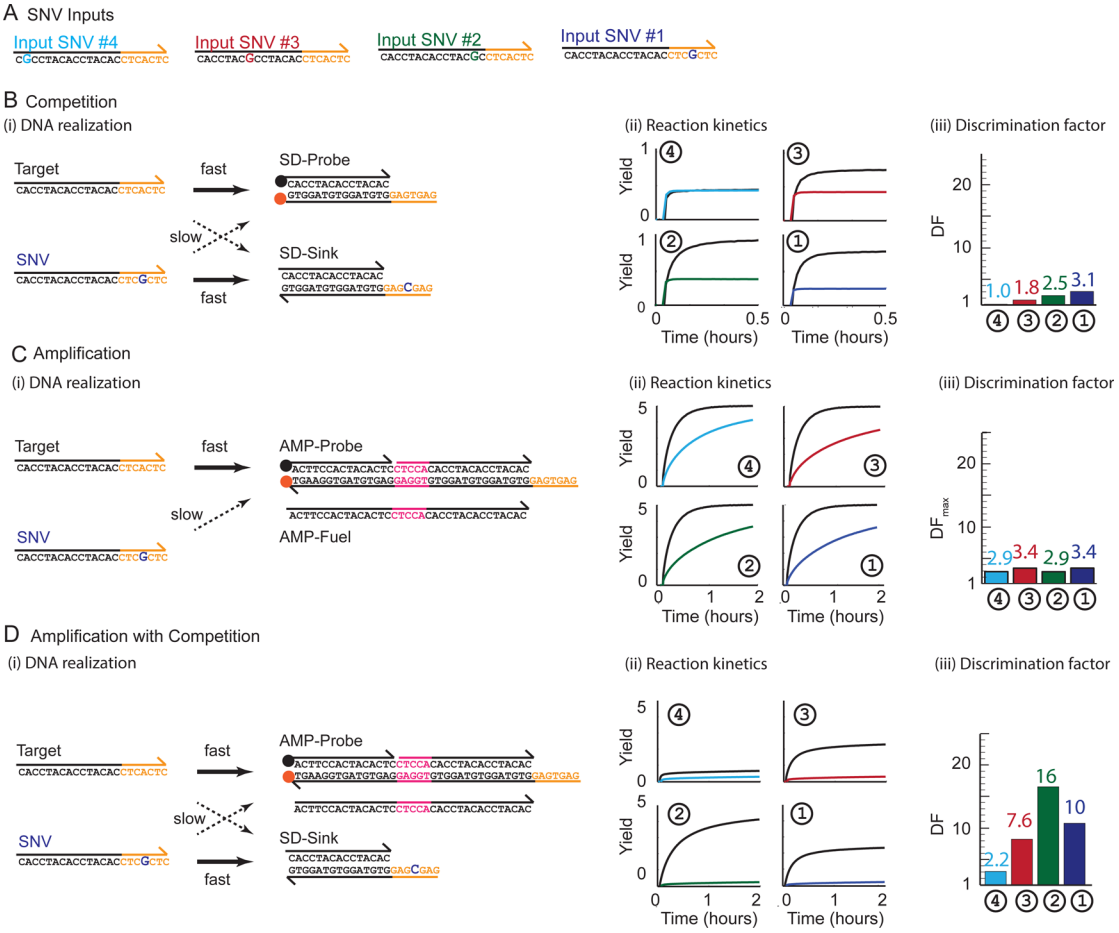


Figure 3. Combining competitive hybridization with amplification achieves higher discrimination than either approach alone. (A) Sequences of the target and SNV inputs. (B) A competitive hybridization circuit using a target-specific probe and an SNV specific sink. (C) An amplification circuit in which target and SNV inputs catalytically generate a fluorescent product. (D) An amplification circuit with a competitive sink. Panel (i) shows the architectures and sequences for the discrimination system. Sinks specific to SNV1 are shown and different, matching sinks are used for experiments with the other SNV inputs. Full thick arrows and dashed thin arrows indicated the fast and slow reaction pathways for a given input species. Panel (ii) shows reaction kinetics traces. Experiments were performed in 1× TAE with 12.5 mM Mg^{2+} at room temperature. Initial concentrations of probe and sink were 10 nM, while the input concentration was 5 nM. Fuel concentration for the amplification reaction was 13 nM. Reaction traces are color-coded to match the colors of the inputs. Panel (iii) shows the discrimination factors calculated from the experimental data at the reaction end point (2 h time point). In the case of the catalytic amplification system the maximal intermediate discrimination factor is shown since there is no end point discrimination.

$[F_{SNV}]$ is completely analogous with k_s instead of k_f and $[I_{SNV}]_0$ in the place of $[I_T]_0$.

We note that at equilibrium, all P_{amp} is converted into F and hence there is no end point discrimination. However, at early time points the different kinetics with I_T and I_{SNV} provide significant discrimination with a discrimination factor

$$DF_{max} \approx k_f/k_s \tag{8}$$

that can be derived from the short time expansion of the exact solution of the equation of motion. Mathematical details are shown in Supplementary Text S1.

