Dynamic DNA Nanotechnology for Probing Single Nucleotide Variants and DNA Modifications

By

Guan Wang

A Thesis Submitted to the Department of Chemistry

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Faculty of Mathematics & Science, Brock University

St. Catharines, Ontario

Abstract

In the last decades, various DNA hybridization probes have been developed that attempt to conquer the challenge of single-nucleotide-variants (SNVs) detection. Even though a powerful toolbox including the toehold-exchange reaction, the dynamic 'sink' design, and the polymerase chain reaction (PCR) has been built, it still faces practical problems. For example, the natural DNA is usually in double-stranded form whereas most hybridization probes aim for single-stranded targets; the concentration of extracted DNA samples is totally unknown thus may lay outside the optimal design of probes/primers. To achieve ultra-high sensitivity and specificity, expensive and sophisticated machines such as digital droplet PCR and next-generation-sequencing may be inapplicable in rural areas. Therefore, the quantitative PCR method is still the gold standard for clinical tests. Thus motivated, my PhD career was mainly focused on the fundamental understanding of the challenges in SNVs discrimination and developing robust, versatile, and user-friendly probes/strategies.

In this thesis, Chapter 1 provides a general introduction of dynamic DNA nanotechnology and its representative applications in discriminating SNVs. Chapter 2 to 4 describe three completed projects that aim to understand the thermodynamic and kinetic properties of strand displacement reactions and to circumvent the challenges of discriminating SNVs through finely tuned probes/assays.

Acknowledgements

I would like to express my sincere gratitude to my mentor, Dr. Feng Li, for offering me this invaluable opportunity to tour in the mysterious world of science, and his unwavering encouragement is crucial for me to finish this journey.

For eight years, I was absent from the traditional Chinese “Lunar New Year” dinner with my family. My parents support me selflessly all along, but still, it has been too long and I really miss them. Now, I’d say it’s time to go home.

In my journey aboard I am not alone, and I really appreciate my friends’ kind help and accompany when the tide against me, especially Yumeng Li, Xiaolong Yang, Junpeng Xu, Peng Yang, Tianyu Dong, Nicolas A. Sardella, Xiaotian Xu, and Jianyuan Feng.

I also want to say thank you to my committee members, Dr. Melanie Pilkington, Dr. Georgii Nikonov, for their insightful opinions to help improve my research skills. It is very lucky for me to spend my Ph.D. career in Brock University.

Table of Contents

List of Figures

List of Tables

List of equation

Equation 1.1: Specificity (DF): $= \frac{\chi_T}{\chi_N} \cong \frac{\frac{[T \cdot Probe]}{[T]_0}}{\frac{[N \cdot Probe]}{[N]_0}}$ $or\ DF = \frac{\chi_T}{\chi_N} \cong \frac{\Delta F_T}{\Delta F_N}$

Equation 1.2: Sensitivity(yield): $= \frac{[T \cdot Probe]}{[T]_0}$

Equation 3.1: $T + CP \rightleftharpoons TC + P, with\ \Delta G^{\circ}_{rxn}$

Equation 3.2: $K_{eq} = e^{-\Delta G^0/RT} = \frac{[TC]_{eq} \cdot [P]_{eq}}{[T]_{eq} \cdot [CP]_{eq}} = \frac{[TC]_{eq} \cdot ([P]_0 + [TC]_{eq})}{([T]_0 - [TC]_{eq}) \cdot ([CP]_0 - [TC]_{eq})}$

Equation 3.3: $K_{eq} = e^{-\Delta G^0/RT} = \frac{[TC]_{eq} \cdot ([P]_0 + [TC]_{eq})/[CP]_0^2}{([T]_0 - [TC]_{eq}) \cdot ([CP]_0 - [TC]_{eq})/[CP]_0^2} = \frac{\eta \cdot (\gamma + \eta)}{(\tau - \eta) \cdot (1 - \eta)}$

$where\ \gamma \coloneqq {[P]_0}/{[CP]_0}\ and\ \tau \coloneqq {[T]_0}/{[CP]_0}\ respectively$

Equation 3.4: $\eta = \frac{K_{eq} + \gamma + K_{eq}\tau - \sqrt{K_{eq}^2\tau^2 - 2K_{eq}^2\tau + K_{eq}^2 + 2K_{eq}\gamma\tau + 2K_{eq}\gamma + 4K_{eq}\tau + \gamma^2}}{2(K_{eq} - 1)}$

$$= \frac{(K_{eq} + \gamma + K_{eq}\tau) - \sqrt{(K_{eq} + \gamma + K_{eq}\tau)^2 - 4K_{eq}\tau(K_{eq} - 1)}}{2(K_{eq} - 1)}$$

$where\ K_{eq} \neq 1.$

Equation 3.5: $\eta = \frac{\tau}{\gamma + \tau + 1},\ when\ K_{eq} = 1,$

Equation 3.6: $DF = \frac{\eta_c}{\eta_s} = \frac{(K_{eq,s} - 1) \cdot \left\{(K_{eq,c} + \gamma + K_{eq,c}\tau) - \sqrt{(K_{eq,c} + \gamma + K_{eq,c}\tau)^2 - 4K_{eq,c}\tau(K_{eq,c} - 1)}\right\}}{(K_{eq,c} - 1) \cdot \left\{(K_{eq,s} + \gamma + K_{eq,s}\tau) - \sqrt{(K_{eq,s} + \gamma + K_{eq,s}\tau)^2 - 4K_{eq,s}\tau(K_{eq,s} - 1)}\right\}}$

Equation 3.7: $DF = \frac{\eta_c}{\eta_s} = \frac{2(K_{eq,s} - 1) \cdot \tau}{(\gamma + \tau + 1) \cdot \left\{(K_{eq,s} + \gamma + K_{eq,s}\tau) - \sqrt{(K_{eq,s} + \gamma + K_{eq,s}\tau)^2 - 4K_{eq,s}\tau(K_{eq,s} - 1)}\right\}}$

Equation 3.8: $\tau = \frac{\eta \cdot (\gamma + \eta)}{K_{eq} \cdot (1 - \eta)} + \eta$

Equation 3.9: $RF = \frac{\tau_{spurious}}{\tau_{correct}} = \frac{\frac{\eta \cdot (\gamma + \eta)}{K_{eq,spurious} \cdot (1 - \eta)} + \eta}{\frac{\eta \cdot (\gamma + \eta)}{K_{eq,correct} \cdot (1 - \eta)} + \eta} = \frac{(1 - \eta) + (\gamma + \eta) \cdot {1}/{K_{eq,spurious}}}{(1 - \eta) + (\gamma + \eta) \cdot {1}/{K_{eq,correct}}}$

Equation 3.10: $RF = \frac{0.5 + (\gamma + 0.5) \cdot {1}/{K_{eq,spurious}}}{0.5 + (\gamma + 0.5)}\ when\ K_{eq} = 1$

Equation 3.11: $Y = \frac{L+x^S\cdot(M-L)}{x^S+E^S}$ *where* M, L, s *and* E *are parameters to be fitted*

Equation 4.1: $W_{(f,r)} + CT_N \rightleftharpoons W_{(f,r)}C + T_N,\ \Delta G^\circ_N$ *for unmodified DNA balance*

Equation 4.2: $W_{(f,r)} + CT_{cis} \rightleftharpoons W_{(f,r)}C + T_{cis},\ \Delta G^\circ_{cis}$ *for cis − cAB DNA balance*

Equation 4.3: $\Delta\Delta G^\circ_{CT} := \Delta G^\circ_N - \Delta G^\circ_{cis} = \left(\Delta G^\circ_{CW} + \Delta G^\circ_{T_N} - \Delta G^\circ_W - \Delta G^\circ_{CT_N}\right)$
$$-\left(\Delta G^\circ_{CW} + \Delta G^\circ_{T_{cis}} - \Delta G^\circ_W - \Delta G^\circ_{CT_{cis}}\right) \approx \Delta G^\circ_{CT_{cis}} - \Delta G^\circ_{CT_N}$$

Equation 4.4: $\Delta G^\circ_{CT_{cis}} = \Delta G^\circ_{CT_N} + \Delta\Delta G^\circ_{CT}$

Equation 4.5: $\eta = \frac{F_{CT}-F_{BG}}{F_{P.C.}-F_{BG}}$

Equation 4.6: $\Delta G^\circ_{iso} = \Delta G^\circ_{trans} - \Delta G^\circ_{cis} = -\mathrm{RT}\cdot\ln(K_2/K_3) \approx -\mathrm{RT}\cdot\ln(k_2/k_3)$

Equation 4.7: $\Delta\Delta G^\circ_{trans} = \Delta G^\circ_N - \Delta G^\circ_{trans} = \left(\Delta G^\circ_N - \Delta G^\circ_{cis}\right) + \left(\Delta G^\circ_{cis} - \Delta G^\circ_{trans}\right)$
$$= \Delta\Delta G^\circ_{cis} - \Delta\Delta G^\circ_{iso}$$

Equation 4.8: $\Delta G^\circ_{CT_{trans}} = \Delta G^\circ_{CT_N} + \Delta\Delta G^\circ_{trans}$

Equation 4.9: $\frac{d[CT_{trans}]}{dt} = -k_{E-f}[CT_{trans}][W] + k_{E-r}[CW][T_{trans}] - k_{iso}[CT_{trans}]$

Equation 4.10: $\frac{d[CT_{cis}]}{dt} = -k_{Z-f}[CT_{cis}][W] + k_{E-r}[CW][T_{cis}] + k_{iso}[CT_{trans}]$

boundary condition $[CT]_{total} = [CT_{trans}] + [CT_{trans}] + [CW]$

Equation 4.11: $\frac{dF}{dt} = \frac{d[CW]}{dt} = k_{E-f}[CT_{trans}][W] + k_{Z-f}[CT_{cis}][W]$
$$-k_{E-r}[CW][T_{trans}] - k_{E-r}[CW][T_{cis}]$$

Equation 4.12: $\frac{d[CT_{trans}]}{dt} = -k_{E-app}[CT_{trans}][W] - k_{iso-app}$

Equation 4.13: $\frac{d[CT_{cis}]}{dt} = -k_{Z-app}[CT_{cis}][W] + k_{iso-app}$

Equation 4.14: $\frac{d[CW]}{dt} = -\frac{d[CT_{trans}]}{dt} - \frac{d[CT_{cis}]}{dt}$
$$= -k_{E-app}[CT_{trans}][W] - k_{Z-app}[CT_{cis}][W]$$

Equation 4.15: $\eta = \frac{F_{CT}-F_{BG}}{F_{e.q.}-F_{BG}} \times K_{e.q.}$

List of Abbreviations

DNA	deoxynucleotide acid
RNA	ribonucleic acid
ssDNA	single-stranded DNA
dsDNA	double-stranded DNA
ctDNA	circulating tumor DNA
SNVs	single-nucleotide-variants
CRNs	chemical reactions networks
PCR	polymerase chain reaction
DF	discrimination factor
DMT	dimethoxytrityl
PAGE	polyacrylamide gel electrophoresis
HPLC	high-performance liquid chromatography
UV	ultra-violet
ITC	Isothermal Titration Calorimetry
DSC	Differential Scanning Calorimetry
T_m	melting temperature
1D	one-dimensional
3D	three-dimensional
AuNP	gold nanoparticle
HBV	*hepatitis B virus*
CHA	catalytic hairpin assembly
3DDN	three-dimensional DNA nanomachine
AE	activation efficiency
WT	wild-type
DF	discrimination factor
RF	robustness factor
DEPs	DNA equalizer probes
DEG	DNA equalizer gate
RM	reaction matrix
STH	soil-transmitted helminth
TT	*trichuris trichiura*
D.R.	drug resistance
PM10	1×PBS buffer with 10 mM $MgCl_2$
TE	1×Tris-EDTA buffer
SDR	strand displacement reaction
TER	toehold-exchange reaction

Chapter 1

Introduction of dynamic DNA nanotechnology and its applications for discriminating SNVs

1.1 General introduction to DNA nanotechnology

With the developments in molecular biology, chemical synthesis, and computer science, a multi-discipline field, DNA nanotechnology, has been growing rapidly. Rather than employing nucleic acids as biological engineering materials, DNA nanotechnology utilizes artificial nucleic acids as building blocks to construct various dynamic reaction networks,[1–4] computation and logic circuits,[5–8] and nanostructures[9–12]. This field can be roughly divided into dynamic and structural DNA nanotechnology. Dynamic DNA nanotechnology uses DNA-based reactions/reaction networks to achieve desired probe or computation architectures; in structural DNA nanotechnology, DNA strands are regarded as 'LEGO' building bricks to construct highly ordered self-assemble nanostructures. In this thesis, we are focusing on the studies of dynamic DNA nanotechnology and its specific application in detection of single-nucleotide-variants (SNVs) in target sequences.

1.1.1 Dynamic DNA nanotechnology

The fundamental building blocks in dynamic DNA nanotechnology are strand displacement reactions. The toehold-mediated strand displacement was first introduced by Yurke et al. (Fig. 1.1a).[13] The term 'toehold' short single-stranded overhangs (Fig. 1.1a,

green sequence domains) that can initiate the hybridization between a single-stranded input and a duplex substrate (Fig. 1.1a, Input A and Duplex X). thereafter, the input strand displaces out the original complementary strand in a 'random walking' manner, called branch migration. As more base pairs are formed, this process is thermodynamically favored. Compared to the toehold binding step, branch migration is relatively much faster. Depending on the length and GC content of toehold region, the overall displacement rate is around the scale of 1×10^6 M^{-1} s^{-1}.[5] It should be pointed out that when mismatches exist between the input and duplex, the kinetics might be significantly slowed.[14]

Inspired by the toehold-mediated strand displacement, alternative types of strand displacement reactions including remote toehold[15] and allosteric toehold[16] designs have been developed. In the remote toehold design, a spacer domain separates the toehold and branch migration domains (Fig. 1.1b). And in the allosteric toehold, the displacement reaction between input and duplex cannot occur in an indirect way. A regulator short strand opens the protected secondary toehold (Fig. 1.1c, orange domain) first, then the input reacts with the three-stranded complex to generate two duplex products (Fig. 1.1c).

Of all toehold designs, toehold-exchange is one of the most powerful and widely used strategies for designing assays and probes that discriminate SNVs (Fig. 1.1d). This reaction was first introduced by Zhang et al.,[17] and it contains two toehold regions which makes reaction reversible. By simply tuning the length of the forward and reverse toehold, one can control the thermodynamics of the reaction with high precision. Moreover, the toehold-exchange principle also effectively decouples the thermodynamics and kinetics of the strand displacement, allowing the reaction to progress rapidly regardless of the thermodynamics.

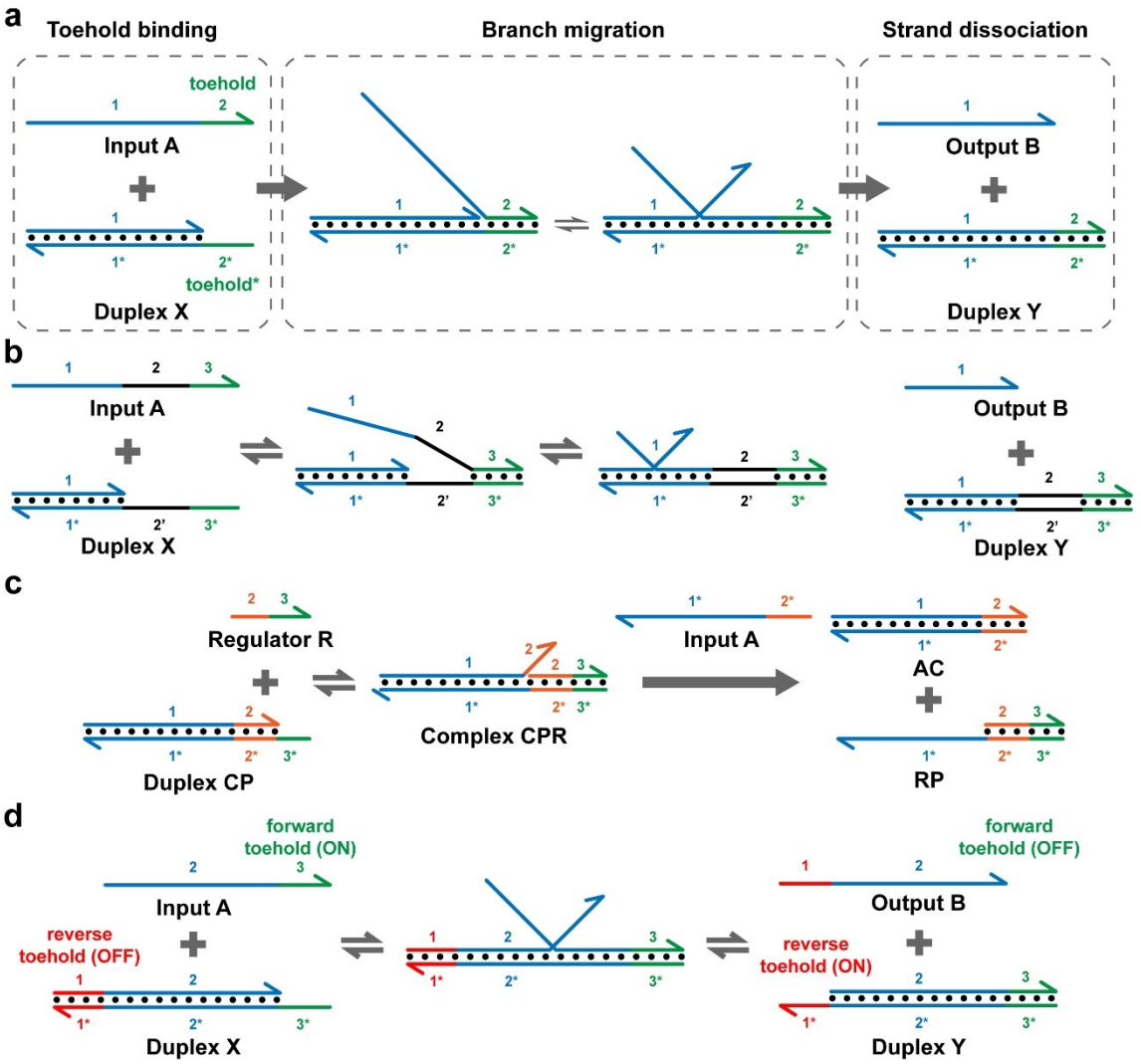


Figure 1.1. (**a**) The Schematic of the toehold-mediated strand displacement reaction. Input A hybridizes with the complementary toehold of complex X to initiate branching migration, and then A displaces B to form a new complex Y. (**b**) Schematic of remote toehold-mediated strand displacement. The toehold and displacement domains on both the duplex substrate and a single-stranded input strand are separated by spacer domains (black domain). Bounding of the duplex substrate and the input strand by hybridization of the toehold domains is followed by an internal diffusion step, initiating the branch migration reaction by which the short strand of the substrate is displaced. (**c**) Schematic of allosteric toehold-mediated strand displacement. A short strand R first reacts with CP to form a reaction intermediate CPR, and the invading region of R exposes a short segment of C that serves as a secondary toehold to drive the strand-exchange between A and CPR to form AC and RP. (**d**) Schematic of the toehold exchange reaction. Hybridization of the duplex X to the input A strand is initiated at the 5′ toehold of X, proceeds through a branch migration process, and is completed via the spontaneous dissociation of the 3′ base pairs of the duplex to release original complementary strand. The toeholds allow the forward and reverse reactions to proceed with fast kinetics.

As the Watson-Crick base pairing is highly predictable and programmable, toehold-mediated DNA strand displacement reactions can be easily simulated *in silico*. Currently, the free software NUPACK[18] and oxDNA[19] can accurately predict the thermodynamic properties of DNA strands. As for the simulation of kinetic process, the strand displacement reactions are always simplified as typical bimolecular reactions. Free software Visual DSD[20] pioneers the work of implementing complex dynamic architectures in a programming-like language. Besides, custom and more specific simulation frames can be accomplished in MATLAB, Python, etc.

With the high similarity between DNA-based reactions and electronic circuits, dynamic DNA nanotechnologies have been employed to construct catalytic amplifiers,[3] arbitrary chemical reactions networks (CRNs),[1,2] complex logic circuits,[5,8] and synthetic biology analogues[4]. Several representative works (mainly by Winfree and Qian's groups) will be discussed here.

In 2006, Winfree et al.[5] designed an enzyme-free nucleic acids logic circuit to mimic the electronic ones (Fig. 1.2). For a basic AND gate, two inputs are required to be present simultaneously to produce the output signal. Reflected in the DNA architecture, the first input G_{in} reacts with the three-stranded complex GFE_{out} to generate an intermediate duplex FE_{out} that possesses an open toehold region to the second input F_{in}. The final output E_{out} is then produced by a strand displacement reaction. By labelling the termini of strands E and F with fluorophore and quencher, the reaction process can be monitored in real time. Only when two input strands G_{in} and F_{in} are added together, the fluorescence representing the output strand E can be generated. All together, this DNA-architecture resembles the

AND logic gate. The NOT gate, thresholding, and series circuits can also be constructed to achieve complex designs.

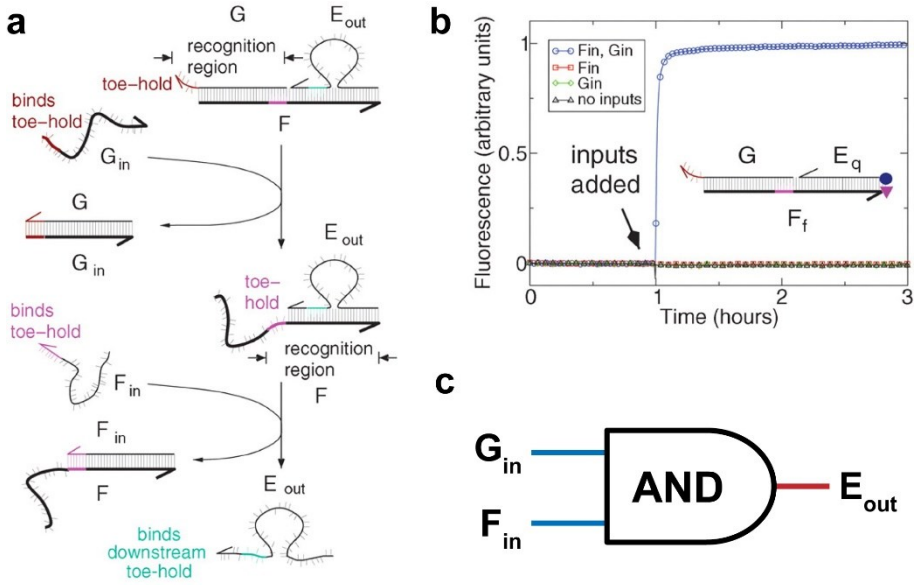


Figure 1.2. The two-input AND gate made of DNA strands. (**a**) The scheme showing the DNA-based reactions within the AND gate. The first input G_{in} reacts with the substrate GFE_{out} by collision of the toehold (red domain). The generated intermediate duplex FE_{out} then reacts with the second input F_{in} to finally produce the output strand E_{out}. (**b**) In fluorescence experiments, carboxytetramethylrhodamine (TAMRA) fluorophore and Iowa Black RQ quencher were labelled at the 3′ and 5′ end of F_f and E_q, respectively. Only when both inputs were present, fluorescence signal can be generated. (**c**) The analog of the DNA circuit to electronical AND circuit.

Signal amplification plays critical roles in biochemical reactions whereas it is difficult to complement for enzyme-free artificial systems. Zhang et al.[3] designed an entropy-driven catalytic circle by displacement reactions in 2007. In this design, the total number of base pairs are constant, but the substrate are consumed to waste (S, W in Fig. 1.3a). In the meanwhile, the fuel strands (F) release the loads (OB and SB, Fig. 1.3a) out in the substrate. Therefore, this process is driven by the entropy gain. Within the cycle, a

short single-stranded DNA initiates the reactions but is restored in the end of circle, thus termed as catalyst strand (C, Fig. 1.3b).

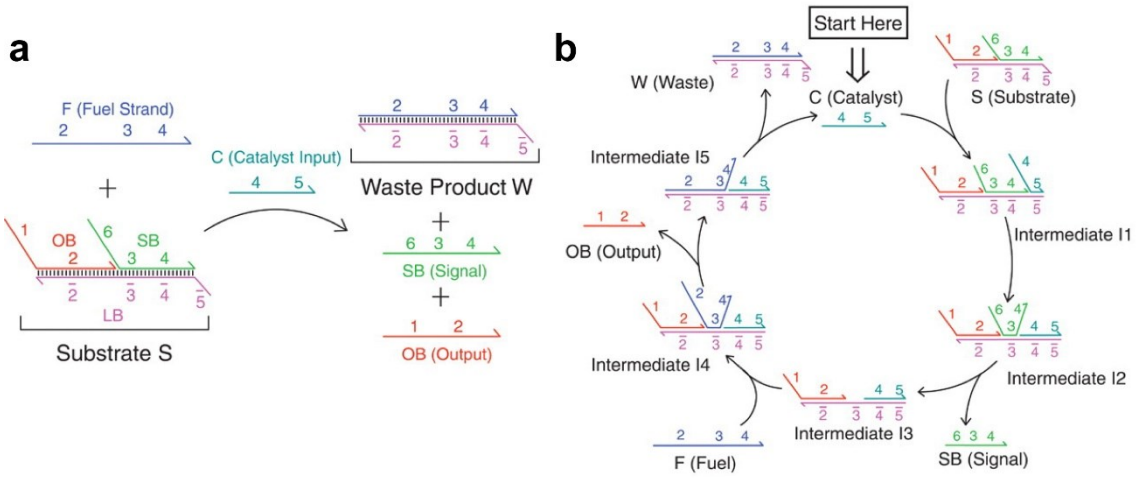


Figure 1.3. The entropy-driven catalytic circle. (**a**) The participating components in the circle. Numerical labels denote the functional domains. (**b**) To start the catalytic circle, the catalyst strand reacts with substrate and generates intermediate I3 and a single-generating strand SB. The fuel strand displaces out the output strand thereafter and the catalyst strand will be restored.

Another showcase of the programmability of dynamic DNA nanotechnology is that DNA can be utilized as a universal substrate for chemical kinetics.[2] That is, an assembly of DNA strand displacement reactions sets can closely approximate the kinetic behavior almost arbitrary coupled chemical reaction networks. Complex kinetic pattens, such as Lotka-Volterra oscillator, chaotic system, and feedback digital logic circuit can be built. Figure 1.4 illustrates the compilation of a mathematically-inspiring whereas chemically not-existing oscillator, the rock-paper-scissors, by DNA reaction modules. It is a good example of the dynamic DNA nanotechnology to fill the gap between mathematics/computer science and the chemistry/biology.

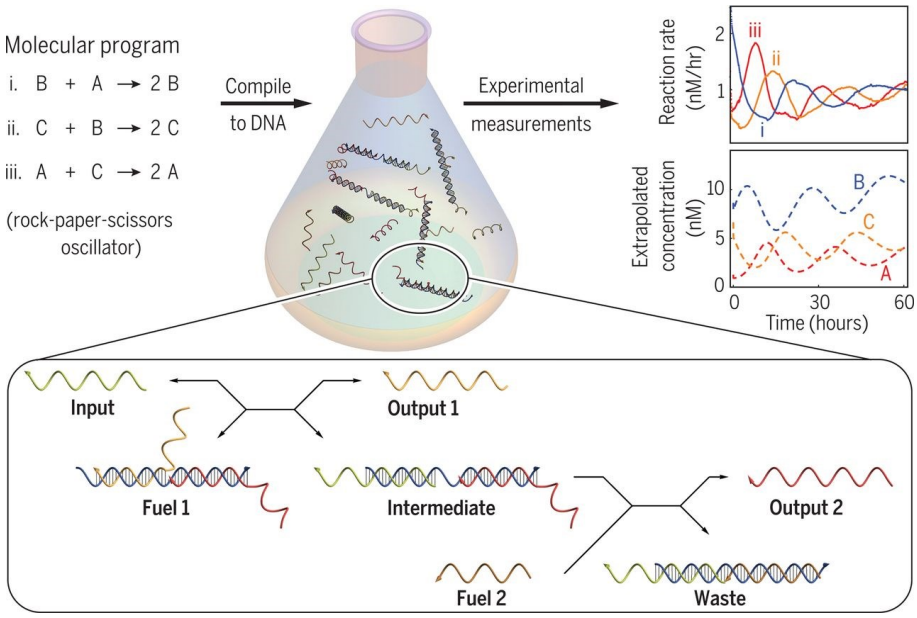


Figure 1.4. Schematic of the mathematical model and corresponding DNA species to implement the rock-paper-scissors oscillation reaction. Oscillations were observed in the concentrations of DNA species that representing formal reactant species (A, B, and C). For a typical reaction module, an input strand initiates a cascade that displaces two output strands from a complex, in the process converting two "fuel molecules" into "waste."

Apart from implementation of chemical reactions, dynamic DNA nanotechnology is also capable of building more advanced biological functions and algorithms. For instance, memory storage is an important function for cells. The synthetic biology techniques can fulfill this function by rewiring the genetic systems in vivo. In contrast, DNA nanotechnology can construct simplified and artificial long-term memory circuits in a more controllable manner. Rondelez et al.[4] constructed switchable circuits of bistable and updatable (push-push) memory patterns. A more fascinating example is the implementation of a winner-take-all neural network with DNA-strand-displacement reactions, by Qian's group.[8] In this design, handwritten digits '1' to '9' are abstracted to 100 bits in 10 × 10 grids, and each grid unit corresponds to a particular DNA molecule. The digits are

characterized by 20 specific DNA molecules. The whole DNA grids set was then trained by picture bank of handwritten digits, through the DNA neural networks. For an unknown digit, the DNA neural network would compute the corresponding unknown DNA grids set with winner-take-all algorithm and predict the most-like digit (Fig. 1.5).

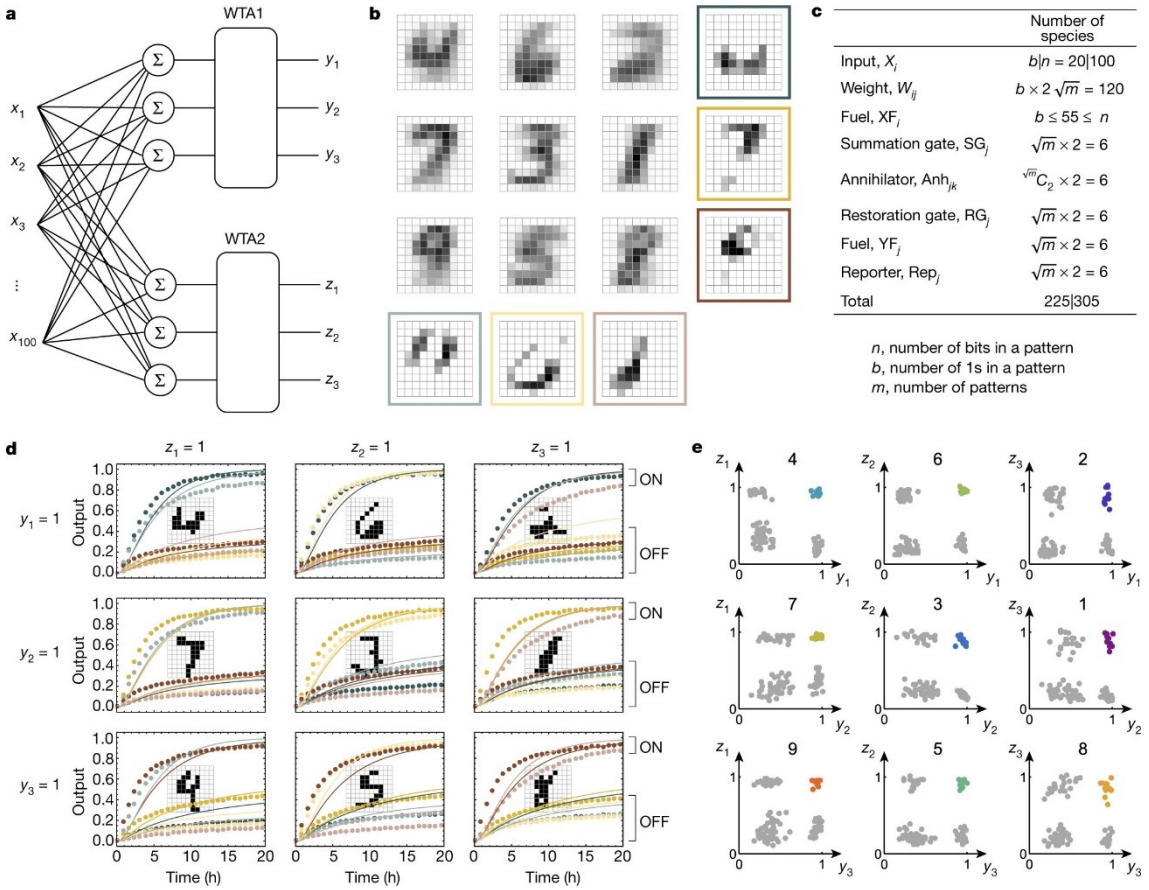


Figure 1.5. (**a**) The circuit diagram for recognizing the '1' to '9' handwritten digits by two winner-take-all functions WTA1 and WTA2. (**b**) The pattern of nine digits are abstracted by two sets of characterization grids that relates to specific DNA strands. (**c**) Number of total DNA species used in this system. (**d**) Fluorescence kinetics data (dotted lines) and simulations (solid lines) of the circuit behaviour with nine representative input patterns (shown in the plots). (**e**) The fluorescence level of each pair of outputs.

1.2 Discrimination of SNVs by DNA hybridization probes

1.2.1 Significance of SNVs detection and traditional methods

A single-nucleotide-variant (SNV) is a genetic variation in a single nucleotide. SNVs occur at a frequency of 1 every 100–300 bases in the human genome, which may have important clinical consequences. Genetic SNV at coding regions are closely related to the causes of many somatic diseases such as cancer, Alzheimer's disease, and many others.[21-24] Therefore, disease-related SNVs found in the blood circulation, such as circulating tumor DNA (ctDNA) and microRNAs, may serve as important disease biomarkers for in vitro diagnosis, and can be used to monitor the development of the disease.[25-29] However, many important disease-related nucleic acids only present in trace levels. For example, the abundance of ctDNA was estimated to be only 1% or even 0.01% of the entire circulating DNA. Discrimination of SNVs in such low abundant nucleic acid markers is often challenged by interference from the large excess of wild-type counterparts. Therefore, techniques capable of discriminating rare SNVs with high sensitivity and specificity in liquid biopsies are highly desired and hold great promise for disease diagnosis and prognosis.

Current techniques for analyzing SNVs rely mainly on polymerase chain reaction (PCR) and nucleic acid sequencing. However, both techniques are challenged by tedious and lengthy operation procedures, the need for expensive infrastructure and special expertise, and high error-rate for discriminating SNVs. The use of nucleic acid hybridization probes, such as molecular beacons and Taqman probes, have helped improve assay speed and accuracy through exquisite Watson–Crick base pairing rules, but are generally difficult to design and operate, and extensive experimental validation and

optimization are required. This is because the thermodynamic difference between a SNV and its wild-type counterpart is only a few kcal mol^{-1}, the discrimination of which requires a delicate balance between hybridization yield and sequence selectivity. Moreover, coexisting nucleic acids with high sequence similarity in liquid biopsies are often at concentrations 100–10,000 times higher than those of target SNVs. As high sequence interference may further comprise the analytical performance of hybridization probes, alternative strategies that are highly programmable and robust are thus urgently needed for the detection of SNVs over high abundance interfering sequences in real clinical samples.

1.2.2 Methods based on dynamic DNA nanotechnology

Facing these challenges, dynamic DNA nanotechnology offers unique solutions for analyzing SNVs with high sensitivity, specificity, programmability and robustness. Specifically, toeholds in DNA strand displacement reactions have remarkable tunability at both thermodynamic and kinetic levels. As such, in silico sequence design becomes possible for generating strand displacement beacons or even more sophisticated strategies capable of isolating or enriching rare SNVs.

To quantitatively indicate the SNV detection performance, the parameters specificity and sensitivity are commonly used. Sensitivity, or detection limit, refers to the lowest analytical concentration detectable by the detection system. Specificity is usually assessed using the DF, indicating the discrimination degree of the method against false targets. The calculation formula of specificity and sensitivity is as follows:[30]

Specificity (DF): $= \frac{\chi_T}{\chi_N} \cong \frac{\frac{[T \cdot Probe]}{[T]_0}}{\frac{[N \cdot Probe]}{[N]_0}}$ $or\ DF = \frac{\chi_T}{\chi_N} \cong \frac{\Delta F_T}{\Delta F_N}$, (**eq. 1.1**)

Sensitivity(yield): $= \frac{[T \cdot Probe]}{[T]_0}$, (**eq. 1.2**)

where T and N represent the target and non-target molecules, respectively. χ is the hybridization yield. $[T]_0$ and $[N]_0$ represent the initial concentration of T and N molecules, respectively. [T · Probe] and [N · Probe] represent the product concentration of T and N molecules reaction, respectively. ΔF is the net increment of the signal.

Ever since the toehold-exchange reaction been developed, it has become the critical tool to discriminate SNVs. The reversible nature of toehold-exchange makes it very sensitive to subtle changes in thermodynamic difference caused by SNVs. Reaction yield and discrimination factors (DFs) can be predicted accurately and tuned by altering the lengths of the forward and reverse toeholds. It is also possible to further improve the specificity for discriminating rare SNVs by introducing competitive "sink" probes.[30] This "sink" strategy takes advantage of both kinetics and thermodynamics to enhance hybridization specificity, making it possible to detect a minute amount of SNVs with a large excess of wild-type interference. Zhang et al.[31] achieved in vitro diagnosis of mutant nucleic acids such as EGFR-L858R by this strategy with a large excess of wild-type interference.

The fundamental basis of discriminating SNVs using toehold-exchange was first established by Zhang and colleagues in 2012 (Fig. 1.6).[31] The standard Gibbs free energy (ΔG°) of a toehold-exchange reaction between an input X and duplex CP was found to be tunable by altering the length of the forward and reverse toehold. When ΔG° is highly negative, this reaction has a high reaction yield but low sequence specificity (DF is close to 1). However, when ΔG° is tuned to be near 0, the reaction yield approaches 50% and the

DF increases dramatically. Hybridization probes making use of toehold exchange were also found to be highly robust, and resistant to changes in assays conditions such as buffer saline and temperature. In addition to toehold length, it is also possible to further fine-tune the toehold-exchange through auxiliary probes such as the protector probe (P) in Fig. 1.6. Because thermodynamic parameters of toehold-exchange probes are readily available, it is possible to achieve the in silico design of such probes without the need for experimental optimization. Such simulation-guided designs and operations of toehold-exchange probes make them highly attractive for practical application in SNV analyses. Zhang et al.[31] detected the let 7 g and other SNV nucleic acids through this design.

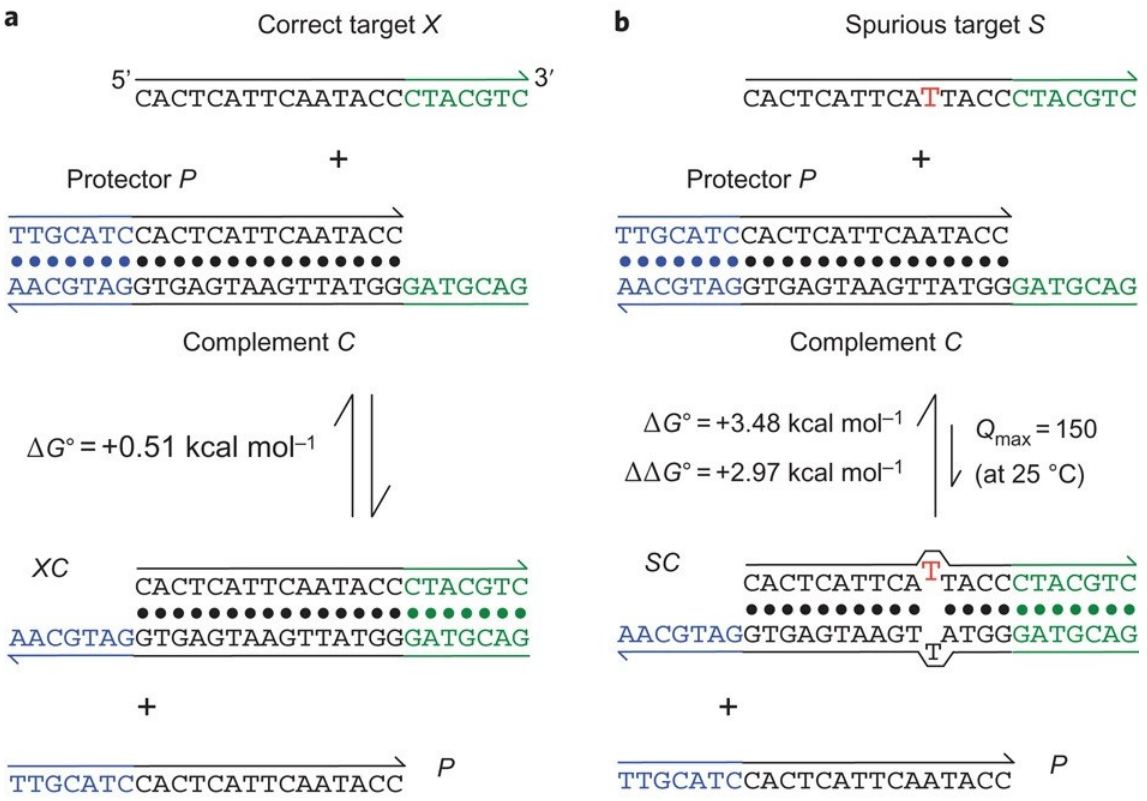


Figure 1.6. Toehold exchange probes for SNV identification. (**a**) The probe reacts with an incorrect target X to release protector strand P. In this process, the standard free energy of the reaction is close to zero. (**b**) Hybridization of a spurious target S with one base change is thermodynamically less favorable by + 2.97 kcal mol^{-1}.

One inherent restriction of typical toehold exchange reaction is that the hybridization probes can only recognize single-stranded targets with mutations sites. Therefore, Zhang et al. further expanded the target to double-stranded nucleic acids through the development of an X-probe (Fig. 1.7).[32] In this design, two different toeholds are created at the same end of each complementary strand in a duplex. Upon binding with the probe, a quadruplex DNA structure will be formed. If the target contained a base pair of SNVs, two mismatch bubbles will be generated in the produced duplexes. Thermodynamically, SNVs in the double-toehold exchange reaction are less favorable than that in a single-toehold counterpart, resulting in the enhancement in specificity.

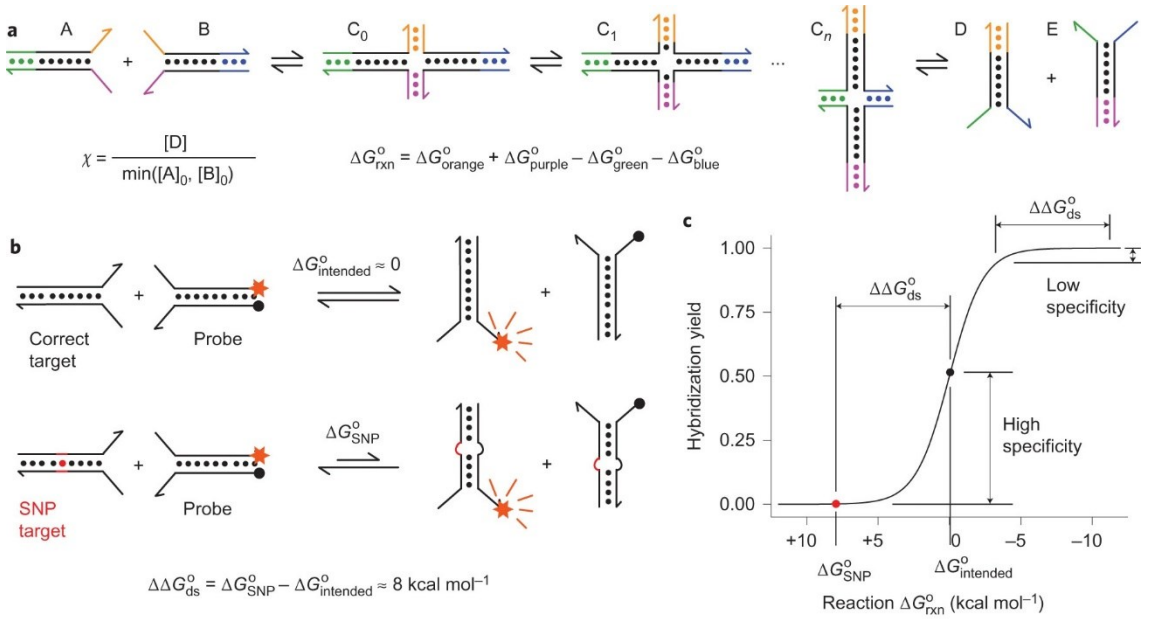


Figure 1.7. Schematic representation of the double-stranded toehold exchange mechanism. (**a**) Reversible branch migration process of double-chain toehold exchange. (**b**) Highly specific variants identification based on double-stranded toehold exchange. The reaction with the intended target has $\Delta G°_{intended}$ is close to zero. The reaction between the probe and the SNP target will result in two mismatch bubbles, and the reaction $\Delta G°_{SNP}$ will be about 8 kcal mol^{-1}. (c) For a specific SNV the difference of reaction free energies is constant. To

optimize the discrimination performance, when the $\Delta G^{\circ}_{intended}$ is around zero the specificity is highest.

Despite excellent performance for discriminating SNVs using toehold-exchange, it remains analytically challenging to detect rare SNVs in the presence of a large excess of somatic nucleic acids. To address the need for discriminating rare mutations, Zhang et al.[30] introduced a "sink" strategy, which dramatically improved the specificity of toehold-exchange probes. Because of the complexity in the reaction network, they built a model to simulate and predict the yield and DF distribution as a function of probe and sink reaction free energy. The basic principle is to reduce the reaction potential between unexpected spurious targets and signal-generating probes by splitting the target-probe hybridization process. The perfect match target-probe pairs have faster kinetics than that with a SNV mismatch. Therefore, unwanted wild-type target will be annihilated by the sink instead of competing the SNV target to generate a signal. Seelig et al.[33] further improve the sensitivity of the Sink design by introducing two fuel strands (Fig. 1.8). Using this strategy, both the sensitivity and specificity can be enhanced. In addition to linear probes, DNA hairpins may also be used as competitive "sink" probes to enhance the sequence specificity.[34] For example, Liu and coworkers[35] combined hairpin competitor and asymmetric PCR amplification with in vitro diagnostics. Eventually, they successfully detected mutant nucleic acids (KRAS-G12D) with only 0.2% abundance.

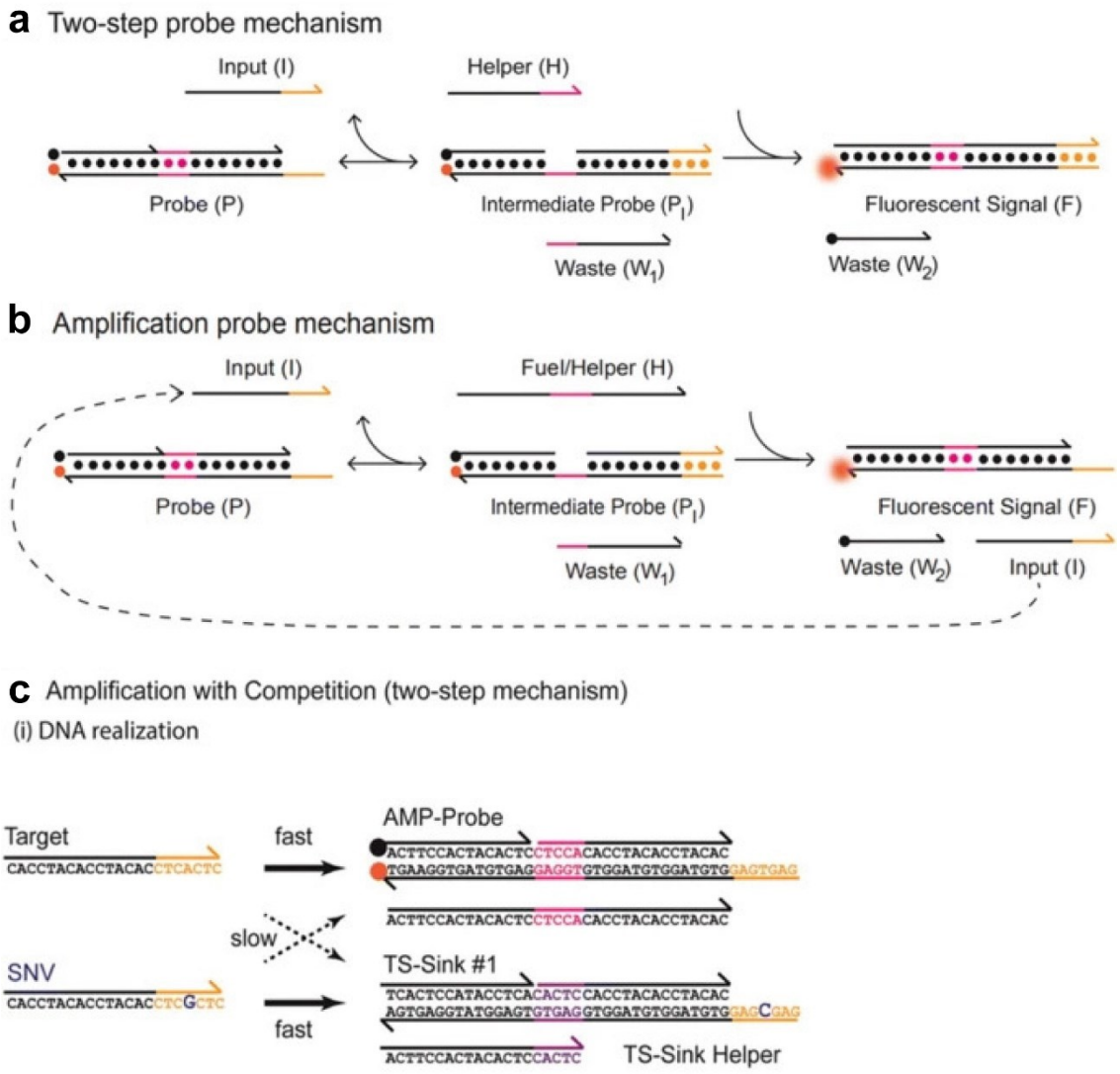


Figure 1.8. Using an engineered two step sink and catalytic probe, SNVs in any position of the input can be reliably identified. (**a**) Two-step reaction mechanism. The first reaction step is a reversible toehold exchange reaction between the input strand (I) and the waste strand (W1). In the second step, an auxiliary helper molecule (H) competes with W1 for binding to the magenta toeholds. Binding of H makes the overall reaction irreversible. (**b**) An amplification probe mechanism. Binding of the input strand is reversible, exactly as for the two step probe. In the second reaction step, the longer helper (or fuel) sequence hybridizes to the bottom strand using the pink toehold and displaces not only the waste strand W2 but also the input irreversibly. After the input is released, it can react with another probe to initiate the next catalytic cycle. (**c**) An amplification circuit with a competitive two-step sink exhibits dramatically increased specificity and sensitivity for SNVs.

Similar to the 'Sink' design, Zhao and Xiao[36] proposed a 'kinetic trap' approach for uniformly sensitive and specific DNA hybridization probes without optimization. For a Standard-probe/Standard-blocker design, the target competes hybridization with the probe against to the blocker strand (Fig. 1.9a). The perfect base-paring between the mutant-type target and probe or the wild-type target and blocker differentiates the thermodynamic difference between correct/spurious targets. However, this standard design needs to be carefully tuned to optimize the specificity. In their strategy, the hairpin probe is replaced by a duplex one (Fig. 1.9b). As such, the target, whichever wild-type or mutant-type, will be kinetically favored to react with the blocker strand first and a 4-way strand exchange reaction leads the signal-generating products (PMT+BS, Fig. 1.9b) back to non-signal state (PS+BMT, Fig. 1.9b). In this '4-way exchange' manner, the specificity is uniformly high for a wide concentration range of blocker. As a demonstration of ultra specificity of this strategy, they successfully detect the *KRAS G13D* and *KRAS G12S* point mutation in low abundance (0.1%).

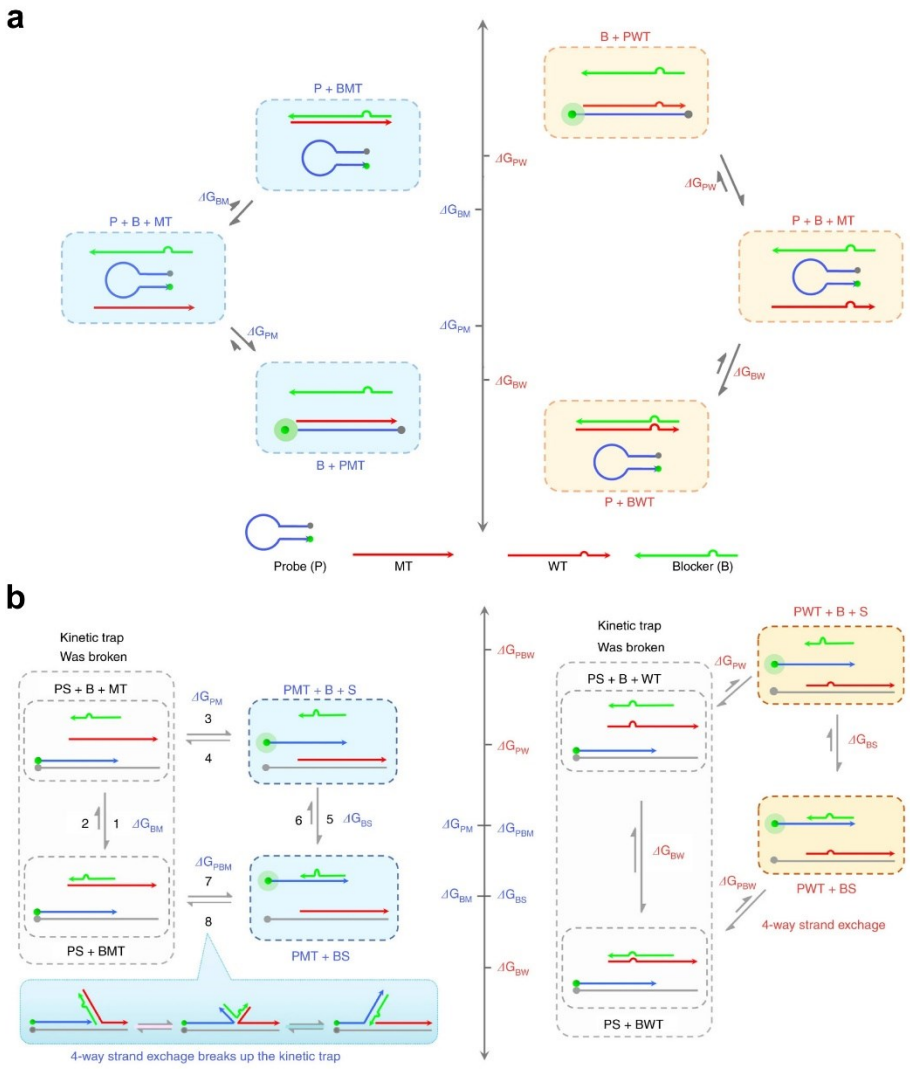


Figure 1.9. (a) The reaction pathways and the levels of associated free energy changes for the Standard probe/standard blocker system. (b) The reaction pathways and the levels of associated free energy changes for the probe/blocker system by using kinetic traps.

1.3 Synthesis and chemical modifications of nucleic acids

The development of DNA nanotechnology is greatly promoted by the chemical synthesis of oligonucleotides. Nowadays, the production of artificial oligonucleotides is based on the phosphoramidite approach (Fig. 1.10). By using the highly optimized synthesis cycles, the coupling yield is more than 99% per cycle and the oligonucleotides can be as long as 200mer.[37] The price per nucleotide has decreased from hundreds of USD to about $0.1.The nucleoside monomer is shown in Fig. 1a, where the 5'-hydroxyl group is protected by dimethoxytrityl (DMT), and a phosphoramidite group on the 3'-hydroxyl group of the deoxyribose sugar. The starting point of a cycle is typically a nucleoside-modified solid support, such as controlled pore glass beads (Fig. 1.10b, blue spheres). DMT in the support-bound nucleoside is removed just before coupling step. A large excess of appropriate nucleoside phosphoramidite is mixed with tetrazole activator to remove the diisopropylamino group. Thereafter, a phosphorus-oxygen bond is formed with the 5'-hydroxyl group of the support-bound nucleoside (Fig. 1.10b, coupling step). For unreacted support-bound nucleoside, the 5'-hydroxyl group is capped by an acetylating reagent thus rendering the unreacted nucleosides inert (Fig. 1.10b, capping step). The formed phosphite-triester is unstable in the following acidic detritylation step and thereby must be converted to a stable phosphotriester form. This step is achieved by the iodine oxidation (Fig. 1.10b, oxidation step). This resultant support-bound oligonucleotide is ready to participate in the next cycle. At the end of synthesis, the newly synthesized oligonucleotide is cleaved from the solid support by ester hydrolysis. The truncated product is then purified by HPLC (40~80 nucleotides) or gel electrophoresis (>80 nucleotides).

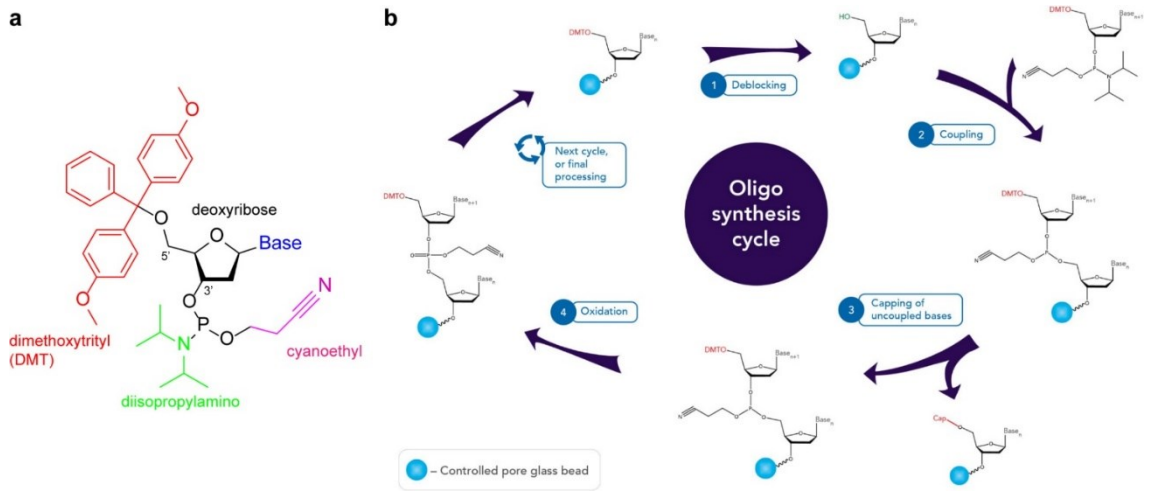


Figure 1.10. Synthesis of oligonucleotide though Phosphoramidite approach. (**a**) A phosphoramidite nucleoside monomer. The reactive 5'-hydroxyl group is usually protected by dimethoxytrityl (DMT); nitrogenous bases, guanine, adenine, cytosine, are protected by *N*(2)-isobutyryl, *N*(6)-benzoyl, *N*(4)-benzoyl, respectively; a diisopropylamino group and a -cyanoethyl group are used to protect the hydroxyl groups on the phosphite. (**b**) The automated oligo synthesis cycle.[28] Activator is 5-(ethylthio)-1H-tetrazole catalyst.

To expand the functionality of canonical nucleic acids, different chemical moieties can be incorporated into the DNA/RNA strands during the phosphoramidite synthesis procedure. For example, introducing a photoswitchable functional group, such as azobenzenes, stilbenes, spiropyrans, can fulfill the modified nucleic acids photoregulation function.[39,40] One of the most widely used photoresponsive chromophores is azobenzene.[41,42] Figure 1.11a shows the azobenzene phosphoramidite monomer synthesized by Komiyama *et al.* in 2007.[43] There exists two isomers of azobenzene: the stable *trans*-isomer (*E*) at room temperature and *cis*-isomer (*Z*) that can be photoswitched by UV light (Fig. 1.11b). The photoisomerization from *trans*- to *cis*-configuration occurs rapidly (femtoseconds) whereas the backward process is relatively slower. Azobenzenes and derivatives have been well studied and incorporated with DNA strands[44,45] to build

light-induced constitutional dynamic network,[46] hydrogel,[47] photo-responsive origami nanostructures,[48] nano/micro-capsules and drug delivery[49,50].

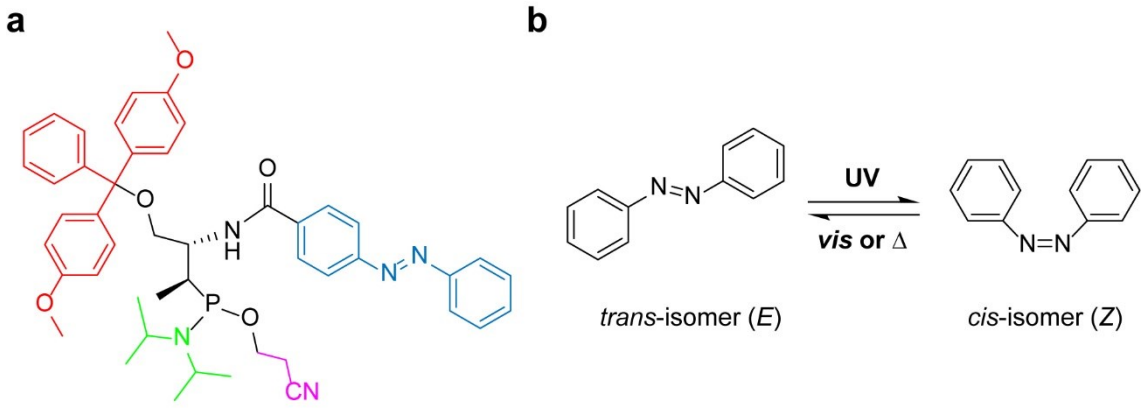


Figure 1.11. (**a**) The azobenzene phosphoramidite monomer. (**b**) The photoisomerization between *trans*-azobenzene(*E*) and *cis*-azobenzene(*Z*).

With diverse modifications have been developed in the chemistry perspective, rare attention has been put on the characterization of the thermodynamic and kinetic properties within the DNA-based reactions. The commonly used methods are heavily dependent on expensive instruments of Isothermal Titration Calorimetry (ITC), Differential Scanning Calorimetry (DSC), and UV melting (T_m). However, those methods can only measure the thermally stable molecules/modifications and the kinetic information cannot be recorded. In this thesis (Chapter 4), we propose a dynamic method to probe both the thermodynamic and kinetic properties of chemically modified DNA-based reactions.

1.4 Objectives and hypothesis

The overall objective is to understand the thermodynamic and kinetic properties of toehold-mediated strand displacement reactions and thereby develop dynamic reaction networks/architectures aiming to discriminate clinically important single-nucleotide-variants (SNVs), and to probe hybridization properties of chemically modified DNA. Hypothesis: combine experimental and *in silico* simulation to develop well-tuned DNA hybridization probes/assays or desired reaction pathways. Three projects are descried here to demonstrate the progress of achieving the overall objective.

The first project was our first theoretical study of the fundamental toehold-mediated strand displacement reaction, especially toehold exchange (Chapter 2). Under the theoretical guidance, we designed a noncovalent DNA catalysis circle and assembled it onto the surface of gold nanoparticles to build a three-dimensional DNA nanomachine (3DDN). Our 3DDN was capable of discriminating SNVs with high sensitivity and specificity.

In the next project, we reinspected the current DNA hybridization probes and summarized the limitations as: 1) most of the probes were targeting single-stranded targets whereas the natural DNA strand preferred more stable double-stranded form; 2) there existed a range of target concentration, which we termed ‘detection window’, where the probes/assays can work properly. To expand the detection window for discriminating SNVs in double-stranded targets, we set out to construct a DNA Equalizer Gate (DEG) that makes the detection event more robust. Complex mathematical models along with simulation were used to guide the experimental design.

After the success of previous two projects, we then realized the fundamental similarity between SNVs and chemical modifications in oligonucleotides. Thus motivated, we intended to profile the thermodynamic and kinetic properties of chemically modified strand displacement reactions. A photoswitchable modification, cyclic-azobenzene (cAB), was chosen as testbed for our hypothesis. A reversible strand displacement DNA mainframe was employed to examine the properties change caused by cAB and a series of short DNA invader strands were used to alter the reaction conditions. Due to the high resemblance with balance scale, we termed the mainframe as DNA balance and invader strands as DNA weights. The thermodynamic changes caused by chemical modification could be weighed by DNA balance.

1.5 Reference

(1) Soloveichik, D.; Seelig, G.; Winfree, E. DNA as a Universal Substrate for Chemical Kinetics. *PNAS* **2010**, *107* (12), 5393–5398.

(2) Srinivas, N.; Parkin, J.; Seelig, G.; Winfree, E.; Soloveichik, D. Enzyme-Free Nucleic Acid Dynamical Systems. *Science* **2017**, *358* (6369).

(3) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science* **2007**, *318* (5853), 1121–1125.

(4) Padirac, A.; Fujii, T.; Rondelez, Y. Bottom-up Construction of in Vitro Switchable Memories. *PNAS* **2012**, *109* (47), E3212–E3220.

(5) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. Enzyme-Free Nucleic Acid Logic Circuits. *Science* **2006**, *314* (5805), 1585–1588.

(6) Qian, L.; Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **2011**, *332* (6034), 1196–1201.

(7) Thubagere, A. J.; Thachuk, C.; Berleant, J.; Johnson, R. F.; Ardelean, D. A.; Cherry, K. M.; Qian, L. Compiler-Aided Systematic Construction of Large-Scale DNA Strand Displacement Circuits Using Unpurified Components. *Nature Communications* **2017**, *8* (1), 14373.

(8) Cherry, K. M.; Qian, L. Scaling up Molecular Pattern Recognition with DNA-Based Winner-Take-All Neural Networks. *Nature* **2018**, *559* (7714), 370–376.

(9) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297–302.

(10) Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M. Self-Assembly of DNA into Nanoscale Three-Dimensional Shapes. *Nature* **2009**, *459* (7245), 414–418.

(11) Liu, X.; Zhang, F.; Jing, X.; Pan, M.; Liu, P.; Li, W.; Zhu, B.; Li, J.; Chen, H.; Wang, L.; Lin, J.; Liu, Y.; Zhao, D.; Yan, H.; Fan, C. Complex Silica Composite Nanomaterials Templated with DNA Origami. *Nature* **2018**, *559* (7715), 593–598.

(12) Aghebat Rafat, A.; Sagredo, S.; Thalhammer, M.; Simmel, F. C. Barcoded DNA Origami Structures for Multiplexed Optimization and Enrichment of DNA-Based Protein-Binding Cavities. *Nature Chemistry* **2020**, *12* (9), 852–859.

(13) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. A DNA-Fuelled Molecular Machine Made of DNA. *Nature* **2000**, *406* (6796), 605–608.

(14) Irmisch, P.; Ouldridge, T. E.; Seidel, R. Modeling DNA-Strand Displacement Reactions in the Presence of Base-Pair Mismatches. *J. Am. Chem. Soc.* **2020**, *142* (26), 11451–11463.

(15) Genot, A. J.; Zhang, D. Y.; Bath, J.; Turberfield, A. J. Remote Toehold: A Mechanism for Flexible Control of DNA Hybridization Kinetics. *J. Am. Chem. Soc.* **2011**, *133* (7), 2177–2182.

(16) Yang, X.; Tang, Y.; Traynor, S. M.; Li, F. Regulation of DNA Strand Displacement Using an Allosteric DNA Toehold. *J. Am. Chem. Soc.* **2016**, *138* (42), 14076–14082.

(17) Zhang, D. Y.; Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **2009**, *131* (47), 17303–17314.

(18) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *Journal of Computational Chemistry* **2011**, *32* (1), 170–173.

(19) Doye, J. P. K.; Ouldridge, T. E.; Louis, A. A.; Romano, F.; Šulc, P.; Matek, C.; Snodin, B. E. K.; Rovigatti, L.; Schreck, J. S.; Harrison, R. M.; Smith, W. P. J. Coarse-Graining DNA for Simulations of DNA Nanotechnology. *Phys. Chem. Chem. Phys.* **2013**, *15* (47), 20395–20414.

(20) Lakin, M. R.; Youssef, S.; Polo, F.; Emmott, S.; Phillips, A. Visual DSD: A Design and Analysis Tool for DNA Strand Displacement Systems. *Bioinformatics* **2011**, *27* (22), 3211–3213.

(21) Xu, Q.; Huang, S.; Ma, F.; Tang, B.; Zhang, C. Controllable Mismatched Ligation for Bioluminescence Screening of Known and Unknown Mutations. *Anal. Chem.* **2016**, *88* (4), 2431–2439.

(22) Gilissen, C.; Hehir-Kwa, J. Y.; Thung, D. T.; van de Vorst, M.; van Bon, B. W. M.; Willemsen, M. H.; Kwint, M.; Janssen, I. M.; Hoischen, A.; Schenck, A.; Leach, R.; Klein, R.; Tearle, R.; Bo, T.; Pfundt, R.; Yntema, H. G.; de Vries, B. B. A.; Kleefstra, T.; Brunner, H. G.; Vissers, L. E. L. M.; Veltman, J. A. Genome Sequencing Identifies Major Causes of Severe Intellectual Disability. *Nature* **2014**, *511* (7509), 344–347.

(23) Poznik, G. D.; Xue, Y.; Mendez, F. L.; Willems, T. F.; Massaia, A.; Wilson Sayres, M. A.; Ayub, Q.; McCarthy, S. A.; Narechania, A.; Kashin, S.; Chen, Y.; Banerjee, R.; Rodriguez-Flores, J. L.; Cerezo, M.; Shao, H.; Gymrek, M.; Malhotra, A.; Louzada, S.; Desalle, R.; Ritchie, G. R. S.; Cerveira, E.; Fitzgerald, T. W.; Garrison, E.; Marcketta, A.; Mittelman, D.; Romanovitch, M.; Zhang, C.; Zheng-Bradley, X.; Abecasis, G. R.; McCarroll, S. A.; Flicek, P.; Underhill, P. A.; Coin, L.; Zerbino, D. R.; Yang, F.; Lee, C.; Clarke, L.; Auton, A.; Erlich, Y.; Handsaker, R. E.; Bustamante, C. D.; Tyler-Smith, C. Punctuated Bursts in Human Male Demography Inferred from 1,244 Worldwide Y-Chromosome Sequences. *Nat. Genet.* **2016**, *48* (6), 593–599.

(24) Khurana, E.; Fu, Y.; Chakravarty, D.; Demichelis, F.; Rubin, M. A.; Gerstein, M. Role of Non-Coding Sequence Variants in Cancer. *Nat. Rev. Genet.* **2016**, *17* (2), 93–108.

(25) Diehl, F.; Schmidt, K.; Choti, M. A.; Romans, K.; Goodman, S.; Li, M.; Thornton, K.; Agrawal, N.; Sokoll, L.; Szabo, S. A.; Kinzler, K. W.; Vogelstein, B.; Diaz Jr, L. A. Circulating Mutant DNA to Assess Tumor Dynamics. *Nat. Med.* **2008**, *14* (9), 985–990.

(26) Bettegowda, C.; Sausen, M.; Leary, R. J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B. R.; Wang, H.; Luber, B.; Alani, R. M.; Antonarakis, E. S.; Azad, N. S.; Bardelli, A.; Brem, H.; Cameron, J. L.; Lee, C. C.; Fecher, L. A.; Gallia, G. L.; Gibbs, P.; Le, D.; Giuntoli, R. L.; Goggins, M.; Hogarty, M. D.; Holdhoff, M.; Hong, S.-M.; Jiao, Y.; Juhl, H. H.; Kim, J. J.; Siravegna, G.; Laheru, D. A.; Lauricella, C.; Lim, M.; Lipson, E. J.; Marie, S. K. N.; Netto, G. J.; Oliner, K. S.; Olivi, A.; Olsson, L.; Riggins, G. J.; Sartore-Bianchi, A.; Schmidt, K.; Shih, le-M.; Oba-Shinjo, S. M.; Siena, S.; Theodorescu, D.; Tie, J.; Harkins, T. T.; Veronese, S.; Wang, T.-L.; Weingart, J. D.; Wolfgang, C. L.; Wood, L. D.; Xing, D.; Hruban, R. H.; Wu, J.; Allen, P. J.; Schmidt, C. M.; Choti, M. A.; Velculescu, V. E.; Kinzler, K. W.; Vogelstein, B.; Papadopoulos, N.; Diaz, L. A. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **2014**, *6* (224), 224ra24-224ra24.

(27) Lu, J.; Getz, G.; Miska, E. A.; Alvarez-Saavedra, E.; Lamb, J.; Peck, D.; Sweet-Cordero, A.; Ebert, B. L.; Mak, R. H.; Ferrando, A. A.; Downing, J. R.; Jacks, T.; Horvitz, H. R.; Golub, T. R. MicroRNA Expression Profiles Classify Human Cancers. *Nature* **2005**, *435* (7043), 834–838.

(28) Calin, G. A.; Croce, C. M. MicroRNA Signatures in Human Cancers. *Nat. Rev. Cancer* **2006**, *6* (11), 857–866.

(29) Esquela-Kerscher, A.; Slack, F. J. Oncomirs — MicroRNAs with a Role in Cancer. *Nat. Rev. Cancer* **2006**, *6* (4), 259–269.

(30) Wang, J. S.; Zhang, D. Y. Simulation-Guided DNA Probe Design for Consistently Ultraspecific Hybridization. *Nature Chemistry* **2015**, *7* (7), 545–553.

(31) Zhang, D. Y.; Chen, S. X.; Yin, P. Optimizing the Specificity of Nucleic Acid Hybridization. *Nature Chemistry* **2012**, *4* (3), 208–214.

(32) Chen, S. X.; Zhang, D. Y.; Seelig, G. Conditionally Fluorescent Molecular Probes for Detecting Single Base Changes in Double-Stranded DNA. *Nature Chemistry* **2013**, *5* (9), 782–789.

(33) Chen, S. X.; Seelig, G. An Engineered Kinetic Amplification Mechanism for Single Nucleotide Variant Discrimination by DNA Hybridization Probes. *J. Am. Chem. Soc.* **2016**, *138* (15), 5076–5086.

(34) Hu, S.; Tang, W.; Zhao, Y.; Li, N.; Liu, F. Ultra-Specific Discrimination of Single-Nucleotide Mutations Using Sequestration-Assisted Molecular Beacons. *Chem. Sci.* **2017**, *8* (2), 1021–1026.

(35) Hu, S.; Li, N.; Liu, F. Combining Cooperativity with Sequestration: A Novel Strategy for Discrimination of Single Nucleotide Variants. *Chem. Commun.* **2018**, *54* (26), 3223–3226.

(36) Chen, X.; Liu, N.; Liu, L.; Chen, W.; Chen, N.; Lin, M.; Xu, J.; Zhou, X.; Wang, H.; Zhao, M.; Xiao, X. Thermodynamics and Kinetics Guided Probe Design for Uniformly Sensitive and Specific DNA Hybridization without Optimization. *Nature Communications* **2019**, *10* (1), 4675.

(37) [15] Chemical Synthesis of Deoxyoligonucleotides by the Phosphoramidite Method. *Methods in Enzymology* **1987**, *154*, 287–313.

(38) Oligo synthesis: Coupling efficiency and quality control | IDT https://www.idtdna.com/pages/education/decoded/article/oligo-synthesis-why-idt-leads-the-oligo-industry (accessed Feb 15, 2021).

(39) Szymański, W.; Beierle, J. M.; Kistemaker, H. A. V.; Velema, W. A.; Feringa, B. L. Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. *Chem. Rev.* **2013**, *113* (8), 6114–6178.

(40) Dohno, C.; Nakatani, K. Control of DNA Hybridization by Photoswitchable Molecular Glue. *Chem. Soc. Rev.* **2011**, *40* (12), 5718–5729.

(41) Deiana, M.; Pokladek, Z.; Olesiak-Banska, J.; Młynarz, P.; Samoc, M.; Matczyszyn, K. Photochromic Switching of the DNA Helicity Induced by Azobenzene Derivatives. *Scientific Reports* **2016**, *6* (1), 28605.

(42) Cabré, G.; Garrido-Charles, A.; Moreno, M.; Bosch, M.; Porta-de-la-Riva, M.; Krieg, M.; Gascón-Moya, M.; Camarero, N.; Gelabert, R.; Lluch, J. M.; Busqué, F.; Hernando, J.; Gorostiza, P.; Alibés, R. Rationally Designed Azobenzene Photoswitches for Efficient Two-Photon Neuronal Excitation. *Nature Communications* **2019**, *10* (1), 907.

(43) Asanuma, H.; Liang, X.; Nishioka, H.; Matsunaga, D.; Liu, M.; Komiyama, M. Synthesis of Azobenzene-Tethered DNA for Reversible Photo-Regulation of DNA Functions: Hybridization and Transcription. *Nature Protocols* **2007**, *2* (1), 203–212.

(44) Yan, Y.; Wang, X.; Chen, J. I. L.; Ginger, D. S. Photoisomerization Quantum Yield of Azobenzene-Modified DNA Depends on Local Sequence. *J. Am. Chem. Soc.* **2013**, *135* (22), 8382–8387.

(45) Zhang, H.; Fu, H.; Shao, X.; Chipot, C.; Monari, A.; Dehez, F.; Cai, W. Conformational Changes of DNA Induced by a Trans-Azobenzene Derivative via Non-Covalent Interactions. *Phys. Chem. Chem. Phys.* **2018**, *20* (35), 22645–22651.

(46) Wang, S.; Yue, L.; Li, Z.-Y.; Zhang, J.; Tian, H.; Willner, I. Light-Induced Reversible Reconfiguration of DNA-Based Constitutional Dynamic Networks: Application to Switchable Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (27), 8105–8109.

(47) Liu, X.; Zhang, J.; Fadeev, M.; Li, Z.; Wulf, V.; Tian, H.; Willner, I. Chemical and Photochemical DNA "Gears" Reversibly Control Stiffness, Shape-Memory, Self-Healing and Controlled Release Properties of Polyacrylamide Hydrogels. *Chem. Sci.* **2019**, *10* (4), 1008–1016.

(48) Kuzyk, A.; Yang, Y.; Duan, X.; Stoll, S.; Govorov, A. O.; Sugiyama, H.; Endo, M.; Liu, N. A Light-Driven Three-Dimensional Plasmonic Nanosystem That Translates Molecular Motion into Reversible Chiroptical Function. *Nature Communications* **2016**, *7* (1), 10591.

(49) Liao, W.-C.; Willner, I. Synthesis and Applications of Stimuli-Responsive DNA-Based Nano- and Micro-Sized Capsules. *Advanced Functional Materials* **2017**, *27* (41), 1702732.

(50) Zhang, Y.; Zhang, Y.; Song, G.; He, Y.; Zhang, X.; Liu, Y.; Ju, H. A DNA–Azobenzene Nanopump Fueled by Upconversion Luminescence for Controllable Intracellular Drug Release. *Angewandte Chemie International Edition* **2019**, *58* (50), 18207–18211.

Chapter 2

Simulation-guided engineering of an enzyme-powered three dimensional DNA nanomachine for discriminating single nucleotide variants

Contribution statement

The contents of Chapter 2 were modified from published paper "Simulation-guided engineering of an enzyme-powered three dimensional DNA nanomachine for discriminating single nucleotide variants". Reprinted with permission from Li, Y.; Wang, T. A.; Mason, S. D.; Yang, X.; Yu, Z.; Tang, Y.; Li, F., Simulation-guided engineering of an enzyme-powered three dimensional DNA nanomachine for discriminating single nucleotide variants. *Chem. Sci.* **2018**, 9, 6434-6439. Copyright 2018 Royal Society of Chemistry.

Yongya Li and I contributed equally on this paper. Li did all the experimental works while I built the mathematical models; we both participated into the experimental design and data analysis.

2.1 Introduction

Nature has evolved highly complicated and hierarchical machines that perform diverse biological functions in living systems with remarkably high efficiency and precision. Inspired by nature, various artificial molecular machines have also been created with DNA as a primary choice of building blocks.[1,2] In particular, DNA-based walking devices that convert chemical energy to mechanical motions hold great promises for smart drug delivery, biocomputing, and diagnostics.[3–22] Despite the extraordinary nanometer precision in transporting payload along well-designed one-dimensional (1D),[3–5] two-dimensional (2D),[6–11] or three-dimensional (3D) tracks,[12] the real-world applicability of current DNA walking devices is challenged by the slow kinetics and low processivity.[13,14] Recent effort in designing stochastic DNA walkers that traverse on 3D tracks made of micro- or nanoparticles has greatly enhanced the walking speed and processivity and thus promoted their uses in biosensing and biodiagnostic applications.[13–22] For example, Ellington group introduced a series of 3D stochastic DNA walkers that could take more than 30 continuous steps by integrating catalytic hairpin assembly (CHA) with microparticles.[13,14] Our group developed a nicking endonuclease-powered stochastic 3D DNA walker that moves rapidly on a 20 nm gold nanoparticle (AuNP).[15] A few other nanoparticle-based high-processive stochastic 3D DNA walkers that were propelled by DNAzymes,[16–18] nucleases,[19–21] and enzyme-free DNA catalysis,[22] have also been recently created and found unique applications in amplified biosensing and imaging. Despite the recent advances for biomolecular analysis,[13–22] assays making use of DNA walking devices often involve iterative empirical sequence-based optimization to achieve the desired analytical performance. So far, there is no simulation tool available to guide the design and

operation of DNA walking devices for bioanalytical applications. Herein, we introduce a simulation-guided strategy that programs the sensitivity and specificity of a 3D DNA walking device for discriminating single nucleotide variants (SNVs).

SNVs are important biomarkers both clinically and biologically, as single base differences in nucleic acid sequences can lead to profound biological and clinical consequences.[23] The discrimination of a single nucleotide mismatch in a given nucleic acid sequence can be technically very challenging. Although the hybridization with a fully complementary sequence is energetically more favourable than a SNV, the thermodynamic gain of many correctly paired bases can easily override the thermodynamic penalty of a single mismatch.[24,25] Similarly, the high processivity of most stochastic 3D DNA walking devices relies on the high energy input through newly paired bases or enzymatic cleavage, and thus are generally not favourable systems to discriminate SNVs.[15] To ensure the sequence specificity, the hybridization has to be performed at or near the melting temperature, where the free energy of the reaction is close to zero.[24,25] However, commonly used tuning strategies including high assay temperatures and chemical denaturation are not practically feasible and difficult to be predicted for complexed dynamic DNA systems, such as DNA walkers. Here, we aim to address this challenge by engineering a 3D DNA nanomachine (3DDN) with frustrated complementary probes that are operated by the toehold-exchange. Such probes hybridize less favourably to their intended targets than the standard hybridization probes and thus can enhance sequence specificity at the cost of the hybridization yield.[24–27] Therefore, guided by our simulation tool, we perform an *in silico* sequence design and optimization to balance the device sensitivity and specificity.

Facilitated by the simulation, we also develop a new in-solution tuning strategy that enhances both sensitivity and specificity of 3DDN through non-covalent DNA catalysis.

2.2 Result and Discussion

2.2.1 Device design and theoretical considerations

As shown in Figure 2.1A, our enzyme-powered 3D DNA nanomachine (3DDN) is engineered by co-conjugating ~20 DNA walkers (D) carrying a nicking recognition site and ~400 fluorescently labelled signal reporters (R) carrying a nicking cleavage site on a single 20 nm AuNP. In the presence of nicking endonuclease, D rapidly walks along the surface of AuNP and cleaves R to generate amplified fluorescence signals. To be used for nucleic acid sensing, D is first deactivated using a protecting DNA (P) through hybridization. P deactivates the 3DDN by partially sequestering the nicking recognition sequence on D, and thus D is unable to hybridize with R (Fig. 2.1B). The target (T) activates the DNA walker by freeing D from DP duplex through a toehold exchange reaction. The reaction can be written as DP + T ↔ TP + D and has a reaction standard free energy ΔG°_{rxn}. DP with more negative ΔG°_{rxn} value binds with T with higher hybridization yield, but it also spuriously binds SNVs. Conversely, DP with less negative ΔG°_{rxn} value binds T with low yield but high selectivity. So far, there are generally two ways to tune the thermodynamics of a toehold-exchange reaction to achieve a reasonable trade-off between hybridization yield (assay sensitivity) and sequence selectivity (assay specificity), including (1) the discrete sequence-level tuning through the length of toehold domains,[24,25] and (2) the continuous in-solution tuning by controlling the stoichiometry between the probe strand and the protector strand.[26] The tuning of both parameters can be performed *in silico*, which has been successfully demonstrated in the design of ultraspecific conditional molecular probes.[24–27] We reason that such powerful simulation tools can also guide the design and applications of more complexed, multi-component, multi-step DNA machinery

systems. Herein, we perform a simulation-guided engineering of the 3DDN for discriminating SNVs. We also introduce a new auxiliary probe, a molecular fuel (F) that tunes both assay sensitivity and specificity in-solution via noncovalent DNA catalysis (Fig. 2.1C).

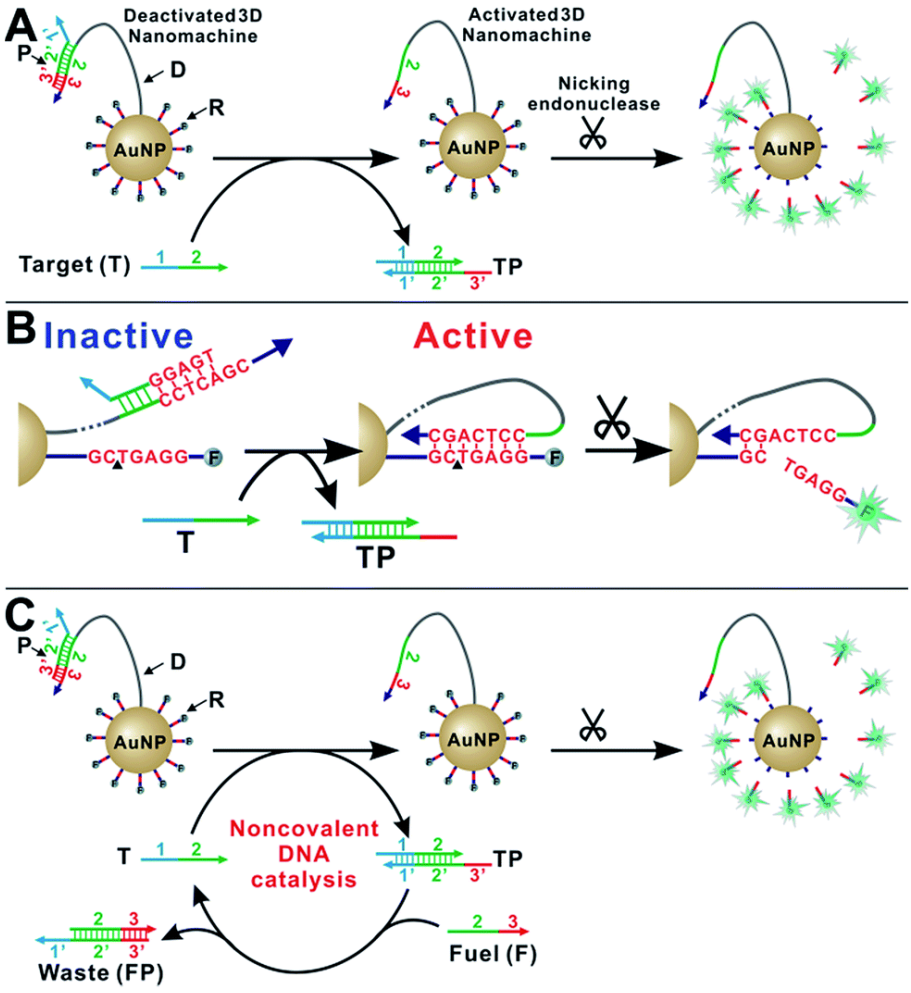


Figure 2.1. (**A**) Schematic illustration of the enzyme-powered 3D DNA nanomachine (3DDN). (**B**) Mechanisms of deactivation of 3DDN by sequestering D with P and activation by T through toehold exchange. (**C**) Schematic illustration of the in-solution tuning of 3DDN via noncovalent DNA catalysis.

2.2.2 Simulation-guided sequence-level tuning of 3DDN for discriminating SNVs

As a proof-of-principle, we designed a 3DDN responsive to a subsequence of the *Mycobacterium tuberculosis rpoB* gene (Fig. 2.2). The sequence design and optimization was guided by simulation, where reaction yield (Fig. 2.2C) or sequence selectivity (Fig. 2.2D) was plotted as a function of ΔG°_{rxn}. ΔG° values for all sequences were determined using NUPACK under a condition (37 °C, 100 mM Na^{+}, 10 mM Mg^{2+}) nearly identical to the actual experiments (Table 2.3). The reaction yield equals to the ratio between [TP] and $[T]_0$, theoretical values of which were estimated using MATLAB (details in Section 2.2.3), where $[T]_0$ was fixed at 1 nM (Fig. 2.2C and Fig. 2.3). The activation efficiency (AE), defined as the ratio between target-activated D over the total, can be further estimated theoretically by plotting $[D]/[DP]_0$ as a function of $[T]_0$ and ΔG°_{rxn} (Fig. 2.4). $[DP]_0$ was estimated to be 2 nM for all experiments, as 100 pM devices were used for all experiments and ~20 DP per device was estimated. The sequence selectivity was determined as the ratio between the reaction yield of a wild type (WT) target and that of a single nucleotide variant at a given ΔG°_{rxn}. By comparing WT with SNV19A and 8 other SNVs (Table 2.5), we determined that the optimal lengths of forward (f) and reverse toeholds (r) were both 10 nt (Fig. 2.2 and Fig. 2.4). Using this sequence design, the reaction yield is ~50% for the WT and less than 10% for all SNVs.

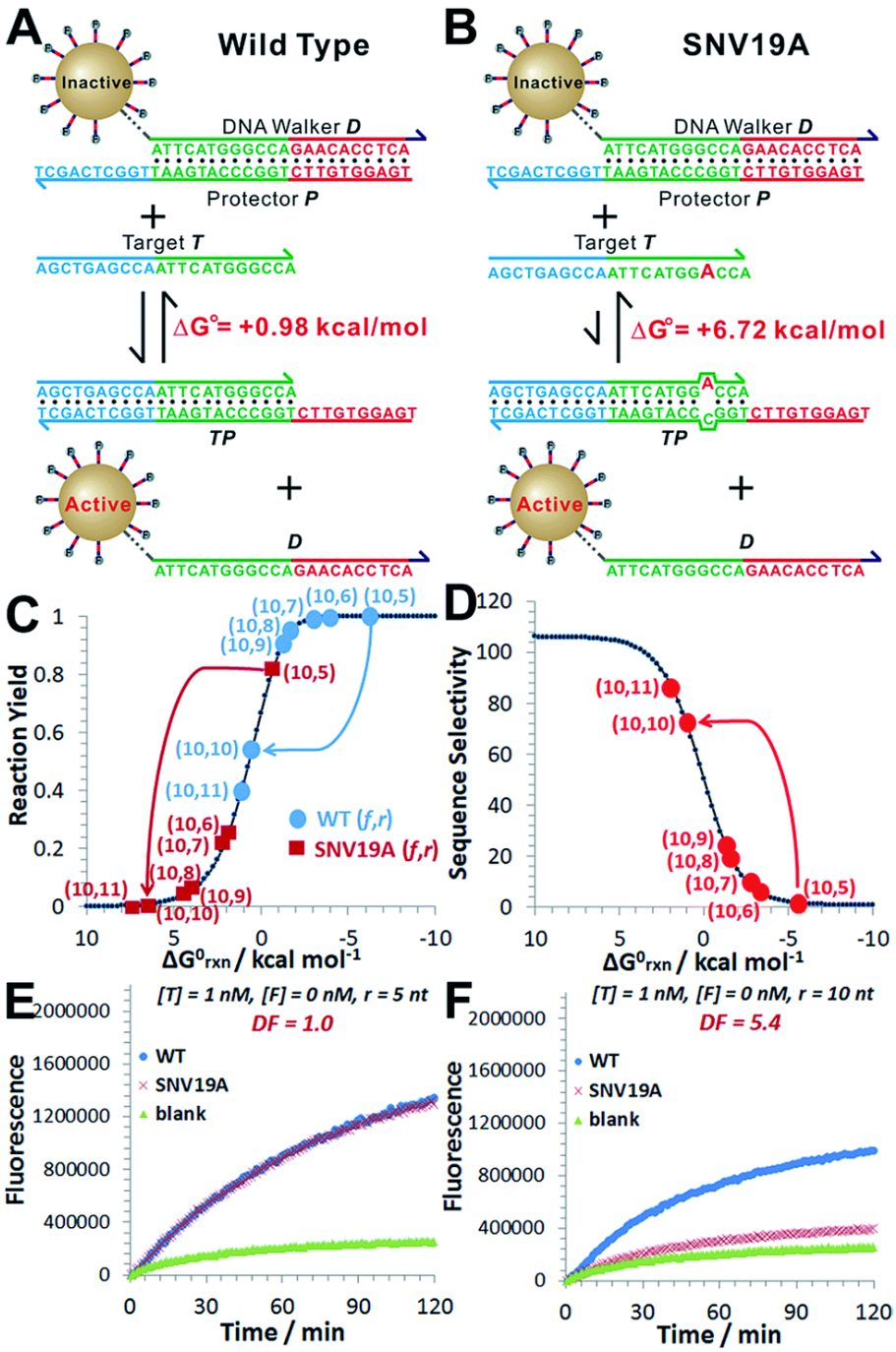


Figure 2.2. Sequence-level tuning of the 3D DNA nanomachine (3DDN) for discriminating SNVs. (**A** and **B**) The toehold-exchange motif (DP) of the 3DDN consists of a DNA walker (D) and pre-hybridized protecting strand (P). This motif can react with an intended target (T) to release D for the subsequent stochastic walking along the AuNP track. Under the optimal sequence design (f = 10 nt, r = 10 nt), the standard free energy (ΔG°) of the forward reaction is +0.98 kcal mol^{-1} for the wild type target (WT) but is +6.72 kcal mol^{-1} for the spurious target (SNV19A). (C and D) In silico analyses of the 3DDN. Reaction yields (**C**) and sequence selectivity (**D**) as a function of ΔG°_{rxn}. The forward

toehold was fixed at f = 10 nt and the lengths of the reverse toehold (r) vary from 5 to 11 nt. Details of the theoretical framework and computer simulation are shown in ESI S2.† (**E** and **F**) experimental time-based fluorescence response of 100 pM 3DDN to 1 nM WT or SNV target. The discrimination factor (DF) was determined using the equation that: DF = $(F_{WT} - F_{blank})/(F_{SNV} - F_{blank})$. All the reactions were performed at 37 °C in 1× NEB CutSmart Buffer containing 20 U *Nb.BbvCI*. Each data point is an average of triplicated analyses.

The *in silico* sequence optimization was further verified experimentally by detecting 1 nM WT and SNV19A using the 3DDN (Fig. 2.2E and F). The discrimination factor (DF) that quantifies the single-base specificity as the ratio of the hybridization yields (net fluorescence) generated by equal concentrations of the intended and SNV targets, was improved by 5.4 times after the sequence optimization (r = 10 nt vs. r = 5 nt). However, this improvement is at the cost of a signal reduction by nearly 1.5 times.

2.2.3 Bimolecular Model and Simulation of the Toehold-Exchange Reaction

Bimolecular Model of the Toehold-Exchange. A typical toehold exchange reaction between T and DP can be written as: $T + DP \rightleftharpoons TP + D$, with a standard Gibbs free energy ΔG°_{rxn}. At equilibrium, the concentration of all species can be expressed in the following equation:

$$K = e^{-\frac{\Delta G^{0}_{rxd}}{RT}} = \frac{[TP]_{eq} * [D]_{eq}}{[T]_{eq} * [DP]_{eq}}$$

For a typical reaction with an initial state $[T]_0$ = 1 nM, $[DP]_0$ = 2 nM, we let [D] = χ; [PT] = χ; [DP] = (2- χ); [T] = (1- χ). The reaction (hybridization) yield (*Y*) is then defined as $\chi/[T]_0$ and activation efficiency (*AE*) of the 3DDN is defined as $\chi/[DP]_0$. By solving the above equation using MATLAB, we are able to plot *Y* or *AE* as a function of ΔG°_{rxn}.

Reaction yield and sequence selectivity as a function of ΔG°_{rxn}. All simulation data using MATLAB were exported into Excel file for subsequent analysis. For example, the sequence analysis for 3DDN with forward toehold length (*f*) fixed at 9 nt and reverse toehold length (*r*) varying from 5 nt to 11 nt is shown in Figure 2.3. The reaction yield as a function of ΔG°_{rxn} was adopted directly from the MATLAB simulation. The sequence selectivity (discrimination factor, *DF*) of WT over SNV19A as a function of ΔG^{0}_{rxd} was calculated as $DF = Y(\Delta G^{\circ}_{rxn}) / Y(\Delta G^{\circ}_{rxn} + 5.74)$. The optimal trade-off between *Y* and *DF* was achieved when $f = 10$ and $r = 10$, where Y ≈ 50% (Figure 2.2, C and D).

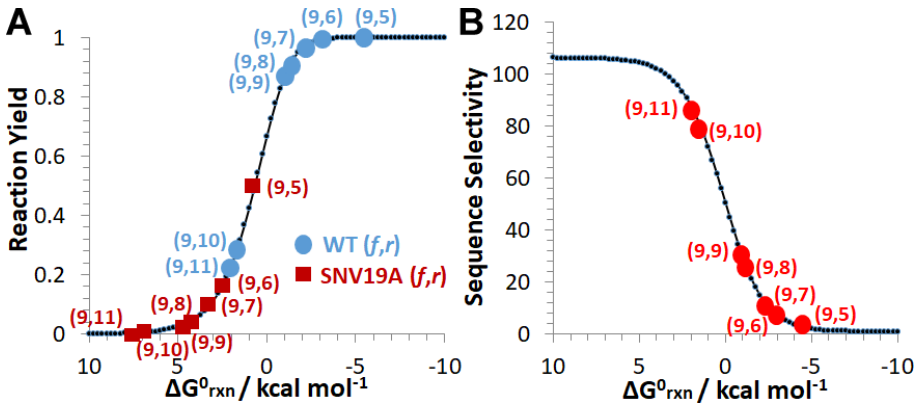


Figure 2.3. Simulation-guided sequence analysis for 3DDN with forward toehold length (*f*) fixed at 9 nt and reverse toehold length (*r*) varying from 5 nt to 11 nt.

Activation Efficiency as a function of ΔG°_{rxn}. To understand how 3DDN performs throughout the concentration range in our experiment, we apply a dimensionless transformation to all DNA molecules, where a new parameter α was introduced as the ratio between $[T]_0$ and $[DP]_0$. We then plotted the *AE* as a function of α and $[T]_0$ (Figure 2.4). Notably, the definition of *AE* ($[D]/[DP]_0$) in our system was different from the hybridization yield (defined as $[D]/\min([T]_0, [DP]_0)$).

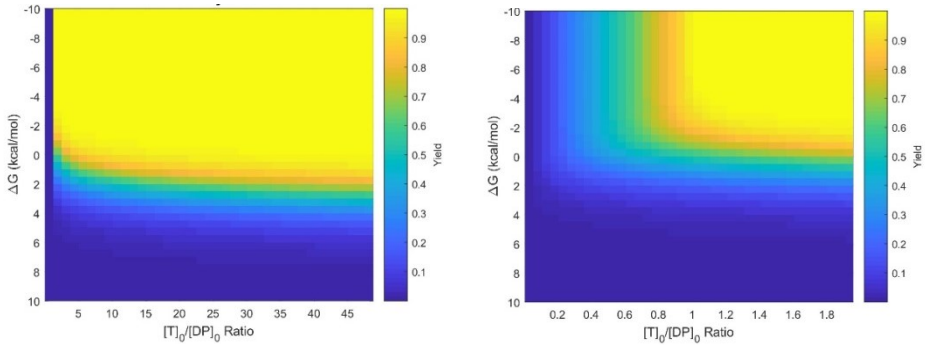


Figure 2.4. Activation efficiency of the 3DDN as a function of α and ΔG°_{rxn}. The range of α in the left panel is from 0 – 50 and is 0 – 2 in the right panel.

2.2.4 Simulation-guided in-solution tuning of 3DDN *via* noncovalent catalysis

We then introduce F as an auxiliary probe for tuning the sensitivity and selectivity of 3DDN in solution (Fig. 2.1C). The addition of F to the system offers an alternative activation pathway through a strand exchange reaction: DP + F ↔ FP + D. This reaction is kinetically inert but can be accelerated by T. As a catalyst, each T activates multiple DNA walkers and thus can enhance the assay sensitivity. Meanwhile, each elementary reaction involving in the DNA catalysis is a toehold exchange with ΔG°_{rxn} to be near zero, and the overall reaction thus remains sensitive to SNVs, ensuring high assay specificity (Fig. 2.5A and B). To quantitatively evaluate the possibility of using F to enhance the discrimination of SNVs, we simulated the catalytic activation of the 3DDN using a two-step reaction model (details in Section 2.2.5). Our simulation results reveal remarkable differences between WT and SNV19A in terms of the level of the intermediate TP (Fig. 2.5C) and the device AE (Fig. 2.5D). In particular, our model suggests that SNV19A and other 8 representative SNVs were much less sensitive to the noncovalent catalysis comparing to WT, evidenced by the low level or absence of the reaction intermediate TP (Fig. 2.5C and Fig. 2.9) and the lack of correlation between T and F in terms of activating the 3DDN (Fig.

2.5D and Fig. 2.10). Therefore, F holds the potential to selectively activate the 3DDN for WT over SNVs and thus can potentially enhance the assay sensitivity without significantly compromising the sequence selectivity. The effectiveness of the F-mediated non-covalent DNA catalysis was further confirmed using polyacrylamide gel electrophoresis (Fig. 2.6). Moreover, our simulation also quantitatively predicts that the catalytic enhancement of 3DDN using F is most effective when $[F]_0/[DP]_0$ ranges from 0 to 5 and $[T]_0/[DP]_0$ ranges from 0 to 10 (Fig. 2.5D).

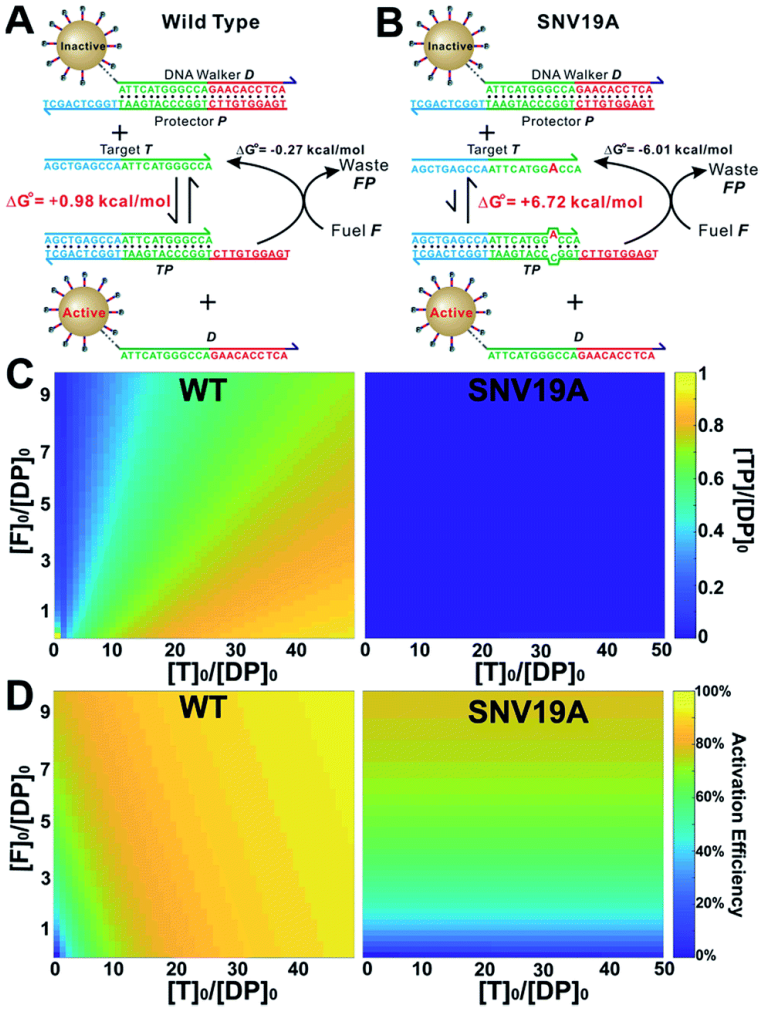


Figure 2.5. In-solution tuning of the 3D DNA nanomachine via noncovalent DNA catalysis. (**A** and **B**) Catalytic enhancement of the toehold-exchange between T and DP using a molecular fuel (F) through noncovalent catalysis. (**C**) Simulation of the level of the

intermediate TP for WT and SNV19A as a function of $[F]_0$ and $[T]_0$. (**D**) Simulation of device activation efficiency for WT and SNV19A as a function of $[F]_0$ and $[T]_0$.