The maximum yield

$$\chi_\infty = [P_{amp}]_0/[I_T] \tag{9}$$

is achieved for $t \to \infty$. In principle, maximum yield is thus limited only by the concentration of P_{amp}, but in practice oligonucleotide synthesis errors typically limit the yield achievable in an experimental system to about 50.[24] Numerical simulations for $[F_T]$, $[F_{SNV}]$ and DF are shown in Figure 2B.

Combining Competition with Amplification. Finally, we will show that combining competition with amplification can result in a system that has quadratically better end point discrimination than a purely competitive system while also being similarly sensitive to a system with amplification only (Figure 2C). We first consider the chemical reaction network for the case where the target input I_T is present:

$$I_T + P_{amp} \xrightarrow{k_f} I_T + F_T \tag{10}$$

$$I_T + S \xrightarrow{k_s} W_T \tag{11}$$

The corresponding set of differential equations for the concentration trajectories can be integrated exactly[30] (see Supplementary Text S1 for details). Here, we are most interested in end point discrimination and we consider only the limit for large times. If, for convenience, we further assume that the initial amounts of catalytic probe and sink are equal we obtain

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

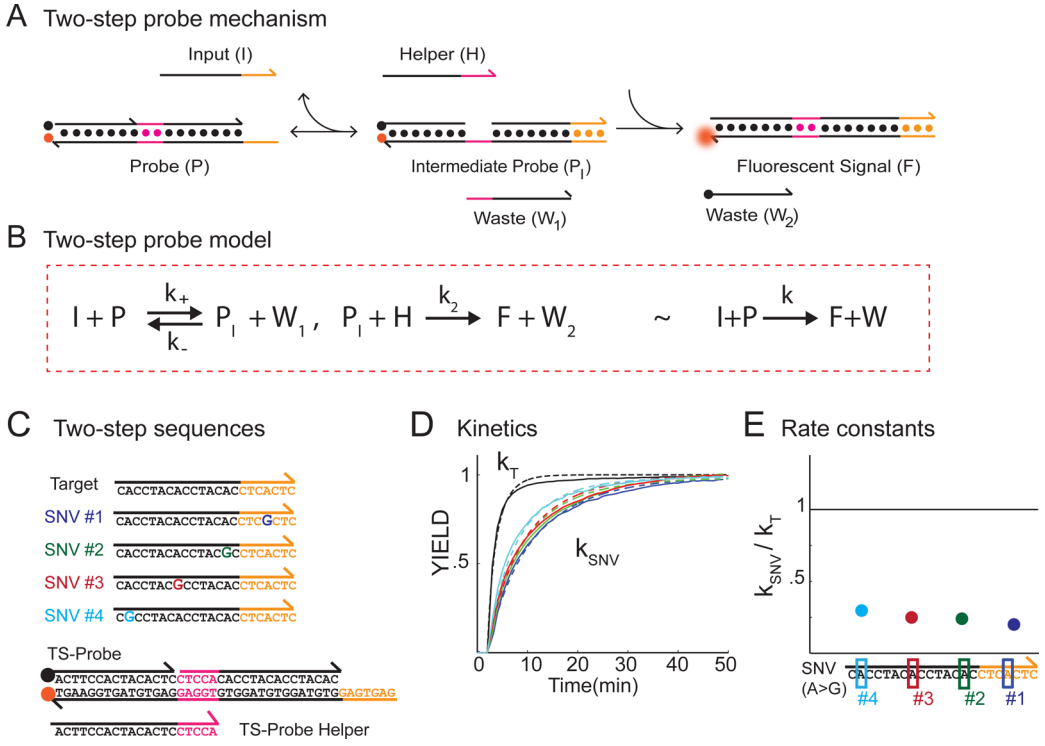


Figure 4. A two-step reaction pathway results in improved kinetic discrimination between target and SNV inputs. (A) Two-step reaction mechanism. The first reaction step is initiated at the orange toeholds and is fully reversible because the displaced waste strand (W_1) can bind back to the newly formed toeholds (magenta) in the reactive probe (P_I). In the second reaction step, an auxiliary helper molecule (H) competes with W_1 for binding to the magenta toeholds. Binding of H makes the overall reaction irreversible. We use a fluorophore (red dot) and quencher (black dot) to follow reaction kinetics experimentally. (B) A mechanistic reaction model where each strand displacement step is approximated by a bimolecular rate constant: k_+ is the rate for initial binding of the input, k_- is the rate for input displacement in the reverse reaction, and k_2 is the rate for binding of helper to the activated probe. Under some conditions the overall reaction is well approximated by a single bimolecular reaction with rate constant k. (C) Input and probe sequences. The probe is designed to exactly match the target input sequence. (D) Kinetics traces for different inputs. Reaction with the target input is the fastest; reactions with all SNV inputs are slower. Unlike for the single step reaction, all traces corresponding to SNV inputs are bunched together. Experiments were performed in 1× TAE with 12.5 mM Mg^{2+} at room temperature. Initial concentrations of probe and input were 10 nM and 5 nM, respectively. Helper concentration was 13 nM. (E) Ratio of reaction rates of each SNV input and target input. Even the reaction rate for SNV4 is clearly slower than that for the target input.

$$[F_T]_\infty = [I_T]_0 k_f/k_s \tag{12}$$

The expression for $[F_{SNV}]_\infty$ is of the same form as that for F_T but with the roles of k_f and k_s interchanged and with $[I_{SNV}]_0$ in place of $[I_T]_0$. Combining these results we find that the discrimination factor is

$$DF_\infty = (k_f/k_s)^2 \tag{13}$$

and thus quadratically better than for either of the two previously introduced systems. The yield at the reaction endpoint is $\chi_\infty = k_f/k_s$. Numerical simulations results for $[F_T]$, $[F_{SNV}]$ and DF are given in Figure 2C.

Experimental Results. Next, we set out to experimentally test our predictions for the three different SNV detection systems. For our experiments we used the same target input and four SNV inputs introduced above (Figure 1C and Figure 3A). The sinks used for all experiments have the same design as a standard strand displacement probe, except that both strands are unlabeled and the sequences match the SNV target used in a given experiment. Sequences for all four sinks used in our experiments are given in Supplementary Table S2.

Discrimination by Competition. Experimentally we found that the level of discrimination strongly depended on the position of the SNV. The best discrimination (DF = 3.1) was achieved for SNV1 corresponding to a mutation in the toehold region. The ability to detect SNVs decreased for SNVs closer to the end of the branch migration region; the system lost specificity (DF = 1) when the mutation occurred near the end of strand displacement region. These results are not surprising given that discrimination is the direct result of differences in binding kinetics between the correct and mutated target but strand displacement rate constants are virtually identical for the target input and the SNV4 input (Figure 1E).

Discrimination by Amplification. To realize a catalytic signal amplification reaction we used the entropy driven catalyst system[24] (Supplementary Figure S1). The amplification system has two components: a single-stranded fuel (AMP-Fuel, Figure 3C) and a three-stranded probe (AMP-Probe, Figure 3C). In the absence of either target or SNV input, the mixture of fuel and probe is metastable and does not react significantly. When the target input is present, a fast pathway for the rearrangement of the probe and fuel strands becomes available. The target first binds to the probe, displacing an auxiliary strand and freeing up a toehold for binding of the fuel. Binding of the fuel results in the displacement of a quencher strand from the probe and thus an increase in fluorescence. Moreover, the fuel also rereleases the target input which is thus free to interact with another probe complex and repeat the catalytic cycle. Our experiments demonstrate that the reaction pathway can be catalyzed not only by the target input but also the SNV input albeit with