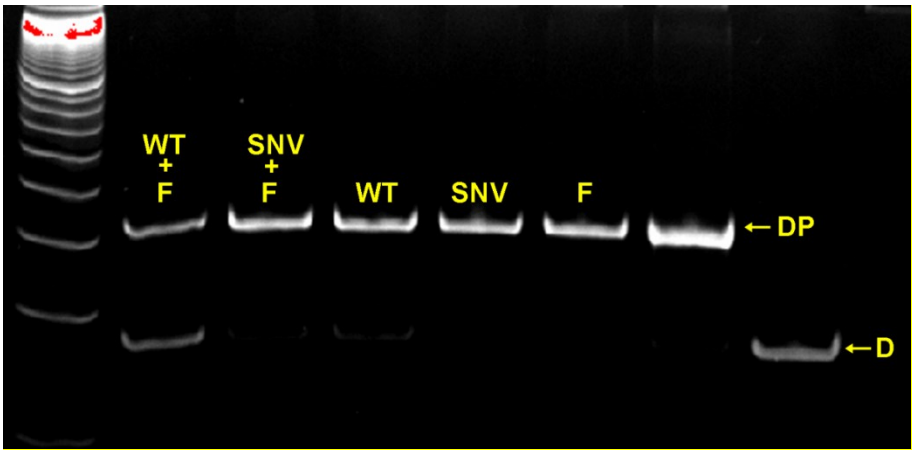


Figure 2.6. Characterization of non-covalent DNA catalysis for enhancing toehold exchange between T and DP using PAGE. The release of active DNA walker motif (D) could only be observed for the wild-type (WT) target (500 nM) when 2 μM fuel (F) was added to 1 μM DP duplex (Lane 2). This result suggests that non-covalent DNA catalysis can not only effectively enhance the assay sensitivity, but also maintain the sequence selectivity of the DNA probes.

We then experimentally examine the effect of F on the 3DDN. The results in Fig. 2.7A show that 10 nM F achieves a 2-fold enhancement for both assay sensitivity and specificity when discriminating 1 nM WT from 1 nM SNV19A. We further challenged our strategy by detecting WT and SNV19A with concentrations ranging from 10 pM to 100 nM. The addition of F with concentrations varying from 0 nM to 10 nM quantitatively shifts the overall detection range of 3DDN towards lower target concentrations by as much as 1 order of magnitude (Fig. 2.7B). Further increases in [F] did not significantly shift the dynamic range (Fig. 2.8). This observation is highly consistent with the theoretical prediction using our two-step reaction model (Fig. 2.5D). SNV19A was also found to be insensitive to F at the concentration range below 5 nM (Fig. 2.7C). Collectively, the non-

covalent catalysis enhances both assay sensitivity (Fig. 2.7B) and specificity (Fig. 2.7D) of the 3DDN at a target concentration range below 5 nM. As the fluorescence signal was fully saturated when target concentration is above 5 nM (Fig. 2.7B), the device becomes less specific at this higher concentration range (Fig. 2.7D).

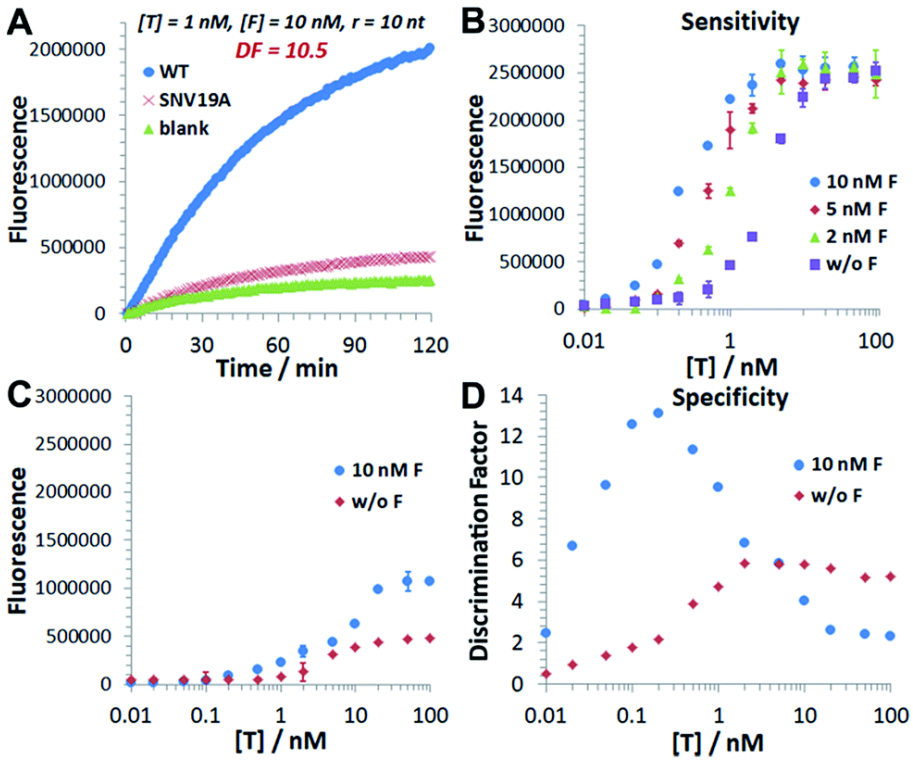


Figure 2.7. Catalytic enhancement of the 3D DNA nanomachine in terms of both assay sensitivity and assay specificity. (**A**) Experimental time-based fluorescence response of 100 pM device to 1 nM WT or SNV19A in the presence of 10 nM F. (**B**) Background-corrected fluorescence as a function of target concentrations in the presence of varying concentrations of F from 0 nM to 10 nM. (**C**) Background-corrected fluorescence as a function of target concentrations in the presence (w F) or absence (w/o F) of 10 nM F. (**D**) Discrimination factor against SNV19A as a function of target concentrations in the presence (w F) or absence (w/o F) of 10 nM F. All the reactions were performed at 37 °C in 1× NEB CutSmart Buffer containing 20 U *Nb.BbvCI* for 1 h. Each error bar represents one standard deviation from triplicated analyses.

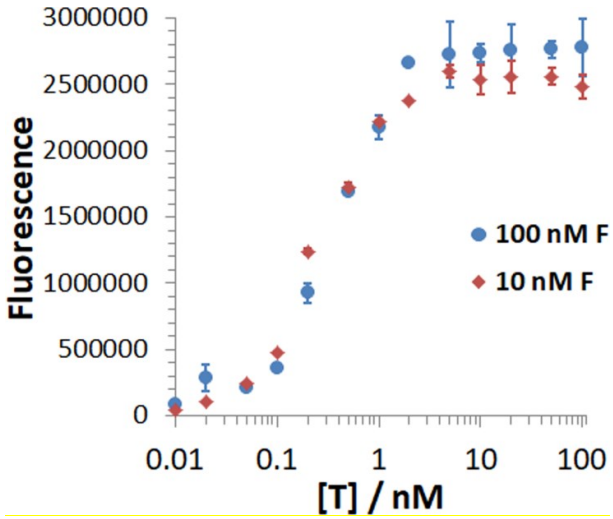


Figure 2.8. Experimental time-based fluorescence response (background corrected) of 100 pM 3DDN to varying concentrations of WT from 10 pM to 100 nM in the presence of 100 nM F and 10 nM F. The reaction mixtures containing varying concentrations of WT from 10 pM to 100 nM, 100 pM 3DDN, 10 nM or 100 nM F, and 20 U nicking endonuclease in 1 × NEB SmartCut buffer were incubated at 37 °C for 1 hr before the fluorescence measurement. Each error bar represents one standard deviation from triplicate analyses.

2.2.5 Two-Step Reaction Model and Simulation of the Non-Covalent DNA Catalysis

The non-covalent catalysis reaction can be simplified as a single overall reaction between F and DP with T serving as a catalyst:

$$F + DP \overset{T}{\Leftrightarrow} FP + D;\ \Delta G_{overall} \approx 0.71\ kcal/mol$$

This simplification is correct only when the intermediate TP does not accumulate significantly over the time. However, this assumption was not met in our system, as the standard free energies for all elementary reactions were very close to 0 and the intermediate TP was maintained at a non-unneglectable level. We therefore simulated the non-covalent catalysis using a two-step reaction model in our system:

$$T + DP \rightleftharpoons TP + D;\ \Delta G_1^0$$

$$F + TP \rightleftharpoons FP + T;\ \Delta G_2^0$$

where $\Delta G_1^\circ = \Delta G_{TP}^\circ + \Delta G_D^\circ - \Delta G_{DP}^\circ - \Delta G_T^\circ$ and $\Delta G_2^\circ = \Delta G_{FP}^\circ + \Delta G_T^\circ - \Delta G_{DP}^\circ - \Delta G_F^\circ$. Values of ΔG_1° and ΔG_2° for WT and SNVs were calculated and listed in Table 2.1.

Table 2.1. Standard Gibbs free energies (ΔG_1° and ΔG_2°) for the two-step reactions involved in the non-covalent DNA catalysis.

Sequence	ΔG^0_1	ΔG^0_2
WT	+0.98	–0.27
SNV19C	+8.42	–7.71
SNV19A	+6.72	–6.01
SNV19T	+6.47	–5.76
SNV12A	+3.98	–3.27
SNV12G	+2.92	–2.21
SNV12C	+4.41	–3.70
SNV7C	+8.30	–7.59
SNV7A	+5.99	–5.28
SNV7T	+5.61	–4.90

The equilibrium concentrations of all DNA species can then be derived by solving a set of equations:

$$K_1 = e^{\frac{-\Delta G_1}{RT}} = \frac{[D] \times [TP]}{[T] \times [DP]};$$

$$K_2 = e^{\frac{-\Delta G_2}{RT}} = \frac{[T] \times [FP]}{[F] \times [TP]};$$

After dimensionless transformation of the equations system we can introduce two tunable variables r and f and two dependent variables y and x ($where\ y = \frac{[D]}{[DP]_0}, x = \frac{[FP]}{[DP]_0}$):

$$K_1 = e^{\frac{-\Delta G_1}{RT}} = \frac{[D] \times [PT]}{[T] \times [DP]} = \frac{[D] \times ([D] - [PF])}{([T]_0 - [D] + [PF]) \times ([DP]_0 - [D])}$$

$$= \frac{[D]/[DP]_0 \times ([D] - [PF])/[DP]_0}{([T]_0 - [D] + [PF])/[DP]_0 \times ([DP]_0 - [D])/[DP]_0}$$

$$= \frac{y \times (y - x)}{(r - y + x) \times (1 - y)};$$

$$Similarly, K_2 = e^{\frac{-\Delta G_2}{RT}} = \frac{x \times (r - y + x)}{(y - x) \times (f - x)};$$

By solving the two equations using MATLAB code, we were able to estimate the level of the intermediate TP (Figure 2.9) and the device activation efficiency (Figure 2.10) as a function of both $[F]_0$ and $[T]_0$ for WT and SNVs.

2.2.6 Matlab simulation results

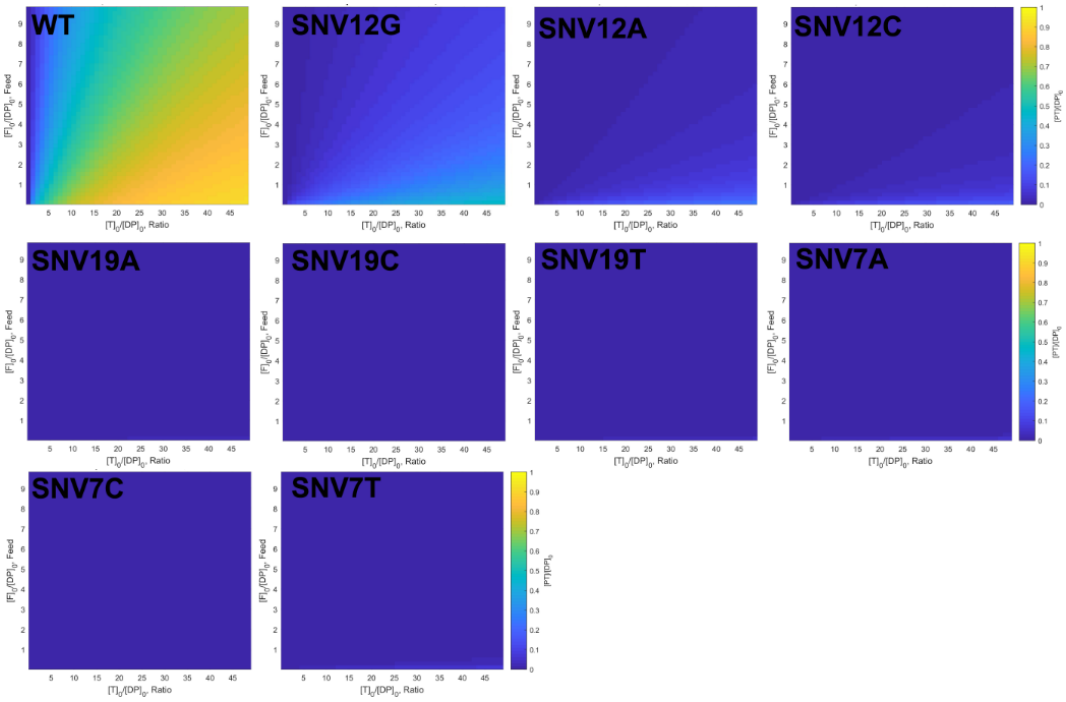


Figure 2.9. The level of the intermediate TP ($[TP]/[DP]_0$) as a function of $[F]_0$ and $[T]_0$.

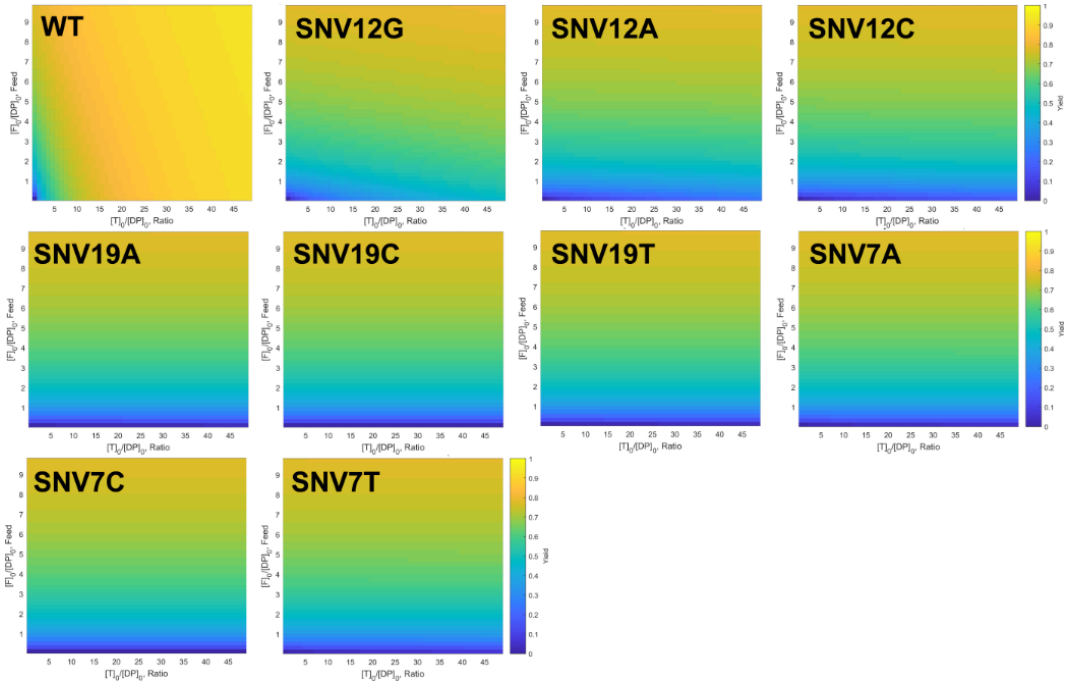


Figure 2.10. The device activation efficiency ([D]/[DP]$_0$) as a function of [F]$_0$ and [T]$_0$.

2.2.7 Predictable discrimination of SNVs using catalytic 3DDN

We next challenged our catalytic 3DDN with nine SNVs of varying base identities and positions (Fig. 2.11A). Our system discriminates all SNVs at 1 nM with DF ranging from 7 to 26, with a median of 12 (Fig. 2.11B and C). The obtained specificity is comparable with commonly used molecular probes, such as molecular beacons,[28] triple-stem DNA probes,[29] and toehold-strand displacement beacons.[27,30,31] However, because of the intrinsic signal amplification capability, our 3DDN improves the detection limit by ~100 times.

It is also possible to predict the reactivity of our 3DDN system towards varying SNVs based on their thermodynamic parameters. As the overall specificity is generally

determined by the toehold-exchange between T and DP, an intuitive way to predict the sequence selectivity can be on the basis of the thermodynamic change (ΔG°_{rxn}) associated with a given single-nucleotide mismatch (Fig. 2.11D). The measured fluorescence responses of our device against three representative SNVs (SNV12G, 12C, and 7C) across a concentration range from 10 pM to 10 μM were consistent with their trends in $\Delta\Delta G^{\circ}_{rxn}$ (Fig. 2.11E). The predictive reactivity of F-enhanced 3DDN was further confirmed using SNVs at different locations and base identities (Section 2.2.8, Fig. 2.12). Here, the tuning of 3DDN enabled by non-covalent DNA catalysis offers an alternative on-the-fly approach to fine-tune the assay performance. This approach is critical in our system, as the ratio between D and P has been locked at 1 : 1 and it is hence not possible to fine-tune the sequence selectivity using the stoichiometry strategy previously described by Wu, et. al.[26]

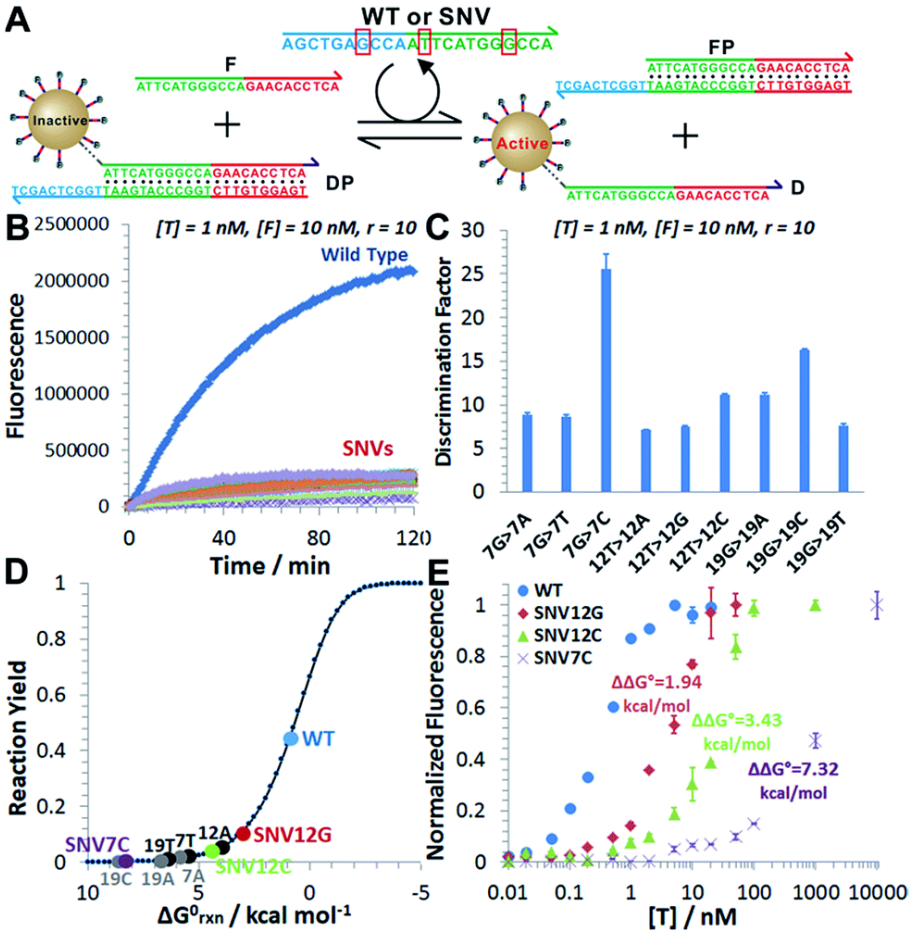


Figure 2.11. (**A**) Sequence of the wild type target (WT) and positions/identities of base changes that lead to the 9 SNV targets. The red circles denote the positions of the mismatch. (**B**) Time-based fluorescence responses of 100 pM device to 1 nM WT or SNVs in the presence of 10 nM F. (**C**) Discrimination factors determined for 9 SNVs at 1 nM concentrations. (**D**) The distribution of the WT and the 9 SNVs on the standard Gibbs free energy landscape. (**E**) Dose-response curves of WT and three representative SNVs obtained by plotting the normalized fluorescence as a function of target concentrations. The fluorescence for each sample was normalized against the background fluorescence for the blank (set to 0) and maximum fluorescence when all DNA walkers are released (set to 1). Each error bar represents one standard deviation from triplicate analyses.

Towards the practical uses of our strategy in biological or clinical samples, we finally challenged the catalytic 3DDN system for discriminating SNVs in complicated sample matrix. To do so, we spiked 1 nM of target DNA (WT or SNV19A) into two types of sample matrixes commonly seen in nucleic acid testing, including high concentrations

of background DNA (0.1 mg mL^{-1} salmon sperm genome fragments) and 10-time diluted human serum samples. As shown in Figure 2.13 the 3DDN remains fully active in both sample matrixes and its ability to discriminate SNVs was not affected either. The noticeable high background fluorescence observed in Figure 2.13B was a result of the autofluorescence of the human serum sample. Another concern for the practical uses of the 3DDN is the long-term colloidal and chemical stability of the DNA-AuNP motif during storage. We further monitored the activity and specificity of the 3DDN over a period of 1 month. As shown in Figure 2.14 no apparent losses in activity were observed for 3DDN during the 1 month storage at 4 °C. Similarly, no significant changes in specificity of the catalytic 3DDN system were observed either. Collectively, these observations suggest that our 3DDN is a chemically robust system for discriminating SNVs against varying sample matrix and long-time storage.

2.2.8 Prediction of the Sequence Selectivity Using $\Delta\Delta G°$

It is possible to predict the reactivity of the catalytic enhanced 3DDN towards SNVs using the thermodynamic parameter $\Delta\Delta G°$. We validated the $\Delta\Delta G°$-based prediction using 7 representative SNVs categorized into three groups, including those of large $\Delta\Delta G°$ differences (Figure 2.11E), those of the same location but with different base identities (Figure 2.12A), and those of the same base identity at 3 different locations (Figure 2.12B).

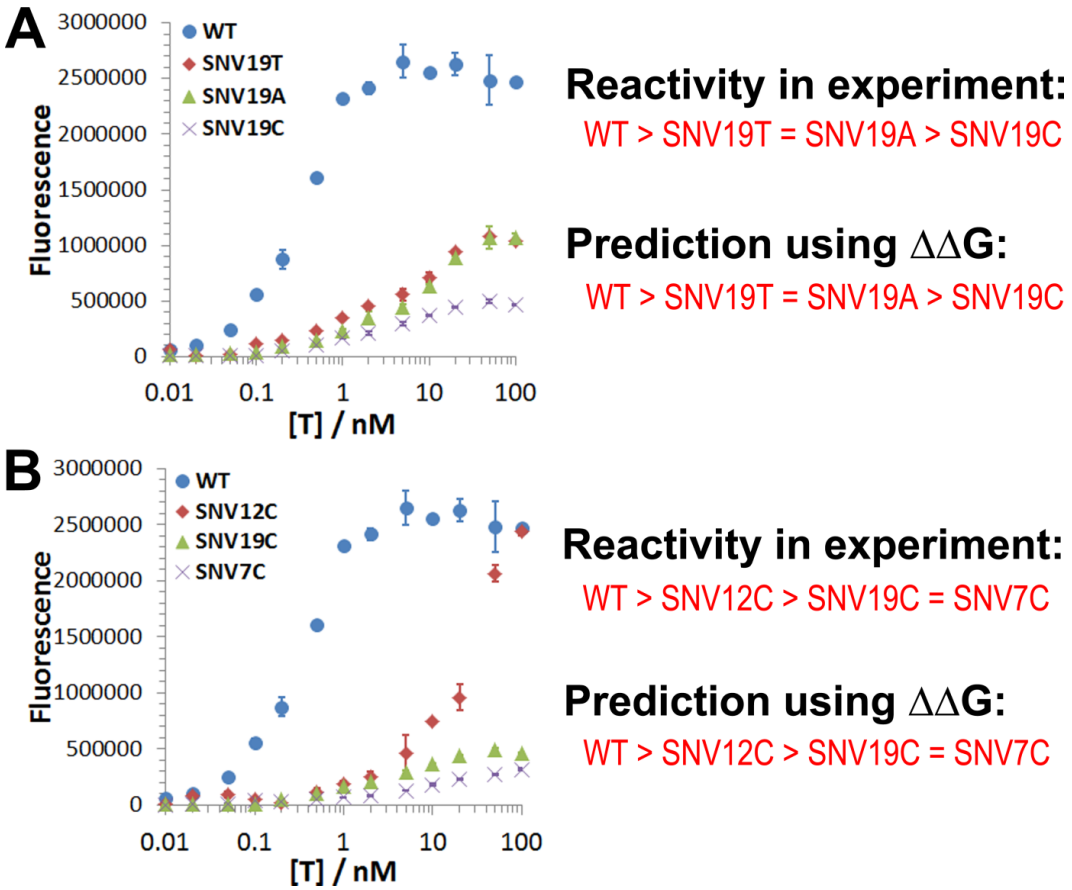


Figure 2.12. Validation of the sequence selectivity of catalytic 3DDN using SNVs at the same location but different base identities (**A**) and SNVs of the same base identity but different locations (**B**). The reaction mixtures containing varying concentrations of SNVs, from 10 pM to 10 μM, 100 pM 3DDN, 10 nM F, and 20 U nicking endonuclease in 1 × NEB SmartCut buffer were incubated at 37 °C for 1 hr before the fluorescence measurement. Each error bar represents one standard deviation from triplicate analyses.

2.2.9 Discrimination of SNV in Complicated Sample Matrix Using Catalytic 3DDN

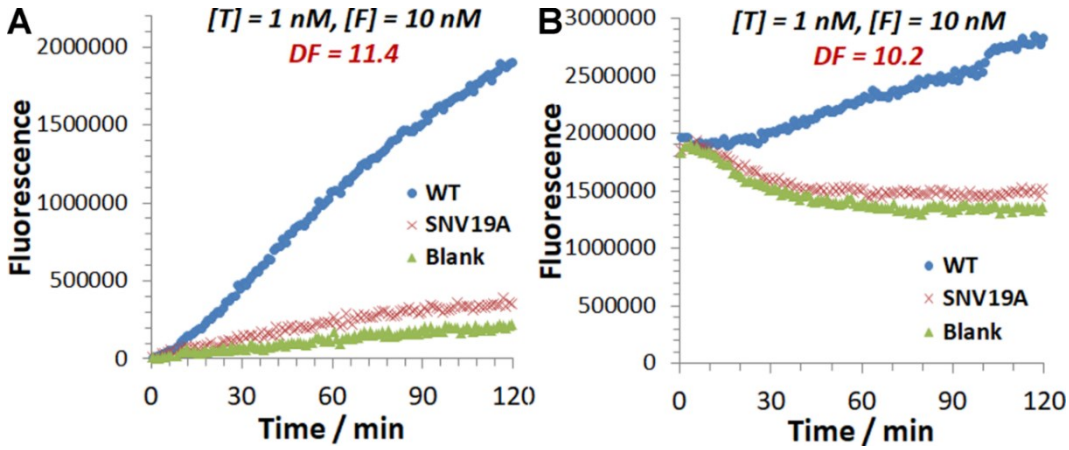


Figure 2.13. (A) Detection of WT and SNV19A using catalytic 3DDN in 0.1 mg/mL salmon sperm genome DNA fragments. (**B**) Detection of WT and SNV19A using catalytic 3DDN in 10-time diluted human serum samples. Each reaction mixture containing 1 nM WT (or SNV19A), 10 nM F, 100 pM 3DDN and 20 U nicking endonuclease in 1 × NEB SmartCut buffer were incubated at 37 °C for 1 hr before the fluorescence measurement. DF was then determined using the equation that: $DF = (F_{WT} - F_{blank})/(F_{SNV} - F_{blank})$. Each error bar represents one standard deviation from triplicate analyses.

2.2.10 Long-Term Stability of the 3DDN System

One important practical concern about the 3DDN system is the long-time colloidal and chemical stability of the DNA-functionalized AuNP motif. As such, we monitored the analytical performance of the same batch of 3DDN over a period of 1 month. Specifically, the activity of the active 3DDN, the stability of the inactive 3DDN (blank), and the ability to discriminate single nucleotide variation (WT against SNV19A) were measured at day 1, 3, 5, 7, 14 and 30 and plotted in Figure 2.14.

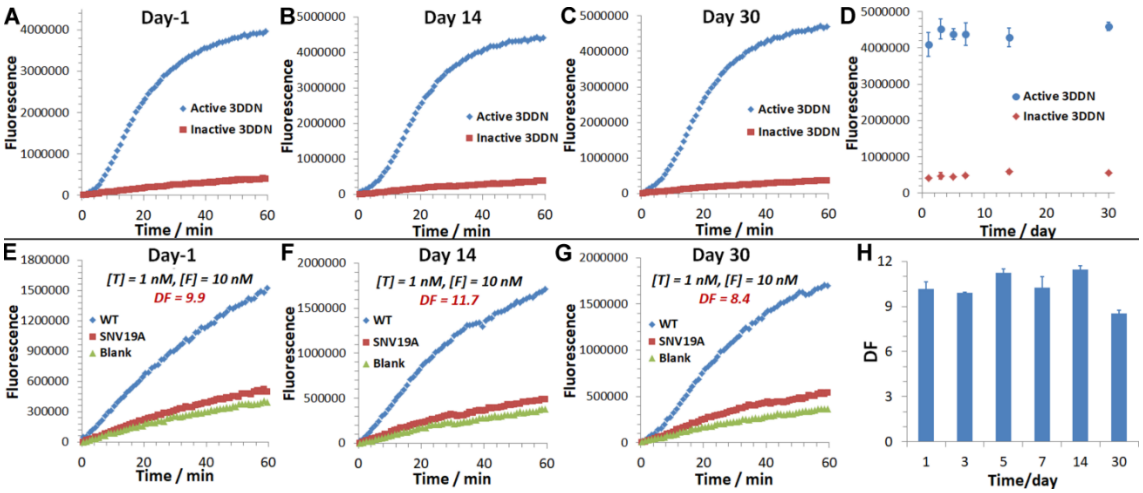


Figure 2.14. (A-D) Long-term activity and the chemical stability of the 3DDN system over a period of 30 days. The active 3DDN was prepared by coconjugating AuNP with R and unprotected D and the inactive 3DDN was prepared by coconjugating AuNP with R and protected D (DP). Each reaction mixture containing 100 pM 3DDN and 20 U nicking endonuclease in 1 × NEB SmartCut buffer were incubated at 37 °C for 1 hr before the fluorescence measurement. **(E-H)** Long-term activity of the 3DDN for discriminating SNVs over a period of 30 days. Each reaction mixture containing 1 nM WT (or SNV19A), 10 nM F, 100 pM 3DDN and 20 U nicking endonuclease in 1 × NEB SmartCut buffer were incubated at 37 °C for 1 hr before the fluorescence measurement. DF was then determined using the equation that: DF = $(F_{WT} - F_{blank})/(F_{SNV} - F_{blank})$. Each error bar represents one standard deviation from triplicate analyses.

2.3 Conclusions

Herein, we have engineered a nicking endonuclease-powered 3D DNA nanomachine for discriminating SNVs with simulation-guided optimization at both sequence-level and operational level. We have also introduced a new in-solution tuning strategy that tunes both assay sensitivity and specificity using a single molecular fuel through noncovalent catalysis. Despite its critical roles in DNA logic circuits (e.g., seesaw gates)[32,33] and for characterizing the thermodynamics of nucleic acid motifs,[34] noncovalent DNA catalysis finds very few applications in regulating DNA devices or sensors. Our work opens the possibility to use this reaction as an on-the-fly approach for tuning the analytical performance of DNA devices or sensors. Collectively, we have developed a detection system that offers the specificity for SNVs comparable to commonly used molecular probes, but a much higher sensitivity. It is also possible to predict the reactivity of our detection system towards different SNVs in silico on the basis of their thermodynamic parameters. As thermodynamic parameters of any hybridization reaction can be easily obtained using DNA analysing software, such as NUPACK, our simulation strategy can be generalized to any nucleic acid target or sequence of interest. As such, our simulation-guide tuning strategy will be a useful addition to the existing simulation tools, such as the coarse-grained simulation recently introduced by Nir and coworkers,[35] for the design and operation of DNA motors and walking devices with better performance and functionality towards biosensing applications.

2.4 Experimental

Materials and Reagents. Solutions of 20-nm gold nanoparticles (AuNPs), TWEEN20, dithiothreitol (DTT), sodium chloride (NaCl), magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$), and 100 × Tris-EDTA (TE) buffer were purchased from Sigma (Oakville, ON, Canada). Nicking endonuclease (*Nb.BvCI*) and 10 × CutSmart Buffer were purchased from New England Biolabs Ltd. (Whitby, ON, Canada). NANOpure H2O (> 18.0 MΩ), purified using an Ultrapure Mili-Q water system, was used for all experiments. All DNA samples were purchased from Integrated DNA Technologies (Coralville, IA) and purified using high-performance liquid chromatography. The DNA sequences and modifications are listed in Table 2.2.

Table 2.2. DNA sequences and modifications.

DNA name		Sequences
DNA probes for constructing the 3D DNA Nanomachine	D (r = 10 nt)	5'-SH-T-50-T-ATT CAT GGG CCA GAACA CCTCAGC-3'-3'
	D (r = 5 nt)	5'-SH-T-50-T-ATT CAT GGG CCA CCTCAGC-3'-3'
	SR	5'-SH- T-10-T- GC*TGA GGAT-FAM-3'(*cleavage site)
	P (r = 10 nt)	5'- TGAGG TGTTC TGG CCC ATG AAT TGGC TCA GCT-3'
	P (r = 5 nt)	5'- TGAGG TGG CCC ATG AAT TGGC TCA GCT-3'
	F	5'- ATT CAT GGG CCA GAACA CCTCA -3'
Target	WT	5'- AGC TGA GCCA ATT CAT GGG CCA-3'
	SNV19A	5'- AGC TGA GCCA ATT CAT GGA CCA-3'
	SNV19T	5'- AGC TGA GCCA ATT CAT GGT CCA-3'

DNA	SNV19C	5'- AGC TGA GCCA ATT CAT GGC CCA-3'
	SNV12A	5'- AGC TGA GCCA AAT CAT GGG CCA-3'
	SNV12G	5'- AGC TGA GCCA AGT CAT GGG CCA-3'
	SNV12C	5'- AGC TGA GCCA ACT CAT GGG CCA-3'
	SNV7A	5'- AGC TGA ACCA ATT CAT GGG CCA-3'
	SNV7C	5'- AGC TGA CCCA ATT CAT GGG CCA-3'
	SNV7T	5'- AGC TGA TCCA ATT CAT GGG CCA-3'

Preparation of the 3D DNA Nanomachine. The 3D DNA nanomachine was prepared by co-conjugating thiolated DNA probes onto the 20-nm AuNPs according to our previously established protocol. Briefly, 2.5 μM DNA walker (D) was first hybridized with equal amounts of the protecting DNA (P) through an annealing process. The obtained DP duplex was then mixed with signal reporter (R) at a ratio of 1 to 20. A 20 μL solution of this mixture containing 2.5 μM DP and 50 μM R was mixed with 1 mL of 1 nM AuNPs. This mixture was incubated at room temperature for 12 hrs and then was slowly mixed with 20 μL of 3 M NaCl solution, followed by 10 s of sonication. This salt aging process was repeated five times with a 1 h interval to maximize the density of oligonucleotide on AuNPs. The solution was then incubated for another 24 hrs. After incubation, the solution was centrifuged at 13,500 rpm for 30 min to separate the DNA-AuNP from the unreacted reagents. The DNA-AuNPs were then washed three times with 1 × TE buffer (pH 7.4) containing 0.01% TWEEN 20 and finally redispersed in TE buffer.

Characterization of the 3D DNA Nanomachine. UV-vis spectrometry was used to characterize the quality of each DNA-AuNP solution and determine the concentration of the device. Typically, a maximum absorbance value of each DNA-AuNP solution was

measured and compared to that of the unconjugated AuNP, whose concentration was provided by the vendor (1.16 nM). A red shift of 4-6 nm of the maximum absorbance could be observed upon the conjugation of DNA to AuNPs. The coverage of R oligonucleotides on each AuNP was then determined by releasing FAM-labeled R from AuNP using 20 mM DTT and measuring fluorescence of the FAM. Fluorescence was measured using a multimode microplate reader (SpectroMax i3, Molecular Devices), and R coverage was quantified by using FAM-labeled R as external standards. The coverage of DP on each AuNP was then estimated according to the R coverage and initial ratio between DP and R.

Detection of WT and SNV Targets Using the 3D DNA Nanomachine. For a typical reaction, a mixture containing 100 pM 3D device and varying concentrations of target DNA (WT or SNVs) were incubated at 37 °C for 10 min. Nicking endonuclease (20 U) was then added to each mixture to initiate the 3D walking. Immediately after the addition of the enzyme, fluorescence was measured every 1 min using a multimode microplate reader with an excitation/emission wavelength of 485/535 nm. To enhance the analytical performance of the 3DDN via noncovalent catalysis, the same protocol was used for target analyses except that 10 nM fuel molecules (F) were added to the mixture of target and DNA nanomachine and followed by incubation at 37 °C for 1 hour before adding the nicking enzyme.

Characterization of non-covalent DNA catalysis using polyacrylamide gel electrophoresis (PAGE). A reaction mixture containing 1 μM DP, 2 μM F, and 500 nM T (WT or SNV19A) was incubated at 37 °C for 1 hr. Reaction mixtures were than loaded onto 12% PAGE gel and a voltage of 110 V was applied. After electrophoresis, the gel was stained with Ethidium Bromide and imaged using Gel Doc XR+ Imager System (BioRad).

Determination of thermodynamic parameters using NUPACK. Standard Gibbs free energies for all DNA probes were estimated using NUPACK. The temperature and saline condition were set to be 37 °C, 100 mM Na^+, and 10 mM Mg^{2+}, which is consistent with our experimental conditions. The standard Gibbs free energy for the toehold-exchange between target (T) and DP was calculated using the following equation: $\Delta G^\circ_{rxn} = \Delta G^\circ_{TP} + \Delta G^\circ_D - \Delta G^\circ_{DP} - \Delta G^\circ_T$, and detailed in Table 2.3 (with forward toehold f = 10 nt) and Table 2.4 (f = 9 nt). The $\Delta\Delta G^\circ$ was then calculated according to the equation: $\Delta\Delta G^\circ = \Delta G^\circ_{rxn}(SNV) - \Delta G^\circ_{rxn}(WT)$, and detailed in Table 2.5.

Table 2.3. Thermodynamic parameters of the enzyme-powered 3D DNA nanomachine for the detection of a wild-type (WT) target and a SNV (SNV19A) target with forward toehold f fixed at 10 nt and reverse toehold r varying from 5 nt to 11 nt.

Reverse toehold *r*	ΔG^0(TP) kcal/mol (WT)	ΔG^0(TP) kcal/mol (SNV19A)	ΔG^0(D) kcal/mol	ΔG^0(DP) kcal/mol	ΔG^0(T) kcal/mol (WT)	ΔG^0(T) kcal/mol (SNV19A)	ΔG^0_{rxn} kcal/mol (WT)	ΔG^0_{rxn} kcal/mol (SNV19A)
5	-29.95	-24.69	-0.28	-23.38	-1.56	-2.04	**-5.29**	**0.45**
6	-29.95	-24.69	-0.48	-25.37	-1.56	-2.04	**-3.5**	**2.24**
7	-29.95	-24.69	-0.43	-25.77	-1.56	-2.04	**-3.05**	**2.69**
8	-29.95	-24.69	0	-26.67	-1.56	-2.04	**-1.72**	**4.02**
9	-29.95	-24.69	-1.52	-28.42	-1.56	-2.04	**-1.49**	**4.25**
10	-29.95	-24.69	0	-29.37	-1.56	-2.04	**0.98**	**6.72**
11	-29.95	-24.69	-1.22	-31.11	-1.56	-2.04	**1.5**	**7.24**

Table 2.4. Thermodynamic parameters of the enzyme-powered 3D DNA nanomachine for the detection of a wild-type (WT) target and a SNV (SNV19A) target with forward toehold f fixed at 9 nt and reverse toehold r varying from 5 nt to 11 nt.

Reverse toehold *r*	ΔG^0(TP) kcal/mol (WT)	ΔG^0(TP) kcal/mol (SNV19A)	ΔG^0(D) kcal/mol	ΔG^0(DP) kcal/mol	ΔG^0(T) kcal/mol (WT)	ΔG^0(T) kcal/mol (SNV19A)	ΔG^0_{rxn} kcal/mol (WT)	ΔG^0_{rxn} kcal/mol (SNV19A)
5	-29.4	-24.13	-0.28	-23.38	-1.56	-2.04	**-4.74**	**1.01**
6	-29.4	-24.13	-0.48	-25.37	-1.56	-2.04	**-2.95**	**2.8**
7	-29.4	-24.13	-0.43	-25.77	-1.56	-2.04	**-2.5**	**3.25**
8	-29.4	-24.13	0	-26.67	-1.56	-2.04	**-1.17**	**4.58**
9	-29.4	-24.13	-1.52	-28.42	-1.56	-2.04	**-0.94**	**4.81**
10	-29.4	-24.13	0	-29.37	-1.56	-2.04	**1.53**	**7.28**
11	-29.95	-24.69	-1.22	-31.11	-1.56	-2.04	**2.05**	**7.8**

Table 2.5. Thermodynamics of single-base changes.

Target	ΔG^0_{rxn} kcal/mol	$\Delta\Delta G^0$ kcal/mol
WT	+0.98	0
SNV19C	+8.42	+7.44
SNV19A	+6.72	+5.74
SNV19T	+6.47	+5.49
SNV12A	+3.98	+3.00
SNV12G	+2.92	+1.94
SNV12C	+4.41	+3.43
SNV7C	+8.30	+7.32
SNV7A	+5.99	+5.01
SNV7T	+5.61	+4.63

2.5 Reference

(1) Abendroth, J. M.; Bushuyev, O. S.; Weiss, P. S.; Barrett, C. J. Controlling Motion at the Nanoscale: Rise of the Molecular Machines. *ACS Nano* **2015**, *9* (8), 7746–7768.

(2) Bath, J.; Turberfield, A. J. DNA Nanomachines. *Nature Nanotechnology* **2007**, *2* (5), 275–284.

(3) Sherman, W. B.; Seeman, N. C. A Precisely Controlled DNA Biped Walking Device. *Nano Lett.* **2004**, *4* (7), 1203–1207.

(4) Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H. A Unidirectional DNA Walker That Moves Autonomously along a Track. *Angewandte Chemie International Edition* **2004**, *43* (37), 4906–4911.

(5) You, M.; Chen, Y.; Zhang, X.; Liu, H.; Wang, R.; Wang, K.; Williams, K. R.; Tan, W. An Autonomous and Controllable Light-Driven DNA Walking Device. *Angewandte Chemie International Edition* **2012**, *51* (10), 2457–2460.

(6) Lund, K.; Manzo, A. J.; Dabby, N.; Michelotti, N.; Johnson-Buck, A.; Nangreave, J.; Taylor, S.; Pei, R.; Stojanovic, M. N.; Walter, N. G.; Winfree, E.; Yan, H. Molecular Robots Guided by Prescriptive Landscapes. *Nature* **2010**, *465* (7295), 206–210.

(7) Wickham, S. F. J.; Bath, J.; Katsuda, Y.; Endo, M.; Hidaka, K.; Sugiyama, H.; Turberfield, A. J. A DNA-Based Molecular Motor That Can Navigate a Network of Tracks. *Nature Nanotechnology* **2012**, *7* (3), 169–173.

(8) Thubagere, A. J.; Li, W.; Johnson, R. F.; Chen, Z.; Doroudi, S.; Lee, Y. L.; Izatt, G.; Wittman, S.; Srinivas, N.; Woods, D.; Winfree, E.; Qian, L. A Cargo-Sorting DNA Robot. *Science* **2017**, *357* (6356).

(9) Tomov, T. E.; Tsukanov, R.; Liber, M.; Masoud, R.; Plavner, N.; Nir, E. Rational Design of DNA Motors: Fuel Optimization through Single-Molecule Fluorescence. *J. Am. Chem. Soc.* **2013**, *135* (32), 11935–11941.

(10) Tomov, T. E.; Tsukanov, R.; Glick, Y.; Berger, Y.; Liber, M.; Avrahami, D.; Gerber, D.; Nir, E. DNA Bipedal Motor Achieves a Large Number of Steps Due to Operation Using Microfluidics-Based Interface. *ACS Nano* **2017**, *11* (4), 4002–4008.

(11) Wang, D.; Vietz, C.; Schröder, T.; Acuna, G.; Lalkens, B.; Tinnefeld, P. A DNA Walker as a Fluorescence Signal Amplifier. *Nano Lett.* **2017**, *17* (9), 5368–5374.

(12) Zhou, C.; Duan, X.; Liu, N. A Plasmonic Nanorod That Walks on DNA Origami. *Nature Communications* **2015**, *6* (1), 8102.

(13) Jung, C.; Allen, P. B.; Ellington, A. D. A Stochastic DNA Walker That Traverses a Microparticle Surface. *Nature Nanotechnology* **2016**, *11* (2), 157–163.

(14) Jung, C.; Allen, P. B.; Ellington, A. D. A Simple, Cleated DNA Walker That Hangs on to Surfaces. *ACS Nano* **2017**, *11* (8), 8047–8054.

(15) Yang, X.; Tang, Y.; Mason, S. D.; Chen, J.; Li, F. Enzyme-Powered Three-Dimensional DNA Nanomachine for DNA Walking, Payload Release, and Biosensing. *ACS Nano* **2016**, *10* (2), 2324–2330.

(16) Chen, J.; Zuehlke, A.; Deng, B.; Peng, H.; Hou, X.; Zhang, H. A Target-Triggered DNAzyme Motor Enabling Homogeneous, Amplified Detection of Proteins. *Anal. Chem.* **2017**, *89* (23), 12888–12895.

(17) Peng, H.; Li, X.-F.; Zhang, H.; Le, X. C. A MicroRNA-Initiated DNAzyme Motor Operating in Living Cells. *Nature Communications* **2017**, *8* (1), 14378.

(18) Wang, K.; He, M.-Q.; Zhai, F.-H.; Wang, J.; He, R.-H.; Yu, Y.-L. Autonomous DNA Nanomachine Based on Cascade Amplification of Strand Displacement and DNA Walker for Detection of Multiple DNAs. *Biosensors and Bioelectronics* **2018**, *105*, 159–165.

(19) Zhang, H.; Lai, M.; Zuehlke, A.; Peng, H.; Li, X.-F.; Le, X. C. Binding-Induced DNA Nanomachines Triggered by Proteins and Nucleic Acids. *Angewandte Chemie International Edition* **2015**, *54* (48), 14326–14330.

(20) Qu, X.; Zhu, D.; Yao, G.; Su, S.; Chao, J.; Liu, H.; Zuo, X.; Wang, L.; Shi, J.; Wang, L.; Huang, W.; Pei, H.; Fan, C. An Exonuclease III-Powered, On-Particle Stochastic DNA Walker. *Angewandte Chemie International Edition* **2017**, *56* (7), 1855–1858.

(21) He, M.-Q.; Wang, K.; Wang, W.-J.; Yu, Y.-L.; Wang, J.-H. Smart DNA Machine for Carcinoembryonic Antigen Detection by Exonuclease III-Assisted Target Recycling and DNA Walker Cascade Amplification. *Anal. Chem.* **2017**, *89* (17), 9292–9298.

(22) Liang, C.-P.; Ma, P.-Q.; Liu, H.; Guo, X.; Yin, B.-C.; Ye, B.-C. Rational Engineering of a Dynamic, Entropy-Driven DNA Nanomachine for Intracellular MicroRNA Imaging. *Angewandte Chemie International Edition* **2017**, *56* (31), 9077–9081.

(23) Kim, S.; Misra, A. SNP Genotyping: Technologies and Biomedical Applications. *Annual Review of Biomedical Engineering* **2007**, *9* (1), 289–320.

(24) Zhang, D. Y.; Chen, S. X.; Yin, P. Optimizing the Specificity of Nucleic Acid Hybridization. *Nature Chemistry* **2012**, *4* (3), 208–214.

(25) Ohta, S.; Glancy, D.; Chan, W. C. W. DNA-Controlled Dynamic Colloidal Nanoparticle Systems for Mediating Cellular Interaction. *Science* **2016**, *351* (6275), 841–845.

(26) Wu, L. R.; Wang, J. S.; Fang, J. Z.; Evans, E. R.; Pinto, A.; Pekker, I.; Boykin, R.; Ngouenet, C.; Webster, P. J.; Beechem, J.; Zhang, D. Y. Continuously Tunable Nucleic Acid Hybridization Probes. *Nature Methods* **2015**, *12* (12), 1191–1196.

(27) Wang, J. S.; Zhang, D. Y. Simulation-Guided DNA Probe Design for Consistently Ultraspecific Hybridization. *Nature Chemistry* **2015**, *7* (7), 545–553.

(28) Tyagi, S.; Bratu, D. P.; Kramer, F. R. Multicolor Molecular Beacons for Allele Discrimination. *Nature Biotechnology* **1998**, *16* (1), 49–53.

(29) Xiao, Y.; Plakos, K. J. I.; Lou, X.; White, R. J.; Qian, J.; Plaxco, K. W.; Soh, H. T. Fluorescence Detection of Single-Nucleotide Polymorphisms with a Single, Self-Complementary, Triple-Stem DNA Probe. *Angewandte Chemie International Edition* **2009**, *48* (24), 4354–4358.

(30) Li, Q.; Luan, G.; Guo, Q.; Liang, J. A New Class of Homogeneous Nucleic Acid Probes Based on Specific Displacement Hybridization. *Nucleic Acids Res* **2002**, *30* (2), e5–e5.

(31) Chen, S. X.; Seelig, G. An Engineered Kinetic Amplification Mechanism for Single Nucleotide Variant Discrimination by DNA Hybridization Probes. *J. Am. Chem. Soc.* **2016**, *138* (15), 5076–5086.

(32) Qian, L.; Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **2011**, *332* (6034), 1196–1201.

(33) Qian, L.; Winfree, E.; Bruck, J. Neural Network Computation with DNA Strand Displacement Cascades. *Nature* **2011**, *475* (7356), 368–372.

(34) Wang, C.; Bae, J. H.; Zhang, D. Y. Native Characterization of Nucleic Acid Motif Thermodynamics via Non-Covalent Catalysis. *Nature Communications* **2016**, *7* (1), 10319.

(35) Khara, D. C.; Schreck, J. S.; Tomov, T. E.; Berger, Y.; Ouldridge, T. E.; Doye, J. P. K.; Nir, E. DNA Bipedal Motor Walking Dynamics: An Experimental and Theoretical Study of the Dependency on Step Size. *Nucleic Acids Research* **2018**, *46* (3), 1553–1561.

Chapter 3

Expanding detection windows for discriminating single nucleotide variants using rationally designed DNA equalizer probes

Contribution statement

The contents of Chapter 3 were modified from published paper "Expanding detection windows for discriminating single nucleotide variants using rationally designed DNA equalizer probes". Reprinted with permission from Wang, G. A.; Xie, X.; Mansour, H.; Chen, F.; Matamoros, G.; Sanchez, A. L.; Fan, C.; Li, F., Expanding detection windows for discriminating single nucleotide variants using rationally designed DNA equalizer probes. *Nat. Commun.* **2020**, 11, 5473. Copyright 2020 Springer Nature.