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

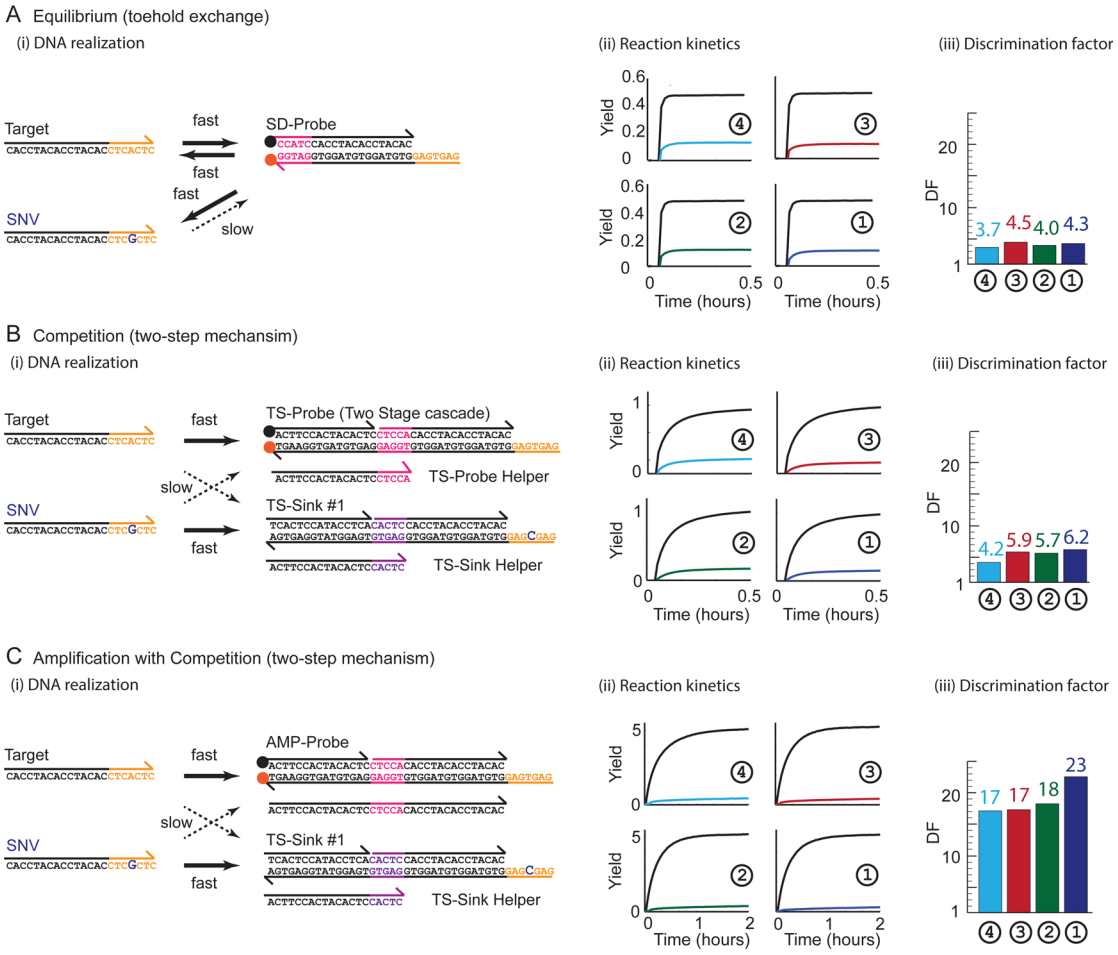


Figure 5. Using an engineered two step sink and catalytic probe, SNVs in any position of the input can be reliably identified. (A) A reversible toehold exchange probe does not exhibit significant SNV position dependence. (B) A competitive network using a two-step probe and sink combines position independence with increased sensitivity. (C) An amplification circuit with a competitive two-step sink exhibits dramatically increased specificity and sensitivity for all SNVs. Panel (i) shows the architectures and sequences for the discrimination system. Sinks specific to SNV1 are shown and different, matching sinks are used for experiments with the other SNV inputs. Full thick arrows and thin dashed arrows indicated the fast and slow reaction pathways for a given input species. Panel (ii) shows reaction kinetics traces. Experiments were performed in 1× TAE with 12.5 mM Mg^{2+} at room temperature. Initial concentrations of probe and sink were 10 nM, while the input concentration was 5 nM. Fuel and helper concentrations were 13 nM. Reaction traces are color-coded to match the colors of the inputs. Panel (iii) shows the discrimination factors calculated from the experimental data at the reaction end point.

slower kinetics (Figure 3C). Because the discrimination factor at the reaction end point is DF = 1, we instead use DF_{max} which we define as the maximum discrimination achieved during the experimental time course. DF_{max} ranges from 3.4 to 2.9 for the four different SNVs and, unlike in the previous case of a competitive system, there is no strong dependence of the discrimination factor on the mutation position. As we will explain in more detail below, this relative insensitivity to SNV position is a direct result of the two step reaction mechanism, specifically the reversibility of the first reaction step.

Combining Competition with Amplification. Next, we set out to test our prediction that a combination of competition with amplification should dramatically enhance probe specificity. For these experiments we combined a strand displacement sink with an entropy-driven catalytic amplification system (Figure 3D). We observed improved discrimination compared to probes that rely on competition or amplification alone. For example, the measured discrimination factor for SNV1 (DF = 10) is approximately quadratically larger than the maximal discrimination factor measured with a competitive probe or an amplifying probe alone (DF = 3.1 and 3.4, respectively).

Although this result demonstrates that amplification in parallel with competition can result in considerably more sensitive and specific probes, the numerical agreement between the model and data is somewhat coincidental. Our model assumes that both the amplification and inhibition reactions can be approximated as a single-step bimolecular reaction but in practice the amplification reaction is a multistep process involving an initial reversible strand exchange step followed by an irreversible strand displacement reaction. Even though such a multistep mechanism can sometimes be reasonably approximated by a bimolecular rate law, the corresponding overall reaction rate will be different from the rate observed for a single strand displacement reaction, making it difficult to quantitatively compare the rates of the amplification and inhibition pathways.

The main shortcoming of the current approach is the strong dependence of the discrimination factor on the exact position of the mismatch. Although all SNVs could be detected, the

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

discrimination factor dropped to about DF = 2.2 for SNV4. Moreover, we would expect that the discrimination factor varies monotonically with SNV position, but observed a non-monotonic behavior in our experiment with optimal discrimination achieved for SNV2 (DF = 16). These discrepancies between model and experiment are artifacts of the asymmetry between the mechanisms for the amplification and inhibition reaction in the experiment. To overcome these issues and create a system that is not only insensitive to the mismatch position but also more easily compared to the model, we next designed a sink that reacts with an input through a multistep reaction pathway that is very similar to that of the amplifying probe.