G.A.W. and F.L. conceived idea and designed all experiments. G.A.W. performed all simulation experiments. G.A.W., X.X., H.M., F.C. performed all wet-lab experiments. G.M. and A.L.S. collected clinical samples and performed sample exaction and analysis. All authors contributed to the data analyses. G.A.W., F.C., and F.L. wrote the manuscript and all authors contributed to the revision of the manuscript.

3.1 Introduction

Hybridization of complementary nucleic acid strands through specific and predictable Watson–Crick base pairs plays central roles in genomics research,[1,2] medical diagnostics,[3-6] and DNA nanotechnology.[7-11] Synthetic nucleic acid hybridization probes and primers have been adopted virtually in all technology platforms to ensure the specific recognition, capture, detection, or assembly of nucleic acid sequences.[12-25] Strategies that combine experimental and simulation approaches to guide the design and operation of nucleic acid hybridization probes are highly effective but remain limited.[11,16-19] Herein, we describe a class of simulation-guided nucleic acid hybridization approach, termed DNA equalizer gate (DEG), which drastically expands detection windows for discriminating single nucleotide variants (SNVs) in double-stranded DNA (dsDNA).

A robust hybridization probe shall be both sensitive and specific. However, discrimination of SNVs is challenged by the intrinsic thermodynamic properties of hybridization reactions, where a trade-off between sensitivity and specificity exists.[16] Therefore, extensive experimental and simulation studies have been focused on obtaining an optimal trade-off between hybridization yield (sensitivity) and sequence selectively (specificity) through the design of various frustrated hybridization probes such as molecular beacons[26,27,28] and toehold-exchange probes,[11,16,29] or through tuning experimental conditions such as temperature and denaturing reagents.[30,31] However, with current approaches, detection signals increase monotonically with increases in concentrations of both correct and spurious targets. For any observed detection signal, it may correspond to a low concentration of a correct target but could also be a result of a high concentration of a spurious target. Therefore, a detection window exists between the

target concentrations that lead to the same level of detection signals (Fig. 3.1a). Current frustrating hybridization probes pursue enlarged detection windows by increasing energy barriers for generating detection signals (Fig. 3.1b). For example, longer stem domains and allosteric inhibitors have been introduced to molecular beacons to expand their detection windows.[32,33,34,35] Enlarged detection windows are also achievable in toehold-exchange probes by elongating the reverse toehold (Fig. 3.2). However, the success of current strategies is at the cost of shifting detection windows towards the higher concentration end, inevitably sacrificing sensitivities at lower concentration range. Different from existing approaches, DEG offers a paradigm to expand the detection window without compromising sensitivity at lower concentration range by transforming the quantitative relationship between the detection signal and target concentrations from a monotonic sigmoidal function to an asymmetric unimodal function (Fig. 3.1c).

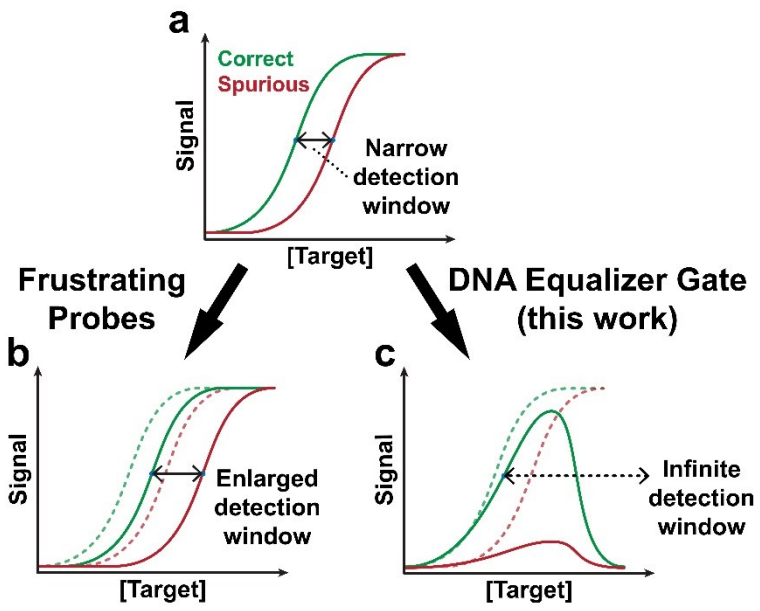


Figure 3.1. (**a**) Typical titration curves for analyzing correct and spurious targets using frustrating nucleic acid probes such as toehold exchange or molecular beacon. Detection signals for both correct and spurious targets increase monotonically with increases in target concentrations. As such, a detection window exists between target concentrations that lead

to the same detection signal. (**b**) Enlarged detection windows may be achieved by shifting the trade-off between sensitivity and specificity through a stronger reverse toehold in a toehold-exchange beacon or a longer stem domain in a molecular beacon. (**c**) DEG expands the detection window by physically transforming the quantitative relationship between the detection signal and target concentrations through a DNA computing module. As a result, detection signals for a spurious target are suppressed throughout the concentration ranges, which ensures the high specificity and robustness for discriminating SNVs.

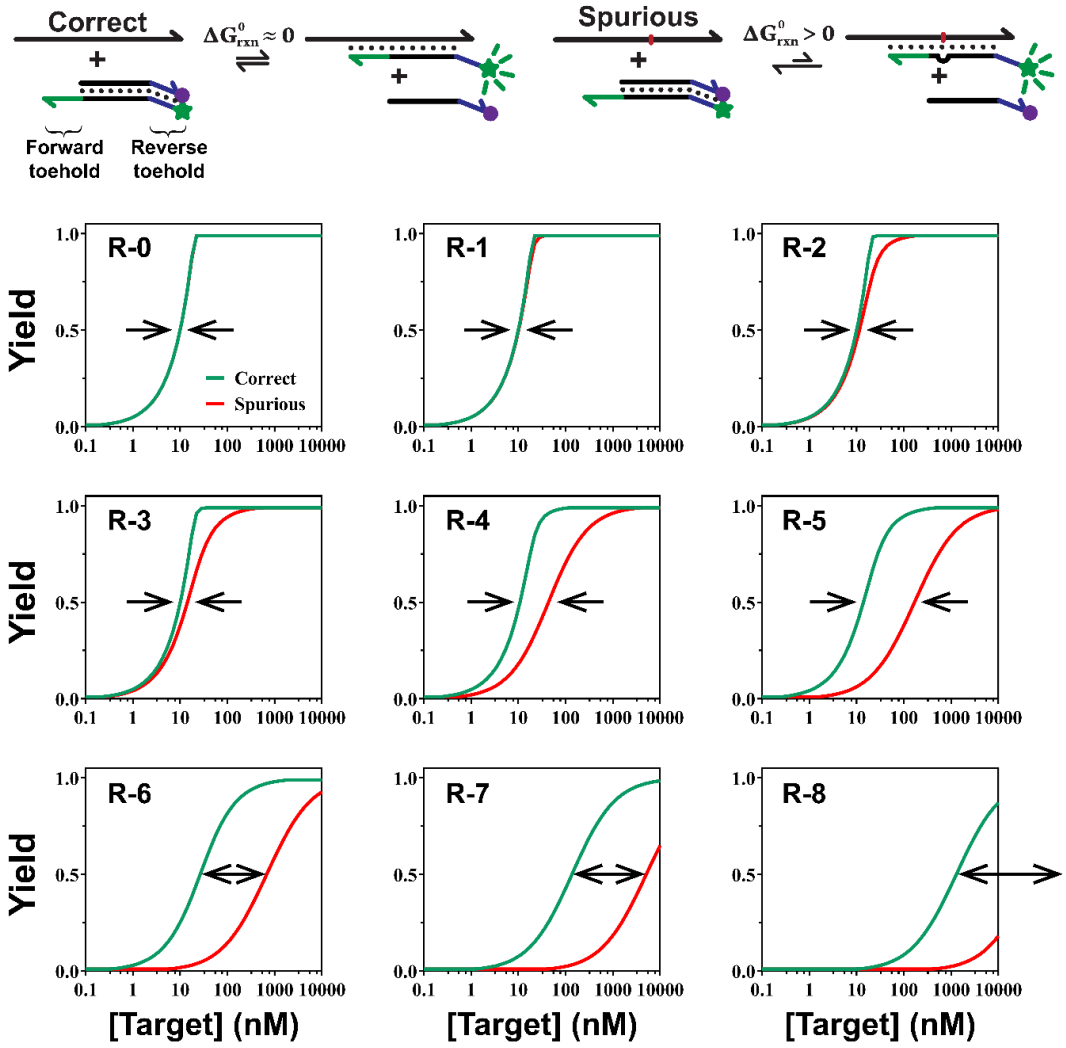


Figure 3.2. Propagation of detection windows in toehold-exchange probes. Simulation demonstrates that the detection windows (indicated by the arrows) for discriminating a pair of correct and spurious targets ($\Delta\Delta G^{\circ}_{rxn}$ = 2.28 kcal/mol) can be enlarged for by increasing energy barriers via the elongation of reverse toeholds of toehold-exchange probes (r = 0 nt to r = 8 nt). The length of forward toehold is fixed at 7 nt. The simulation also reveals that

the enlarged detection window is also a cost of shifting the entire window to the higher concentration end.

DEG is a DNA computing module that acts on dsDNA and allows user-definable transformation of the quantitative relationship between detection signals and target concentrations (Fig. 3.3a). Through the transformation, DEG drastically expands detection windows for discriminating SNVs in dsDNA to as much as infinite. DEG also possesses an intrinsic self-competing mechanism that further improves the sequence selectivity. A thermodynamic-driven mathematical model was constructed, where the detection window, reaction yield, and sequence specificity can be precisely simulated and predicted in silico. The practical usefulness of DEG was established through the integration of polymerase chain reaction (PCR) for the simultaneous infection detection and drug-resistance screen in clinical parasitic worm samples collected from school-age children residing in endemic rural areas of Honduras.

3.2 Results

3.2.1 Design principle

The goal of designing DEG is to maximize detection windows for discriminating single nucleotide variants by suppressing the detection signals for spurious targets through the transformation of the quantitative relationship between detection signals and target concentrations. To quantitatively describe the detection window, we introduce a Robustness Factor (RF) that is defined as the ratio of concentrations between a spurious and a correct target generating the same level of detection signal, $RF = [T]_{spurious}/[T]_{correct}$ (Fig. 3.3b). As such, the greater the RF value, the wider the detection window. Although DEG acts on dsDNA, the detection of single-stranded DNA (ssDNA) can also be considered as a special case of DEG, where the concentrations of DNA Equalizer Probes (DEPs) approach infinite.

The workflow and principle of DEG are illustrated in Fig. 3.3. A double-stranded input AB produces a single-stranded target A and its complementary sequence B through a rapid heating at 95 °C and then snap cooling to 0 °C in a thermal protocol. B is then consumed by DEPs that are of the identical sequences with A forming three-stranded complex BCD (Fig 3.3b). The yield (η) of A is thus determined quantitatively by the concentration of DEPs. When the concentration of AB is less than those of DEPs, A is the predominant product, although a competition exists between A and DEPs for hybridizing to B (Fig. 3.3c). When the concentration of AB is greater than those of DEPs, unconsumed B will rehybridize with A in renaturation process (Fig. 3.3d). Therefore, a maximum yield exists when the concentration of AB equals to those of DEPs. Finally, the net A is quantified using a toehold-exchange reporter which is designed to be sensitive to SNVs.

As each DEP is designed to only contain the sequence of either the toehold domain or the branch migration domain of the reporter, no fluorescence signal can be produced in the absence of the target (Fig. 3.4). Through DEG, a conventional sigmoid detection curve of hybridization probes is transformed into an asymmetric unimodal one (Fig. 3.3d).

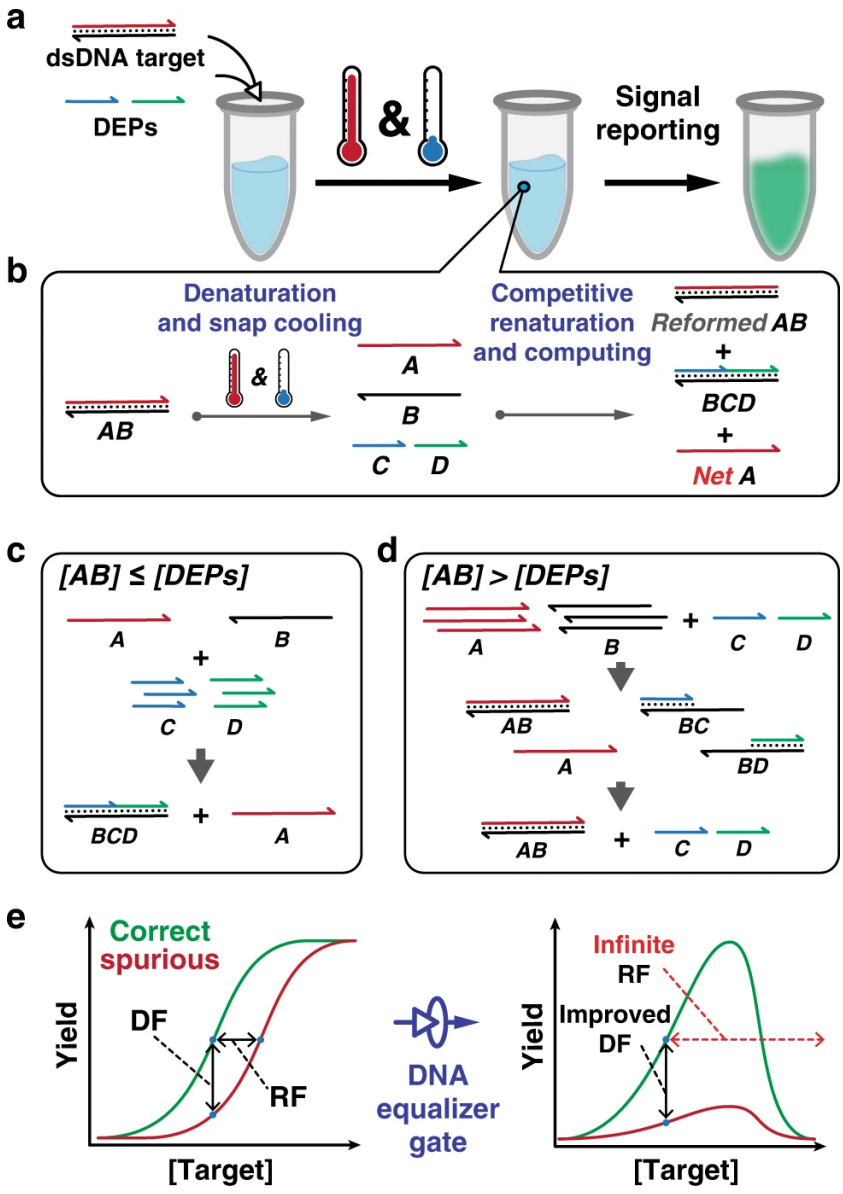


Figure 3.3. (**a**) The overall workflow for quantifying dsDNA using DEG. A mixture of target dsDNA and DNA Equalizer Probes (DEPs) is heated and rapidly cooled to produce ssDNA outputs with well-controlled quantity using an autonomous molecular computation in the test tube. Fluorescence signals are then generated via a reporter probe. (**b**)

Mechanistically, the dsDNA target (AB) is denatured into A and B during a heating and snap cooling procedure. A competition between DEPs (C and D) and A then occurs for hybridizing with B during renaturation. The net amount of the ssDNA output (A) is quantitatively determined by an autonomous computing process that compares the initial concentrations between the target and DEPs. (**c**) When [AB] ≤ [DEPs], the reaction between B and DEPs (i.e. the formation of BCD) is thermodynamically favored, which maximizes the production of A. (**d**) When [AB] > [DEPs], BC and BD are generated as intermediates, which then consumes A through strand displacement. (**e**) Through this computing process, DEG transforms the quantitative relationship between the detection signal and target concentrations from a typical sigmoidal function to a unimodal function. As such, detection signals for a spurious target is significantly suppressed, enabling a much-enlarged detection window and improved discrimination factor (DF).

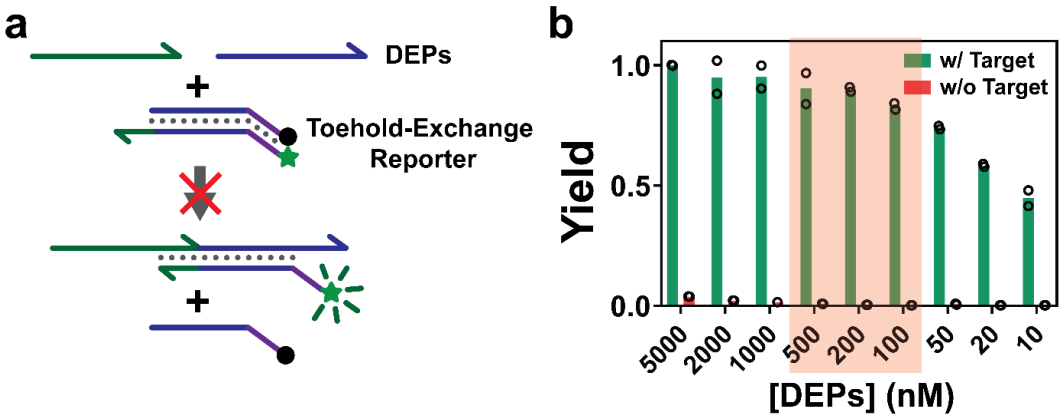


Figure 3.4. Estimation of signal leakage from DEPs. A possible source of fluorescence background is the signal caused by the interaction of DEPs and the reporter probe. Therefore, we estimated the signal leakage at varying DEP concentrations. (**a**) Schematic illustration of the signal leakage caused by the interaction among DEPs and the reporter. (**b**) Estimated leakage (w/o Target) as a function of DEP concentrations, which is also compared to the target-specific fluorescence (w/ Target). The target concentration is fixed at 10 nM, concentrations of DEPs are varied from 10 nM to 5 μM. No apparent fluorescence signal was observed when treating the fluorescence probe with up to 1 μM DEPs, suggesting that there was no cross-reaction between DEPs and the probe and thus there was no competition between DEPs and ssDNA output for the probe. Bars represent the mean of individual replicates (circles, n=2).

Comparing to existing strategies *via* manipulation of energy barriers of frustrating probes, transformation of the quantitative relationship between detection signals and target

concentrations from a sigmodal function to an assymetric unimodal one offers three distinct advantages. First, the transformation only suppresses detection signals at higher concentration end. As such, it allows the dramatic expansion of detection windows without compromising the sensitivity at the lower concentration end (Fig. 3.5). Second, the manipulation of detection window is user-definable and can be achieved at any target concentration. In principle, correct and spurious targets achieve maximum yields simultaneously in DEG, both of which are governed by the concentration of DEPs. As such, a detection window is definable and tunable by simply altering the concentration of DEPs. Moreover, the detection signal for a correct target remains to be much higher than that of a spurious one throughout concentration ranges (RF = ∞, Fig. 3.3e right), whereas the detection window for a conventional probe is much narrower (Fig. 3.3e. left, and Fig. 3.5). Third, as detection signals for the spurious target is significantly suppressed, discrimination factor (DF) is significantly enhanced through a wide concentration range (Fig. 3.3e right). At molecular level, B serves as a molecular sink that competitively consumes A regardless the identity or the position of the mutation, which is significantly different from existing strategies harnessing molecular sinks or reservoirs[36,38] that are designed specifically for known mutations. To quantitatively simulate and predict the effectiveness of DEG for expanding the detection window and for improving sequence selectivity, a theoretical model was established and detailed in next section.

3.2.2 Comparison between DEG and increase in energy barrier for expanding the detection window

An already established detection window for discriminating a single nucleotide mismatch (left in Fig. 3.5) can be enlarged either through increase in energy barriers for activating the probe (right, bottom in Fig. 3.5) or using our DEG approach (right, up in Fig. 3.5). As demonstrated by the simulation results in Fig. 3.5, our DEG approach works better in both the degree of expansion (essentially to infinite) and more sensitive at low concentration range.

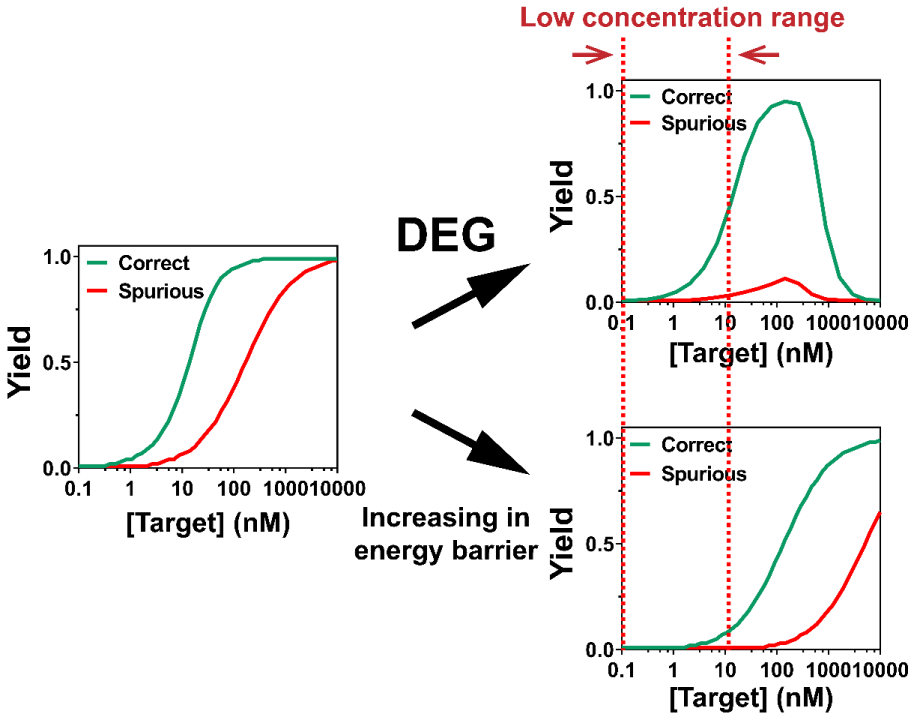


Figure 3.5. Simulation results of enlarged detection windows achieved through DEG (Top) and the increase in the energy barrier for activating the toehold-exchange probe (Bottom). Increase in the energy barrier is achieved by elongating the reverse toehold by 2 bp.

3.2.3 Experimental validation and optimization of DEG

Having re-examined the toehold-exchange probe, we developed the DNA Equalizer Gate method. To study and optimize the design, we have tested various experimental conditions/variables, as shown in this section.

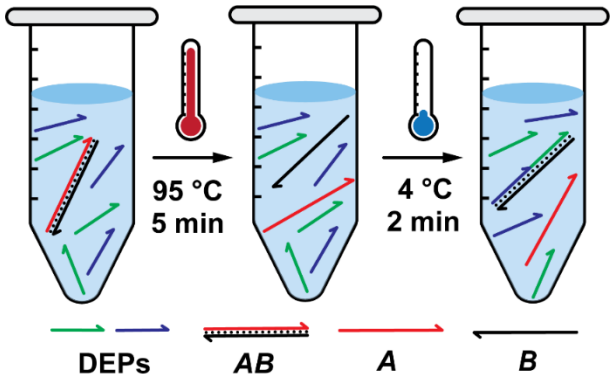
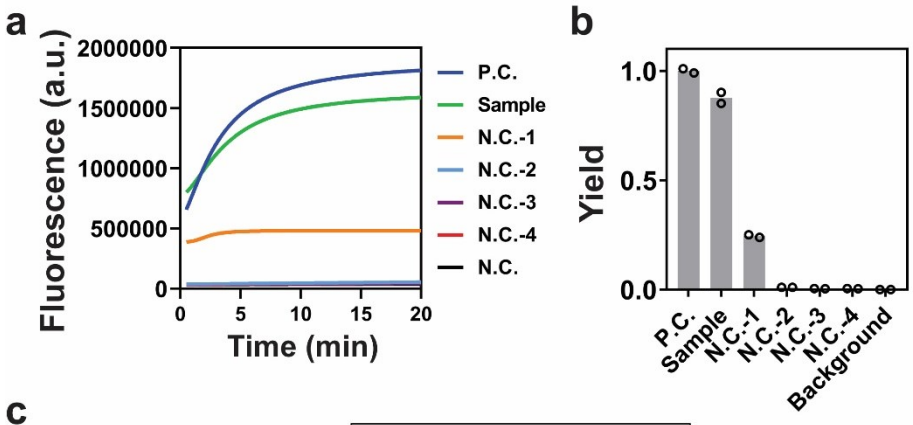


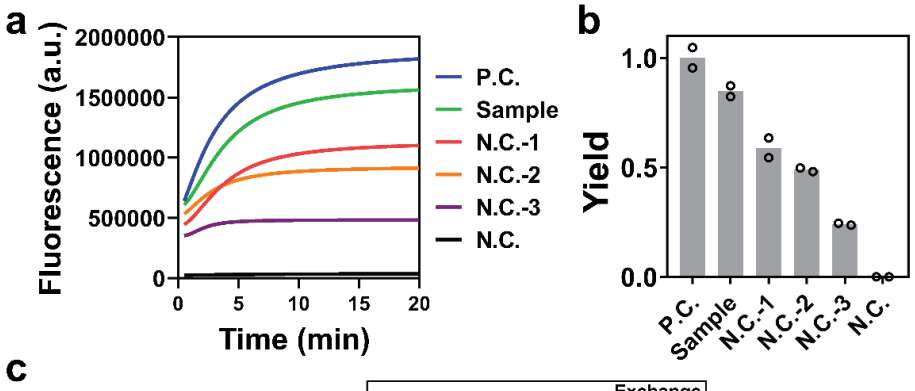
Figure 3.6. Scheme illustration of the experimental procedures and DNA reactions in DEG. Double-stranded DNA (dsDNA) target AB is mixed with DEPs at a proper ratio. The solution was then heated to 95 °C, where AB is denatured into A and B. The reaction is then quenched using a snap cooling protocol that cools the reaction to 4 °C within 2 min. DEPs then competitively bind to B, which generates A as the output for the subsequent quantitative analysis.



	Target	DEPs	Thermal Protocol	Exchange Reporter
P. C.	ssDNA			✓
Sample	dsDNA	✓	✓	✓
N.C. 1	dsDNA	✗	✓	✓
N.C. 2	dsDNA	✓	✗	✓
N.C. 3	—	✓	✓	✓
N.C. 4	—	✓	✗	✓
N.C.	—			✓

Figure 3.7. Characterization the effect of DEPs and thermal protocols on the performance of DEG. Both DEPs and thermal protocols are critical to ensure the high yield of DEG for producing single-stranded output A. (**a**) Real-time fluorescence monitoring the kinetics of the reporter probe for measuring A produced by DEG. (**b**) Yield of each reaction in terms of producing A measured by endpoint fluorescence at 20 min. The yield was calculated by

setting fluorescence of the positive control as 1. Bars represent the mean of individual replicates (circles, n=2). (**c**) Detailed reactants and experimental procedures in each sample or control. Each sample containing 10 nM AB, 20 nM reporter, and 200 nM DEPs in 1 × Tris buffer containing 1 mM Mg^{2+} and 0.1% of Tween 20 (v/v) was incubated at 37 °C.



c

	Target	DEP-1	DEP-2	Exchange Reporter
P. C.	ssDNA			✓
Sample	dsDNA	✓	✓	✓
N.C. 1	dsDNA	✓	✗	✓
N.C. 2	dsDNA	✗	✓	✓
N.C. 3	dsDNA	✗	✗	✓
N.C.	—			✓

Figure 3.8. Effect of DEPs on the performance of DEG. Each DEP probe was found to partially complement with B and thus facilitate the conversion of AB to A. However, maximum yield can only be achieved when both DEPs are present in the reaction. (**a**) Real-time fluorescence monitoring the kinetics of the reporter probe for measuring A produced by DEG. (**b**) Yield of each reaction in terms of producing A measured by endpoint fluorescence at 20 min. Bars represent the mean of individual replicates (circles, n=2). (**c**) Detailed reactants and experimental procedures in each sample or control.

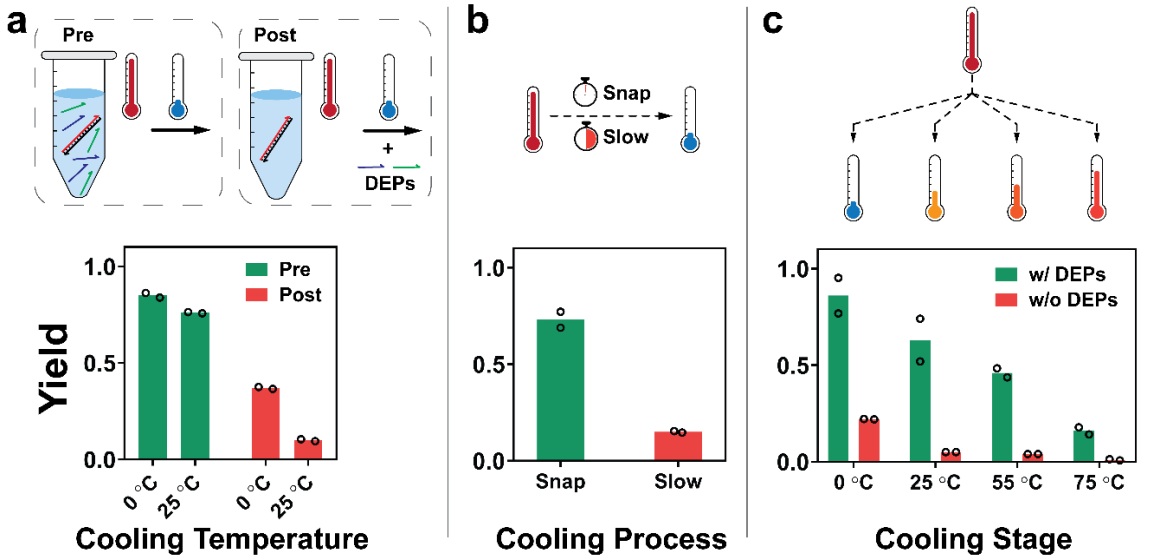


Figure 3.9. Optimization of the thermal protocol for the denaturation and renaturation processes. Maximal yield is established when all DNA species pre-mixed in the same test tube and undergo a heating followed by a snap cooling step to 4 °C. (**a**) The addition of DEPs before (pre) or after (post) the thermal protocol was found to significantly affect the yield of A. A premixing of DEPs and the target was chosen as the optimal procedure, as it both improves the reaction yield and simplifies the operation (**b**) A snap cooling step was also found to be critical to ensure the high yield of A. (**c**) Maximal yield was achieved when a snap cooling to 4 °C was used as a final temperature. Increase final temperature to 25, 55 and 75 °C (blue, yellow, orange and red thermometer, respectively) were found to gradually reduce the yield of the reaction. Bars represent the mean of individual replicates (circles, n=2).

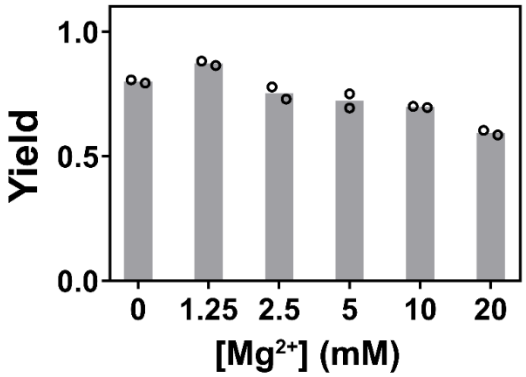


Figure 3.10. Effect of Mg^{2+} on the performance of DEG. The DEG was found to be robust to Mg^{2+} in range of 0 to 10 mM. A slight decrease in reaction yield was found when increasing Mg^{2+} concentration to 20 mM, as high concentration of Mg^{2+} may favor the

formation of AB by accelerating renaturation. Bars represent the mean of individual replicates (circles, n=2).

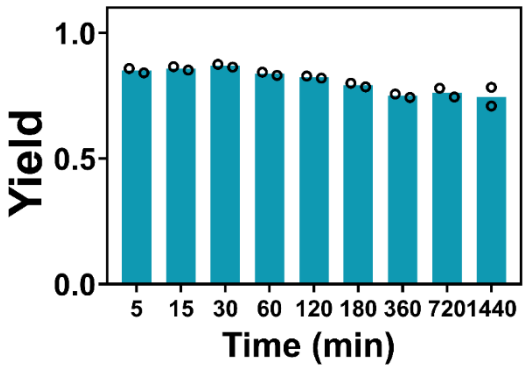


Figure 3.11. Stability of the output ssDNA *A* produced by DEG. Once generated by DEG, B is blocked by DEPs and thus unable to react with A through renaturation. To accurately quantify AB through DEG and the reporter, it is critical to ensure the stability of free A in the reaction mixture. We monitored the concentration of A in the solution after DEG reaction at 5 min, 15min, 30 min, 1 hr, 2 hr, 3 hr, 6 hr, 12 hr, and 24 hr, respectively, at room temperature. Our result demonstrates that A is highly stable, with no apparent losses in the first 2 hrs. Practically, we analyze A using the reporter probe within the first 30 min. Bars represent the mean of individual replicates (circles, n=2).

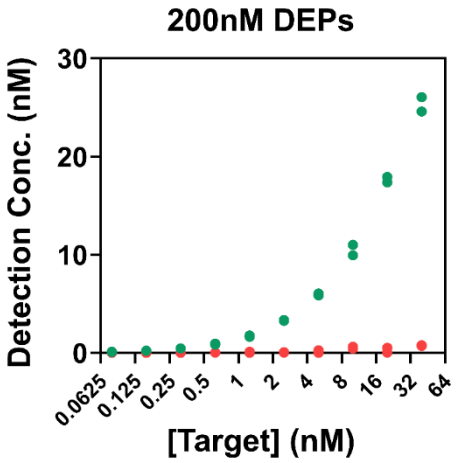


Figure 3.12. Estimated the LOD for detecting *AB* using DEG. The LOD was estimated to be 0.5 nM when 200 nM DEPs were used to generate single-stranded output A. Values of individual replicates (n=2) are shown as dots.

3.2.4 Theoretical model

Here, a mathematical model was introduced to quantitatively profile DEG by taking all possible reactions into consideration (Fig. 3.13a). To derive the yield of each DNA species in this reaction network as a function of both sequence design ($\Delta\Delta G^{\circ}$) and equalizer probe concentrations, a set of eight equilibrium equations need to be solved. However, we found that these equations were coupled to one another, which was mathematically difficult to solve. Therefore, a stoichiometric matrix RM was introduced to help simplify the calculation (Fig. 3.13a), where the first four rows were ranked to be essential (details in Section 3.2.5). This essential set of equilibrium equations was then solved by a numerical approach, where distributions of A and AB were solved as a function of the target concentration and plotted in Fig. 3.13c.

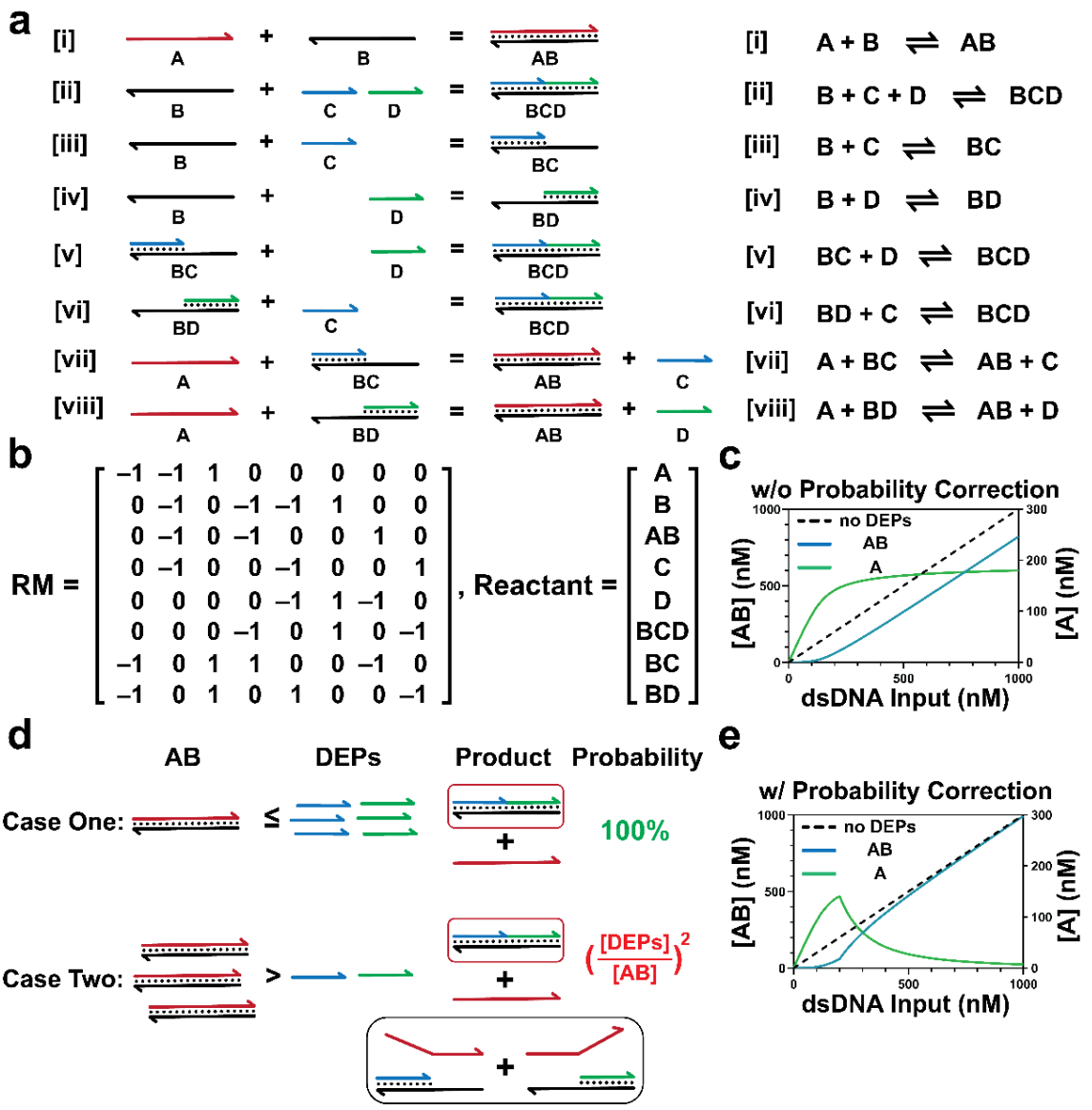


Figure 3.13. Theoretical model of DNA Equalizer Gate (DEG). (**a**) Schematic illustration of all possible elemental reactions occurring in the DEG. (**b**) Linearization of the complex reaction network in DEG into $0 = \mathbf{RM} \cdot \mathbf{Reactant}$ to extract the independent equations, where RM is the stoichiometric matrix and Reactant represents DNA species. The rank of RM was determined to be 4, indicating that four independent equilibrium equations need to be solved. As such, reactions [i - iv] were chosen to build mathematical model. (**c**) In silico prediction of the yields of A and AB as a function of the concentration of dsDNA target without performing the probability correction. (**d**) Schematic illustration the need for probability correction when [AB] > [DEPs]. As each DEP binding is an independent event, the multiplicity rule was applied here. When [AB] ≤ [DEPs], the probability that two DEPs bind to the same B is 100 %. However, when [AB] > [DEPs], the probability becomes dependent on the ratio between the initial concentrations of AB and DEPs, where

$\text{Probability} = ([\text{DEPs}]_0/[\text{AB}]_0)^2$ (**e**) In silico prediction of the yields of A and AB as a function of the concentration of dsDNA target with the probability correction.

The thermodynamic-driven model successfully predicted the distribution of A and AB at the concentration range, where [AB] > [DEPs] (Fig. 3.13c). However, it failed to simulate the thermodynamic behavior of DEG when [AB] ≤ [DEPs]. We found that a probability function that took the possible distributions of DEPs on B was necessary to correctly reflect the final equilibrium distribution of each DNA species (Fig. 3.13d and Fig. 3.17). Mathematically, the probability for the successful formation of a BCD complex is $([\text{DEPs}]_0/[\text{AB}]_0)^2$ (Fig. 3.13d). The combination of the thermodynamic-driven model with probability correction leads to a characteristic asymmetric unimodal curve (Fig. 3.13e), which was also confirmed experimentally (details in the next section).

3.2.5 Theoretical framework and mathematical simulation

The detection window reflects the performance of nucleic acid hybridization probes across wide concentration ranges. Concentration is a vital variable in evaluating the sensitivity and specificity of DNA hybridization probes, yet the concentration dependency of yield, discrimination factor (DF), and robustness factor (RF) remains unexplored in a systematic manner. We first set out to profile the concentration-dependency of hybridization yield and sequence specificity across a wide concentration range in silico. Toehold-exchange probes are chosen as a testbed in our system. To highlight the numerical relationship between variables, we applied a dimensionless transformation to all concentrations prior to derivations.

Concentration dependence and robustness. A toehold exchange reaction can be simplified to a bimolecular reversible reaction (**eq. 3.1**).

$$T + CP \rightleftharpoons TC + P, with\ \Delta G^{\circ}_{rxn};\ \text{(eq. 3.1)}$$

where *T* is the target, *C* is the partial complementary strand to *T* and *P* is the protecting strand for *C*. Thermodynamics of a toehold-exchange probe is tunable by altering the lengths of forward and reverse toehold or by controlling the stoichiometry between *CP* and *P*. The free energy of each reactant and product can be calculated using NUPACK software.

Zhang and colleagues have previously defined the yield of the reaction as $\eta^* = \frac{[TC]}{\min([T]_0,[CP]_0)}$, where $[T]_0$ and $[CP]_0$ are the initial concentrations of *T* and *CP*, respectively.[2] We reason that this definition is useful for guiding the sequence design of toehold-exchange probes but not suitable for predicting the analytical behaviors of probes at specified experimental conditions, because the initial target concentration is often an unknown variable in the system. Practically, the concentration of *CP* is fixed, and *T* is a variable, so we defined the reaction yield as $\eta = \frac{[TC]}{[CP]_0}$. We also chose $[CP]_0$ as the characteristic concentration for dimensionless transformations.

For a typical reversible reaction, the equilibrium constant can be derived from the reaction free energy ΔG° (**eq. 3.2**) and the concentration of all nucleic acid species follows the rule of mass conservation.

$$K_{eq} = e^{-\Delta G^0/RT} = \frac{[TC]_{eq}\cdot[P]_{eq}}{[T]_{eq}\cdot[CP]_{eq}} = \frac{[TC]_{eq}\cdot([P]_0+[TC]_{eq})}{([T]_0-[TC]_{eq})\cdot([CP]_0-[TC]_{eq})};\ \text{(eq. 3.2)}$$

We further dimensionless the formula converting concentrations into numeric values, where the target concentration in its dimensionless form is denoted as τ and that of $[P]_0$ denoted as γ, respectively (**eq. 3.3**).

$$K_{eq} = e^{-\Delta G^0/RT} = \frac{[TC]_{eq}\cdot([P]_0+[TC]_{eq})/[CP]_0^2}{([T]_0-[TC]_{eq})\cdot([CP]_0-[TC]_{eq})/[CP]_0^2}$$

$$= \frac{\eta\cdot(\gamma+\eta)}{(\tau-\eta)\cdot(1-\eta)};\ \textbf{(eq. 3.3)}$$

$$where\ \gamma := {[P]_0}\Big/{[CP]_0}\ and\ \tau := {[T]_0}\Big/{[CP]_0}\ respectively.$$

The concentration dependence of η is solved by equation S4:

$$\eta = \frac{K_{eq}+\gamma+K_{eq}\tau-\sqrt{K_{eq}^2\tau^2-2K_{eq}^2\tau+K_{eq}^2+2K_{eq}\gamma\tau+2K_{eq}\gamma+4K_{eq}\tau+\gamma^2}}{2(K_{eq}-1)}$$

$$= \frac{(K_{eq}+\gamma+K_{eq}\tau)-\sqrt{(K_{eq}+\gamma+K_{eq}\tau)^2-4K_{eq}\tau(K_{eq}-1)}}{2(K_{eq}-1)};\ \textbf{(eq. 3.4)}$$

$$where\ K_{eq} \neq 1.$$

When the $K_{eq} = 1$, $\eta = \frac{\tau}{\gamma+\tau+1}$. **(eq. 3.5)**

To quantitatively describe and compare sequence specificity, discrimination factor (DF) is commonly employed, where $DF = \eta_{correct}/\eta_{spurious}$. In a general case where neither correct ($K_{eq,c}$) nor spurious ($K_{eq,s}$) target has equilibrium constant as 1, the discrimination factor (DF) is expressed as:

$$DF = \frac{\eta_c}{\eta_s} = \frac{(K_{eq,s}-1)\cdot\left\{(K_{eq,c}+\gamma+K_{eq,c}\tau)-\sqrt{(K_{eq,c}+\gamma+K_{eq,c}\tau)^2-4K_{eq,c}\tau(K_{eq,c}-1)}\right\}}{(K_{eq,c}-1)\cdot\left\{(K_{eq,s}+\gamma+K_{eq,s}\tau)-\sqrt{(K_{eq,s}+\gamma+K_{eq,s}\tau)^2-4K_{eq,s}\tau(K_{eq,s}-1)}\right\}};\ \textbf{(eq. 3.6)}$$

For a well-designed probe where $K_{eq,c}$ is tuned to 1, the DF formula can be simplified as:

$$DF = \frac{\eta_c}{\eta_s} = \frac{2(K_{eq,s}-1)\cdot\tau}{(\gamma+\tau+1)\cdot\left\{(K_{eq,s}+\gamma+K_{eq,s}\tau)-\sqrt{(K_{eq,s}+\gamma+K_{eq,s}\tau)^2-4K_{eq,s}\tau(K_{eq,s}-1)}\right\}}; \quad \textbf{(eq. 3.7)}$$

Practically, the DF value for a pair of correct and spurious targets is a function of both sequence design (ΔG and K_{eq}) and target concentration τ (**eq. 3.6, 3.7**).

The mathematical prediction of DF for a correct target against five spurious targets as a function of target concentration (τ) is shown in Fig. 3.14a. DF values for all spurious targets containing single nucleotide mutations decrease monotonically with increasing target concentration τ. The simulated DF values maximize at the range of low target concentrations when the yield of the spurious target approaches extreme small values. However, when the concentration of the target is approach or below the limit of detection (LOD) of a specified analytical technique, the numeric values of yield and DF become meaningless. Despite the high DF, our simulation reveals that the absolute differences in yield between correct and spurious targets become much less at the concentration range $\tau < 0.6$ and are thus difficult to be resolved experimentally (Fig. 3.14b). Therefore, we corrected our mathematical model by introducing the LOD of the analytical method into the simulation. The LOD can be defined arbitrary as a minimal yield that allows the experimental differentiation of signals generated by the correct or spurious targets from background. Figure 3.14c shows the simulation of DF using corrected model when the spurious becomes non-detectable (LOD was set to 1% yield). Practically, we can also set the LOD as the minimal detectable yield for the correct target.

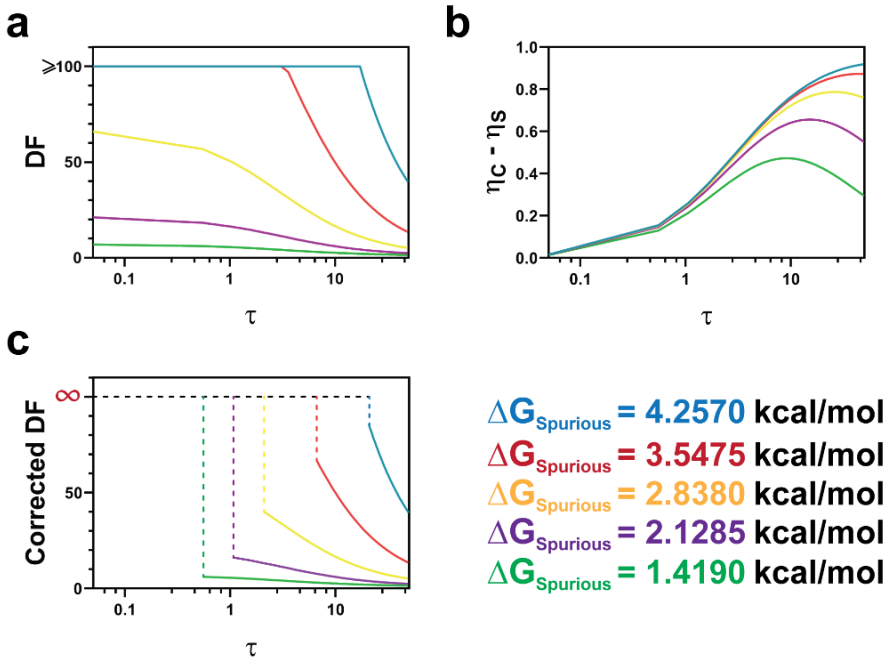


Figure 3.14. The theoretical concentration dependency of discrimination factor (DF). (**a**) Theoretical DF as a function of target concentration for a correct target against five theoretical spurious targets defined by their thermodynamic parameters. Reaction free energies of spurious targets are shown with the same colors with DF curves. (**b**) The differences in yield between a pair of correct and spurious targets as a function of target concentration. (**c**) Simulation of DF by including the correction using LOD.

Robustness factor. To quantitatively describe the detection window for discriminating SNVs, here we mathematically define a robustness factor (RF) which is the concentration ratio of a pair of spurious and correct targets, when their yields are the same. To do so, we first derived target concentration τ as a function of yield and equilibrium constant (**eq. 7**). RF can then be mathematically derived using **eq. 8**.

$$\tau = \frac{\eta \cdot (\gamma + \eta)}{K_{eq} \cdot (1 - \eta)} + \eta; \quad \text{(eq. 3.8)}$$

$$RF = \frac{\tau_{spurious}}{\tau_{correct}} = \frac{\frac{\eta \cdot (\gamma + \eta)}{K_{eq,spurious} \cdot (1 - \eta)} + \eta}{\frac{\eta \cdot (\gamma + \eta)}{K_{eq,correct} \cdot (1 - \eta)} + \eta} = \frac{(1 - \eta) + (\gamma + \eta) \cdot {}^{1}/_{K_{eq,spurious}}}{(1 - \eta) + (\gamma + \eta) \cdot {}^{1}/_{K_{eq,correct}}}; \quad \text{(eq. 3.9)}$$

Under an optimal trade-off between sensitivity and specificity, where $K_{eq,correct} = 1$and the yield is 50%, a practically useful RF can be simplified as:

$$RF = \frac{0.5+(\gamma+0.5)\cdot {}^{1}/_{K_{eq,spurious}}}{0.5+(\gamma+0.5)}; \quad \textbf{(eq. 3.10)}$$

The theoretical RF values increase linearly as a function of η (Fig. 3.15a). However, this is significantly deviated from the experimental observations because the absolute concentration differences ($\tau_S - \tau_C$) between spurious and correct targets become much less significant when the yields of hybridization approach 0 or 100% (Fig. 3.15b). To better reflect the analytical performance, we corrected our model by taking both LOD and limit of linearity (LOL) into consideration. Figure 3.15e shows a corrected RF simulation by setting LOD to be 1% yield and LOL to 95% yield. To understand the concentration-dependency of RF, we further convert the x-axis from the yield η to the concentration of the target τ _correct (Fig. 3.15c-e).

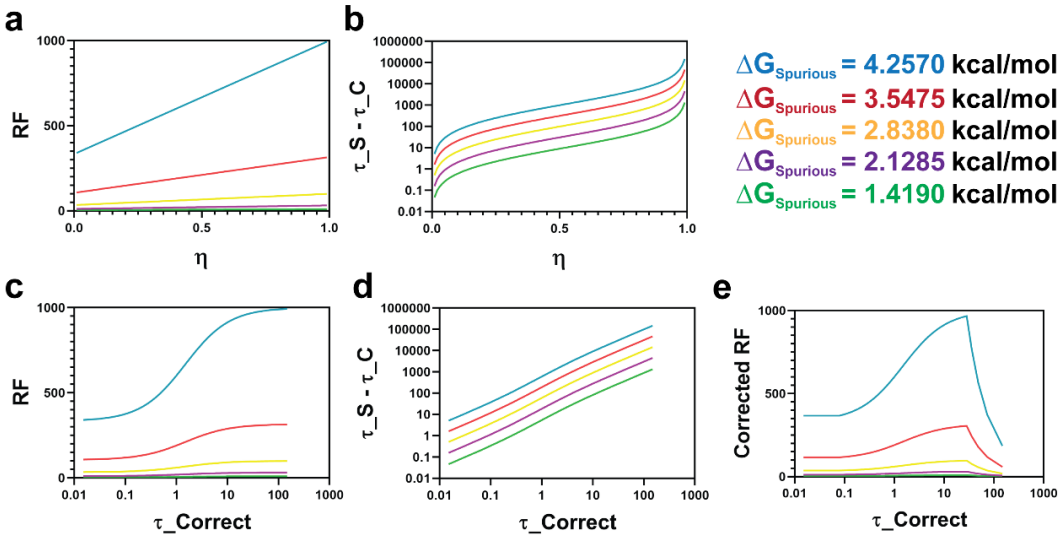


Figure 3.15. The dependence of RF on reaction yield and target concentration. (**a**) The theoretical prediction of RF values as functions of reaction yield η. (**b**) The absolute concentration differences between spurious and correct targets as a function of yield. (**c**)

RF as a function of target concentration. (**d**) Absolute concentration differences at the same yield for the spurious and correct targets as a function of target concentration. (**e**) Corrected RF using LOD and LOL.