Position Dependence. *Two Stage Probe.* To build a discrimination system that enables a direct comparison of the model to the experiment we created a two-step strand displacement probe. The reaction mechanism for this system is almost identical to the mechanism of the entropy driven catalyst but uses a truncated fuel species, referred to as a "helper" strand (Figure 4A). Consequently, unlike in the catalytic system, the input is not released when the helper binds; instead, binding of the helper results in an irreversible sequestration of the input by the probe (Figure 4A). As above, quencher and fluorophore modifications are used to follow reaction kinetics. The corresponding reaction model is given in Figure 4B; every strand displacement step is modeled as a bimolecular reaction. Sequences for all probe strands used in our experiments are listed in Supplementary Table S3.

Next, we experimentally characterized the reaction kinetics for this system with the target input and all four SNV variants (Figure 4C). Example reaction kinetics traces are shown in Figure 4D. To obtain the rate constants (Figure 4E), we fitted the experimental data to a second order rate law, rather than the more detailed multistep mechanism, since these second order rate constants provide a measure of reaction speed that is easy to interpret. Unlike what we observed previously for a probe using only a single strand displacement step (Figure 1), we here find that mutations in any location within the target result in slower kinetics than the correct target. In fact, the measured slowdown is almost identical for all mutations (Figure 4D,E). The reversible nature of the initial toehold exchange step is responsible for making the discrimination rates independent of SNV position, as explained in more detail in the Supplementary Text S2.

Toehold Exchange Probe. Zhang et al.[12] recently demonstrated that a toehold exchange probe alone (i.e., without the second irreversible step introduced above) could serve as a highly specific SNV discrimination system (Figure 5A). As explained above, the equilibrium signal in such a system will be different for the target and SNV input. As shown in ref 12. and detailed in Supplementary Text S2, the discrimination factor for this system is given by

$$\mathrm{DF} = \frac{1}{2}\left(1 + \sqrt{\frac{k_-^{\mathrm{SNV}}}{k_+^{\mathrm{SNV}}}}\right) \tag{14}$$

The forward and reverse reaction rate constants for the reaction with the SNV input, k_+^{SNV} and k_-^{SNV}, will vary depending on the location of the SNV but the ratio remains fixed at k_+^{SNV} and $k_-^{\mathrm{SNV}}- = e^{-\Delta G_{\mathrm{SNV}}/R\tau}$. Here we also assume that the reaction free energy for binding of the target input is close to zero (balanced toeholds) and rate constants for the forward and reverse reactions are equal, $k_+ \approx k_-$.

For comparison to our own kinetic discrimination probes, we experimentally tested the equilibrium toehold exchange probe with our target and SNV inputs (Figure 5A). As expected, the discrimination factor is similar for all mutations, though not very high. Moreover, the yield is very low for this type of probe and is in fact inversely correlated with the level of discrimination.

Competition with Two Step Probe and Sink. To maintain the position independence we get from a reversible reaction and improve the final yield, we created a competitive kinetic discrimination system using our two step sink architecture. Combining a probe using this two step reaction mechanism and with a competitive sink using the same mechanism we obtain the system shown in Figure 5B; as was the case for the one step competitive system introduce above, here the probe matches the target input, and the sink matches the SNV input.

In Supplementary Text S2, we derive an approximate expression for the discrimination factor. If we design our system such that $k_+ = k_-$, we find that

$$\mathrm{DF} = \sqrt{k_-^{\mathrm{SNV}}/k_+^{\mathrm{SNV}}} \tag{15}$$

Here, k_-^{SNV} and k_+^{SNV} are the reverse and forward rate for binding of the SNV input in the reversible reaction step. This result is very similar to the result we found for a fully reversible probe, but we note that the two-step mechanism results in slightly better discrimination than the former. Importantly, addition of the irreversible second reaction step also results in a higher yield for the reaction.

Experimentally, we find that the end point discrimination factor for this system ranges from 4.2 (SNV4) to 6.2 (SNV1). Thus, compared to one step competition, the position dependence is much reduced. Moreover, the yield in this reaction is considerably higher than for the toehold exchange system while the discrimination factor is comparable or slightly higher.

Combining Competition with Amplification. Next, we combine a sink based on a two step mechanism with an amplification reaction (Figure 5C). Again, we first derive an analytical expression for the discrimination factor (Supplementary Text S2) and find that

$$\mathrm{DF} = k_-^{\mathrm{SNV}}/k_+^{\mathrm{SNV}} \tag{16}$$

As for the simpler one-stage mechanism introduced above, the amplification reaction results in quadratically better discrimination, suggesting that combining amplification with competition results in a quadratic performance improvement independently of the details of the underlying reaction mechanism. Results of a numerical simulation are in agreement with the analytical derivation and are shown in Supplementary Figure S2.

In our experiments we found that the discrimination factors measured for this system are indeed approximately equal to the product of the discrimination factors measured independently for a two-step competitive system or a catalytic system alone. Discrimination factors range from DF = 17 for SNV4 to DF = 23 for SNV1. Moreover, because of the amplification reaction, the yield for this system is about 5-fold higher than for the system without amplification and at least 10-fold better than for a simple toehold exchange probe.