Detection window of a toehold-exchange probe. To demonstrate the concentration-dependency of detection window for a toehold-exchange probe, we next simulated the η, DF, and RF using a numeric approach through MATLAB. A 42-nt synthetic DNA (Fig. 3.28) was used as a model target and a single T to A mutation was introduced to create the spurious target. Standard Gibbs free energy (ΔG^0) of each DNA species can be calculated using NUPACK software and ΔG°_{rxn} for each toehold-exchange reaction can thus be calculated as $\Delta G^0_{rxn} = \Delta G^0(TC) + \Delta G^0(P) - \Delta G^\circ(T) - \Delta G^\circ(CP)$. The thermodynamical differences between a pair of correct and spurious targets can be quantified using $\Delta\Delta G^\circ$, where $\Delta\Delta G^\circ = \Delta G^\circ_{rxn}(spurious) - \Delta G^\circ_{rxn}(correct)$. In our model system, $\Delta\Delta G^\circ$ is determined to be at 2.30 kcal/mol. The yield (Fig. 3.16a), sequence selectivity (Fig. 3.16b), and concentration robustness (Fig. 3.16c) can then be predicted for this pair of synthetic sequences in silico. As expected, all three parameters are strongly concentration-dependent (Fig. 3.16a-c), indicating that a well-designed and -optimized toehold-exchange probe may only excel in a certain concentration range.

By further including $\Delta\Delta G^\circ$ as a variable in our model, we were able to simulate the concentration-dependency of a toehold-exchange probe to all possible mutations that are mathematically reflected as varying $\Delta\Delta G^\circ$ values (Fig. 3.16d-f). Our simulation results quantitatively reflect that the detection window is inversely related to the difficulty for discriminating a certain mutation: the smaller the $\Delta\Delta G^\circ$ value, the narrower the concentration robustness range that allows effective discrimination (Fig. 3.16e).

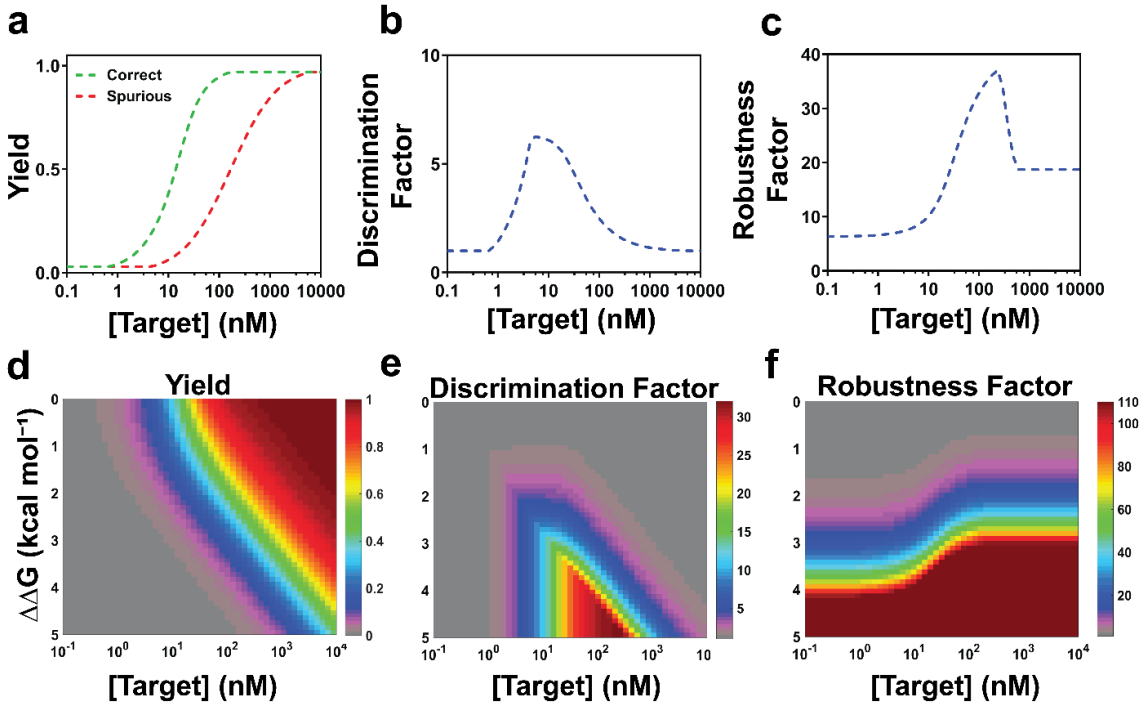


Figure 3.16. Detection window of nucleic acid hybridization probe. (**a-c**) Theoretical prediction of the reaction yield, DF, RF of a pair of correct and spurious targets with $\Delta\Delta G^{\circ} = 2.30\ kcal/mol$ through simulation. (**d-f**) In silico profiling the yield, DF, and RF of all possible mutations with $\Delta\Delta G^{\circ}$ between 0 and 5 kcal/mol.

DNA Equalizer Gate. DEG is designed to convert a dsDNA target into a ssDNA output in a quantitative manner with well-defined detection window. To simulate this process, we consider that all reactions are thermodynamically driven, and all DNA species are in their thermodynamic stable states. Under this assumption, a set of equilibrium equations could be used to predict the concentration distribution of newly formed DNA species (Fig. 3.13a). However, only independent equations need to be solved otherwise meaningless answers will be generated. To help determine independent equilibrium equations, we extract a numerical reaction matrix (RM) from the reaction system:

$$\begin{bmatrix} -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 & 0 & -1 \end{bmatrix};$$

In RM, each row represents a possible chemical reaction and Reactant is the column of all DNA species. The rank of RM is 4 (validated by Matlab), which is less than the dimension of RM. As such, only 4 independent equations are existing in this reaction system and we choose the first four reactions in our model. All ΔG°_{rxn} values are predicted using NUPACK and the equilibrium equations are shown below:

$$\frac{[AB]_{eq}}{[A]_{eq}\cdot[B]_{eq}} = K_{eq,i} = e^{-\Delta G^{\circ}_{i}/RT}$$

$$\frac{[BCD]_{eq}}{[B]_{eq}\cdot[C]_{eq}\cdot[D]_{eq}} = K_{eq,ii} = e^{-\Delta G^{\circ}_{ii}/RT}$$

$$\frac{[BC]_{eq}}{[B]_{eq}\cdot[C]_{eq}} = K_{eq,iii} = e^{-\Delta G^{\circ}_{iii}/RT}$$

$$\frac{[BD]_{eq}}{[B]_{eq}\cdot[D]_{eq}} = K_{eq,iv} = e^{-\Delta G^{\circ}_{iv}/RT}$$

$$where\ [A]_0 = [A]_{eq} + [AB]_{eq};\ [B]_0 = [B]_{eq} + [BCD]_{eq} + [BC]_{eq} + [BD]_{eq};$$

$$[C]_0 = [C]_{eq} + [BC]_{eq};\ [D]_0 = [D]_{eq} + [BD]_{eq};$$

$$[A]_0, [B]_0, [C]_0, and\ [D]_0\ are\ initial\ concentrations;$$

Standard reaction free energies are calculated at 4 °C according to the experimental condition of DEG. When $[\text{Target}] > [\text{DEP}]$, a probability function is introduced to the

model to quantitatively describe the probabilistic binding occurring between DEPs and complementary strand.

Figure 3.17 shows the mathematical prediction of the yield of each DNA species at varying target concentrations. The yield of the output DNA (A) decreases linearly as a function of the concentration of the input target without the correction using the probability function (dashed line in Fig. 3.17a and 3.17b). With probability correction, a sharp transition occurs when [target] equals [DEP], which was also confirmed experimentally. This transition is determined exclusively by the concentration of DEP and thus allows defining the detection window in Fig. 3.17c and suppressing the signals of spurious targets as shown in Fig. 3.17d.

The combination of DEG model with a classic toehold-exchange model allows us to precisely simulate the yield and discrimination factor for the correct target and any given mutation. To simulate RF in DEG system, a build-in mathematical reverse function in Matlab was used to first convert the reaction yield to the concentration of the ssDNA output using the toehold-exchange model and then convert the concentration of ssDNA to that of the dsDNA target using the DEG model.

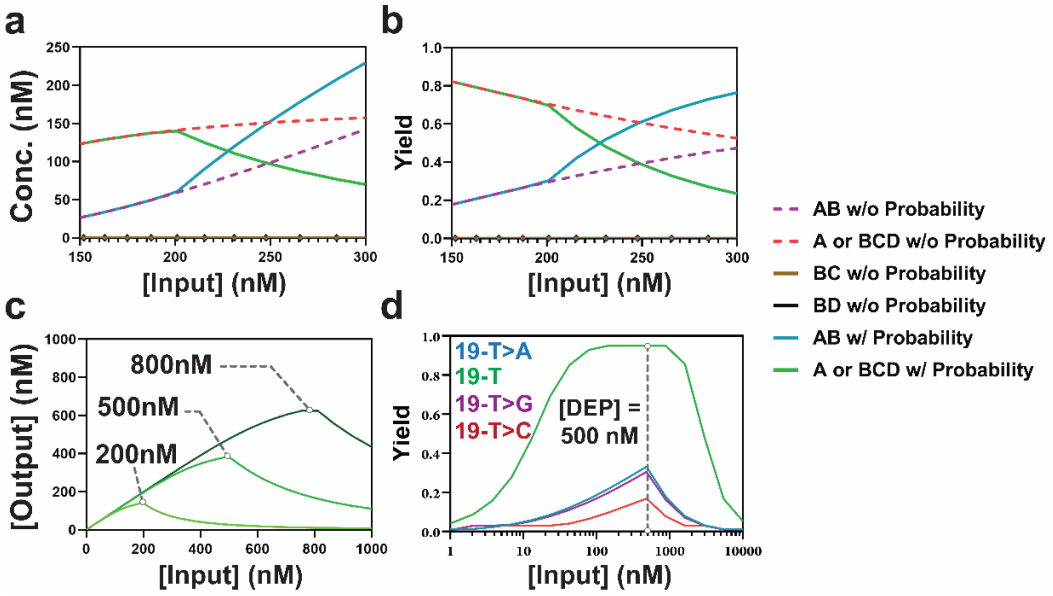


Figure 3.17. Theoretical predictions of yields for producing ssDNA output A from dsDNA input AB using DEG. (**a**) The output ssDNA (A) and input dsDNA (AB) concentrations as a function of the initial input concentration. The probability function plays critical roles to ensure the accuracy of the model. (**b**) The yield of ssDNA (A) as a function of the initial input concentration. (**c**) Effect of DEPs on the performance of DEG for producing ssDNA output. A maximum yield exists for each titration curve, where [Input] = [DEP]. **(d)** Titration curve for analyzing a correct dsDNA target against three single nucleotide mutations. [DEP] = 500 nM.

3.2.6 Parameter correction and fitting

Correction of ΔG°_{rxn}. It was previously found by Zhang and colleagues that corrections to ΔG°_{rxn} values predicted using NUPACK software are necessary to improve the agreement between theoretical prediction and experimental observation. A similar correction was also performed in our study to improve the accuracy of the mathematical prediction (Fig. 3.18). By comparing theoretically predicted and experimentally determined yields at varying ΔG°_{rxn}, a correction of 1.575 kcal/mol was determined and applied through out this study.

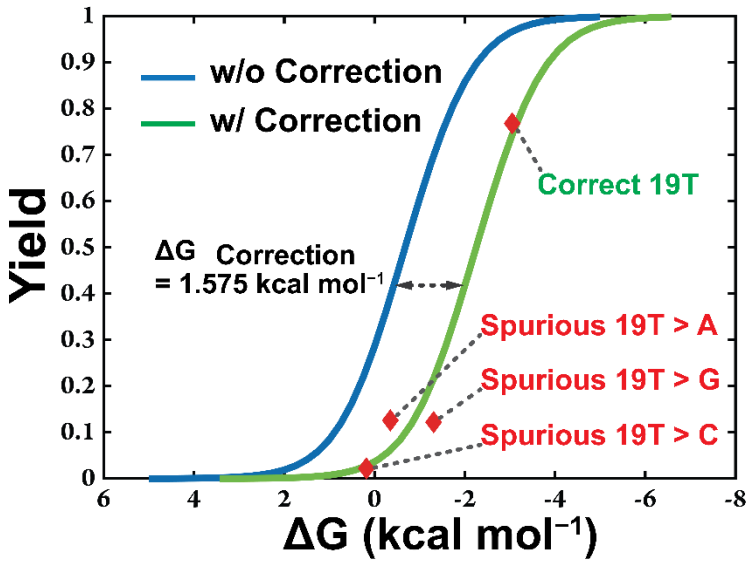


Figure 3.18. Correction of ΔG°_{rxn}. The correction was performed by measuring a correct target and three single nucleotide mutations using a toehold-exchange reporter. The yield can be predicted using predicted ΔG for each DNA species (blue curve). To fit the theoretical curve with experimental results, a correction of -1.575 kcal/mol was found to be necessary. This correction was applied to all simulations throughout this work. [AB] = 10 nM, [reporter] = 20 nM, theoretical ΔG for all DNA species were predicted using NUPACK software.

Determination of experimental RF through fitting. As both calibration curves for the correct and spurious targets were established using scattered data spots, it is not possible to determine the experimental RF directly. Therefore, we combined experimental fitting and mathematical conversion to address this issue (Fig. 3.19). A 4-parameter nonlinear fitting was used first to fit the experimental results. A set of four parameters including *M*, *L*, *s* and *E*, will be determined through the fitting (**eq. 11**). *M* and *L* represents the highest and lowest signals in the curve; *E* represents the concentration of target that gives halfway between maximum and minimum limits; and ***s*** represents the steepness of the fitting curve. Once established this mathematical model through fitting, we were able to convert any yield in a toehold-exchange reaction into a corresponding concentration of either a correct

or spurious target. Experimental RF can then be determined. For a DEG system, a calibration curve needs to be first split into two parts: $[Target] =< [DEP]$ and $[Target] > [DEP]$ (Fig. 3.19).

Nonlinear Model:

$$Y = \frac{L+x^S\cdot(M-L)}{x^S+E^S} \text{ (eq. 3.11)}$$

$$where\ M, L, s\ and\ E\ are\ parameters\ to\ be\ fitted.$$

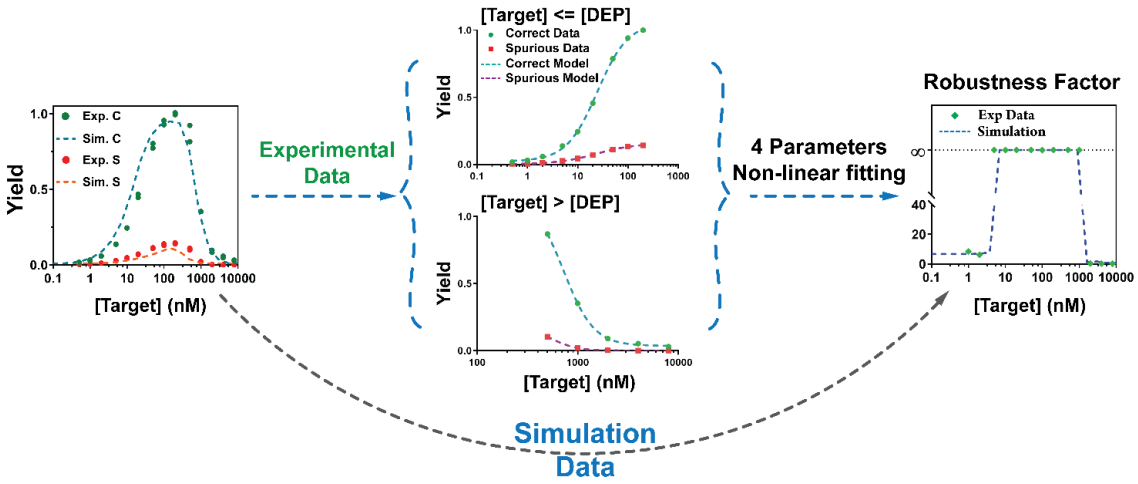


Figure 3.19. Determination of theoretical and experimental RF. The theoretical RF is determined by extracting the concentrations of a pair of correct and spurious targets producing the same yield using Matlab software. The experimental RF is determined by first fitting experimental data in a calibration curve using a 4-parameter non-linear model and then extracting the concentrations of a pair of correct and spurious targets using Matlab. RF at each concentration is then calculated using eq. 6 and plotted as a function of target (correct target) concentrations. Experimentally measured yields are shown in empty dots (n=2).

3.2.7 In silico prediction and experimental validation

Using our theoretical model, η, DF, and RF were firstly quantitatively profiled in silico against three critical factors in DEG, including the target concentration, the sequence

design ($\Delta\Delta G^{\circ}$), and the detection window defined by DEPs. The detection of ssDNA may also be described in our model by setting the concentration of DEPs to be infinite, where the yield for producing A is 100%. Simulation results in Fig. 3.20 depict the theoretical transitions from the detection of ssDNA ([DEPs] = ∞) to the detection dsDNA with varying concentrations of DEPs at 50, 100, 200, and 500 nM. Unlike conventional frustrating probes (toehold exchange or molecular beacon) where η is saturated beyond a certain target concentration (Fig. 3.20a), a maximum η exists in DEG at a single target concentration that is defined exclusively by DEP ($[T]_{max}$ = [DEPs]) and is sequence-independent (Fig. 3.20b). The simulation results also reveal a significant expansion of the detection window where highly specific discrimination of single nucleotide mutations can be achieved (Fig. 3.20d). The level of improved DF is also definable by the concentration of DEP (Fig. 3.20d). As η for high concentrations of SNVs has been suppressed exclusively, a remarkable transition of RF is observed from finite values (Fig. 3.20e) to infinite (Fig. 3.20f).

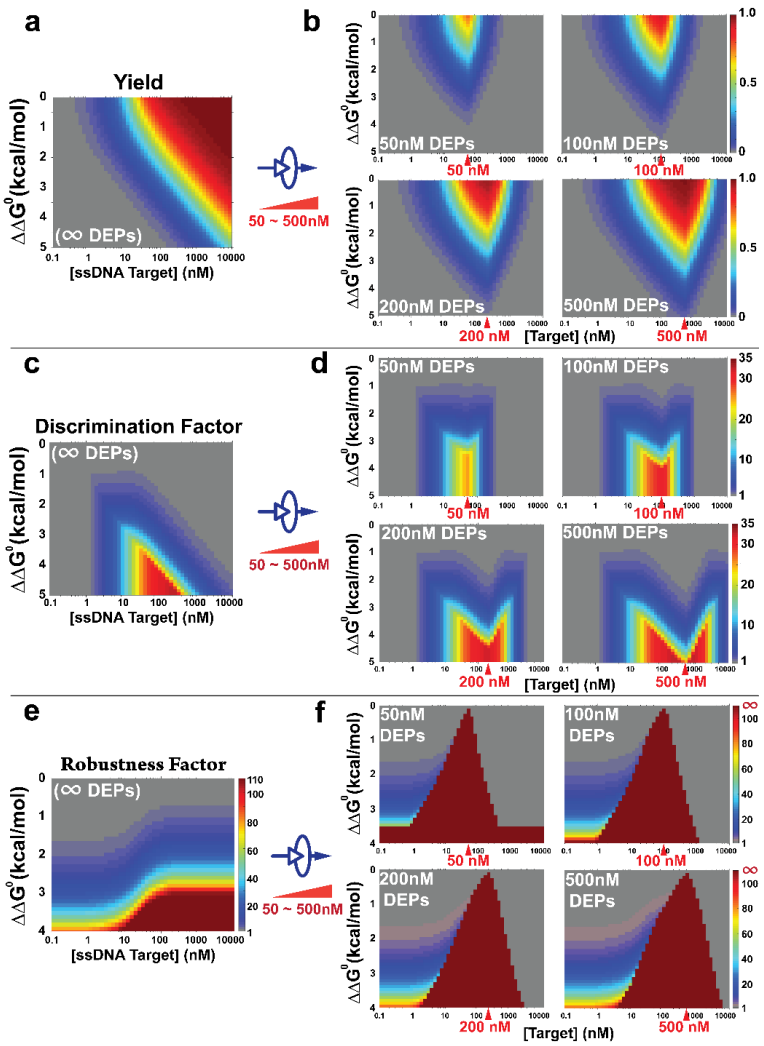


Figure 3.20. Simulation results of DNA Equalizer Gate (DEG). In silico prediction of the reaction yield as a function of both target concentration and $\Delta\Delta G^{\circ}$ for classic toehold-exchange (**a**) and DEG of varying DEP concentrations at 50, 100, 200, and 500 nM (**b**) The classic toehold-exchange can be considered as a special case of DEG, where [DEPs] = ∞. Maximum yields exist for DEG, where [AB] = [DEPs]. Yields of spurious targets are significantly suppressed across wide concentration ranges, which can help improve the specificity and expand the detection window. In silico prediction of discrimination factors for classic toehold-exchange (**c**) and DEGs (**d**). The detection window for discriminating SNVs is tunable by altering the concentrations of DEPs. In silico prediction of robustness factors for classic toehold-exchange **e** and DEGs **f**. The use of DEG dramatically increases RF values from a finite value to infinite.

The experimentally measured η, DF, and RF at varying concentrations of a synthetic dsDNA target are plotted in Fig. 3.21 for comparison with those predicted in silico. A correction of ΔG°_{rxn} by +1.58 kcal/mol was found to significantly improve the agreement between experimental observation and in silico prediction (Fig. 3.18). η and DF at a specific target concentration were calculated directly using fluorescence readout from the reporter. Consistent with in silico prediction, maximum η were observed for both correct and spurious targets, which was defined strictly by the concentration of DEP (Fig. 3.21a). As theoretically predicted, η for the spurious target is significantly suppressed by DEG. As a result, improved DF was also observed, which also agreed well with simulation (Fig. 3.21b). RF was measured indirectly by first fitting a calibration curve using a non-linear model and then calculated according to definition (Fig. 3.19). Again, infinite RF was determined across wide concentration ranges (Fig. 3.21c). The effectiveness and flexibility of DEG were further verified experimentally against varying types and locations of single nucleotide mutations (Fig. 3.21d and 3.21e), varying length of dsDNA targets (Fig. 3.22-3.24), and finally 9 sets of clinically important SNVs (Fig. 3.25-3.27). DEG works well for all sets of targets except when mutation occurs at the very edge of the dsDNA (Fig. 3.21e).

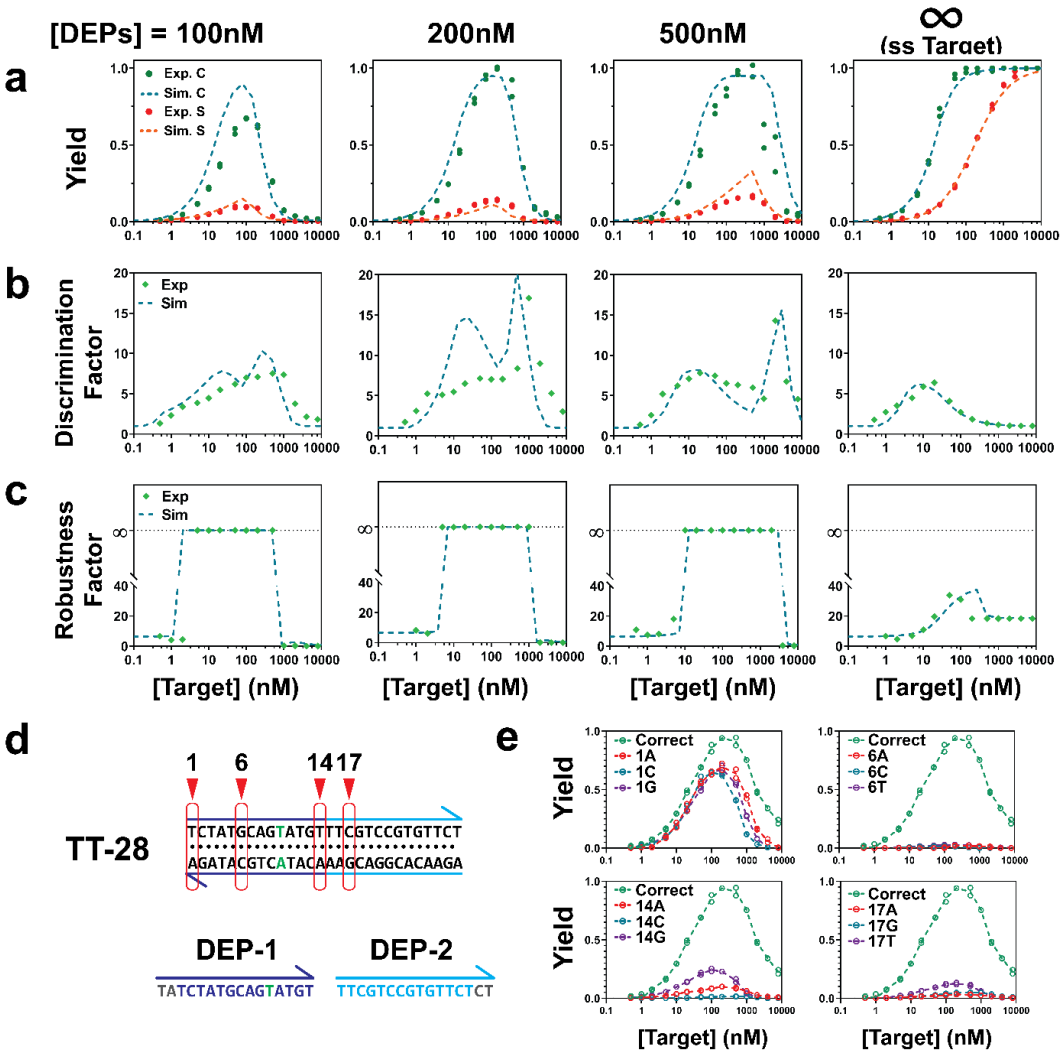


Figure 3.21. Experimental validation of DNA Equalizer Gate (DEG). (**a**) Experimentally determined yields (Exp) plotted against target concentrations for DEGs with varying DEP concentrations and compared to simulation (Sim). The classic toehold-exchange can be considered as a special case of DEG, where [DEPs] = ∞. Individual replicates (n = 2) are shown as dots. (**b**) Experimentally determined discrimination factors using a pair of synthetic correct and spurious targets ($\Delta\Delta G^{\circ}$ = 2.29 kcal/mol) plotted against target concentrations and compared to those predicted in silico. (**c**) Robustness factor plotted against target concentrations and compared to in silico prediction. (**d**) Schematic illustration of sequences of the target and DEPs. Single nucleotide mutations were made to the target at positions 1, 6, 14, and 17. **e** Experimentally determined yields plotted against target concentrations for correct and spurious targets carrying mutations at four designated positions. Individual replicates (n = 2) are shown as circles. All experiments were run at 37 °C in 1 × PBS buffer with 1 mM Mg^{2+} and 20 nM of toehold-exchange beacons.

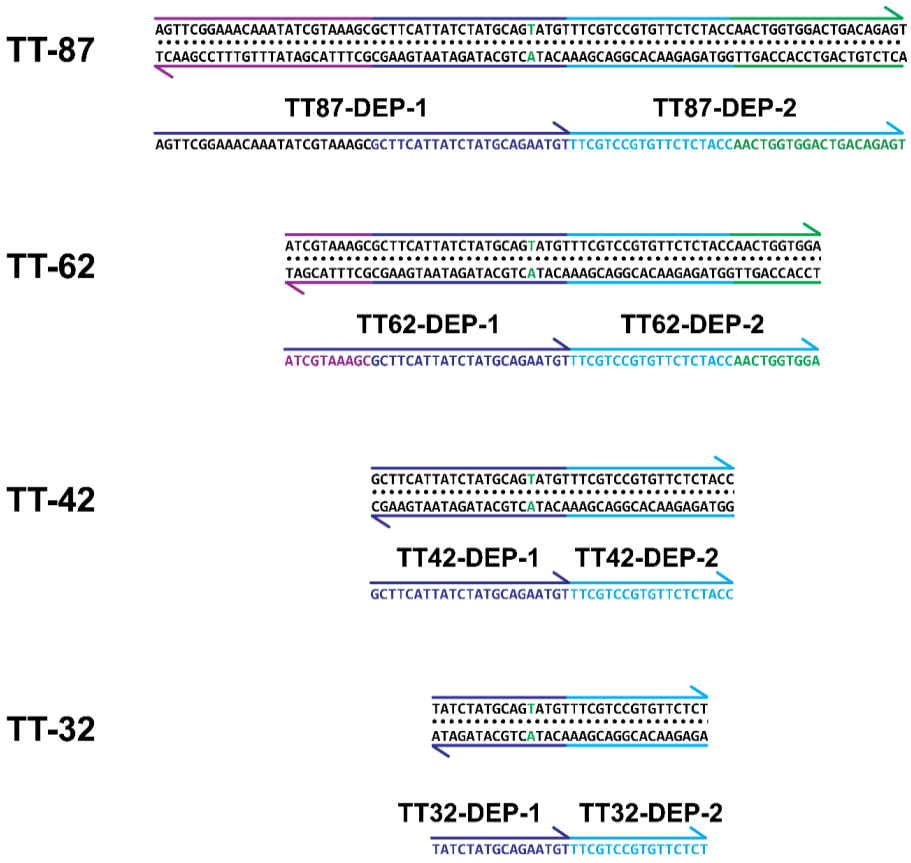


Figure 3.22. Sequence for varying length of TT targets and corresponding DEPs to validate the length effect of target/DEPs.

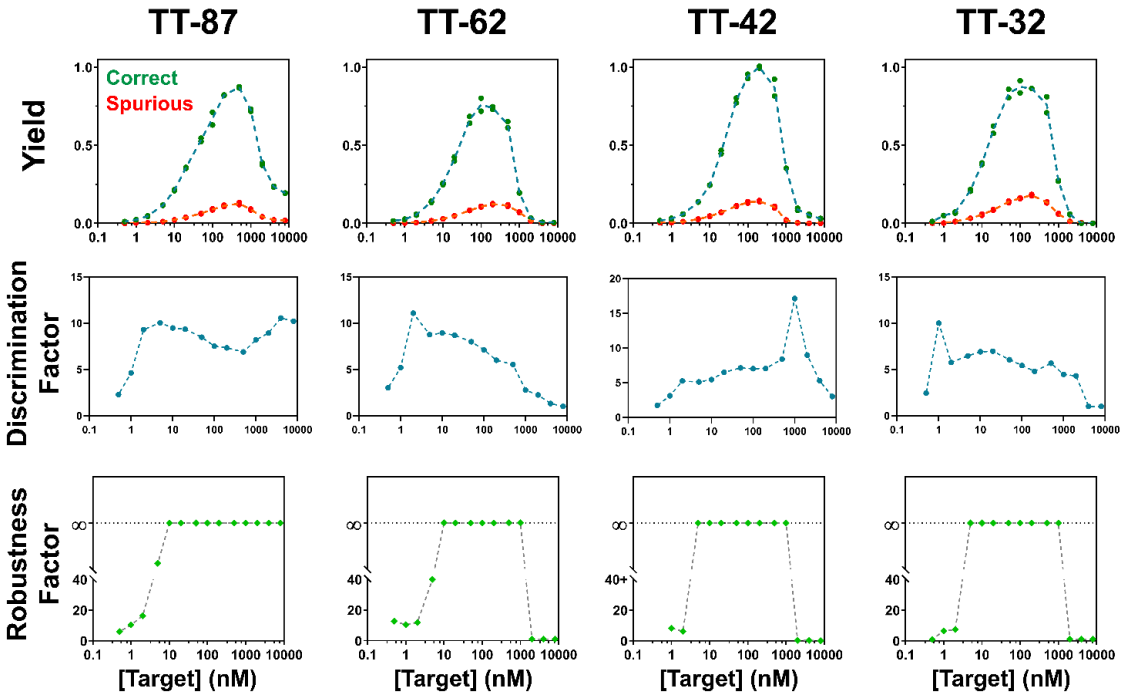


Figure 3.23. Experimentally measured yields, DFs, and RFs for varying length of dsDNA targets from 87 bp to 32 bp using corresponding DEPs at concentrations of 200 nM. These results suggest that our DEG approach is workable for targets with varying length ranges with minimal impact to the analytical performance. Individual replicates (n=2) of yield are shown as dots, and lines represent the mean values (correct and spurious yield curves).

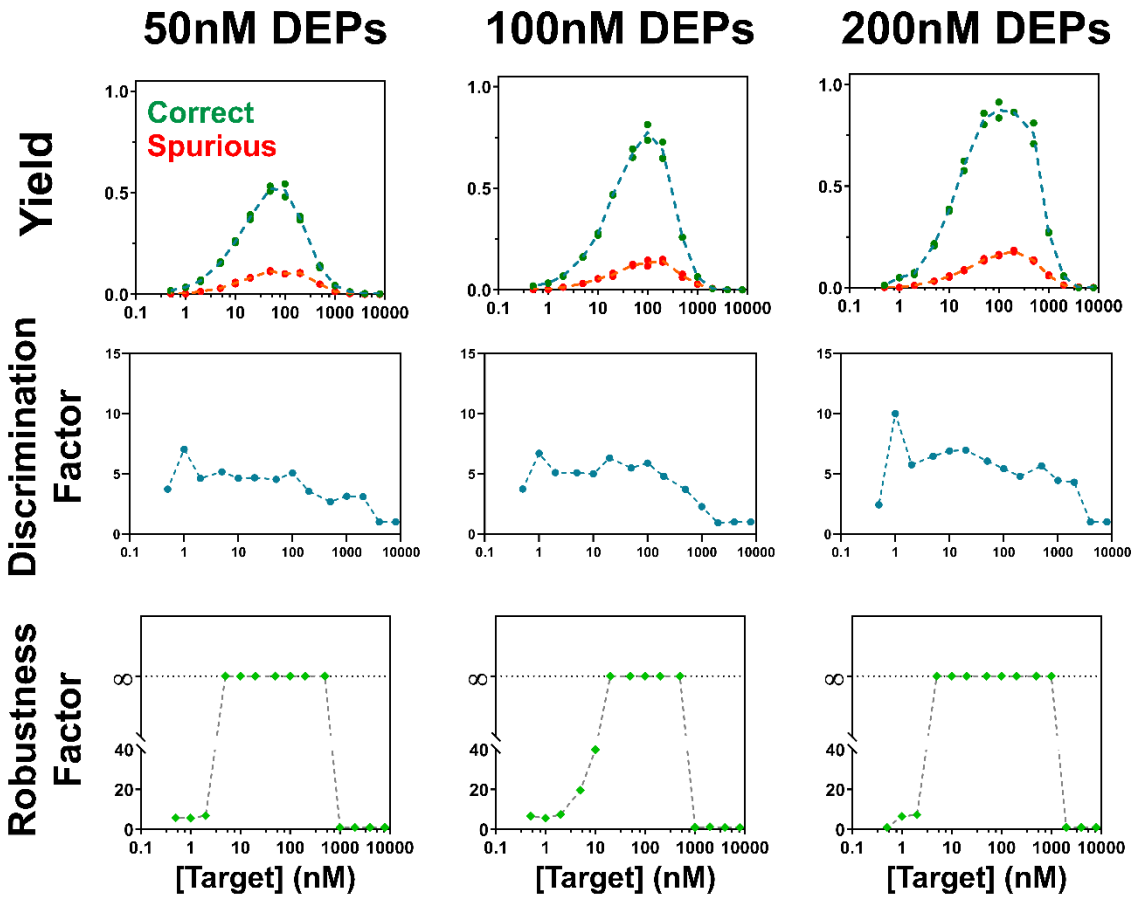


Figure 3.24. Experimentally measured yields, DFs, and RFs for target TT-32 using DEPs at concentration of 50, 100, and 200 nM, respectively. Individual replicates (n=2) of yield are shown as dots, and lines represent the mean values (correct and spurious yield curves).

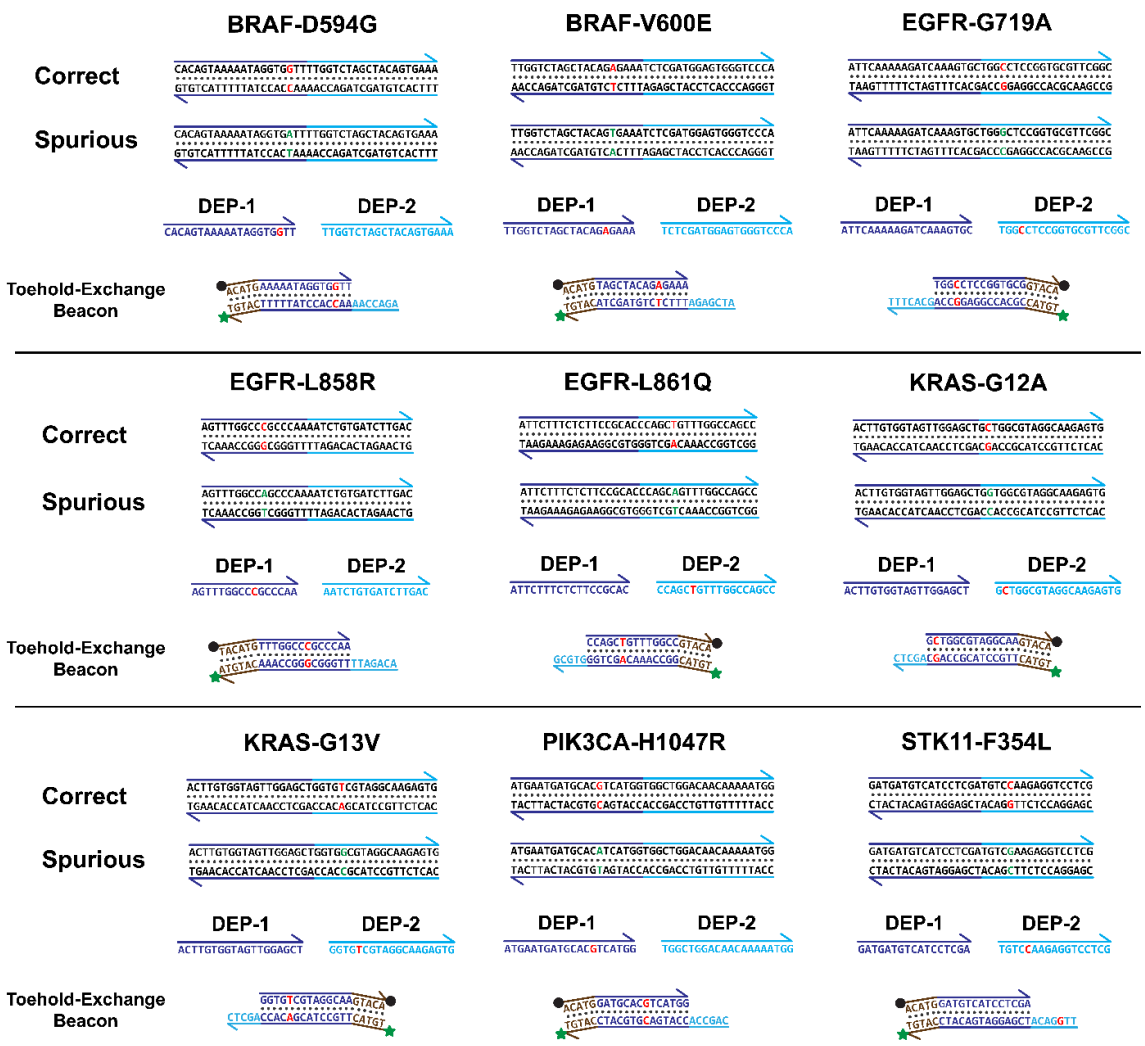


Figure 3.25. Designs and sequences for 9 sets of clinically important single nucleotide variants frequently encountered in cancer.

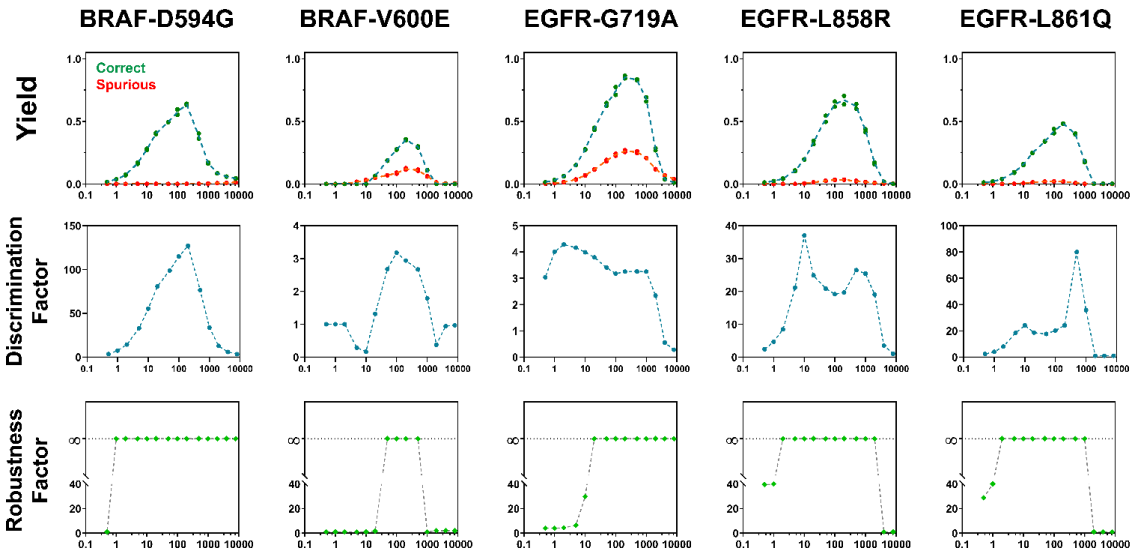


Figure 3.26. Experimentally measured yields, DFs, and RFs for analyzing BRAF-D594G, BRAF-V600E, EGFR-G719A, EGFR-L858R and EGFR-L861Q. The concentration of DEG is fixed at 200 nM. Values of individual replicates (n=2) of yield are shown as dots, and lines represent the mean values (correct and spurious curves).

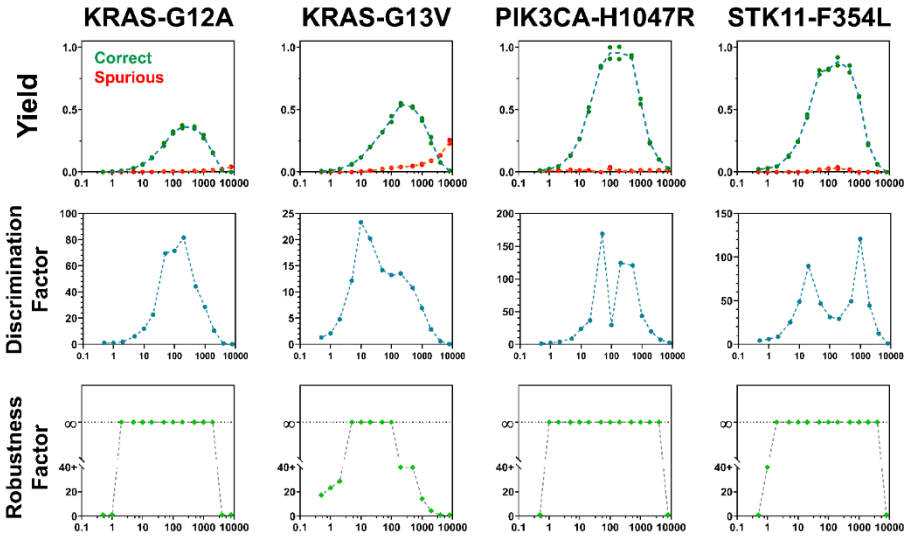


Figure 3.27. Experimentally measured yields, DFs, and RFs for analyzing KRAS-G12A, KRAS-G13V, PIK3CA-H1047R and STK11-F354L. The concentrations of DEPs were fixed at 200 nM. Values of individual replicates (n=2) of yield are shown as dots, and lines represent the mean values (correct and spurious curves).

3.2.8 Detection of varying single nucleotide mutations using DEG

We examined the analytical performance and versatility of DEG for discriminating single nucleotide variants using two sets of synthetic targets: a set of subgenomes (28 to 87 bp) from a β-tubulin gene of a parasitic worm, *Trichuris trichiura* (TT) (Fig. 3.28-3.34) and a 44 bp subgenomic sequence from *Hepatitis B Virus* (HBV) S gene (Fig. 3.35-3.38). Both diseases are major threat to human health worldwide. Varying types of mutations and indels were tested using our DEG detection platform. Synthetic rare mutation environment was used to represent the robustness and clinical application potential of our DEG method (Fig. 3.39). We also demonstrated the possibility of multiplexed DEG by mixing the two sets of targets and corresponding DEPs into the same test tube (Fig. 3.40).

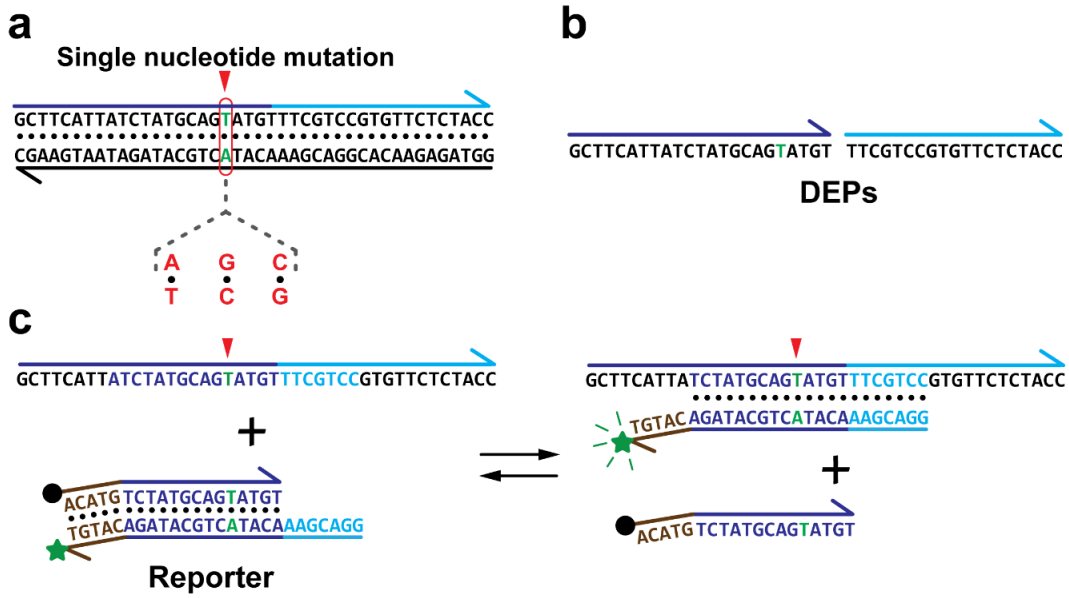


Figure 3.28. Schematic illustration of analyzing single nucleotide mutations of a subgenome of *Trichuris trichiura* (TT). (**a**) Sequences and point mutations of the TT target. (**b**) Sequences of a pair of DEPs designed for the TT target. (**c**) The sequence design of the reporter probe operated by the toehold-exchange.

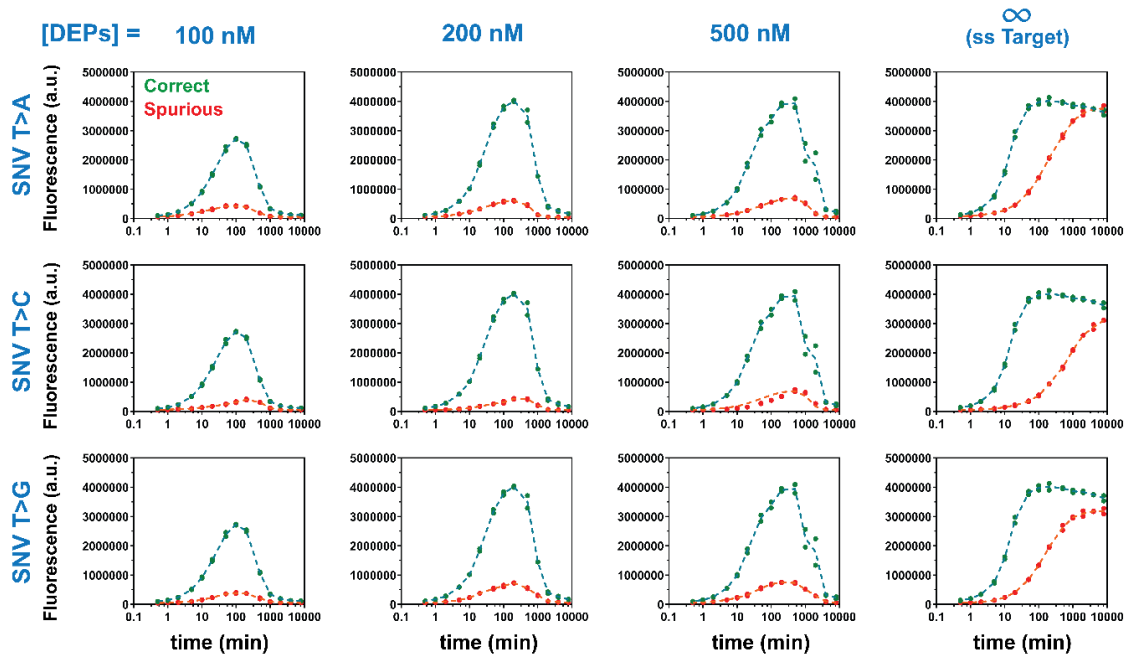


Figure 3.29. Experimental validation of DEG for discriminating single nucleotide T > A, T > G, and T > C mutations in the 42-bp dsDNA TT target. Raw fluorescence signals for detecting correct TT target and three single nucleotide mutations at DEP concentrations from 100 nM, to 200 nM, and to 500 nM. Single-stranded TT target was also analyzed directly using the reporter probe, which equivalent to a DEG system with infinite DEPs. Individual replicates (n=2) of yield are shown as dots, and lines represent the mean values (correct and spurious yield curves).

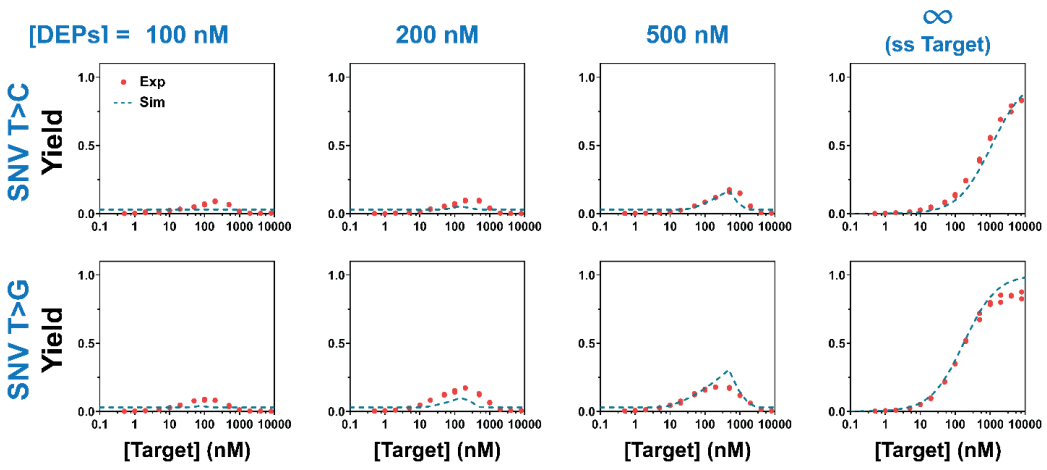


Figure 3.30. Comparison of experimentally determined yields with those predicted through simulation for analyzing T>G and T>C mutations to the double-stranded TT

targets using DEG. Values of individual replicates (n=2) are shown as dots, and lines represent the simulation prediction.

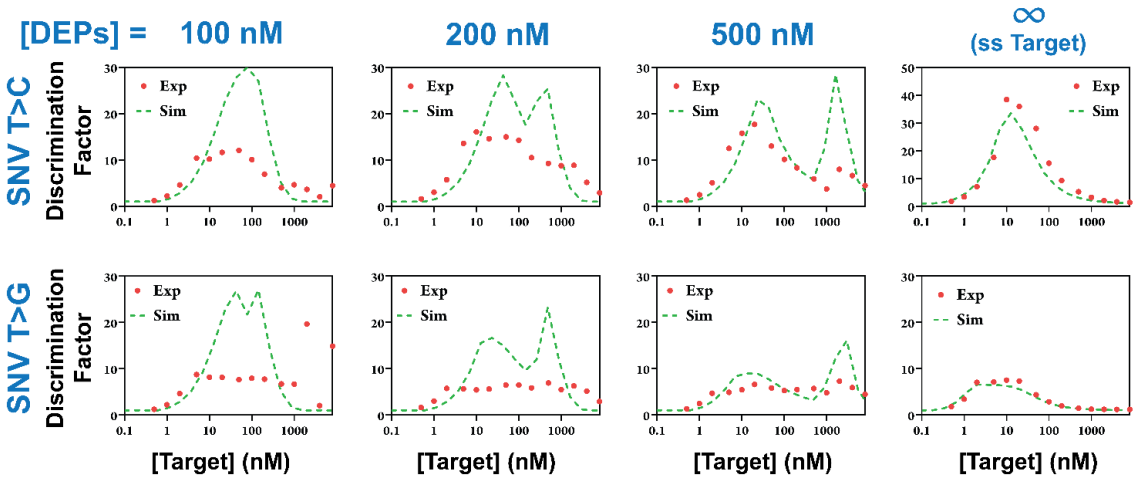


Figure 3.31. Comparison of experimentally measured and simulated DF for T > G and T > C mutations in the double-stranded TT target using DEG. Experimentally measured DF values are shown as dots, and lines represent the simulation prediction.

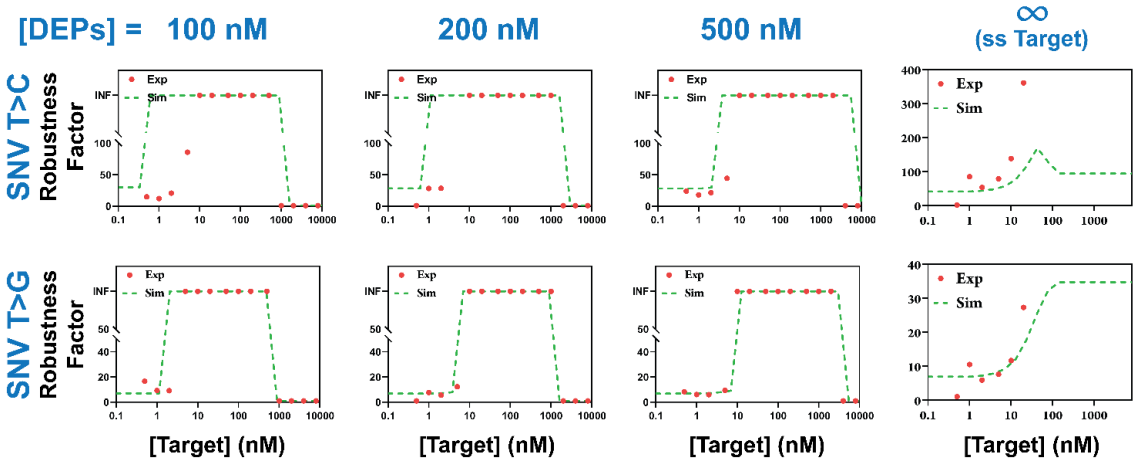


Figure 3.32. Comparison of experimentally measured and simulated RF for T > G and T > C mutations in the double-stranded TT target using DEG. Experimentally measured RF values are shown as dots, and lines represent the simulation prediction.

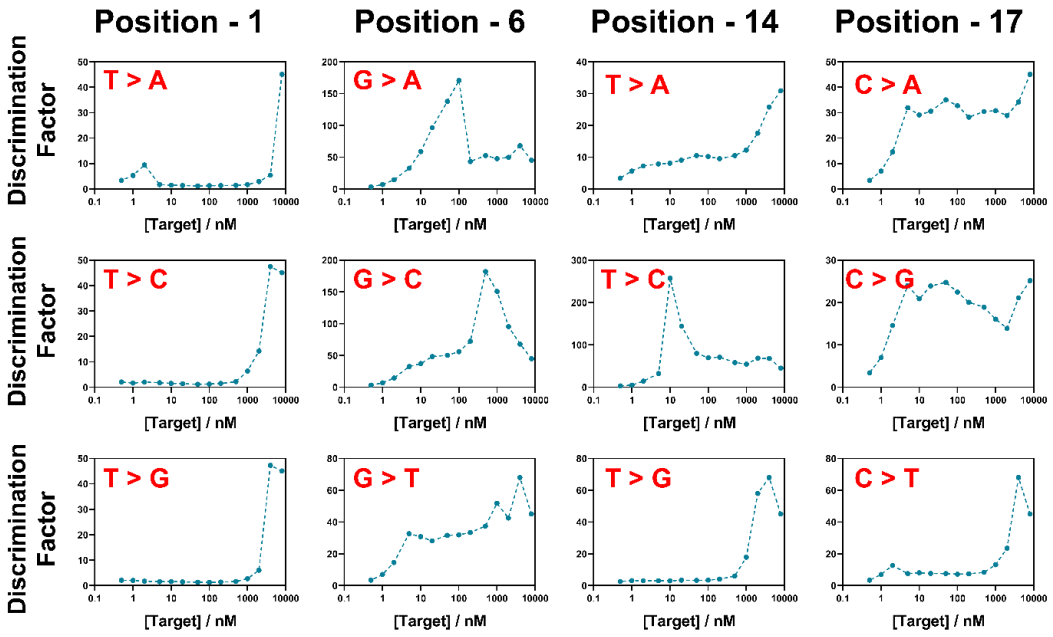


Figure 3.33. Experimentally measured DF of target TT-28 with different mutations.

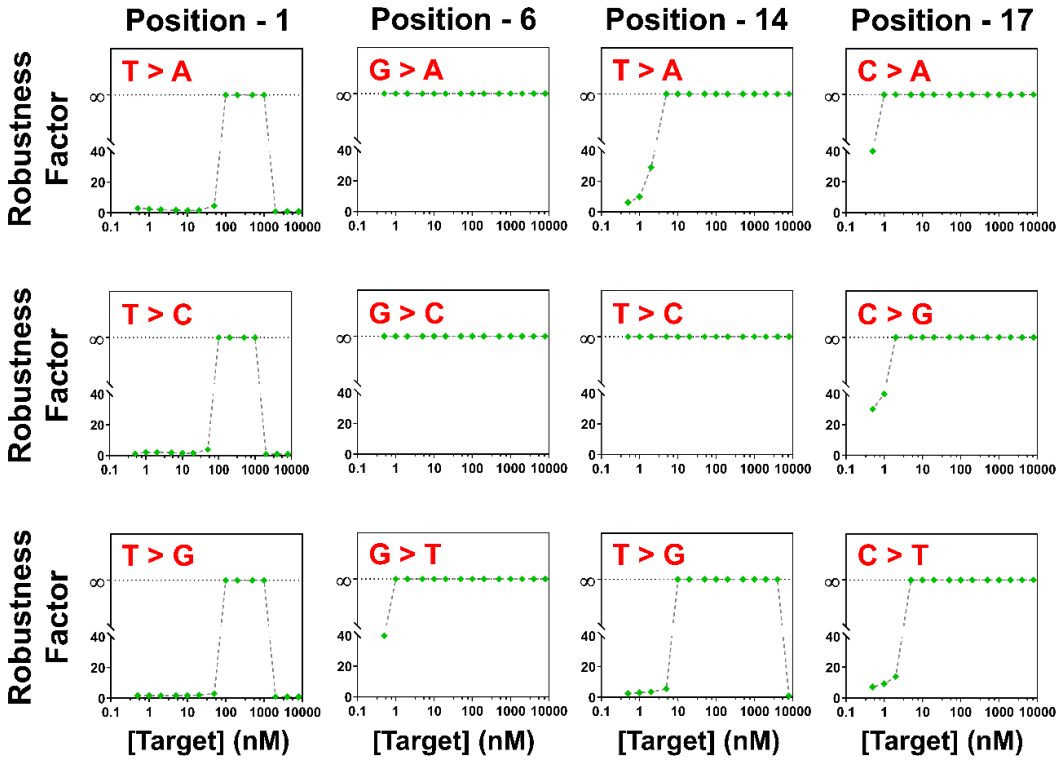


Figure 3.34. Experimentally measured RF of target TT-28 with different mutations.

3.2.9 Detection of single nucleotide mutations in HBV

The goal of designing the double-stranded synthetic HBV S-gene target is to test the versatility of our DEG approach. As shown in Fig. 3.35, a pair of DEPs and a reporter probe were designed for this synthetic target. Both single nucleotide mutations and base insertion/deletion were introduced and tested in this system using DEG. To verify the DEG approach for discriminating challenging single nucleotide mutations, we intentionally introduced an A to G mutation, a well-known challenging SNV because of the formation of G-T wobble that reduces differences in free energies between correct and spurious targets. We found that our DEG approach has effectively improved the specificity and concentration robustness for analyzing this challenging SNV comparing to the direct analysis using the toehold-exchange beacon (highlights in Fig. 3.37 and 3.38).

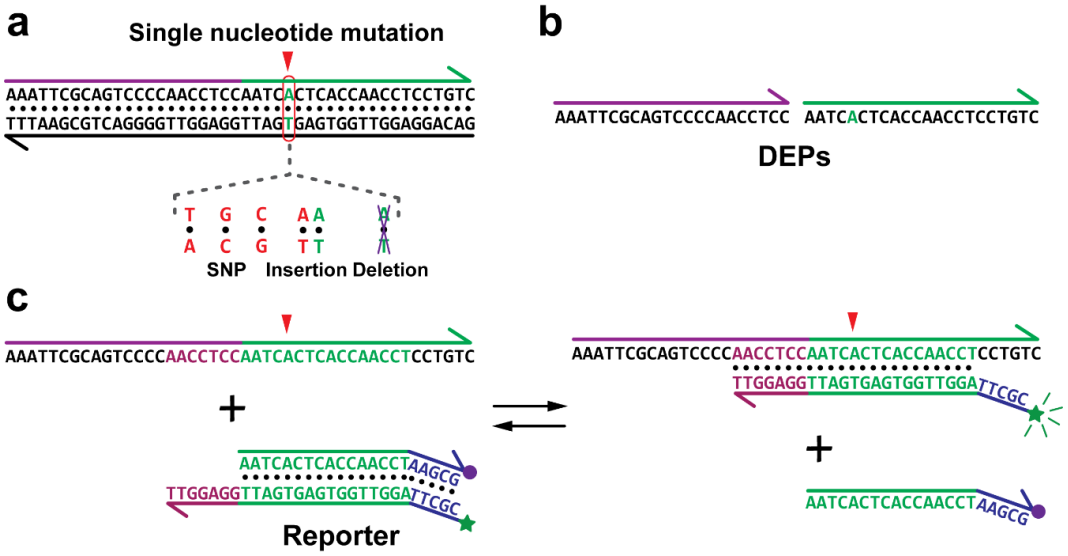


Figure 3.35. Schematic illustration of analyzing single nucleotide mutations of a subgenome of *HBV S gene*. (**a**) Sequences and point mutations of the target. (**b**) Sequences of a pair of DEPs designed for the HBV target. (**c**) The design of the reporter probe operated by the toehold-exchange.

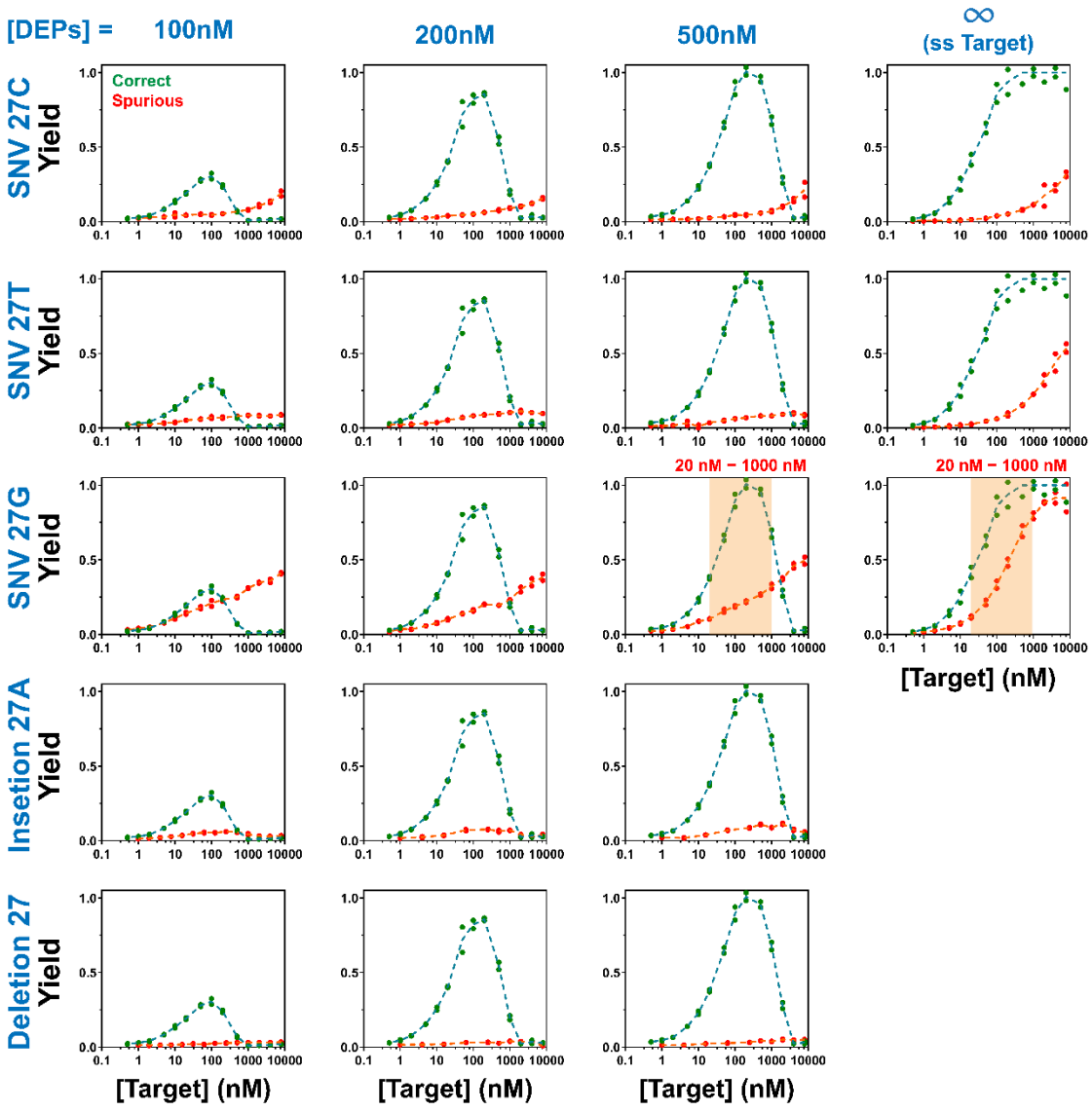


Figure 3.36 Detection of the synthetic HBV target with varying mutations and indels using DEG. Five SNVs, including three single nucleotide mutations, one insertion (INS) and one deletion (DEL), were tested using DEG with varying DEP concentrations from 100 nM, to 200 nM, and to 500 nM. Detection single nucleotide mutations in single-stranded HBV target was also included for mutations at 27C, 27T, and 27G. Shaded area indicates the detection window where DEG outperforms toehold-exchange beacon in discriminating the most challenge SNV27G. Individual replicates (n=2) of yield are shown as dots, and lines represent the mean values.

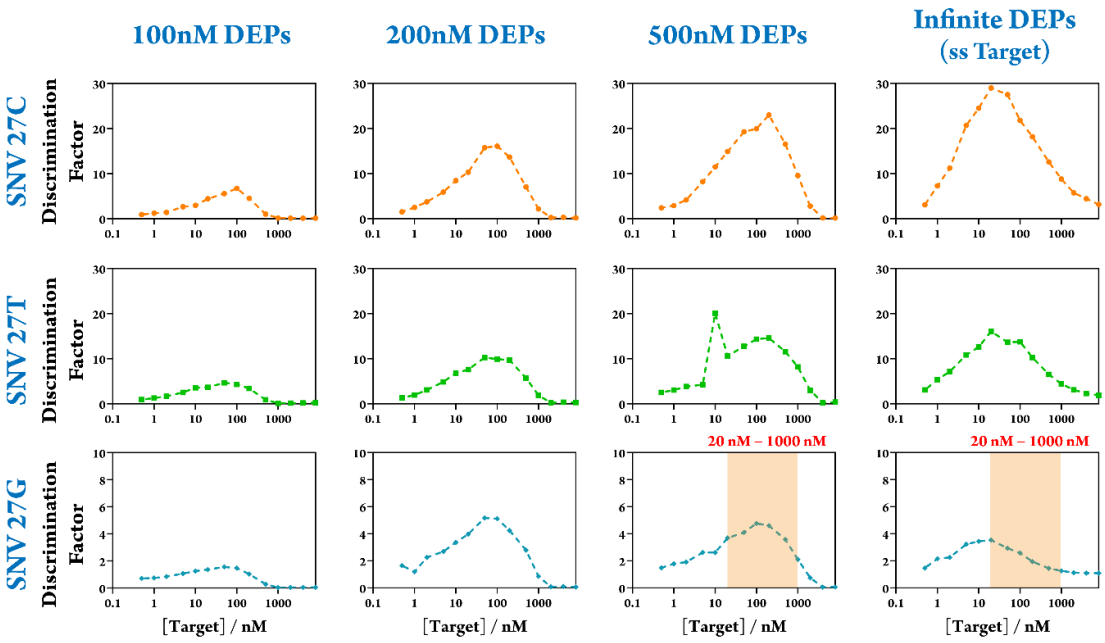


Figure 3.37. Experimentally measured DF values for three single nucleotide mutations (SNV27C, SNV27T, and SNV27G) in the 44 bp HBV target using DEG. Shaded area indicates the detection window where DEG outperforms toehold-exchange beacon in discriminating the most challenge SNV, 27G.

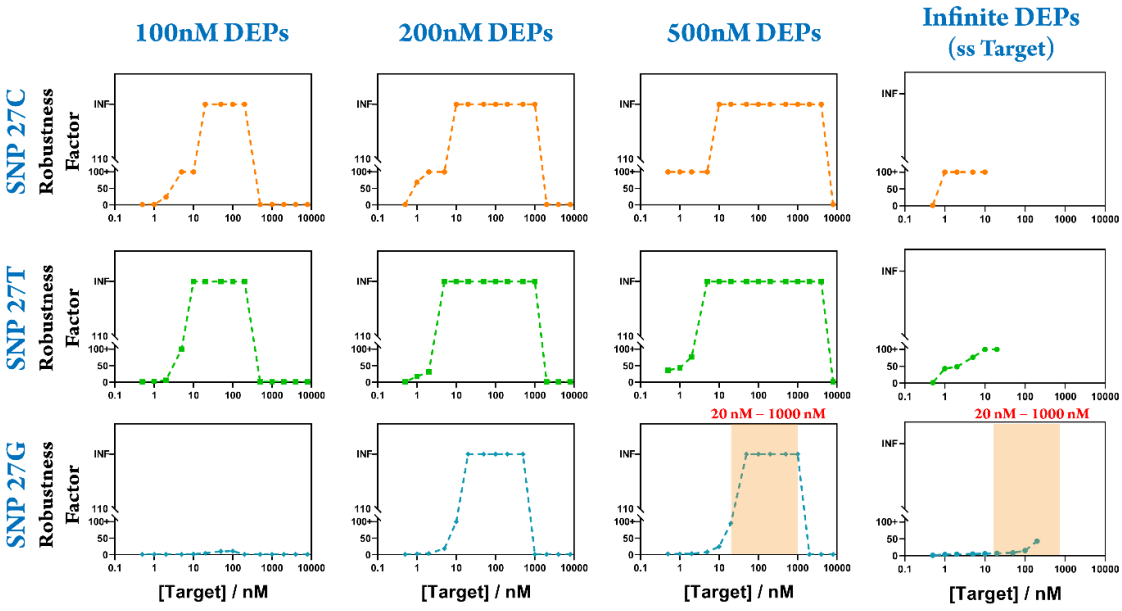


Figure 3.38. Experimentally measured RF values for single nucleotide mutations in the 44 bp HBV target using DEG.

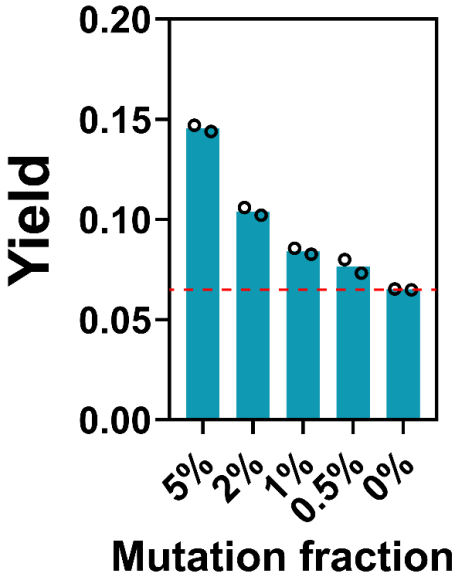


Figure 3.39. As low as 0.5% mutated targets in the background of high concentrations of unmutated sequences can be detected effectively using DEG. Bars represent the mean values of individual replicates (circles, n=2).

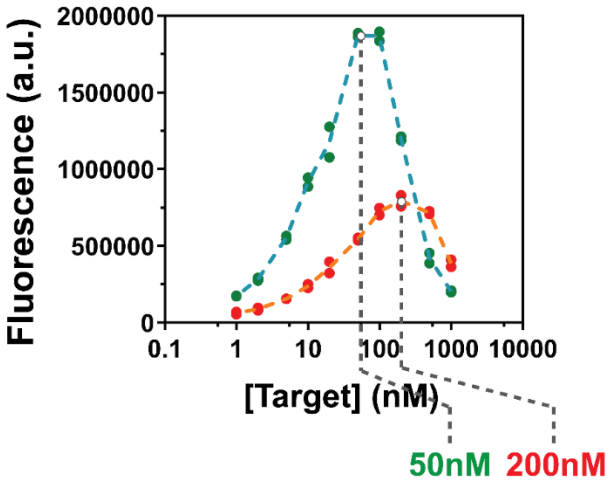


Figure 3.40. Simultaneous manipulation of TT and HBV targets using two sets of DEPs in the same test tube. The characteristic detection curves were observed for each target with detection window controlled by their corresponding DEPs ([DEP] = 50 nM for TT and 200 nM for HBV). This experiment demonstrates that multiplexed DEG can be performed for the independent control of multiple strand displacement reactions in the same test tube.

Values of individual replicates (n=2) of yield are shown as dots, and lines represent the mean values.

3.2.10 Integration of DEG with PCR

A practically applicable DNA hybridization probe shall compatible with commonly used nucleic acid amplification techniques, such as PCR. As DEG acts directly on dsDNA, it is an ideal probe for analyzing dsDNA amplicons. Therefore, we next verified the adaptivity of DEG to PCR. As a proof-of-principle, a set of four DEPs were designed for a representative 87bp dsDNA amplicon (Fig. 3.41a), which was shown to be fully compatible with DEG (Fig. 3.41b, Fig. 3.43, 3.44). To avoid potential cross-reactions, two outer DEPs were designed intentionally to be identical as the PCR primers (Fig. 3.41c and Fig. 3.42).

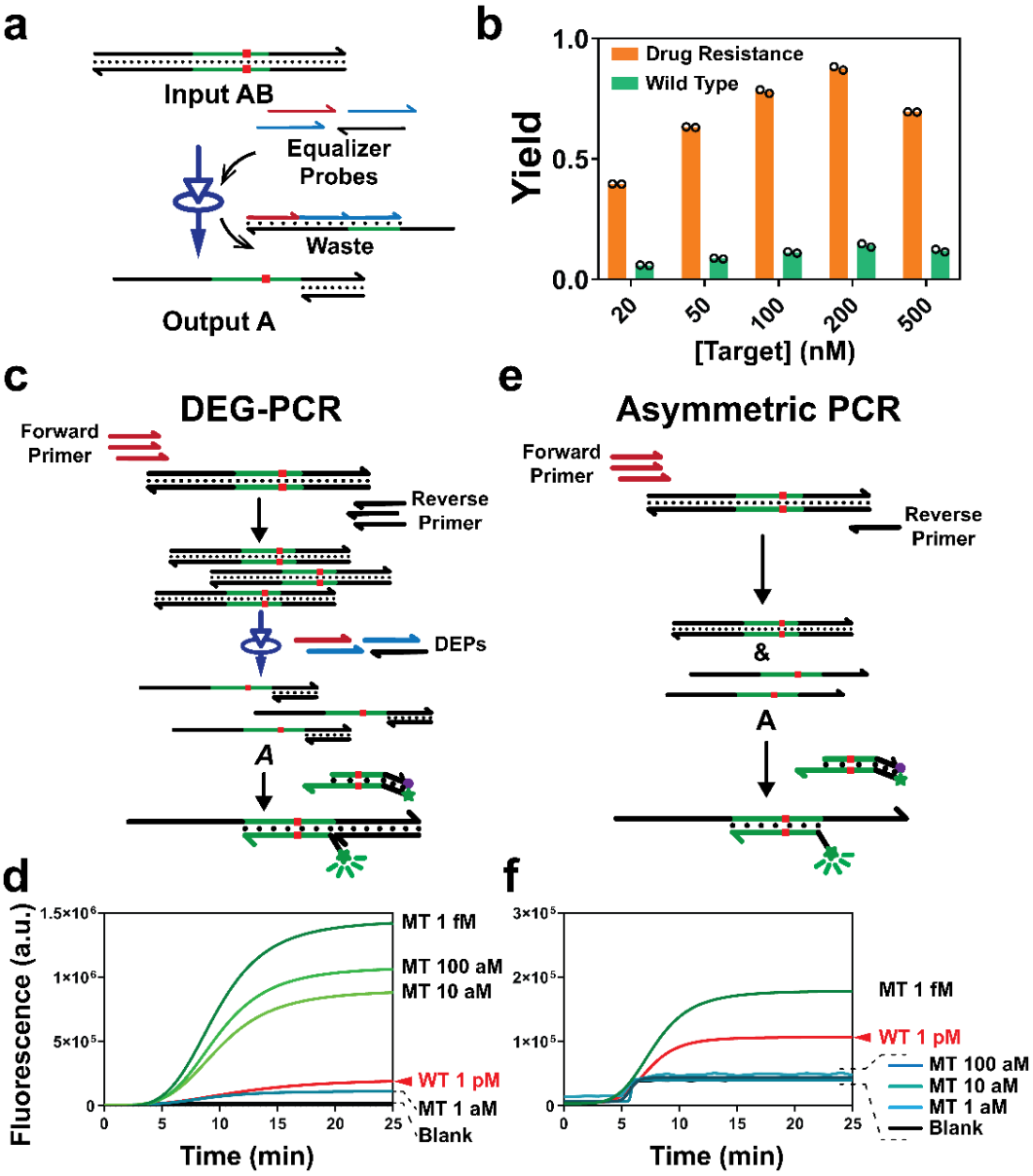


Figure 3.41. Integration of DNA Equalizer Gate (DEG) with polymerase chain reaction (PCR). (**a**) Schematic illustration of analyzing dsDNA using a 4-DEP design. (**b**) Experimental validation of the 4-DEP design for the detection of an 87bp dsDNA as a mimic of PCR amplicon. The concentration of outer DEPs was fixed at 500 nM and that of inner DEPs was set to be 200 nM. Individual replicates (n=2) are shown as circles. (**c**) Schematic illustration of DEG-PCR using a 4-DEP design. The outer two DEPs (red and black) are designed to be identical with PCR primers. (**d**) Real-time monitoring of DEG-PCR using a toehold-exchange reporter. A wide detection window was achieved, where as low as 10 aM of correct template could be clearly discriminated from 1 pM spurious target containing a single nucleotide mutation. (**e**) Schematic illustration of the asymmetric PCR followed by the detection using a toehold-exchange reporter. (**f**) Real-time monitoring the

detection of asymmetric PCR amplicon revealing a much narrower detection window than DEG-PCR, where correct discrimination could be made only above 1 fM.

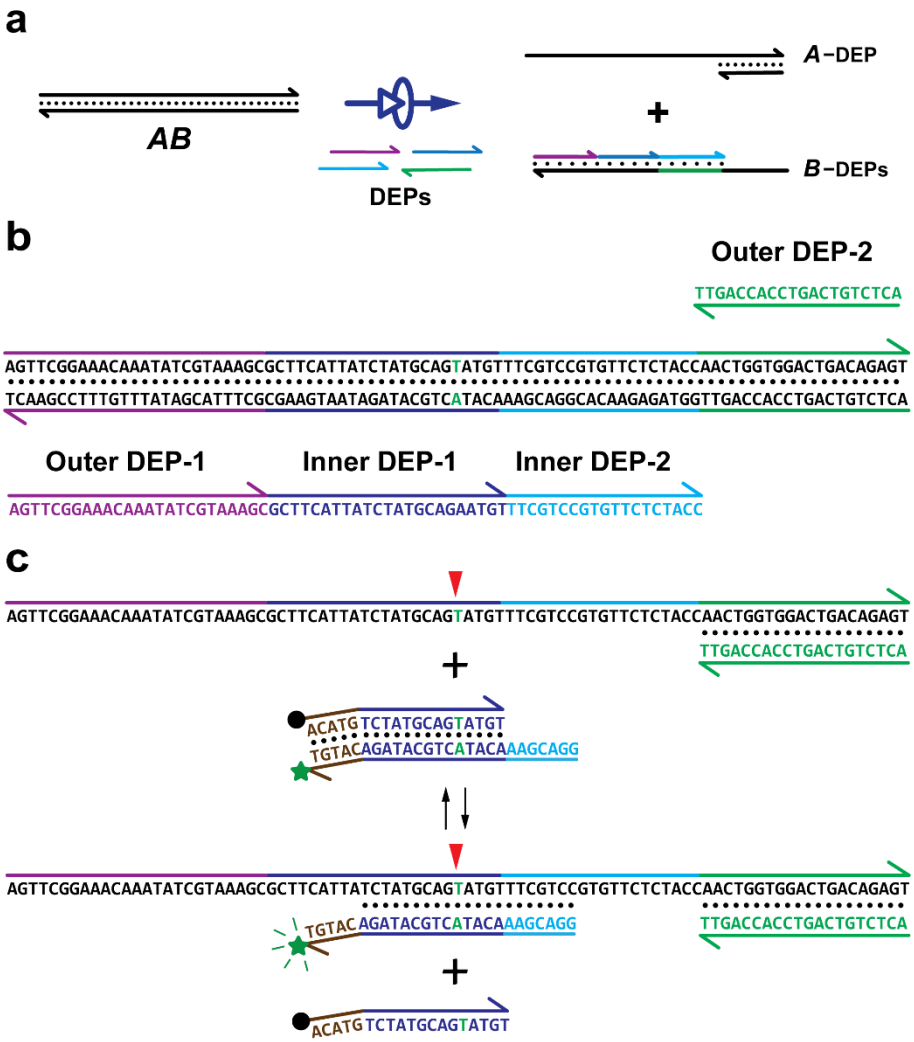


Figure 3.42. Schematic illustration of the sequence design for a set of four DEPs that target an 87 bp amplicon (AB) for the detection of the hotspot for drug resistance in TT worm. (**a**) Scheme showing the 4-DEP design for the PCR amplicon. (**b**) The two inner DEPs were designed to expose ssDNA domains that can be detected by the reporter probe. The two outer DEPs were designed to facilitate the generation of ssDNA output (A) the sequence of which were identical with a pair of forward and reverse PCR primers. (**c**) The design of the reporter probe operated by the toehold-exchange.

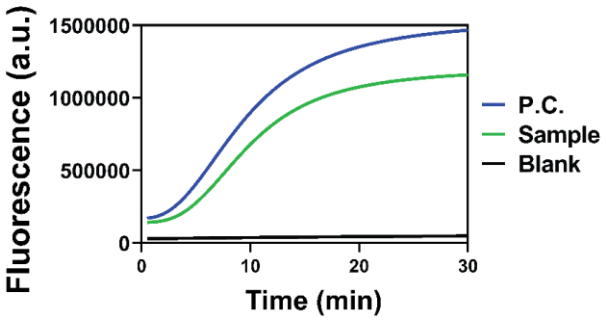


Figure 3.43. Validation of DEG for the detection of ssDNA output A *via* the 4-DEP design. Real-time fluorescence monitoring the kinetics of the reporter for measuring A produced by 4-DEPs equalizer gate. Concentrations of all DEPs were fixed at 200 nM for detection of 20 nM of target *AB* with drug resistance mutation (mutant), generating 80% of fluorescent yield. Our DEG approach allows the discrimination of single nucleotide mutation with high sequence specificity.

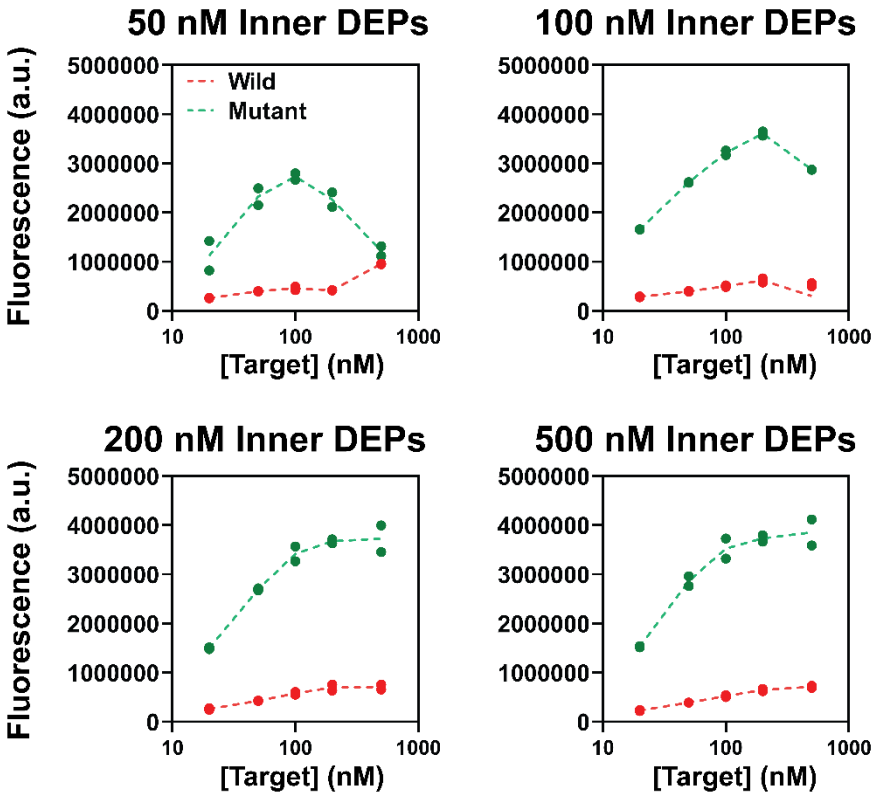


Figure 3.44. Analyzing PCR amplicons using DEG with varying concentrations of inner DEPs from 50 nM to 500 nM (500nM outer DEPs). DEG allows the detection of the double-stranded PCR amplicon meanwhile discrimination of single nucleotide mutation

through wide concentration ranges. Lines represent the mean values of individual replicates (red and green dots, n=2).

Results in Figure 41d demonstrate that the DEG-PCR is both highly sensitive and specific. As low as 1 aM synthetic DNA templates were detectable. More importantly, fluorescence signal for 1 pM spurious template containing a single nucleotide mutation is significantly suppressed using DEG, which is much lower than that of 10 aM of the correct template (Fig. 3.41d). By contrast, a much narrower detection window (above 1 fM) was observed when asymmetric PCR was used to generate detectable ssDNA amplicon followed by the readout using the same toehold-exchange reporter (Fig. 3.41e, 3.41f and Fig. 3.45).

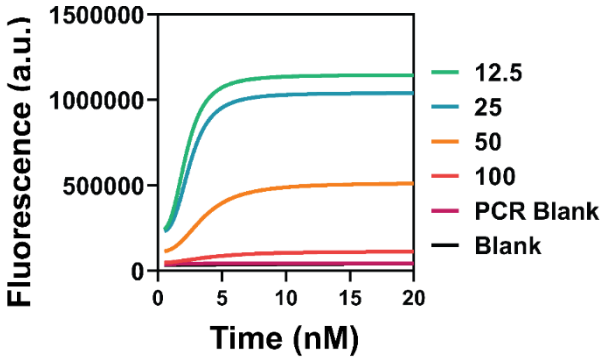


Figure 3.45. Optimization of the asymmetric PCR using varying ratios between the concentrations of forward and reverse primers. Kinetic curves showing the measurement of ssDNA output generated by asymmetric PCR using the reporter probe that is operated using toehold-exchange.

3.2.11 Clinical validation of DEG-PCR

We finally employed DEG-PCR for the diagnosis of soil-transmitted helminth (STH) infections with clinical samples collected from school-age children living in highly endemic rural areas in Honduras. STH infections are global health issue, affecting more than 1.5 billion world's population.[39] The extensive drug usage (e.g., Albendazole) for treating STH infections in endemic countries or regions has created issues of drug resistance.[40,41] As such, an ideal diagnostic test for STH infection shall allow simultaneous detection of STH infection and screen for drug resistance (D.R.).

Thus motivated, we employed DEG-PCR for the detection of STH infections, meanwhile screening for drug resistance in the same assay (Fig. 3.46a). Two fluorescence reporters were designed to test codon 196 to 203 and codon 206 to 213 of the β-tubulin gene of *Trichuris trichiura* (Fig. 3.46b). A single-nucleotide A to T mutation at the 200th codon of β-tubulin is a well-established genetic variant for drug-resistance screening (Fig. 3.47).[41] The toehold-exchange reporter testing this domain (codon 196 to 203) was thus designed to be highly sensitive to this SNV by including a 5-nt reverse toehold, whereas no reverse toehold was designed for the reporter targeting codon 206 to 213. The two reporters were labeled with spectrally distinct fluorescent dye (FAM and Cy5) and thus operated simultaneously in solution (Fig. 3.48, 3.49). Synthetic DNA standards of varying concentrations and 13 clinical TT samples with negative resistance were first tested using the dual channel DEG-PCR (Fig. 3.42, 3.43) and plotted in Fig. 3.46c, where three regions (error eclipse with 99% confidence) were definable representing positive infection and positive resistance (D.R.+), positive infection but negative resistance (D.R.–), and no detectable infection (N.C.). Six clinical parasitic specimens expelled by patients who

received Albendazole treatment in Honduras were tested and found to be TT positive but no drug resistance (Fig. 3.46d). Two clinical Ascaris worm specimens serving as negative control were also tested and found to be TT negative. All results were consistent with diagnostic tests using microscopy (Kato-Katz), post-PCR gel analysis (Fig. 3.54) and DNA sequencing (Fig. 3.55).

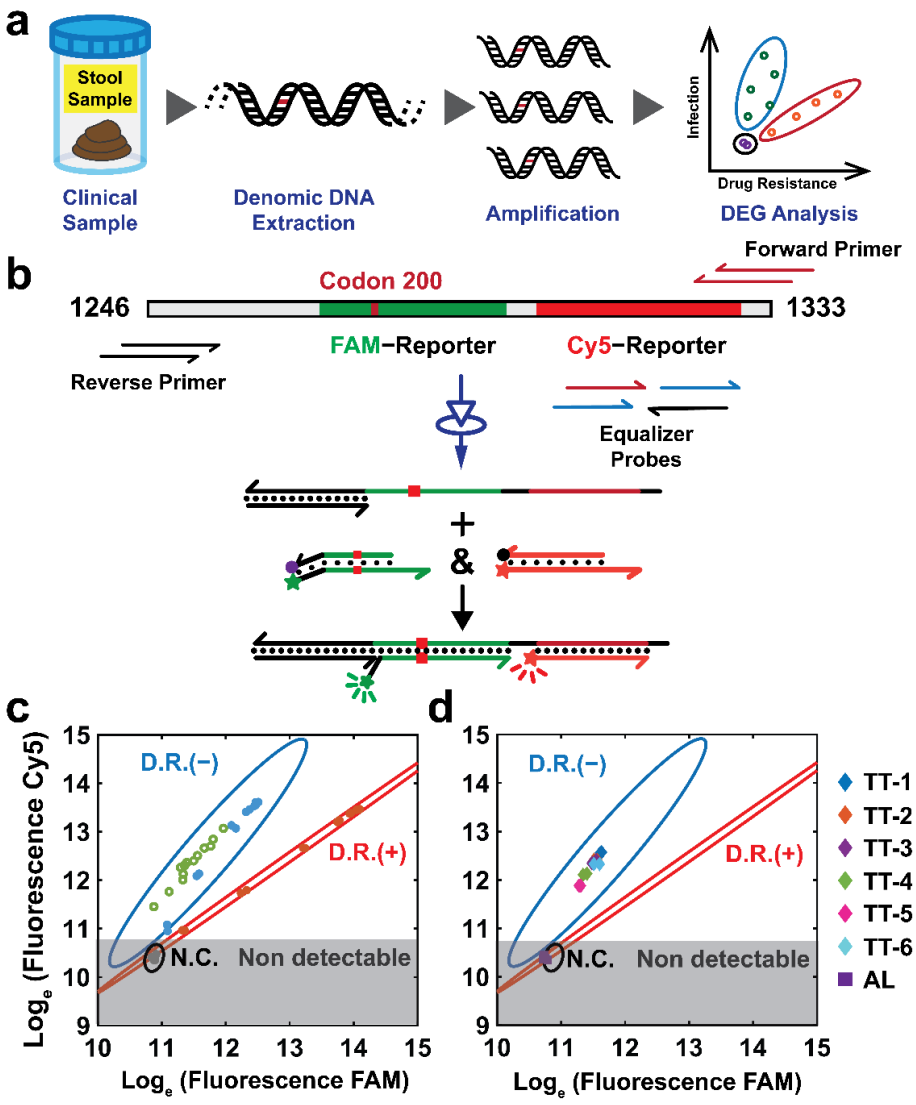


Figure 3.46. Application of DEG-PCR to analyzing clinical parasitic worm samples. (**a**) A typical workflow for analyzing parasitic worm (*Trichuris trichiura,* TT) specimens collected from stool samples of school-age children in the rural areas of Honduras followed by the detection using DEG-PCR. (**b**) Simultaneous detection of parasitic infection and

screening for drug-resistance was achieved using a dual-channel design (FAM- and Cy5-Reporter). PCR primers were deigned to amplify nucleotide 1246-1333 in the β-tubulin gene, containing the 200th codon. A single nucleotide A to T mutation of this codon is a hotspot for drug resistance screening. A toehold-exchange reporter (FAM-reporter, green duplex) labeled with FAM was used to discriminate this point mutation, whereas a strand-displacement reporter with no reverse toehold (Cy5-reporter, red duplex) was employed to detect a conservative region near codon 200. Experimental tests of the dual channel DEG-PCR using synthetic DNA standards (blue and red dots) and 13 (D.R.–) clinical samples (green circles) as a training set (**c**) and 8 unknown clinical parasitic worm samples (**d**). Test results are classified into three areas defining the positive infection and drug resistance (D.R.+), positive infection and no drug resistance (D.R.–), and negative infection (N.C.). Error eclipses with 99% confidence interval and 2-degrees of freedom (two fluorescence channels) were used to define D.R.+ and D.R.–. Eight clinical worm specimens including six *Trichuris trichiura* worms (TT-1 to TT-6) and two *A. lumbricoides* worms (AL, as negative controls) were tested and plotted in **d**.

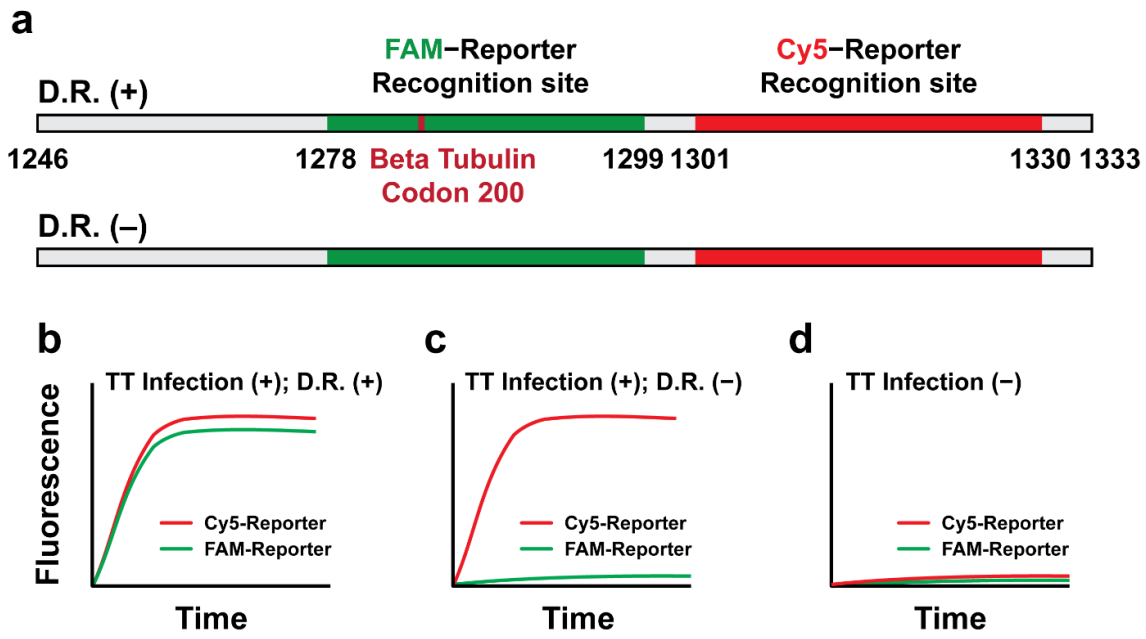


Figure 3.47. The sequence design of DEG-PCR for analyzing clinical parasitic worm samples. (**a**) A pair of primers that amplify β-tubulin gene from 1246 to 1333 bp are designed through primer-design software BLAST. The FAM reporter (green) is designed to detect the specific A to T mutation at codon 200 at the 1278th to 1299th bp of the β-tubulin gene. The Cy5-reporter (red) is designed to analyze the 1301st to 1330th bp of the β-tubulin gene. (**b**) The representative fluorescence kinetic curves that indicate positive infection (red) and positive drug resistance (green, D.R.+). (**c**) The representative fluorescence kinetic curves that indicate positive infection (red) but negative drug

resistance (green, D.R.–). (**d**) The representative fluorescence kinetic curves that indicate no infection and no drug resistance.

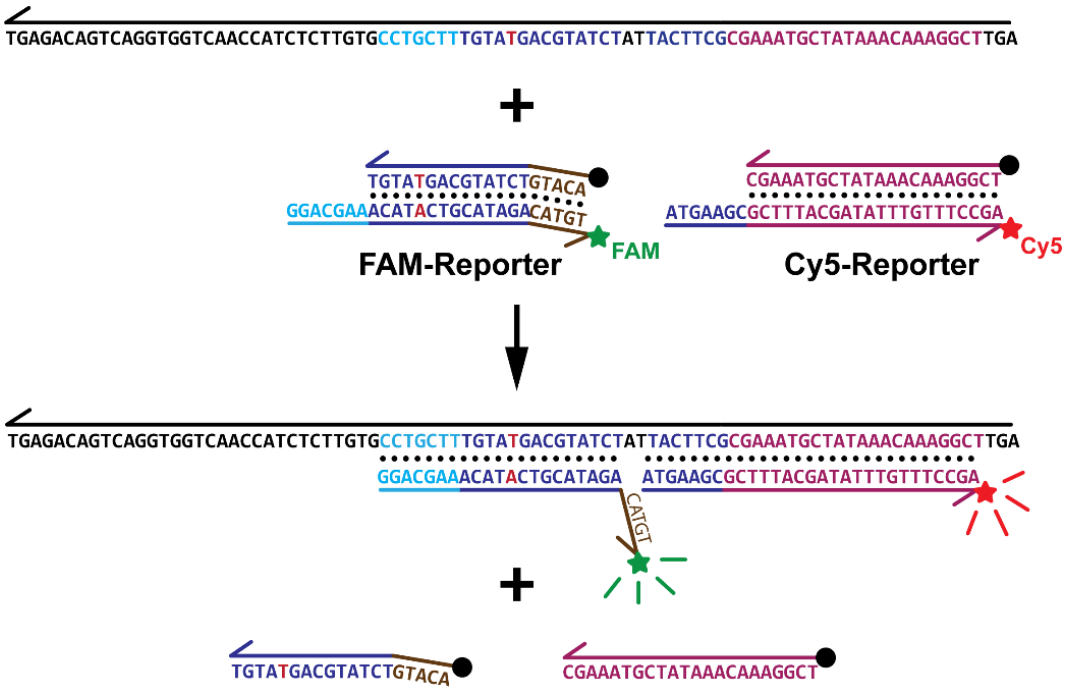


Figure 3.48. Schematic representation of the sequence designs for the DEPs and two reporters. The first reporter (FAM-reporter) that is operated through the principle of toehold-exchange is designed to a specific A to T mutation at the 200th codon of β-tubulin, which is a well-established hotspot of TT for resistance to benzimidazole (BZ, drug). As such, the fluorescence of this reporter (FAM) is turned on only when drug resistance occurs in TT infection (D.R.+ plus TT infection). The second reporter (Cy5-reporter) that is operated by toehold-mediated strand displacement is designed to detect TT infection. The reporter has a reverse toehold of 0 and is thus not sensitive to single-nucleotide mutations, which ensures the detection of infections regardless of the existence of SNPs. Simultaneous detection of the two fluorescence channels (FAM and Cy5) allows the detection of infection and screen for drug resistance in a single test.

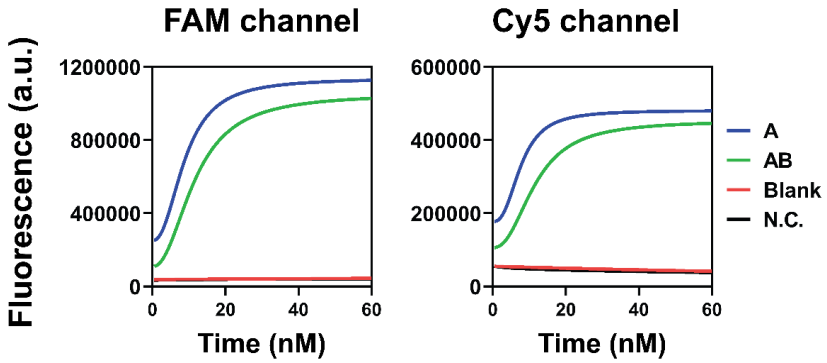


Figure 3.49. Validation of the 4-DEP, dual reporter DEG for the detection of the double-stranded TT target with the A to T mutation at β-tubulin codon 200. Real-time fluorescence monitoring the kinetics of both FAM- and Cy5-reporter for measuring the double-stranded TT target (AB) at two distinct sites as indicated in Fig. 3.53. Rapid increases in fluorescence was observed in both channels when detecting 20 nM T using a set of 4 DEPs with 200 nM for each.