Let-7 MicroRNA Discrimination. So far, we only tested our engineered discrimination system with DNA inputs, but discrimination between single-stranded RNA is of more likely

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

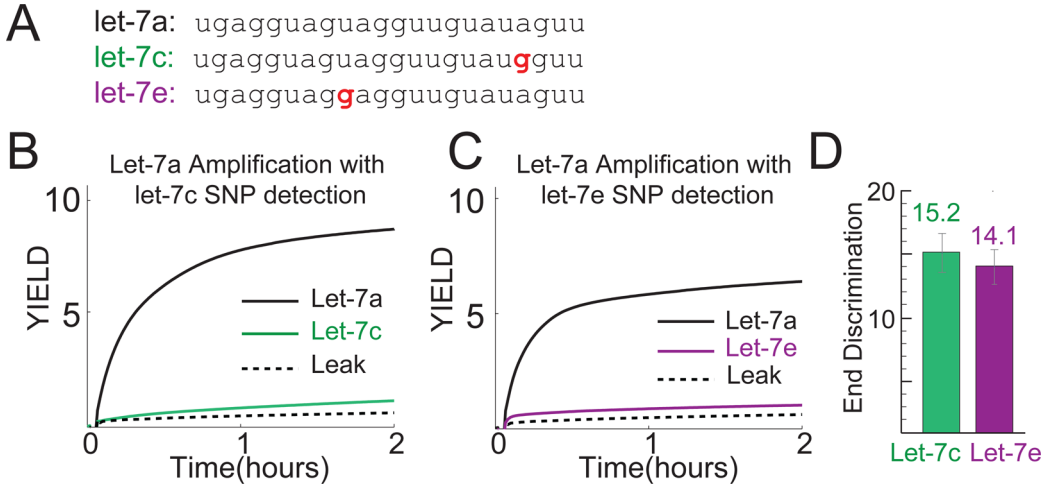


Figure 6. Detection of SNVs in microRNA. (A) Sequences of the let-7a, let-7c, and let-7e microRNA. Both let-7c and let-7e differ from let-7a by only a single nucleotide. Synthetic RNA sequences are used in the experiment. (B) Time-course fluorescence data for the competitive amplification system targeting let-7a and suppressing let-7c. (C) Time-course fluorescence data for the competitive amplification system targeting let-7a and suppressing let-7e. Let-7a interacts differently with the let-7e sink than with the let-7c sink resulting in slightly different kinetics traces for the let-7a input. (D) Summary of observed discrimination factor values. Error bars show the standard deviation across two repeats.

interest for practical applications. As a proof-of-concept demonstration of RNA detection we designed a system that could distinguish between different members of the let-7 family of microRNAs (miRNAs).[31] Let-7 family miRNA have important roles in development and at least eight different family members are found in humans. For our experiments, we used synthetic RNA oligonucleotides with the same sequences as the human miRNA let-7A, let-7C and let-7E. Specifically, we aimed to discriminate between let-7A and the other two sequences, which are both only one nucleotide removed from let-7A (Figure 6A).

We found that a discrimination system combining competition with amplification could distinguish let-7A from both let-7C (Figure 6B) and let-7E (Figure 6C) with high specificity (DF ≈ 15) and sensitivity. It is particularly encouraging that this experiment did not require any specific modifications to the amplification system or sink, even though DNA−RNA interaction parameters, and thus the strand displacement rate constants, can be fairly different from those measured for DNA−DNA interactions.[32]

■ DISCUSSION

We showed that a rational design approach based on strand displacement reactions can be used to create a sensitive and specific discrimination system. First, we mathematically analyzed and compared the specificity of three different systems, a first discrimination system based on competitive strand displacement, a second system based on amplification alone, and a third system that integrated competition with amplification. We found that amplification in parallel with competition results in quadratically better discrimination between a target input and an SNV input, than either amplification or competition alone. We then tested these predictions using existing components, namely an entropy-driven catalytic amplifier operated in parallel with a strand displacement sink. Although, this system performed as expected for mutations near the toehold domain, it was not capable of identifying mutations further away from the toehold. To overcome these limitations we created a competitive sink including a reversible pre-equilibrium step. The improved system combining this two step sink with a catalytic amplification mechanism was capable of discriminating agains SNVs occurring anywhere in the input with similarly high specificity. Importantly, unlike equilibrium detection approaches based on reversible strand displacement which had high specificity but low sensitivity, our system has at least 20-fold better sensitivity because of the built in amplification step while also having quadratically better specificity. Together, these results demonstrate that rational reaction engineering can be used to design specific and sensitive SNV detection systems.