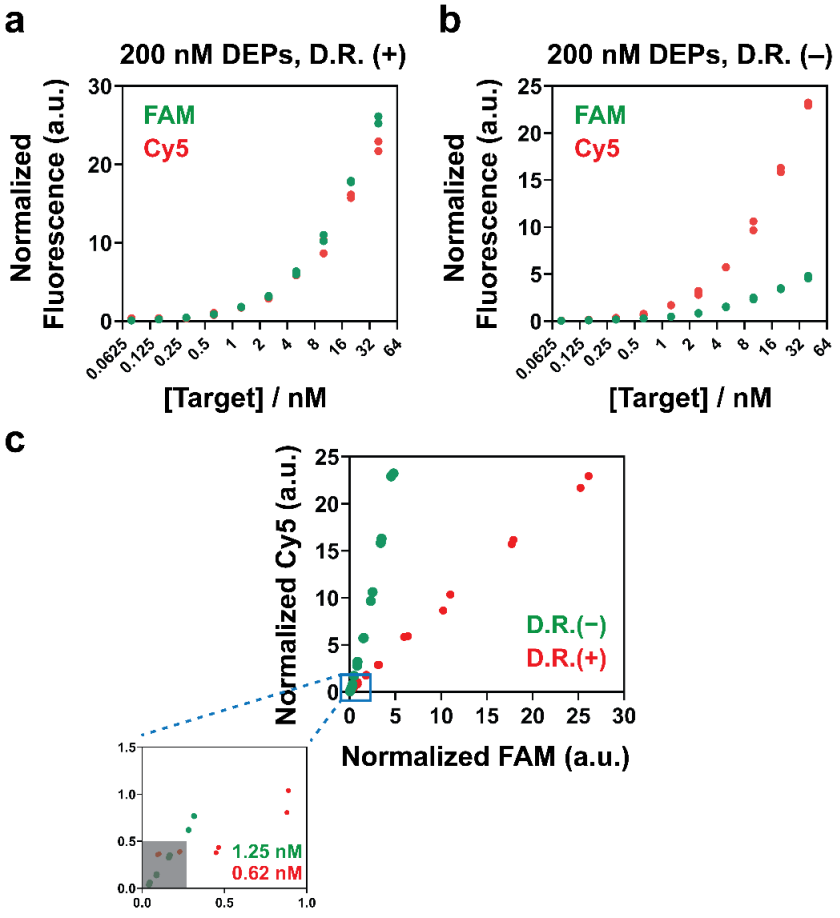


Figure 3.50. The detection limit of the 4-DEP, dual reporter DEG for analyzing synthetic DNA targets either drug resistant mutant that is drug resistant positive (D.R.+) or a wild-type that is drug resistant negative (D.R.–). (**a**) Normalized fluorescence in both FAM and Cy5 channels as a function of target concentrations for the detection of drug resistant positive mutant. (**b**) Normalized fluorescence in both FAM and Cy5 channels for the detection of a wild-type target. (**c**) The dual-channel fluorescence distribution map for targets (either D.R.+ or D.R.–) spanning over 0.16, 0.31, 0.62, 1.25, 2.5, 5, 10, 20, 40, and 80 nM. The detection limit was found to be 0.62 nM for drug resistant positive target while 1.25 nM for drug resistant negative target. Gray shading area in the subplot indicating the fluorescence distribution that cannot distinguish D.R+ and D.R.– targets. Values of individual replicates (n=2) are shown as dots.

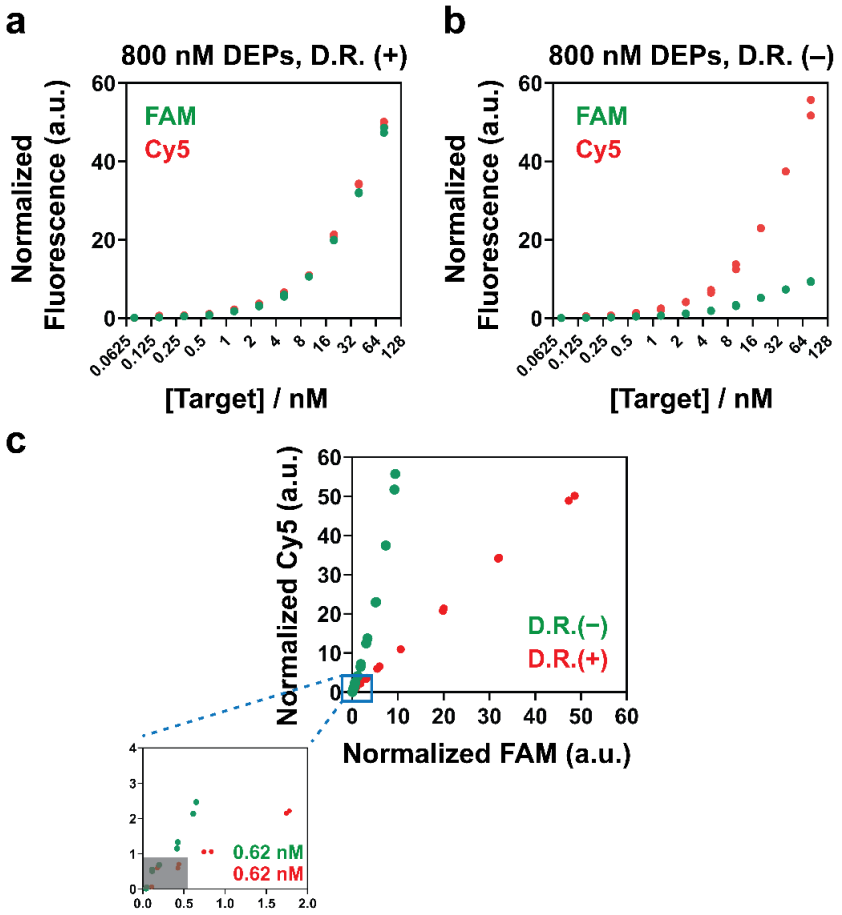


Figure 3.51. The detection limit of the 4-DEP, dual reporter DEG for analyzing synthetic DNA targets with 800 nM DEPs. (**a**) Normalized fluorescence in both FAM and Cy5 channels as a function of target concentrations for the detection of drug resistant positive mutant. (**b**) The fluorescence distribution of FAM and Cy5 channels as a function of mutant-type target concentration. (**c**) The dual-channel fluorescence distribution map for targets (either D.R.+ or D.R.–) spanning over 0.16, 0.31, 0.62, 1.25, 2.5, 5, 10, 20, 40, and

80 nM. The detection limit was found to be 0.62 nM for both drug resistant positive and drug resistant negative targets. Gray shading area in the subplot indicating the fluorescence distribution that cannot distinguish D.R+ and D.R.– targets. Values of individual replicates (n=2) are shown as dots.

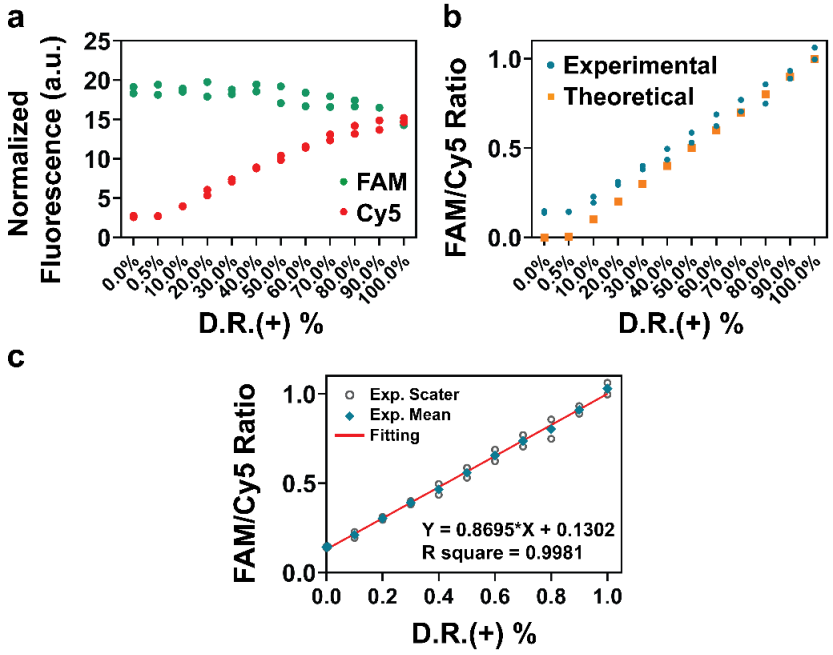


Figure 3.52. Detection of drug resistant mutant in the presence of varying concentrations of wild-types. (**a**) Normalized fluorescent intensities of FAM and Cy5 channels as a function of the percentage of spiked mutant in the wild-type control. The total target concentration was fixed at 20 nM. Values of individual replicates (n=2) are shown as dots. (**b**) Experimental (blue dots, n=2) and theoretical (yellow squares) calibration curves using FAM/Cy5 ratio as a readout. (**c**) Linear regression of the experimental calibration mean values. This result suggests that our 4-DEP, dual reporter DEG approach was able to discriminate clinically important single nucleotide variants present at low abundance.

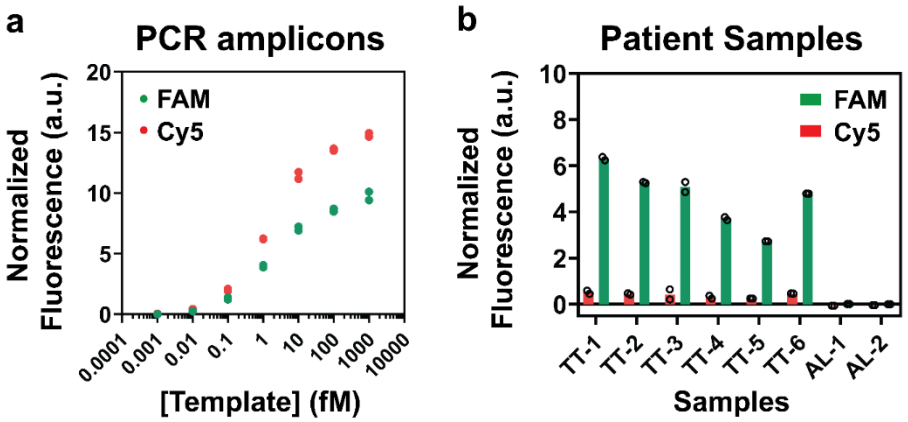


Figure 3.53. Deployment of the dual reporter DEG-PCR for analyzing clinical parasitic worm samples. (**a**) Normalized fluorescent intensities of FAM and Cy5 channels as a function of the original concentrations of the synthetic DNA template prior to PCR amplification. This template is of the same subgenomic sequence as the drug-resistant mutate. Values of individual replicates (n=2) are shown as dots. (**b**) Normalized fluorescence intensities of FAM and Cy5 channels for clinical parasitic worm samples, including 6 *Trichuris trichiura* (TT) and 2 *Ascaris Lumbricoides* (AL) worms. All the TT samples show positive infection but negative drug-resistance, while two AL samples are showing negative TT infection. The negative fluorescence intensity indicates that fluorescence signals for AL are lower than that for the blank. Bars represent the mean values of individual replicates (circles, n=2).

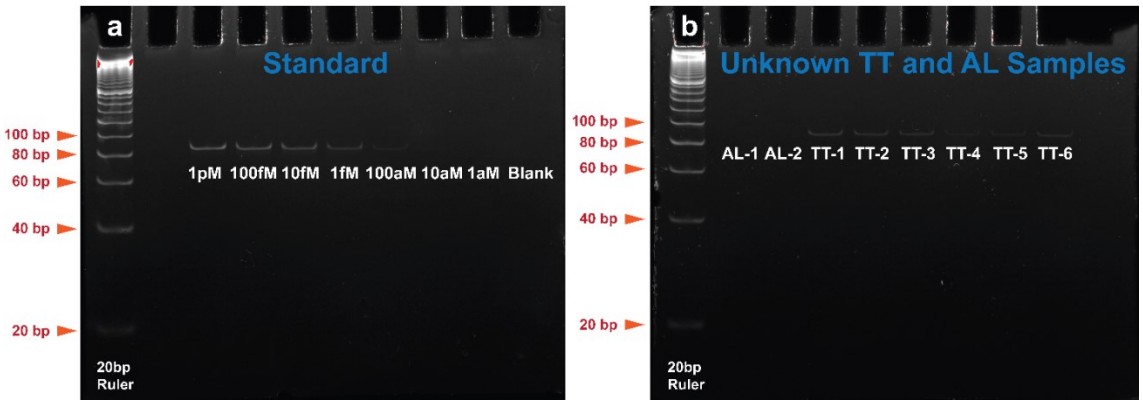


Figure 3.54. Detection of clinical parasitic worm samples using standard PCR followed by polyacrylamide gel electrophoresis (PAGE) analysis. (**a**) Representative PAGE analysis of PCR amplicons of standard synthetic DNA templates ranging from 1 aM to 1 pM. (**b**) Representative PAGE analysis of PCR amplicons of eight clinical parasitic worm samples. Each experiment was repeated twice to verify the reproducibility. A 5-μL solution containing PCR amplicons was mixed with loading buffer and then loaded onto 6% PAGE gel. A voltage of 110 V was applied for driving the electrophoresis. After electrophoresis,

the gel was stained with Ethidium Bromide and imaged using Gel Doc XR+ Imager System (BioRad).

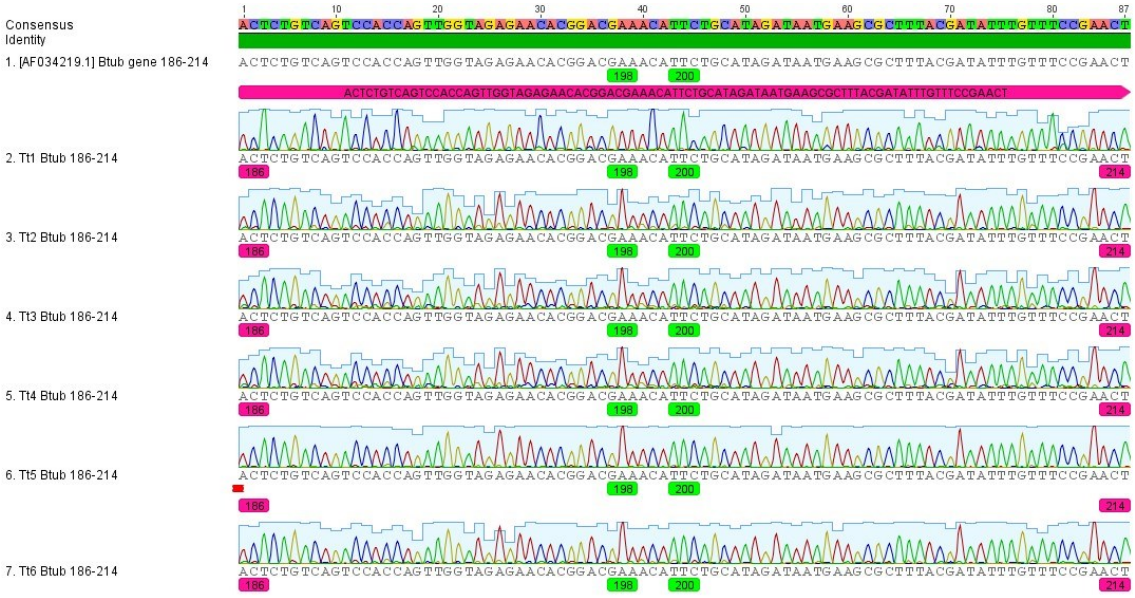


Figure 3.55. Genome sequencing data of clinical parasitic worms. The first row illustrates the DNA sequence of codon 186 to codon 214 for wild-type *Trichuris trichiura* β-tubulin gene. Codon 198 and 200 are highlighted as drug-resistance mutation hotspots. Six worm specimens extracted from patients were of the same sequences with the wild-type, which was well consistent with the diagnostic results measured using DEG-PCR.

3.3 Discussion

We have introduced DEG, a class of nucleic acid hybridization probes for the direct analysis of dsDNA with user-definable expansion of detection windows and improved sequence selectivity. Using DEG, the quantitative relationship between the detection signal and target concentrations was transformed from a sigmoidal function to an asymmetric unimodal one, where maximum yield exists at a single target concentration that is defined exclusively by DEP ($[AB]_{max}$ = [DEPs]) and is sequence-independent. Therefore, unlike conventional hybridization probes where the detection signal of a spurious target will eventually catch up that of a correct one,[33] signals for spurious targets remain to be suppressed in DEG despite the increases in target concentrations and RF will eventually become infinite. Because of the mathematical transformation, the same detection signal may correspond to two concentrations of the same correct target (Fig. 3.56). This will not cause any issue for discriminating single nucleotide mutations, as the detection signal remains to be higher than any concentration of the mutated targets. For further quantifying the correct target using DEG, we found that the inclusion of a dilution step would effectively narrow the target concentration (Fig. 3.56).

The ability to transform quantitative relationship between the detection signal and target concentrations sets the DEG approach apart from existing nucleic acid hybridization probes. Prior to our study, the concentration dependency of DNA hybridization probes has been emphasized and tested. For example, Zhang and colleagues have introduced an R value that was experimentally defined by the horizontal distance between the calibration curves of a correct and that of a spurious target at 50% yield.[17] Ricci and colleagues have experimentally examined and tuned useful dynamic ranges of molecular beacons in

response to varying allosteric designs.[33-35] However, existing solutions focus on manipulating reaction energy barriers rather than altering the monotonic quantitative relationship between detection signals and target concentrations.

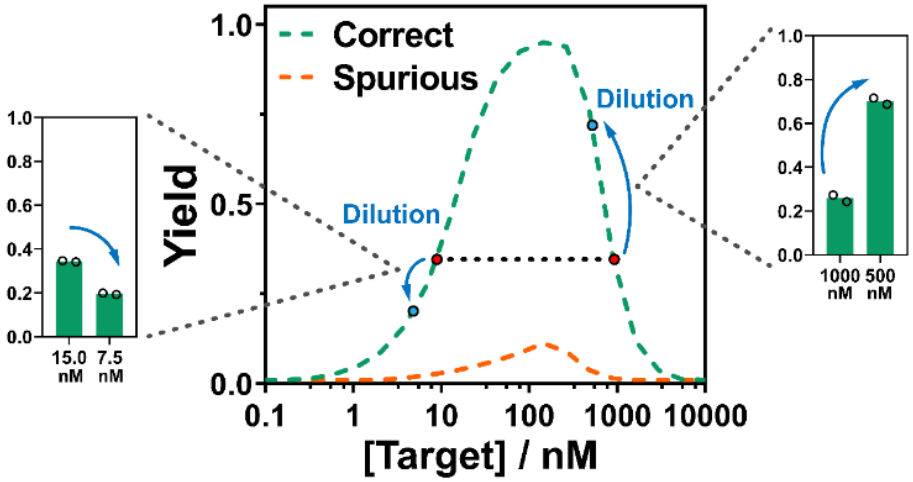


Figure 3.56. A dilution strategy to narrow the concentration range for quantifying target nucleic acid. Experimental demonstration of a representative scenario, where the same detection signal corresponds to two possible concentrations of the same target. The yield of 0.35 corresponds to either 15 nM or 1 µM of the correct dsDNA target. After a 2 × dilution step, the original 1 µM target concentration (500 nM after dilution) produced an even higher detection signal (0.8) whereas the signal of the original 15 nM target concentration (7.5 nM after dilution) decreased to 0.2. Individual replicates (n=2) are shown as circles.

Another advantage of DEG is the fact that it acts on dsDNA using a simple, one-step, wash-free, and enzyme-free procedure but produces ssDNA output in a highly quantitative and predictable manner. This differs significantly with existing strategies where enzymes and tedious procedures were often used to generate toehold domains for dsDNA[17, 43] or to remove antisense strands to produce ssDNA[44]. The unique design of DEPs makes them fully compatible with upstream nucleic acid amplification techniques such as PCR, and downstream detection probes, such as toehold-exchange beacons, with

minimal signal leakages (Fig. 3.4), which eliminates the need for the removal of antisense ssDNA through enzymatic degradation or denaturing followed by separation. As such, the principle of DEG can be employed to develop diverse assays for point-of-care applications.

More importantly, instead of treating the antisense ssDNA as a waste[30], our DEG system takes it as a molecular sink that competitively consumes the sense ssDNA once the mutation exists. Unlike existing molecular sink or reservoir created to enhance sequence selectivity,[36,38] the design of which requires prior knowledge of the specified mutation, the DEG splits a dsDNA and thus produces a corresponding sink for any mutation. We demonstrated that DEG approach is particularly effective for discriminating challenging mutations, such as a single nucleotide A to G mutation (because of the formation of G-T wobble), where both DF and RF have been drastically improved within a wide detection window (Fig. 3.36-3.38).

Collectively, our DEG approach demonstrates remarkable analytical performance for analyzing mutations in genetic markers (Fig. 3.26, 3.27). When comparing to asymmetric PCR, a widely used strategy to produce ssDNA amplicons or to prepare double-stranded toehold-exchange probes, our DEG-PCR shows significantly better sensitivity, improved specificity and wider detection window. The practical usefulness of DEG has also been successfully verified using clinical STH parasitic worm samples collected at endemic regions. The capability of simultaneous detection of parasitic infection and drug-resistance screening will make our strategy an idea tool for genetic analysis in diverse clinical settings.

Notably, the analytical performance of our DEG approach has not only been demonstrated experimentally, but also been quantitatively and precisely predicted through simulation. Our success in combining thermodynamic parameters calculated using NUPACK software and numerical approaches using MATLAB further echoes the programmable and predictable nature of the Watson-Crick base pairing rule. The ability to make accurate mathematical predictions for systems involving complexed reaction pathways, such as the DEG system in this work, reveals again the power of in silico tools to help guide the rational design of nucleic acid hybridization probes and DNA-mediated biosensors. Therefore, we anticipate our effort in developing DEG approach may benefit both fundamental research in DNA nanotechnology and practical uses of nucleic acid hybridization probes to real-world applications.

3.4 Experimental

3.4.1 Methods

DNA oligonucleotides. The DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Fluorophore (FAM- and Cy5-) and quencher (Iowa Blank) modified DNA oligonucleotides were purified by IDT using high-performance liquid chromatography (HPLC). Other DNA species were used as purchased without further purification. Sequences and modifications of all oligonucleotides are listed in Supplementary Table 3.1.

Buffer conditions. DNA oligonucleotides were re-suspended by dissolving oligonucleotides using 1 × tris-EDTA (TE) buffer (10 mM Tris-HCl, pH 8.0, 1M EDTA, purchased from Sigma as 100 × stock) and then stored at −20 °C. Unless indicated otherwise, 1 × TE buffer containing 10 mM $MgCl_2$ and 0.5% (v/v) TWEEN 20 (Sigma) was used as the molecular reporter buffer. 1 × PBS (pH 7.4, purchased as 10 × PBS stock from Sigma) buffer containing 1 mM $MgCl_2$ and 0.5% (v/v) TWEEN 20 was used as the reaction buffer. TWEEN 20 was used to prevent the potential loss of DNA oligonucleotides during dilution and pipetting.

Preparation of fluorescent reporters. All strand-displacement (SDR) and toehold-exchange (TER) reporters were annealed using a BioRad T100 thermocycler in molecular reporter buffer. The samples (typically at a final concentration of 5 μM) were heated to 95 °C for 5 min, and subsequent gradually cooled to room temperature at a constant rate over a period of 40 min. Particularly, the quencher to fluorophore concentration ratio used for SDR was 1.5 and that of TER was 3. Prepared reporter solutions were stored in 4 °C until use.

Mathematical model building. Free energy of DNA strands and complexes were estimated by NUPACK. For thermodynamic parameters setting of DEG, the temperature was set to 4 °C (in ice-water bath), concentration of Na^+ was 0.1 M, and Mg^{2+} was 0.001 M; whereas temperature setting for DNA species in toehold-exchange reaction was 37 °C. Other parameters were used set as default setting.

Analytical solutions of concentration-dependence equations for η, DF, and RF were calculated through symbolic approach in MATLAB (2019a, MathWorks). Matrix (RM) analysis and solving equilibrium equations system were performed in the same platform. Particularly, numerical computing approach is necessary due to highly coupled variables in the equation system (3^{rd} order). Boundary conditions were restricted to real value and reasonable answers (for example, yield needs to be larger than 0 yet smaller than 1). To calculate theoretical RF values, two reverse functions were adopted: first one was to convert yield (normalized) to the concentration of ssDNA target; second one was to reverse the ssDNA target back to corresponding concentration of dsDNA with the consideration of probability function. Experimental RF values were calculated by fitted non-linear curve functions (see details in Section 3.4.2). Two-dimensional curves were plotted in GraphPad Prism 8.0.1, and three-dimensional heat maps were drawn in MATLAB.

Protocols for using DNA Equalizer Gate. An ice-water cooling bath was prepared (4 °C) in advance. The dsDNA target and DEPs with user-defined concentration were mixed in a 0.2 mL PCR tube, adjusting volume to 100 μL. Sample tubes were then placed in a thermal cycler (Bio-Rad T100™) and heated up to 95 °C for 5 min (setting as 10 min with redundancy for next step). While samples were kept hot in thermal cycler, tubes were rapidly transferred and immersed inside ice-water bath for 2 min (Fig. 3.44). 90 μL of the

samples were translocated to microplate (Corning) and warmed for 5 min within a microplate reader (Molecular Devices) which was set as 37 °C. Thereafter, 10 μL of 200 nM toehold-exchange reporter was added to trigger the reaction.

DEG-PCR. In a typical PCR protocol, 4 μL of DNA template, 20 μL Taq 2 × Master Mix, and proper concentration (typically 500 nM) of forward and reverse primers (Table. 3.1. Forward and Reverse primer) were mixed to a final volume of 40 μL. PCR was initiated by incubation at 94 °C for 3 min then followed by 35 cycles (denaturation at 94 °C, annealing at 52 °C, and extension at 72 °C for 30 sec each) and a final extension at 72 °C for 5 min in a Bio-Rad T100^{TM} thermal cycler. The thermal protocol of asymmetric PCR remained the same, whereas with unbalanced primer concentration (500 nM forward primer and 40 nM revere primer, Fig. 3.45). The PCR amplicons were then mixed with 4 DEPs and adjusted volume to 90 μL. To avoid potential side reactions, the outer DEPs (identical with primers) were set to 500 nM and inner two DEPs were 200 nM. A typical DEG protocol was followed, and dual reporters (separate FAM and Cy5 fluorescent channels) were added to embark reaction.

Time-based fluorescence studies. SpectraMax i3 microplate reader (Molecular Devices) was used to acquire real-time fluorescence data and analyzed in Excel 2016. Temperature was set to 37°C and fluorescence was monitored in a frequency of 1 data point per minute for 1h. The excitation/emission wavelength for FAM channel was set 485 nm/515 nm and that of Cy5 channel was 640 nm/675 nm. Fluorescence data were normalized and converted into apparent hybridization yield according to formula $\eta = (F - F_b)/(F_m - F_b)$, where F is the sample fluorescence readout at equilibrium, F_m denotes the maximum fluorescence observed for 50-fold excess of correct ssDNA target to strand-displacement

beacon, and F_b is the background fluorescence generated by protected beacon only. For practical purpose, equilibrium fluorescence data was collected at around 20~30 min when reaction roughly reached to equilibrium.

Analyzing STH clinical samples using DEG-PCR. STH worm specimens were recovered from eight school-age children in the rural region of La Hicaca located in the northwestern area of Honduras.[42] All STH-infected children were treated by the health center's registered nurse as per national guidelines. A subgroup of eight children harboring infections of heavy and moderate intensity were invited to receive a special deworming treatment with the aim of recovering adult parasite specimens. The treatment schedule was administered by the health center's nurse as described previously.[42] The eight participants received a treatment scheme based on pyrantel pamoate and oxantel pamoate (Conmetel) during the first three days and Albendazole during a fourth day. The adult worms expelled in feces were washed with saline solution and stored in 70% ethanol. Following the recovery of specimens, DNA was extracted using the Automate Express DNA Extraction System (Thermo Fisher Scientific Inc.) with the commercial kit PrepFiler Express BTA, according to the manufacturer's protocol. Thereafter, a typical DEG-PCR procedure (250 nM of each PCR primer; 200 nM of each DEP) was followed to detect these clinical DNA samples.

Two batches (duplicates in each batch) of synthetic DNA templates from 1 aM to 1 pM (containing D.R.(–) and D.R.(+)) using the same DEG-PCR protocol as well as 13 clinical parasitic worm samples to build the fluorescence distribution map. Error eclipses with 99% confidence interval and 2-degrees of freedom (two fluorescence channels) were used to define D.R.(+) and D.R.(–).

Ethics statement. This study received clearance from the Bioscience Research Ethics Board of the Brock University (file number 13-195), as well as from the Research Ethics Board of the master's Program in Infectious and Zoonotic Diseases (MEIZ), School of Microbiology, National Autonomous University of Honduras (UNAH). Informed consent was obtained from the parents or legal guardians and was documented with research participants' signatures or fingerprints on the consent forms that had been fully explained to them. Upon parents or guardians' consent, children were invited to enroll in the study and those willing to participate provided verbal assents that were documented on a child assent form with the signature of a third-party witness.

Polyacrylamide gel electrophoresis. A 5 µL PCR amplicon solution was mixed with loading buffer (Bio-Rad) and then loaded on 8 % native PAGE gel to verify and estimate the PCR procedure. A voltage of 110 V was used to drive the electrophoresis. Thereafter, the gel was stained with Ethidium Bromide and imaged using Gel Doc XR+ Imager System (Bio-Rad).

3.4.2 DNA Sequences and Modifications

Table 3.1. DNA sequences information.

DNA Names		Sequences
Synthetic target	Correct Target	5′– GC TTC AT TA TC TAT GCA GTA TGT TTC GTC C GT GTT CTC TAC C –3′
	Spurious Target	5′– GC TTC AT T A TC TAT GCA GAA TGT TTC GTC C GT GTT CTC TAC C –3′
	DEP-1	5′– GC TTC AT TA TC TAT GCA GTA TGT –3′

	DEP-2	5′– TTC GTC C GT GTT CTC TAC C –3′
	Reporter-F	5′– G GAC GAA ACA TAC TGC ATA GA CATGT–FAM –3′
	Reporter-Q	5′– Iowa Blank FQ– ACATG TC TAT GCA GTA TGT –3′
HBV44	**Correct Target**	5′– AA ATT CGC AGT CCC CAA CCT CC AATC ACT CAC CAA CCT CCT GTC –3′
	HBV-DEP-1	5′– AA ATT CGC AGT CCC CAA CCT CC –3′
	HBV-DEP-2	5′– A ATC ACT CAC CAA CCT CCT GTC –3′
	SNV27T	5′– AA ATT CGC AGT CCC C AA CCT CC AATC TCT CAC CAA CCT CCT GTC –3′
	SNV27G	5′– AA ATT CGC AGT CCC C AA CCT CC AATC GCT CAC CAA CCT CCT GTC –3′
	SNV27C	5′– AA ATT CGC AGT CCC C AA CCT CC AATC CCT CAC CAA CCT CCT GTC –3′
	INS27A	5′– AA ATT CGC AGT CCC C AA CCT CC AATC AACT CAC CAA CCT CCT GTC –3′
	INS27C	5′– AA ATT CGC AGT CCC C AA CCT CC AATC CACT CAC CAA CCT CCT GTC –3′
	DEL27	5′– AA ATT CGC AGT CCC C AA CCT CC AATC CT CAC CAA CCT CCT GTC –3′
	Reporter-F	5′– FAM– CGCTT AGG TTG GTG AGT GATT GG AGG TT –3′
	Reporter-Q	5′– A ATC ACT CAC CAA CCT AAGCG –Iowa Black FQ –3′
TT42	**Correct Target**	5′– GC TTC AT TA TC TAT GCA GTA TGT TTC GTC C GT GTT CTC TAC C –3′
	SNV19A	5′– GC TTC AT T A TC TAT GCA GAA TGT TTC GTC C GT GTT CTC TAC C –3′

	SNV19G	5′– GC TTC AT T A TC TAT GCA GGA TGT TTC GTC C GT GTT CTC TAC C –3′
	SNV19C	5′– GC TTC AT T A TC TAT GCA GCA TGT * TTC GTC C GT GTT CTC TAC C –3′
TT42	TT-DEP-1	5′– GC TTC AT TA TC TAT GCA GTA TGT –3′
	TT-DEP-2	5′– TTC GTC CGT GTT CTC TAC C –3′
TT84	Correct Template (T^+)	5'– AGT TCG GAA ACA AAT ATC GTA AAG C GC TTC AT TA TC TAT GCA GAA TGT TTC GTC C GT GTT CTC TAC C AA CTG GTG GAC TGA CAG AGT –3'
	Spurious Template	5'– AGT TCG GAA ACA AAT ATC GTA AAG C GC TTC AT TA TC TAT GCA GTA TGT TTC GTC C GT GTT CTC TAC C AA CTG GTG GAC TGA CAG AGT –3'
	Forward Primer (TT-DEP-3)	5'– AGT TCG GAA ACA AAT ATC GTA AAG C –3'
	Reverse Primer (TT-DEP-4)	5'– ACT CTG TCA GTC CAC CAG TT –3'
Dual Channel TT Reporters	Cy5-reporter-F	5′– AT GAA GC G CTT TAC GAT ATT TGT TTC CGA –Cy5–3′
	Cy5-reporter -Q	5′– Iowa Blank RQ – TCG GAA ACA AAT ATC GTA AAG C –3′
	FAM-Reporter-F	5′– G GAC GAA ACA TAC TGC ATA GA CATGT–FAM –3′
	FAM-Reporter-Q	5′– Iowa Blank FQ– ACATG TC TAT GCA GTA TGT –3′

TT87	Long-DEP-1	5'–AGT TCG GAA ACA AAT ATC GTA AAG C GC TTC AT TA TC TAT GCA GTA TGT–3'

	Long-DEP-2	5′–TTC GTC CGT GTT CTC TAC C AAC TGG TGG ACT GAC AGA GT–3′
TT62	**Correct Target**	5'–ATC GTA AAG C GC TTC AT TA TC TAT GCA GTA TGT TTC GTC C GT GTT CTC TAC C AAC TGG TGG A–3'
	Spurious Target	5'–ATC GTA AAG C GC TTC AT TA TC TAT GCA GAA TGT TTC GTC C GT GTT CTC TAC C AAC TGG TGG A–3'
	Long-DEP-1	5'–ATC GTA AAG C GC TTC AT TA TC TAT GCA GTA TGT–3'
	Long-DEP-2	5′–TTC GTC CGT GTT CTC TAC C AAC TGG TGG A–3′
TT32	**Correct Target**	5'–TA TC TAT GCA GTA TGT TTC GTC C GT GTT CTC T–3'
	Spurious Target	5'–TA TC TAT GCA GAA TGT TTC GTC C GT GTT CTC T–3'
	DEP-1	5'–TA TC TAT GCA GTA TGT–3'
	DEP-2	5′–TTC GTC CGT GTT CTC T–3′
TT28	**SNV 1A**	5'–AC TAT GCA GTA TGT TTC GTC C GT GTT CT–3'
TT28	**SNV 1C**	5'–CC TAT GCA GTA TGT TTC GTC C GT GTT CT–3'
	SNV1G	5'–GC TAT GCA GTA TGT TTC GTC C GT GTT CT–3'
	SNV 6T	5'–TC TAT TCA GTA TGT TTC GTC C GT GTT CT–3'
	SNV 6C	5'–TC TAT CCA GTA TGT TTC GTC C GT GTT CT–3'
	SNV 6A	5'–TC TAT ACA GTA TGT TTC GTC C GT GTT CT–3'

	SNV 14A	5'-TC TAT GCA GTA TGA TTC GTC C GT GTT CT-3'
	SNV 14C	5'-TC TAT GCA GTA TGC TTC GTC C GT GTT CT-3'
	SNV 14G	5'-TC TAT GCA GTA TGG TTC GTC C GT GTT CT-3'
	SNV 17T	5'-TC TAT GCA GTA TGT TTT GTC C GT GTT CT-3'
	SNV 17A	5'-TC TAT GCA GTA TGT TTA GTC C GT GTT CT-3'
	SNV 17G	5'-TC TAT GCA GTA TGT TTG GTC C GT GTT CT-3'
BRAF-D594G	Correct Target	5'- CAC AGT AAA AAT AGG TGG TT | TTG GTC TAG CTA CAG TGA AA -3'
	Spurious Target	5'- CAC AGT AAA AAT AGG TGA TT | TTG GTC TAG CTA CAG TGA AA -3'
	DEP-1	5'- CAC AGT AAA AAT AGG TGG TT| -3'
	DEP-2	5'- TTG GTC TAG CTA CAG TGA AA| -3'
BRAF-D594G	TEB-F	5'- *AGA CCA A* | AA CCA CCT ATT TTT * *CATGT* -3'-36-FAM
	TEB-Q	5IABkFQ-5'- *ACATG* * AAA AAT AGG TGG TT| -3'
BRAF-V600E	Correct Target	5'- TTG GTC TAG CTA CAG AGA AA | T CTC GAT GGA GTG GGT CCC A -3'
	Spurious Target	5'- TTG GTC TAG CTA CAG TGA AA | T CTC GAT GGA GTG GGT CCC A -3'
	DEP-1	5'- TTG GTC TAG CTA CAG AGA AA| -3'
	DEP-2	5'- |T CTC GAT GGA GTG GGT CCC A -3'

	TEB-F	5'- *ATC GAG A* | TT TCT CTG TAG CTA * *CATGT* -3'-36-FAM
	TEB-Q	5IABkFQ -5'- *ACATG* * TAG CTA CAG AGA AA| -3'
EGFR-G719A	Correct Target	5'- ATT CAA AAA GAT CAA AGT GC | T GGC CTC CGG TGC GTT CGG C -3'
	Spurious Target	5'- ATT CAA AAA GAT CAA AGT GC | T GGG CTC CGG TGC GTT CGG C -3'
	DEP-1	5'- ATT CAA AAA GAT CAA AGT GC| -3'
	DEP-2	5'- |T GGC CTC CGG TGC GTT CGG C -3'
	TEB-F	56-FAM -5'- *TGTAC* * CGC ACC GGA GGC CA | *G CAC TTT* -3'
	TEB-Q	5'- |T GGC CTC CGG TGC G * *GTACA* -3'-3IABkFQ
EGFR-L858R	Correct Target	5'- AGT TTG GCC CGC CCA A | AA TCT GTG ATC TTG AC -3'
EGFR-L858R	Spurious Target	5'- AGT TTG GCC AGC CCA A | AA TCT GTG ATC TTG AC -3'
	DEP-1	5'- AGT TTG GCC CGC CCA A| -3'
	DEP-2	5'- |AA TCT GTG ATC TTG AC -3'
	TEB-F	5'- *ACA GAT T* | TT GGG CGG GCC AAA * *CATGT A* -3'-36-FAM
	TEB-Q	5IABkFQ -5'- *T ACATG* * T TTG GCC CGC CCA A| -3'
	Correct Target	5'- ATT CTT TCT CTT CCG CAC | CCA GCT GTT TGG CCA GCC -3'

EGFR-L861Q	Spurious Target	5'– ATT CTT TCT CTT CCG CAC | CCA GCA GTT TGG CCA GCC –3'
	DEP-1	5'– ATT CTT TCT CTT CCG CAC | –3'
	DEP-2	5'– |CCA GCT GTT TGG CCA GCC –3'
	TEB-F	56-FAM –5'– *TGTAC* * GGC CAA ACA GCT GG | *G TGC G* –3'
	TEB-Q	5'– CCA GCT GTT TGG CC * *GTACA* –3'–3IABkFQ
KRAS-G12A	Correct Target	5'– ACT TGT GGT AGT TGG AGC TGC TGG CGT AGG CAA GAG TG –3'
	Spurious Target	5'– ACT TGT GGT AGT TGG AGC TGG TGG CGT AGG CAA GAG TG –3'
	DEP-1	5'– ACT TGT GGT AGT TGG AGC T| –3'
	DEP-2	5'– |GCT GGC GTA GGC AAG AGT G –3'
	TEB-F	56-FAM –5'– *TGTAC* * TTG CCT ACG CCA GC | *A GCT C* –3'
KRAS-G12A	TEB-Q	5'– GCT GGC GTA GGC AA * *GTACA* –3'–3IABkFQ
KRAS-G13V	Correct Target	5'– ACT TGT GGT AGT TGG AGC TGG TGT CGT AGG CAA GAG TG –3'
	Spurious Target	5'– ACT TGT GGT AGT TGG AGC TGG TGG CGT AGG CAA GAG TG –3'
	DEP-1	5'– ACT TGT GGT AGT TGG AGC T| –3'
	DEP-2	5'– |GGT GTC GTA GGC AAG AGT G –3'
	TEB-F	56-FAM –5'– *TGTAC* * TTG CCT ACG ACA CC | *A GCT C* –3'

	TEB-Q	5'– GGT GTC GTA GGC AA * *GTACA* –3'-3IABkFQ
PIK3CA-H1047R	**Correct Target**	5'–ATG AAT GAT GCA CGT CAT GGT GGC TGG ACA ACA AAA ATG G –3'
	Spurious Target	5'–ATG AAT GAT GCA CAT CAT GGT GGC TGG ACA ACA AAA ATG G –3'
	DEP-1	5'– ATG AAT GAT GCA CGT CAT GG –3'
	DEP-2	5'– TGG CTG GAC AAC AAA AAT GG –3'
	TEB-F	5'– *CA GCC A* * CC ATG ACG TGC ATC * *CATGT* –3'-36-FAM
	TEB-Q	5IABkFQ –5'– *ACATG* * GAT GCA CGT CAT GG –3'
STK11-F354L	**Correct Target**	5'– GAT GAT GTC ATC CTC GAT GTC CAA GAG GTC CTC G –3'
	Spurious Target	5'– GAT GAT GTC ATC CTC GAT GTC GAA GAG GTC CTC G –3'
	DEP-1	5'– GAT GAT GTC ATC CTC GA –3'
STK11-F354L	**DEP-2**	5'– TGT CCA AGA GGT CCT CG –3'
	TEB-F	5'– *TTG GAC A* * TC GAG GAT GAC ATC * *CATGT* –3'-36-FAM
	TEB-Q	5IABkFQ –5'– *GTACA* * GAT GTC ATC CTC GA –3'

3.5 Reference

1. M. W. Schmitt, S. R. Kennedy, J. J. Salk, E. J. Fox, J. B. Hiatt, L. A. Loeb, Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. USA*. **109**, 14508 (2012).
2. H. Zhang, J. Chao, D. Pan, H. Liu, Y. Qiang, K. Liu, C. Cui, J. Chen, Q. Huang, J. Hu, L. Wang, W. Huang, Y. Shi, C. Fan, DNA origami-based shape IDs for single-molecule nanomechanical genotyping. *Nat. Commun*. **8**, 1–7 (2017).
3. S. Kim, A. Misra, SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. of Biomed. Eng*. **9**, 289–320 (2007).
4. M. Stroun, P. Anker, J. Lyautey, C. Lederrey, P. A. Maurice, Isolation and characterization of DNA from the plasma of cancer patients. *Europ. J. Cancer Clin. Oncol*. **23**, 707–712 (1987).
5. M. Murtaza, S.-J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S.-F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. **497**, 108–112 (2013).
6. A. M. Newman, S. V. Bratman, J. To, J. F. Wynne, N. C. W. Eclov, L. A. Modlin, C. L. Liu, J. W. Neal, H. A. Wakelee, R. E. Merritt, J. B. Shrager, B. W. Loo, A. A. Alizadeh, M. Diehn, An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med*. **20**, 548–554 (2014).
7. B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, J. L. Neumann, A DNA-fuelled molecular machine made of DNA. *Nature*. **406**, 605–608 (2000).
8. J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*. **309**, 1728–1732 (2005).
9. P. Yin, H. M. T. Choi, C. R. Calvert, N. A. Pierce, Programming biomolecular self-assembly pathways. *Nature*. **451**, 318–322 (2008).
10. S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W. M. Shih, Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*. **459**, 414–418 (2009).
11. D. Y. Zhang, G. Seelig, Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem*. **3**, 103–113 (2011).
12. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*. **270**, 467–470 (1995).
13. Q. Li, G. Luan, Q. Guo, J. Liang, A new class of homogeneous nucleic acid probes based on specific displacement hybridization. *Nucleic Acids Res*. **30**, e5–e5 (2002).
14. Y. Xiao, K. J. I. Plakos, X. Lou, R. J. White, J. Qian, K. W. Plaxco, H. T. Soh, Fluorescence Detection of Single-Nucleotide Polymorphisms with a Single, Self-Complementary, Triple-Stem DNA Probe. *Angew. Chem. Int*. **48**, 4354–4358 (2009).
15. D. Y. Zhang, E. Winfree, Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).

16. D. Y. Zhang, S. X. Chen, P. Yin, Optimizing the specificity of nucleic acid hybridization. *Nat. Chem*. **4**, 208–214 (2012).
17. S. X. Chen, D. Y. Zhang, G. Seelig, Conditionally fluorescent molecular probes for detecting single base changes in double-stranded DNA. *Nat. Chem*. **5**, 782–789 (2013).
18. L. R. Wu, J. S. Wang, J. Z. Fang, E. R. Evans, A. Pinto, I. Pekker, R. Boykin, C. Ngouenet, P. J. Webster, J. Beechem, D. Y. Zhang, Continuously tunable nucleic acid hybridization probes. *Nat. Methods*. **12**, 1191–1196 (2015).
19. J. S. Wang, Y. H. Yan, D. Y. Zhang, Modular probes for enriching and detecting complex nucleic acid sequences. *Nat. Chem*. **9**, 1222–1228 (2017).
20. R. K. Saiki, D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, H. A. Erlich, Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. **239**, 487–491 (1988).
21. C. Fan, K. W. Plaxco, A. J. Heeger, Electrochemical interrogation of conformational changes as a reagentless method for the sequence-specific detection of DNA. *Proc. Natl. Acad. Sci*. **100**, 9134–9137 (2003).
22. J. Das, K. B. Cederquist, A. A. Zaragoza, P. E. Lee, E. H. Sargent, S. O. Kelley, An ultrasensitive universal detector based on neutralizer displacement. *Nat. Chem*. **4**, 642–648 (2012).
23. S. O. Kelley, C. A. Mirkin, D. R. Walt, R. F. Ismagilov, M. Toner, E. H. Sargent, Advancing the speed, sensitivity and accuracy of biomolecular detection using multi-length-scale engineering. *Nat. Nanotechnol*. **9**, 969–980 (2014).
24. X. Yang, Y. Tang, S. D. Mason, J. Chen, F. Li, Enzyme-Powered Three-Dimensional DNA Nanomachine for DNA Walking, Payload Release, and Biosensing. *ACS Nano*. **10**, 2324–2330 (2016).
25. Y. Li, G. A. Wang, S. D. Mason, X. Yang, Z. Yu, Y. Tang, F. Li, Simulation-guided engineering of an enzyme-powered three-dimensional DNA nanomachine for discriminating single nucleotide variants. *Chem. Sci.* **9**, 6434–6439 (2018).
26. N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, E. Winfree, On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res*. **41**, 10641–10658 (2013).
27. S. Tyagi, F. R. Kramer, Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat. Biotechnol*. **14**, 303–308 (1996).
28. S. Tyagi, D. P. Bratu, F. R. Kramer, Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol*. **16**, 49–53 (1998).
29. S. Tyagi, Imaging intracellular RNA distribution and dynamics in living cells. *Nat. Methods*. **6**, 331–338 (2009).
30. J. Das, I. Ivanov, E. H. Sargent, S. O. Kelley, DNA Clutch Probes for Circulating Tumor DNA Analysis. *J. Am. Chem. Soc.* **138**, 11009–11016 (2016).
31. C. B. Markegard, C. P. Gallivan, D. D. Cheng, H. D. Nguyen, Effects of Concentration and Temperature on DNA Hybridization by Two Closely Related Sequences via Large-Scale Coarse-Grained Simulations. *J. Phys. Chem. B*. **120**, 7795–7806 (2016).
32. A. Vallée-Bélisle, F. Ricci, K. W. Plaxco, Engineering Biosensors with Extended, Narrowed, or Arbitrarily Edited Dynamic Range. *J. Am. Chem. Soc.* **134**, 2876–2879 (2012).

33. F. Ricci, A. Vallée-Bélisle, A. J. Simon, A. Porchetta, K. W. Plaxco, Using Nature's "Tricks" To Rationally Tune the Binding Properties of Biomolecular Receptors. *Acc. Chem. Res.* **49**, 1884–1892 (2016).
34. F. Ricci, A. Vallée-Bélisle, A. Porchetta, K. W. Plaxco, Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *J. Am. Chem. Soc.* **134**, 15177–15180 (2012).
35. A. Vallée-Bélisle, F. Ricci, K. W. Plaxco, Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proc. Natl. Acad. Sci*. **106**, 13802–13807 (2009).
36. J. S. Wang, D. Y. Zhang, Simulation-guided DNA probe design for consistently ultraspecific hybridization. *Nat. Chem*. **7**, 545–553 (2015).
37. J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, N. A. Pierce, NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem*. **32**, 170–173 (2011).
38. S. X. Chen, G. Seelig, An Engineered Kinetic Amplification Mechanism for Single Nucleotide Variant Discrimination by DNA Hybridization Probes. *J. Am. Chem. Soc.* **138**, 5076–5086 (2016).
39. R. L. Pullan, J. L. Smith, R. Jasrasaria, S. J. Brooker, Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites. Vectors*. **7**, 37 (2014).
40. M. Albonico, D. Engels, L. Savioli, Monitoring drug efficacy and early detection of drug resistance in human soil-transmitted nematodes: a pressing public health agenda for helminth control. *Int. J. Parasitol*. **34**, 1205–1210 (2004).
41. N. Rashwan, M. Scott, R. Prichard, Rapid Genotyping of β-tubulin Polymorphisms in Trichuris trichiura and Ascaris lumbricoides. *PLOS Negl. Trop. Dis*. **11**, e0005205 (2017).
42. G. Matamoros, M. M. Rueda, C. Rodríguez, J. A. Gabrie, M. Canales, G. Fontecha, A. Sanchez, High Endemicity of Soil-Transmitted Helminths in a Population Frequently Exposed to Albendazole but No Evidence of Antiparasitic Resistance. *Int. J. Infect. Dis*. **4**, 73 (2019).
43. Khodakov, D. A., Khodakova, A. S., Linacre, A. & Ellis, A. V. Toehold-Mediated Nonenzymatic DNA Strand Displacement as a Platform for DNA Genotyping. *J. Am. Chem. Soc.* **135**, 5612–5619 (2013).
44. Ren, B., Zhou, J.-M. & Komiyama, M. Straightforward detection of SNPs in double-stranded DNA by using exonuclease III/nuclease S1/PNA system. *Nucleic Acids Res*. **32**, e42–e42 (2004).
45. Wang, G. A. Expanding detection windows for discriminating single nucleotide variants using rationally designed DNA equalizer probes, DNA-Equalizer-Gate, DOI: 10.5281/zenodo.4059727, 2020.

Chapter 4

Weighing the thermodynamics and kinetics of chemically modified DNA using a rationally designed DNA balance

Contribution statement

The contents of Chapter 4 were done by Guan Alex Wang, Junpeng Xu, Sarah M. Traynor, Fatma Eljabu, and Xinghong Wu. S. M. T. and G. A. W. did the wet lab works and analyzed the experimental data. G. A. W. , J. X. and X. W. did the mathematical model building and fitting works. F. E. in Dr. Yan's team synthesized the cyclic-azobenzene modified oligonucleotides. Dr. Li conceived the idea and designed all the experiments.

4.1 Introduction

Besides its biological roles for storing and transmitting genetic information, DNA has also emerged into a class of highly programmable engineering material for the construction of diverse nanostructures and nanodevices. The research field of DNA nanotechnology has been revolutionary in the ability to control molecular self-assembly and thus made remarkable impact not only on material science[1–4] but also on molecular computing,[5–9] biosensing,[10–15] diagnostics,[16–20] and therapeutics[21–25]. One of the biggest advantages that separate DNA from any other engineering material is the predictability based on the unique Watson-Crick base-pairing rule. Important thermodynamic parameters such as standard Gibbs free energy (ΔG°) can be predicted accurately using *in silico* tools such as NUPACK,[26] which plays critical roles in designing nucleic acid hybridization probes and primers. More advanced uses of thermodynamic parameters have further enabled simulation-guided development of ultraspecific hybridization probes.[11,27,28] Critical kinetic parameters, such as rate constants, have also been determined for toehold-mediated strand displacement reactions,[10] a class of important building blocks of dynamic DNA nanotechnology, making it possible to simulate complex DNA reaction networks using *in silico* tools such as Visual DSD.[29]

To further expand the chemical diversity and functionality of DNA, numerous chemical modifications have been introduced with the ability to increase the nuclease resistance,[30–32] regulate gene expression,[33,34] or allow spatial/temporal control of hybridization events.[35–37] While drastically diversifies the chemical nature of DNA, accurate prediction of hybridization kinetics and thermodynamics becomes difficult because of the changes made by the chemical modifications. Critical thermodynamic

parameters such as ΔG° and Tm can be measured using Isothermal Titration Calorimetry (ITC), Differential Scanning Calorimetry (DSC), and DNA melting analysis. However, these methods often require tedious operations and heating procedures thus are not applicable to thermally unstable modifications such as many photoswitchable molecules including azobenzenes, spiropyrans, diarylethenes and hemithioindigos.[35] Moreover, these techniques do not offer kinetic analysis as only the static equilibrium states are recorded. Therefore, a simple, versatile, and rapid method that measures both the thermodynamic and kinetic properties of chemically modified DNA at its native form is highly desirable but remains unavailable.