Intriguingly, the detection system we developed here can easily be integrated with existing DNA logic gates and other sensors to create complex diagnostic assays for multianalyte detection and analysis. Moreover, because of the robustness of DNA strand displacement reactions to temperature and buffer conditions such system could become important components for diagnostic applications in low resource settings. Our work thus provides a foundation for the design of a new class of customizable, rationally designed diagnostics for the point of care.[33,34]

Above, we derived the discrimination factors for the different SNV detection systems from the equations of motion. However, the observed quadratic improvement for a system that combines competition and amplification can also be understood from a more intuitive argument (see also Supplementary Figure S3). For a given target input, the probability to bind to a catalytic probe rather than a sink is $k_f/(k_f + k_s) = x$, if we assume equal concentrations for probe and sink. For this discussion, we further assume that the concentrations of the probe and sink are constant. In an initial "round" of catalysis we thus expect a fraction $x[I_T]$ of the target strands to react productively in the catalytic reaction producing a fluorescent signal $F_T = x[I_T]$. In a second round, any member of this initially successful fraction can again produce a signal with probability x, resulting in a cumulative amount of signal given by $F_T = (x + x^2)[I_T]$. Assuming that this process continues, we can estimate that the final signal produced by I_T will be $F_T = (x + x^2 + x^3 + ...)[I_T] = x/(1 - x)[I_T]$. Similarly, the probability that an SNV input results in a productive reaction is $k_s/(k_f + k_s) = 1 - x$ and thus the fluorescent signal

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

after a single round is $F_{SNV} = (1 - x)[I_{SNV}]$. Applying the same reasoning as for the target input, we find that the asymptotic fluorescence signal is $F_{SNV} = (1 - x)/x[I_{SNV}]$. Thus, if we assume $[I_T] = [I_{SNV}]$ and calculate the discrimination factor DF = F_T/F_{SNV} after a single reaction round, i.e., the case of a simple competitive discrimination system, we find DF = $x/(1 - x) = k_f/k_s$. Conversely, after many rounds of catalysis the result is DF = $(x/(1 - x))^2 = (k_f/k_s)^2$ and discrimination is thus quadratically improved in the system with catalysis.

Although the details of the mechanism are very different, our results are reminiscent of kinetic proof-reading[35] or kinetic amplification,[36] mechanisms used by biological enzymes to discriminate with high specificity between productive and nonproductive substrates. Hopfield[35] and Ninio[36] argued that a multistep reaction pathway including an irreversible proof-reading step coupled to energy expenditure could explain the discrepancy between the actual fidelity of protein synthesis and that expected from the thermodynamics of tRNA binding alone. Here, we similarly show that the ability of a DNA hybridization probe to discriminate between a fully complementary target and a target containing a single nucleotide variant can be quadratically better for a system that is irreversible and consumes energy.

In conclusion, our work demonstrates that systematic, model-guided reaction engineering can provide a valuable path toward improving the performance of biotechnological systems.

METHODS

DNA Oligonucleotides. DNA oligonucleotides were purchased from Integrated DNA Technology (IDT). Fluorophore- and quencher-labeled oligonucleotides were HPLC purified; all other oligonucleotides were PAGE purified. DNA oligonucleotides were resuspended to 100 μM and stored in EB buffer (10 mM Tris·HCl, pH 8.5; Qiagen).

Probe Preparation. Probe and Sink molecules consist of two or three distinct strands. Strands were mixed with a 1.2× excess of the short top strands and then thermally annealed (Biorad T100), cooling uniformly from 98 to 25 °C over the course of 73 min. Probe and sink molecules were then gel purified to ensure stoichiometry using a 10% polyacrylamide gel.

Purification gel solutions were prepared from 40% 19:1 acrylamide:bis(acrylamide) stock (J.T. Baker Analytical) in 1× tris-acetate-EDTA buffer (TAE)/Mg^{2+} solution, and cast between 20 cm by 20 cm glass plates with 1.5 mm spacers. Samples were loaded with 80% glycerol to achieve 10% glycerol concentration by volume. The gel was run at room temperature using Hoefer SE600 chamber at 140 V for 4 h. Gel bands were visualized using Entela UL3101 UV light, using a fluorescent backplate (Whatman UV254 Polyester 4410222), and then cut out and eluted into 1 mL buffer.

Time Course Fluorescence Studies. Kinetic fluorescence measurements were performed using a Horiba Fluoromax 3 spectrofluorimeter and Hellma Semi-Micro 114F spectrofluorimeter cuvettes. For kinetics experiments, probes were labeled with the ROX fluorophore (excitation 584 nm, emission 603 nm). Slit sizes were set at 5 nm for all monochromators. An external temperature bath maintained a designated reaction temperature (25 ± 1 °C). A four-sample changer was used, so that time-based fluorescence experiments were performed in groups of 4. Each data point represents the integrated fluorescence over 10 s per minute of reaction.

Fluorescence Normalization. All fluorescence values were normalized and converted to concentrations using the following formula:

$$[F] = \frac{F - F_b}{F_s - F_b}[I]$$

where F is observed fluorescence, F_b is background fluorescence observed before input addition, and F_s is the saturated fluorescence observed after triggering the probe with 10-fold excess of the input.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b00277.

Details of the mathematical models (Text S1 and S2), additional results and system descriptions (Figures S1−S3) and sequence information (Tables S1−S3). (PDF)

AUTHOR INFORMATION

Corresponding Author

*gseelig@uw.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank David Yu Zhang and David Soloveichik for insightful discussion and helpful manuscript preparation suggestions. This work was funded by NSF CAREER Award CBET 0954566, NSF Award CCF 1317653, and DARPA Young Faculty Award YFA N66001-12-1-4225 to G.S.

REFERENCES

(1) Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. *Science* **1998**, *239*, 487−491.
(2) Kim, S.; Misra, A. *Annu. Rev. Biomed. Eng.* **2007**, *9*, 289−320.
(3) Mardis, E. R. *Annu. Rev. Genomics Hum. Genet.* **2008**, *9*, 387−402.
(4) Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. *Chem. Commun.* **1998**, *4*, 455−456.
(5) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 1895−1897.
(6) Simeonov, A.; Nikiforov, T. T. *Nucleic Acids Res.* **2002**, *30* (17), e91.
(7) Jacobsen, N.; Bentzen, J.; Meldgaard, M.; Jakobsen, M. H.; Fenger, M.; Kauppinen, S.; Skouv, J. *Nucleic Acids Res.* **2002**, *30* (19), e100.
(8) Komiyama, M.; Ye, S.; Liang, X.; Yamamoto, Y.; Tomita, T.; Zhou, J.-M.; Aburatani, H. *J. Am. Chem. Soc.* **2003**, *125*, 3758−3762.
(9) Tyagi, S.; Kramer, F. R. *Nat. Biotechnol.* **1996**, *14*, 303−308.
(10) Li, Q.; Luan, G.; Guo, Q.; Liang, J. *Nucleic Acids Res.* **2002**, *30* (2), e5.
(11) Xiao, Y.; Plakos, K. J. I.; Lou, X.; White, R. J.; Qian, J.; Plaxco, K. W.; Soh, H. T. *Angew. Chem., Int. Ed.* **2009**, *48*, 4354−4358.
(12) Zhang, D. Y.; Chen, S. X.; Yin, P. *Nat. Chem.* **2012**, *4*, 208−214.
(13) Chen, S. X.; Zhang, D. Y.; Seelig, G. *Nat. Chem.* **2013**, *5*, 782.
(14) Zhang, D. Y.; Seelig, G. *Nat. Chem.* **2011**, *3*, 103−114.
(15) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605−608.
(16) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585−1588.
(17) Soloveichik, D.; Seelig, G.; Winfree, E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5393−5398.
(18) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196−1201.
(19) Chen, Y.-C.; Dalchau, N.; Srinivas, N.; Phillips, A.; Cardelli, L.; Soloveichik, D.; Seelig, G. *Nat. Nanotechnol.* **2013**, *8*, 755.
(20) Chirieleison, S. M.; Allen, P. B.; Simpson, Z. B.; Ellington, A. D.; Chen, X. *Nat. Chem.* **2013**, *5*, 1000−1005.
(21) Turberfield, A. J.; Mitchell, J. C.; Yurke, B.; Mills, A. P., Jr; Blakey, M. I.; Simmel, F. C. *Phys. Rev. Lett.* **2003**, *90*, 118102.
(22) Seelig, G.; Yurke, B.; Winfree, E. *J. Am. Chem. Soc.* **2006**, *128*, 12211−12220.
(23) Bois, J. S.; Venkataraman, S.; Choi, H. M. T.; Spakowitz, A. J.; Wang, Z. G.; Pierce, N. A. *Nucleic Acids Res.* **2009**, *33*, 4090−4095.

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086

(24) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121−1125.
(25) Yurke, B.; Mills, A. P., Jr. *Genet. Program. Evol. M.* **2003**, *4*, 111−122.
(26) Zhang, D. Y.; Winfree, E. *J. Am. Chem. Soc.* **2009**, *131*, 17303−17314.
(27) Srinivas, N.; Ouldridge, T. E.; Sulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P. K.; Winfree, E. *Nucleic Acids Res.* **2013**, *41*, 10641−10658.
(28) Machinek, R. R. F.; Ouldridge, T. E.; Haley, N. E. C.; Bath, J.; Turberfield, A. J. *Nat. Commun.* **2014**, *5*, 5324.
(29) SantaLucia, J.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415−440.
(30) Zhang, D. Y.; Seelig, G. In *DNA Computing and Molecular Programming*; Satoshi, M., Satoshi, K., Eds; Springer: Kyoto, 2011; pp 176−186.
(31) Roush, S.; Slack, F. J. *Trends Cell Biol.* **2008**, *18*, 505−516.
(32) Sugimoto, N.; Nakano, S. I.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. *Biochemistry* **1995**, *34*, 11211−11216.
(33) Niemz, A.; Ferguson, T. M.; Boyle, D. S. *Trends Biotechnol.* **2011**, *29*, 240−250.
(34) Jung, C.; Ellington, A. D. *Acc. Chem. Res.* **2014**, *47*, 1825−1835.
(35) Hopfield, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4135−4139.
(36) Ninio, J. *Biochimie* **1975**, *57*, 587−595.

DOI: 10.1021/jacs.6b00277
J. Am. Chem. Soc. 2016, 138, 5076−5086