Thus motivated, we propose a rationally designed DNA balance capable of accurately measuring both thermodynamic and kinetic parameters for challenging DNA modifications at constant room temperature (Fig. 4.1a). As an analogue to real-world balance, our DNA balance weighs the thermodynamic and kinetic parameters of a chemically modified DNA by comparing it with a panel of DNA weights made of unmodified DNA with pre-determined thermodynamics and kinetics. A photoswitchable modification, cyclic-azobenzene (cAB), was chosen as a testbed to demonstrate the feasibility of our strategy (Fig. 4.1b). Photoswitchable molecules such as azobenzenes, are common gears to construct light-responsive/controllable materials or systems through photo-isomeriazation.[38,39] Azobenzenes and derivatives have been well studied and incorporated into DNA to build light-responsive constitutional dynamic network, hydrogels, origami nanostructures, and drug delivery capsules.[36–44] As the trans(E)-to-cis(Z) switching of azobenzene is often limited by the UV light, efforts were made to tune the light into visible region and isomerization at room temperature for *in vivo* applications.[45–

[47] Introducing cyclic structure pulls the absorption bands into visible region and produces unusual cis(Z)-to-trans(E) photoisomerization.[48–53] The unstable trans(E)-configuration can be switched back to cis(Z)-isomer through ambient heat or irradiation of green light (520 nm, Fig. 4.1c). In this study, we have employed the DNA balance to examine the thermodynamic and kinetic changes caused by the cAB-modifications in DNA. Static equilibrium states are used to calculate the Gibbs free energy shift ($\Delta\Delta G^{\circ}_{rxn}$) of stable cis(Z)-cAB modification. The real-time kinetic analysis were used to determine rate constants in strand displacement reactions and to estimate the $\Delta\Delta G^{\circ}_{rxn}$ caused by the unstable trans(E)-cAB.

4.2 Results

4.2.1 Design principle and theoretical basis

As illustrated in Figure 4.1a, our DNA balance was designed as a duplex DNA (CT) with one strand labeled with a fluorophore (C) and the other labeled with a quencher (T). It also contains a sticky end known as DNA toehold, so that it could react rapidly with a panel of single-stranded DNA (ssDNA) weights (W) through the principle of toehold-exchange (Fig. 4.1b). The "weight" of each probe, $W_{(f,\,r)}$, is defined by the length of the forward (*f*) and reverse toehold (*r*). A 'heavier' DNA weight is designed to contain a longer forward toehold and/or shorter reverse toehold than a 'lighter' one and thus yields a higher reaction yield when reacting with the DNA balance. The standard free energy, yield, and rate constant of toehold-exchange reaction upon interrogating the DNA balance with each DNA weight can be quantitatively determined *in silico* through mathematical model (Section 4.2.3) or experimentally by monitoring the fluorescence signals in real-time as well as at equilibrium (Section 4.2.5). To measure chemical modifications, a "taring" step is performed first, where kinetic and thermodynamic landscapes of unmodified DNA balance (CT_N) is established by reacting with a panel of DNA weights with varying *f* and *r* (equation 1). Modification is then placed on the DNA balance by chemically incorporating it to the quencher-labeled DNA strand. The DNA balance carrying the chemical modification (CT_{cis}) is interrogated again using the same panel of DNA weights to obtain both thermodynamic and kinetic changes induced by the chemical modification (equation 4.2-4.4).

$$W_{(f,r)} + CT_N \rightleftharpoons W_{(f,r)}C + T_N,\ \Delta G_N^{\circ}\ for\ unmodified\ DNA\ balance;\ \textbf{(eq. 4.1)}$$

$$W_{(f,r)} + CT_{cis} \rightleftharpoons W_{(f,r)}C + T_{cis},\ \Delta G_{cis}^{\circ}\ for\ cis-cAB\ DNA\ balance;\ \textbf{(eq. 4.2)}$$

$$\Delta\Delta G^{\circ}_{CT} := \Delta G^{\circ}_{N} - \Delta G^{\circ}_{cis} = \left(\Delta G^{\circ}_{CW} + \Delta G^{\circ}_{T_N} - \Delta G^{\circ}_{W} - \Delta G^{\circ}_{CT_N}\right) - \left(\Delta G^{\circ}_{CW} + \Delta G^{\circ}_{T_{cis}} - \Delta G^{\circ}_{W} - \Delta G^{\circ}_{CT_{cis}}\right) \approx \Delta G^{\circ}_{CT_{cis}} - \Delta G^{\circ}_{CT_N};\ \textbf{(eq. 4.3)}$$

$$\Delta G^{\circ}_{CT_{cis}} = \Delta G^{\circ}_{CT_N} + \Delta\Delta G^{\circ}_{CT};\ \textbf{(eq. 4.4)}$$

In this study, cyclic azobenzene (cAB) was employed as a testbed to validate the effectiveness of our DNA balance strategy. cAB was chemically incorporated into the phosphate backbone of the quencher labeled probe T during the phosphoramidite synthesis (Fig. 4.1c). The thermally stable cis-cAB isomer can be photo-switched to trans-cAB upon irradiation at 390 nm, and the reverse isomerization occurs spontaneously under ambient temperature or irradiation at 520 nm (Fig. 4.1d). Introduction of the cis-cAB destabilizes a DNA duplex while the trans-cAB stabilizes the duplex through a more favored planar conformation (Fig. 4.1e). We chose cAB as testbed, because it is a representative class of photoswitchable chemical modifications to DNA, where direct probing of thermodynamics and kinetics using conventional techniques is challenging.

Mathematically, ΔG° of the cis-cAB modified duplex CT_{cis} ($\Delta G^{\circ}_{CT_{cis}}$) can be derived as ($\Delta G^{\circ}_{CT_N} + \Delta\Delta G^{\circ}_{CT}$) (equation 4), and $\Delta G^{\circ}_{CT_N}$ can be readily calculated using software tools such as NUPACK. Therefore, to determine $\Delta G^{\circ}_{CT_{cis}}$, all one needs to measure is $\Delta\Delta G^{\circ}_{CT}$ (equation 3). In principle, $\Delta\Delta G^{\circ}_{CT}$ can be measured using a single DNA weight, which is nearly impossible and unreliable for a chemical modification with unknown thermodynamics. By contrast, the panel of DNA weights were designed to cover a much wider range of reaction free energies (Fig. 4.2a). As such, we were able to easily establish standard thermodynamic curves (yield as a function of ΔG°_{rxn}) before and after

placing the chemical modification on the DNA balance (Fig. 4.2b). $\Delta\Delta G^{\circ}_{CT}$ could then be accurately determined as the shift of the standard curves using MATLAB code.

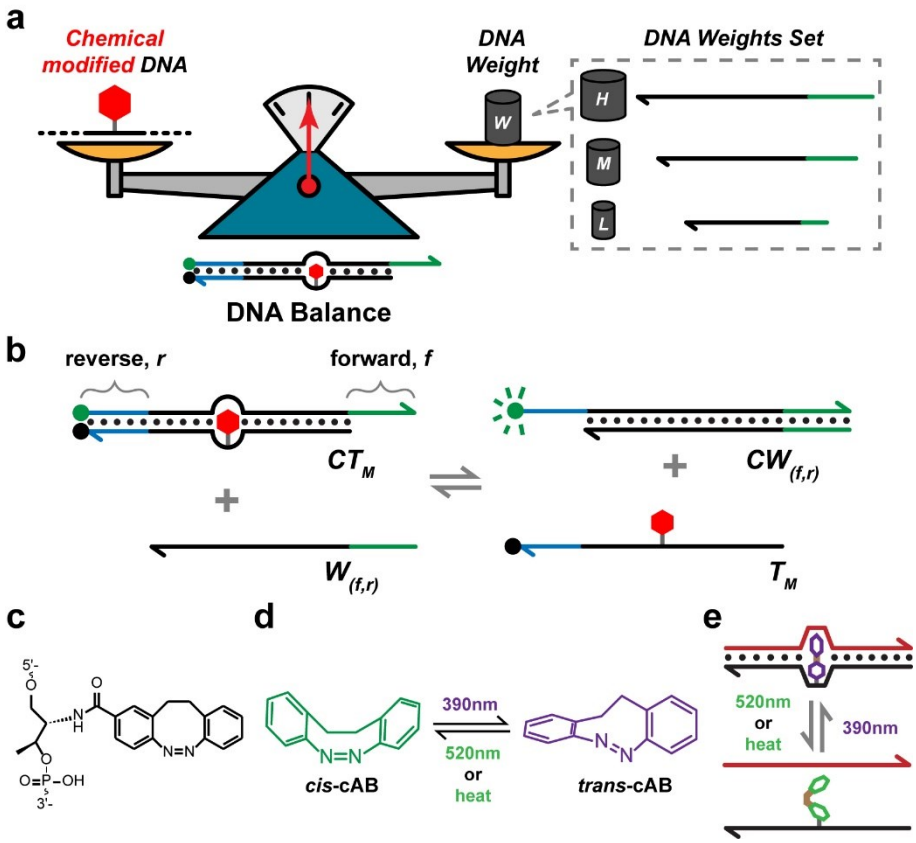


Figure 4.1. (**a**) Schematic showing the principle of DNA balance. The target DNA strands containing chemical modifications are weighed by a series of competing strands termed as DNA weights (W). The lengths of toeholds in DNA weights characterize the competitivity or weight of the strands. (**b**) Incorporation of chemical modification into the phosphate backbone of oligonucleotides (T_M). (**c**) Photoisomerization of cyclic-azobenzene. Thermally stable cis-cAB isomerized to unstable trans-cAB by irradiation of 390 nm light and the backward process is automated under ambient heat. (**d**) The cis-cAB modification destabilizes DNA duplex structure whereas trans-cAB configuration stabilizes DNA duplex. (**e**) Scheme of the reaction in DNA balance that weighing the cAB-modification. Green and blue domains represent the forward and reverse toeholds, respectively. Each DNA weight $W_{(f,r)}$ contains a specific pair of toehold lengths.

For kinetic analysis, rate constant of strand displacement reactions with cis-cAB modified DNA can be directly measured by monitoring fluorescence signals in real-time.

However, for strand displacement reactions involving trans-cAB, direct experimental measurement is impossible. This is because trans-cAB spontaneously relaxes to cis-cAB and thus leads to a much more complex kinetic behavior. Using the DNA balance, we demonstrated that it was possible to address this challenge and accurately determine rate constants for both strand displacement and the transition from trans- to cis- isomers by combining experimental measurement and mathematical modeling. Once obtaining the critical kinetic parameters, we were able to further derive standard Gibbs free energy for the transition between trans-cAB and cis-cAB.

4.2.2 Weighing the thermodynamic of cis(Z)-cAB modified DNA

We first set out to weigh the thermodynamics of cis-cAB modified DNA using the DNA balance. To do so, a panel of 16 DNA weights were designed by altering the length of the forward toehold from $f = 4$ nt to $f = 9$ nt and the length of reverse toehold from $r = 6$ nt to $r = 10$ nt. The combination of f and r, as well as the predicted ΔG°_{rxn} for each W is listed in Table 4.1. Each DNA weight was then denoted by the numeric values of f and r (Fig. 4.2a, left). For example, a DNA weight with $f = 9$ nt and $r = 4$ nt is denoted as $W_{(9,4)}$. The standard thermodynamic curve of unmodified DNA balance was then established by plotting the experimental determined yield against predicted ΔG°_{rxn} for each DNA weight. The experimentally determined calibration curve was found to be highly consistent with *in silico* prediction with a slight shift of +0.1 kcal·mol^{-1} (Fig. 4.3), suggesting that the thermodynamic landscape could be accurately established experimentally using our DNA balance.

Cis-cAB was then placed onto the DNA balance for the thermodynamic measurement. Three single-cAB modifications were chemically incorporated into the DNA balance but at different locations: near the forward toehold ($TC_{cis\text{-}1N}$, Fig. 4.2c), or in the middle ($TC_{cis\text{-}1M}$, Fig. 4.2d), or on the reverse toehold ($TC_{cis\text{-}1F}$, Fig. 4.2e). We also varied the number of modifications from one to two (Fig. 4.2f) and three (Fig. 4.2g). Each green dot in Figure 4.2c-4.2g represents the experimental measurement of a specific DNA weight. By fitting the experimental results of a panel of DNA weights against the theoretical model (Section 4.2.3) using least-square algorithm, we were able to establish thermodynamic calibration curves for cis-cAB modified DNA balance (the red dashed curves in Fig. 4.2c-4.2g). Positive shifts of the ΔG°_{CT} were observed for all cis-cAB modifications, which is consistent with previous reports that cis-cAB destabilizes DNA duplex.[54] Quantitatively, a single cis-cAB modification led to an average shift of +1.6 kcal·mol^{-1} and the location of modification had little impact on the numeric value of the energy shift. As expected, increasing the number of cis-cAB modifications promptly enlarges $\Delta\Delta G^{\circ}_{CT}$ (Fig. 4.2e-4.2f). The observation that the modification with two cis-cAB led to a much greater $\Delta\Delta G^{\circ}_{CT}$ (+6.3 kcal·mol^{-1}) than the simple addition of two single modifications (~ +3.2 kcal·mol^{-1}), suggesting that multiple modifications may work coordinately to destabilize a DNA duplex.

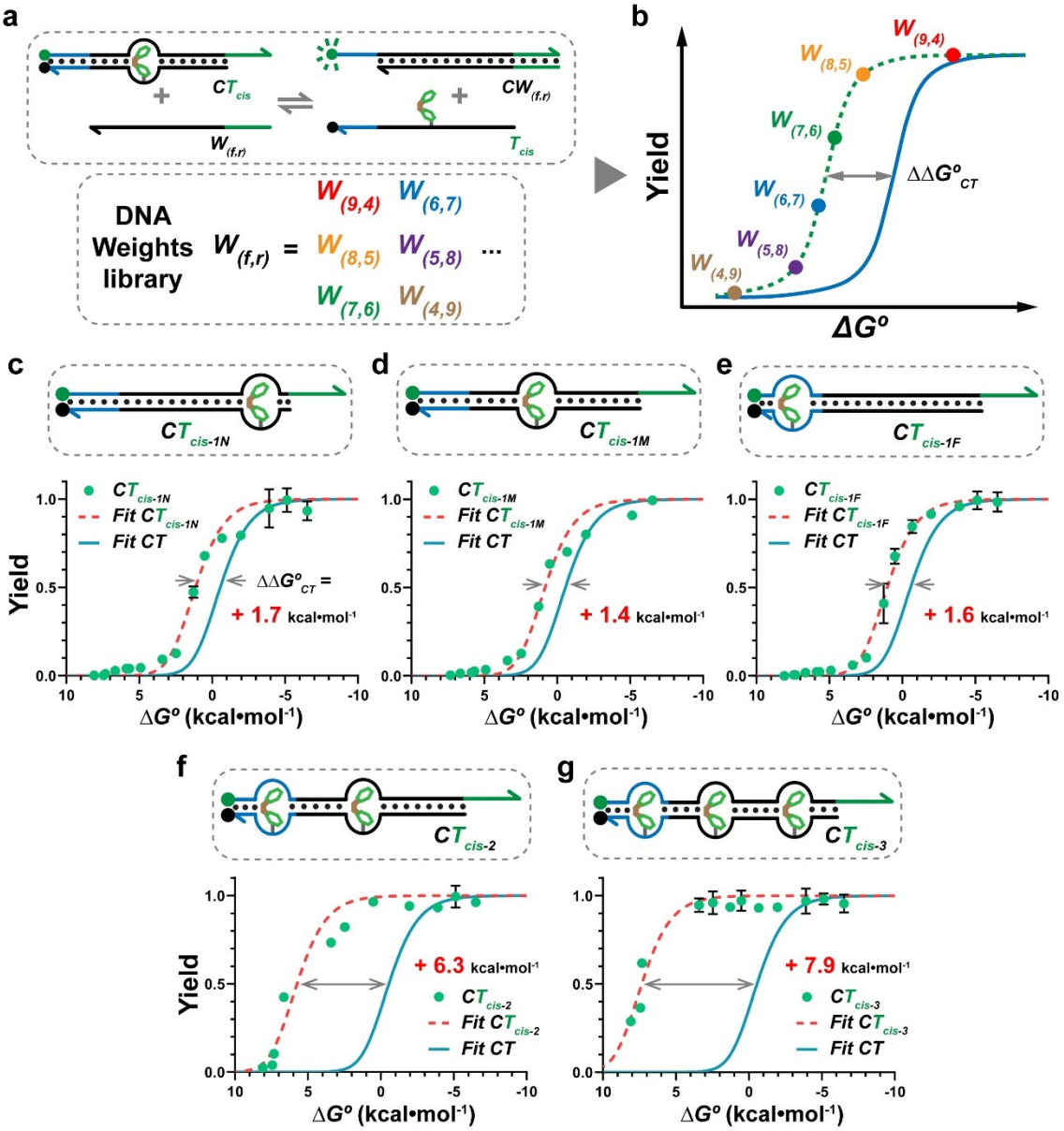


Figure 4.2. Thermodynamic study of the cis-cAB modified DNA balance. (**a**) Schematic illustration of the DNA weights library that are used to chart the reaction yield-versus-energy(ΔG°_{rxn}) standard curve (**b**). The blue solid line represents the standard curve of an 'empty' DNA balance, while green dotted line represents that of weighing the cos-cAB modification. Free energy shift ($\Delta\Delta G^{\circ}_{CT}$) is derived by the mathematical model (**eq. 1-3**) and the 'empty' DNA balance ($\Delta G^{\circ}_{CT_N}$) possesses energy of –25.56 kcal·mol^{-1}. (**b-f**) Experimental yields of DNA weights panel (green dots) are fitted to standard curves according to thermodynamic model (Section 4.2.3), and the $\Delta\Delta G^{\circ}_{CT}$ values (red) are derived thereafter. For single cAB modification at different positions ($TC_{cis\text{-}1N}$, $TC_{cis\text{-}1M}$, $TC_{cis\text{-}1F}$), the close values (+1.7, +1.4, and +1.6 kcal·mol^{-1}, respectively) suggest little impact of modification position. Two and three cAB-modifications ($TC_{cis\text{-}2}$, $TC_{cis\text{-}3}$) possess larger $\Delta\Delta G^{\circ}_{rxn}$ values of +6.3 and +7.9 kcal·mol^{-1}, respectively.

Table. 4.1. The complete 16 DNA weights used in our study and corresponding reaction free energies (ΔG°_{rxn}) for DNA balances $CT_{cis\text{-}0}$, $CT_{cis\text{-}1N}$, $CT_{cis\text{-}1M}$, $CT_{cis\text{-}1F}$, $CT_{cis\text{-}2}$, and $CT_{cis\text{-}3}$.

DNA Weight	$\Delta G^{\circ}_{CT_{cis-0}}$	$\Delta G^{\circ}_{CT_{cis-1N}}$	$\Delta G^{\circ}_{CT_{cis-1M}}$	$\Delta G^{\circ}_{CT_{cis-1F}}$	$\Delta G^{\circ}_{CT_{cis-2}}$	$\Delta G^{\circ}_{CT_{cis-3}}$
$W_{(4,10)}$	3.7410	3.6374		3.0298	2.2000	-0.1250
$W_{(5,10)}$	3.2781	2.6444	3.1836	3.0618	1.4576	-1.1420
$W_{(4,9)}$	3.0555	2.0555	2.3363	2.3158	0.3889	-2.2166
$W_{(5,9)}$	2.8737	1.8279	2.0986	2.1796	0.1768	-2.1613
$W_{(3,8)}$	2.8879	3.6548		5.2418	2.2149	0.1855
$W_{(4,8)}$	3.4200	1.9333	2.4792	2.8985	0.5579	-3.7024
$W_{(5,8)}$	2.6379	1.7530	1.9257	2.0067	-0.1213	-6.6800
$W_{(6,8)}$	2.3382	1.2380	1.2876	1.5429	-1.0096	-3.8282
$W_{(7,8)}$	1.9988	0.9890	0.9897	1.1612	-1.3152	-4.1904
$W_{(8,8)}$	0.3573	-1.4236	-1.2100	-1.4160	-2.4264	-4.5954
$W_{(9,8)}$	-1.9600	-2.0833	-2.1300	-3.2691	-2.9235	-3.5800
$W_{(7,7)}$	1.1385	-0.5467	-0.2337	-0.2999	-2.5810	-3.5922
$W_{(8,7)}$	-2.5140	-3.8467	-2.8594	-4.1694	-4.6771	-4.5068
$W_{(9,7)}$	-0.6300	-1.9817	-1.5453	-2.4788	-2.3831	-3.5183
$W_{(9,6)}$	-5.1300	-3.5692	-3.1708	-5.3313	-5.1300	-5.1300
$W_{(9,5)}$	-3.0935	-6.5200	-6.5200	-6.5200	-3.0608	-4.0536

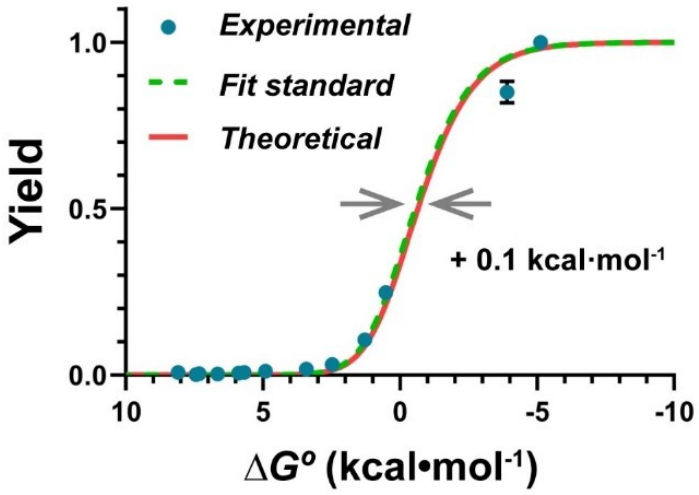


Figure 4.3. The yield-versus-ΔG°_{rxn} thermodynamic standard curve. The theoretical standard curve was plotted against the ΔG°_{rxn} values estimated from NUPACK and corresponding theoretical yields were calculated by **eq. 4.5**. A panel of DNA weights reacted with the non-modified DNA balance to generate experimental yields (blue dots) that were fitted to a reference standard curve (green dotted line). A correction of +0.1 kcal·mol^{-1} was derived according to the least-square error algorithm.

Having achieved thermodynamic characterization of cis-cAB, our next goal is to measure the thermodynamics of trans-cAB and the transition between the two isomers. However, we found it impossible for the direct measurement of thermodynamic parameters of trans-cAB, due to the spontaneous transition from trans- to cis-isomers (Fig. 4.4). As the DNA balance was dominated by kinetics in this scenario, we next measure critical kinetic parameters of the trans-cAB modified DNA. Thermodynamic parameters can then be derived once the kinetic behavior of the cAB was fully characterized.

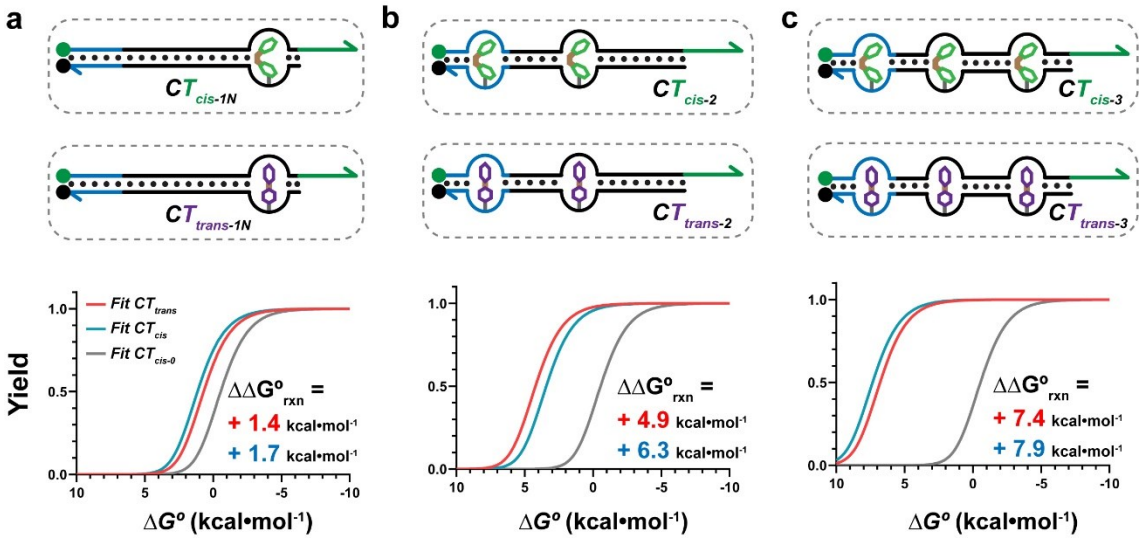


Figure 4.4. The yield versus reaction free energy ΔG°_{rxn} landscapes for DNA balances $CT_{cis\text{-}cAB}$ and $CT_{trans\text{-}cAB}$. (**a**) Schematic of DNA balances with single cAB-modification. The grey standard curve denotes the non-modified DNA balance, and the similar values of energy shift for cis- and trans-cAB modifications (+1.7 and +1.4 kcal·mol^{-1}, respectively) suggests the thermodynamic properties of $CT_{trans\text{-}cAB}$ cannot be distinguished from static reaction equilibrium states. It is same for the balances bearing two (**b**) and three (**c**) cAB modifications.

4.2.3 Thermodynamic standard curve

The principle of DNA balance is using toehold-exchange reaction to compare the binding affinities to object chemical modification with a series of DNA weights. The reaction yield (η) is derived as:

$$W + CT \rightleftharpoons WC + T, with\ \Delta G^{\circ}_{rxn};$$

$$\eta = \frac{(K_{eq}+\gamma+K_{eq}\tau) - \sqrt{(K_{eq}+\gamma+K_{eq}\tau)^2 - 4K_{eq}\tau(K_{eq}-1)}}{2(K_{eq}-1)}, where\ K_{eq} \neq 1;$$

$$\text{or } \eta = \frac{\tau}{\gamma+\tau+1}, where\ K_{eq} = 1;$$

where W is the DNA weight, C is the partial complementary strand (DNA beam) to the target T; the equilibrium constant $K_{eq} = e^{-\Delta G^{\circ}_{rxn}/RT}$; the reaction yield $\eta = [W_{(f,r)}C]/[CT_{cAB}]_0$; the experimental parameters $\gamma \coloneqq [T_{cAB}]/[CT_{cAB}]_0$ and $\tau \coloneqq [W_{(f,r)}]/[CT_{cAB}]_0$ (subscript 0 represents initial state). The free energy ΔG°_{rxn} of each DNA species can be calculated using NUPACK software. When the experimental conditions of τ and γ are constant, the only changing variable in DNA balance is the reaction free energy ΔG°_{rxn} corresponding to DNA weights. This ΔG°_{rxn}-versus-yield landscape is reflected as a sigmoidal standard curve.

To obtain the dimensionless reaction yield values from fluorescence data, we normalized the fluorescence values by following formula:

$$\eta = \frac{F_{CT} - F_{BG}}{F_{P.C.} - F_{BG}}; \text{ (eq. 4.5)}$$

where the F_{CT}, $F_{P.C.}$, and F_{BG} represent the equilibrium fluorescence signal generated from the DNA balance, the background level from CT duplex, and the maximum signal level produced by C only, respectively.

4.2.4 Weighing kinetics of cAB modified DNA.

Kinetic characterization of cis-cAB modified DNA in toehold-mediated strand displacement reaction is straightforward using our DNA balance (Fig. 4.5-4.7). The kinetic plots in Figure 4.7 demonstrate that the incorporation of cis-cAB modification into CT accelerated kinetics of the strand displacement with several DNA weights, including $W_{(7, 8)}$, $W_{(7, 7)}$, $W_{(9, 8)}$, and $W_{(8, 7)}$. This is to be expected, because cis-cAB destabilized the CT

duplex. For other 'heavier' and 'lighter' weights, no observable kinetic difference was found (Fig. 4.8b-4.8f, 4.8i-4.8j), as the overall reactions became much faster or slower. Collectively, these observations suggest that some of the DNA weights were kinetically more sensitive than others in chemical modifications, which were better suited for probing the kinetics of chemically modified DNA.

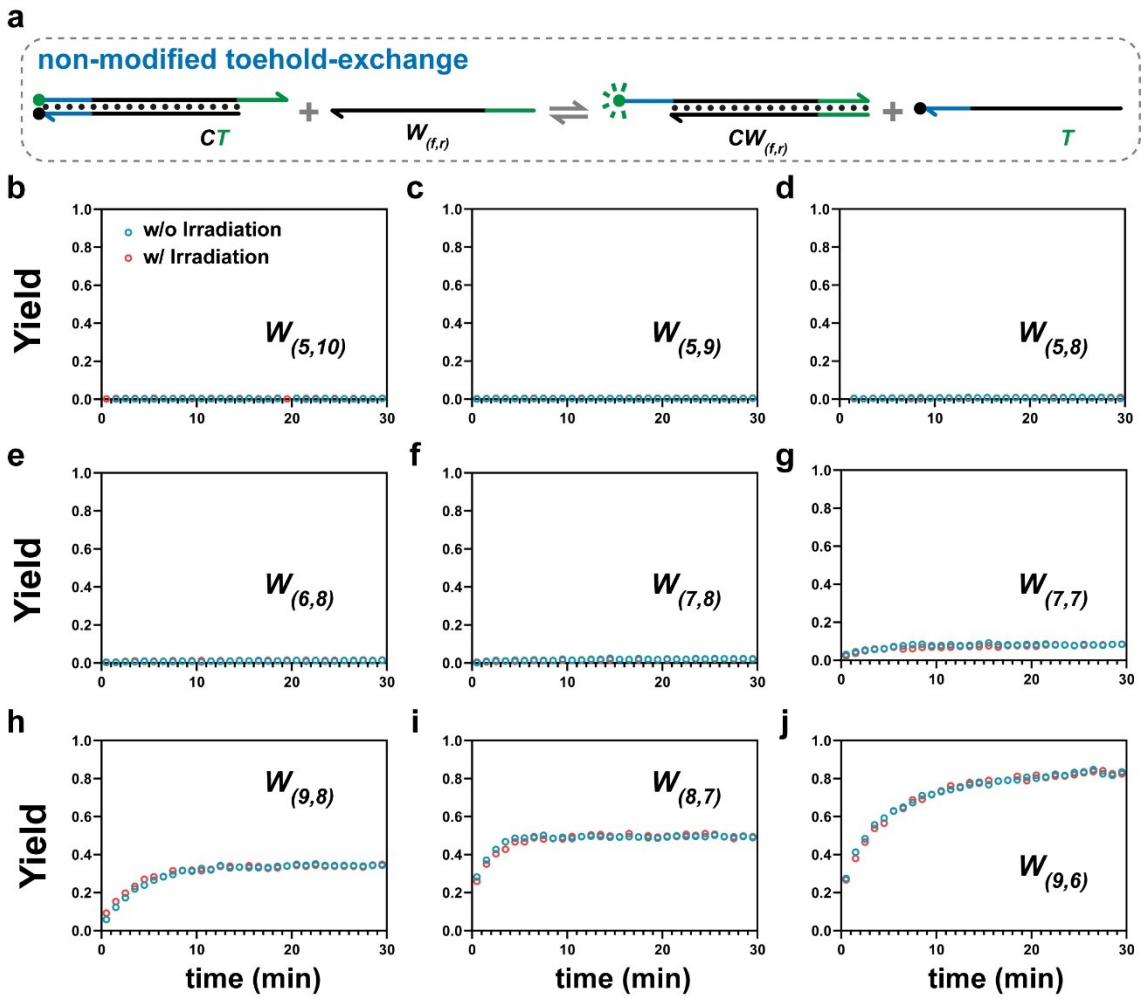


Figure 4.5. Kinetics of non-modified DNA balance with/without irradiation. (**a**) The toehold-exchange reaction occurred in a non-modified DNA balance. (**b-j**) Real-time normalized yield curves for a panel of DNA weights with forward toehold length $f = 5$ to 9 and reverse toehold length $r = 6$ to 10. No difference was observed for DNA balance $CT_{0\text{-}cAB}$ with or without irradiation.

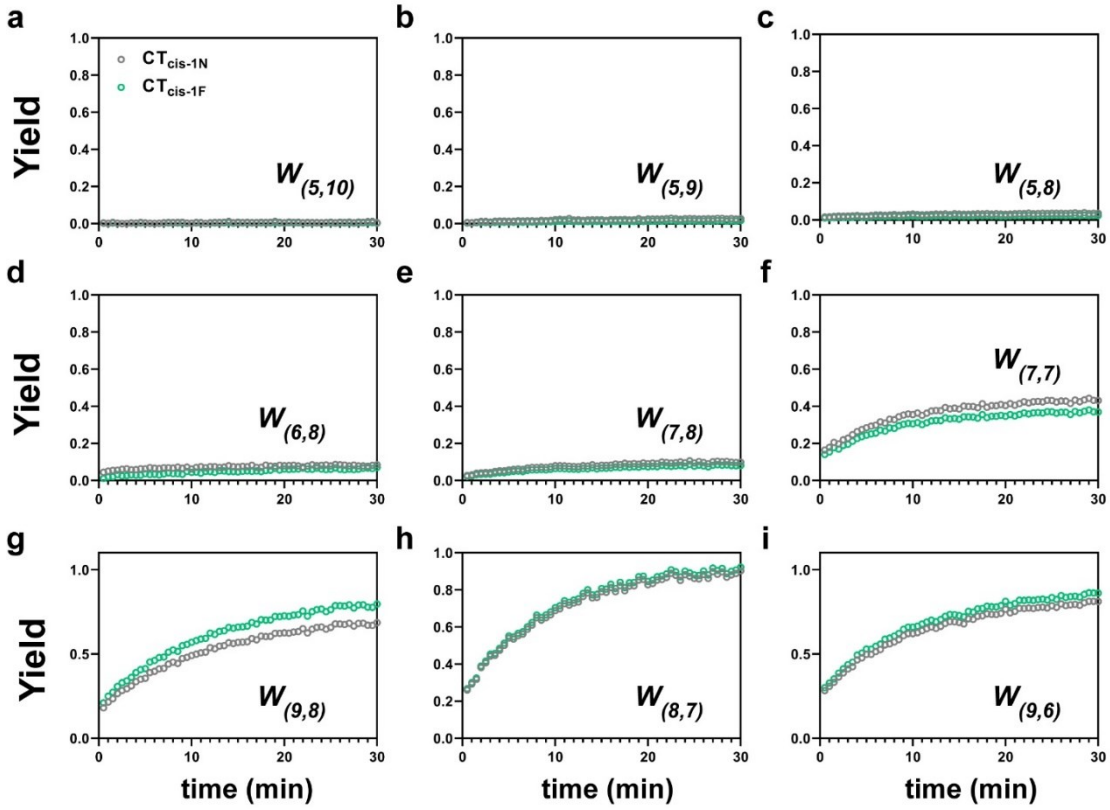


Figure 4.6. Real-time normalized yield curves of DNA balances that carrying a single cis-cAB modification $CT_{cis\text{-}1N}$ and $CT_{cis\text{-}1F}$. The little difference between the two distinct modification sites suggests the modification position has trivial impact on the performance of DNA balances.

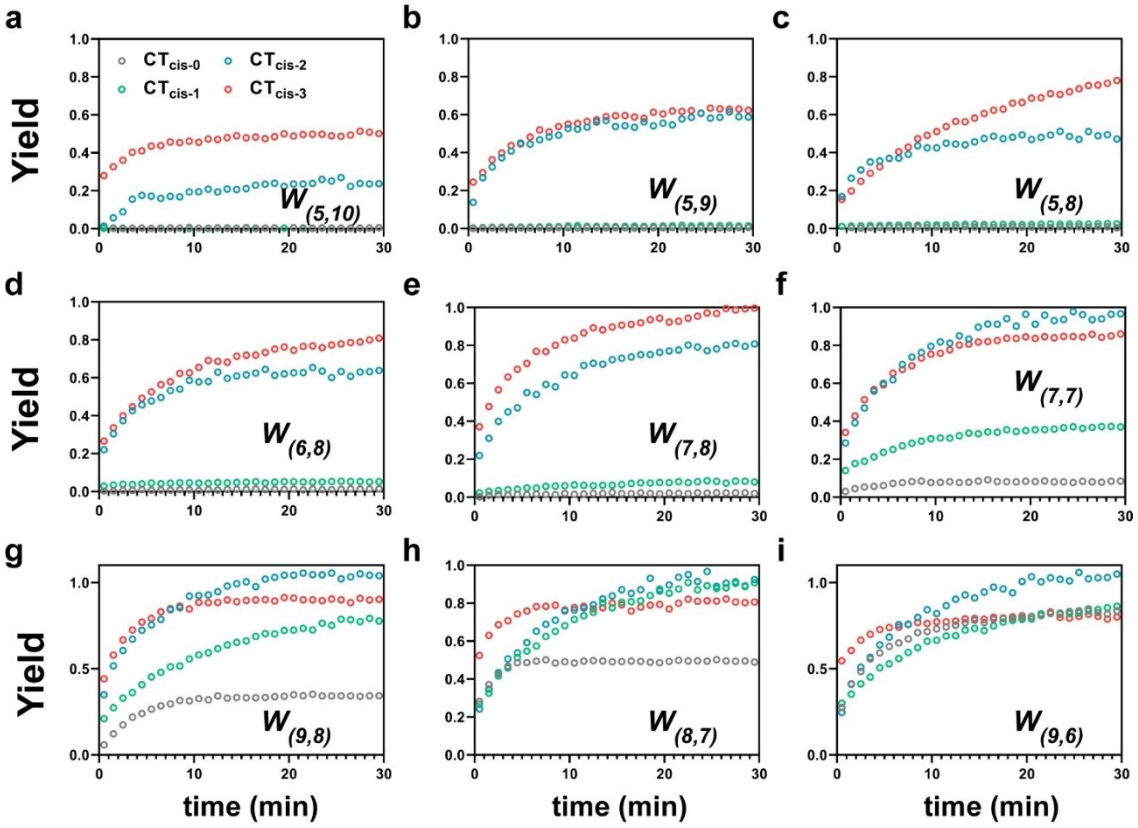


Figure 4.7. The real-time normalized yield curves of DNA balances containing different number of cAB modifications. Incorporation of cAB modification into the DNA balances can accelerate the kinetics. For single modification, only proper DNA weights including $W_{(7, 8)}$, $W_{(7, 7)}$, $W_{(9, 8)}$, and $W_{(8, 7)}$ can accelerate the kinetics . More cAB modifications (two and three) can accelerate the kinetics for light DNA weights (**a-h**).

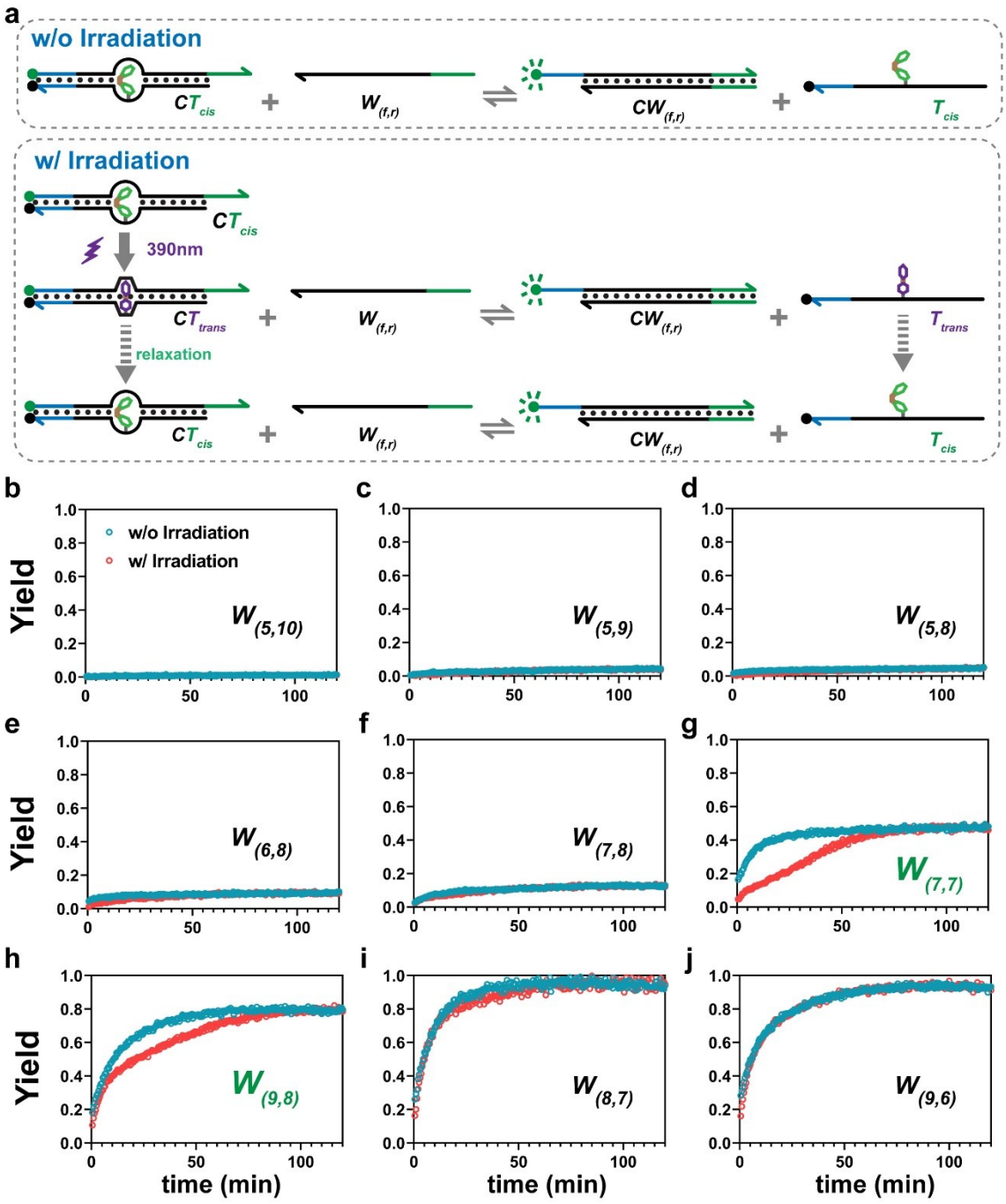


Figure 4.8. Kinetic study of the cAB modified DNA balance coupling with the trans-cis isomerization. (**a**) Proposed reaction pathways for DNA balances weighing cis-cAB (upper panel) and trans-cAB (lower panel) modification. The cis-isomer was photoswitched to trans-configuration by irradiation of 390 nm light while isomerized back to cis-form under ambient temperature. (**b-j**) Normalized real-time yield curves for different DNA weights. Heavier weights (**i**, **j**) exhibit higher yields and faster kinetics, and lighter weights (**a-f**) produced lower yields. However, the weight $W_{(7,7)}$ (**g**) and $W_{(9,8)}$ (**h**) exhibit obvious

sensitivity for the trans-cis isomerization process, which represented as the difference in kinetic curves (blue and red).

Our next goal is to screen DNA weights that are kinetically more sensitive to DNA balance carrying the trans-cAB modification. To do so, a cis-cAB modified DNA duplex CT_{cis} was irradiated at 390 nm for 30 min to ensure the complete transition to obtain the trans-cAB modification (CT_{trans}). Immediately after irradiation, the DNA balance bearing trans-cAB was interrogated with a panel of DNA weights to initiate toehold-exchange reactions (Fig. 4.8a). Meanwhile, a spontaneous relaxation from the trans- to cis-isomers occurred in parallel, which converted CT_{trans} back to CT_{cis} (Fig. 4.8a). By comparing kinetic curves with or without photoirradiation (Fig. 4.8b-4.8j), we found that most DNA weights were either too 'light' (Fig. 4.8b-4.8f) or too 'heavy' (Fig. 4.8i-4.8j) to reflect the kinetic difference induced by photoisomerization. By contrast, $W_{(7,7)}$ and $W_{(9,8)}$ revealed drastic changes in kinetics in the presence of trans-cAB (Fig. 4.8g and 4.8h). As expected, the trans-cAB stabilized the CT duplex and thus significantly slow down the rate of strand displacement. However, because of the spontaneous relaxation, a linear kinetic domain was observed before reaching equilibrium. Interestingly, we also found that the kinetically most sensitive DNA weight, $W_{(7,7)}$, led to a final reaction yield near 50% (Fig. 4.8g). Similar observations were also made for CT_{trans} with two and three cAB modifications (Fig. 4.9, 4.10).

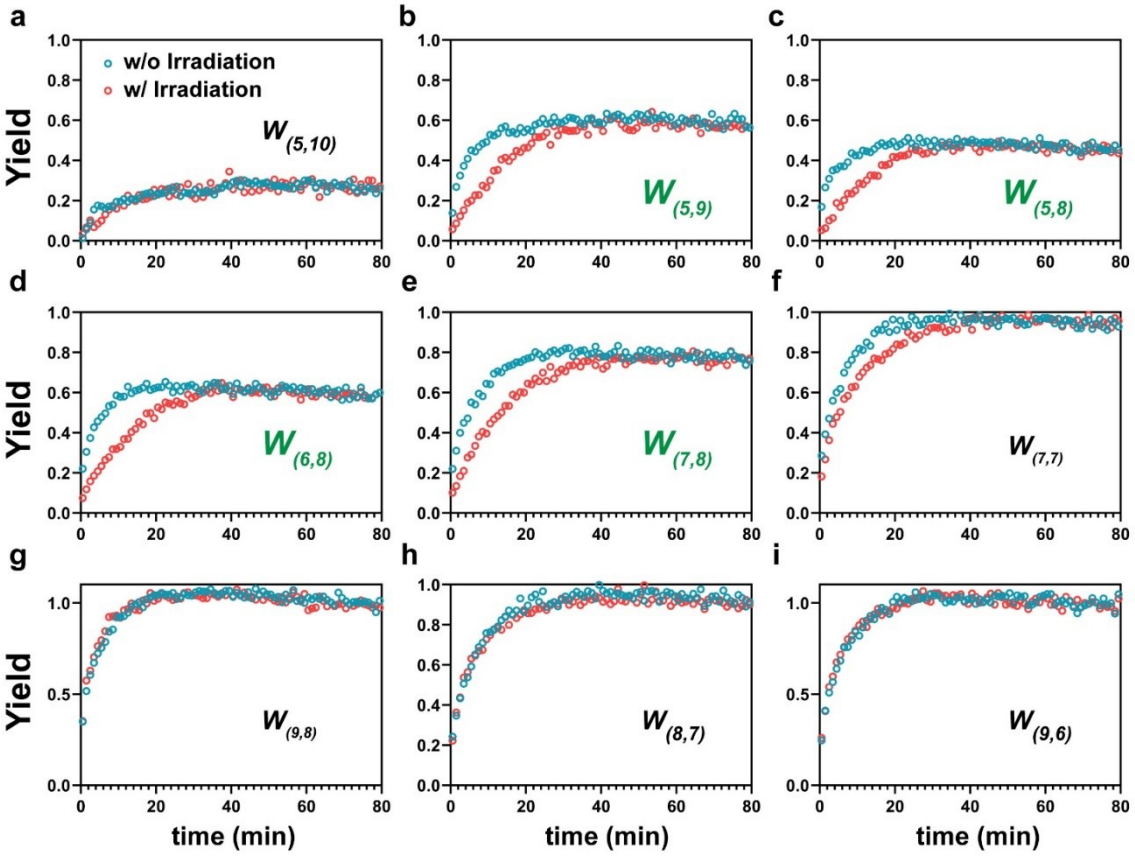


Figure 4.9. The real-time normalized yield curves of DNA balances containing two cAB-modifications ($CT_{cis\text{-}2}$). The DNA weights $W_{(5,9)}$ (**b**), $W_{(5,8)}$ (**c**) $W_{(6,8)}$ (**d**) and $W_{(7,8)}$ (**e**) show the kinetic sensitivity between with (red empty dots) and without (blue empty dots) of 390 nm irradiation.

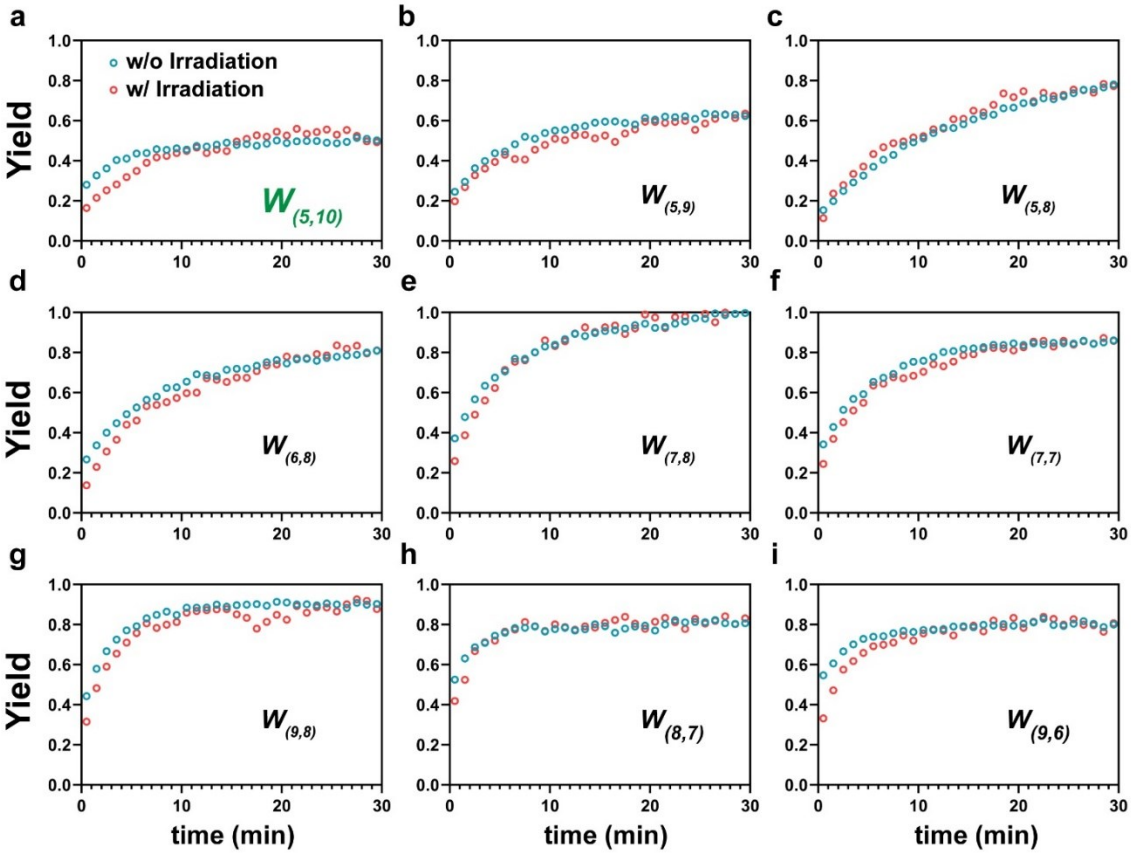


Figure 4.10. The real-time normalized yield curves of DNA balances containing three cAB-modifications ($CT_{cis\text{-}3}$). Only the DNA weights $W_{(5,10)}$ (**a**) shows kinetic sensitivity between with (red empty dots) and without (blue empty dots) of 390 nm irradiation.

Because of the complexity of the reaction network (Fig. 4.11a), direct measurement of kinetic parameters for trans-cAB modified DNA balance is impossible. As such, we propose to address this challenge by combining *in silico* simulation with experimental validation (Fig. 4.11b). Specifically, we first build a mathematical model capable of quantitatively describing the kinetic behaviors of kinetically-sensitive weights (Section 4.2.5). Our model suggests that the observed linear region in Figure 4.8g is a result of the spontaneous isomerization from CT_{trans} to CT_{cis} and the value of the rate constant for the isomerization (k_1) can be derived by the slope of the linear region (Fig. 4.11c). The

observed rate constant for cis-cAB modified DNA balance (k_3) could be determined experimentally using the DNA balance without photoirradiation (Fig. 4.11d). Once numeric values of k_1 and k_3 were measured, k_2 could then be determined using the workflow outlined in Figure 4b. Although varying kinetic profiles could be established *in silico* using our mathematical model (Fig. 4.11e), only when the correct k_2 value was reached, the predicted kinetic profile could be superimposed with the experimental measurement.

For the DNA balance with a single cAB modification ($CT_{cAB\text{-}1}$, Fig. 4.11f), the DNA weight $W_{(7,7)}$ was kinetically most sensitive to the trans-cAB. By fitting experimental observation with mathematical model using the workflow outlined in Figure 4.11b, rate constants k_1, k_2, and k_3 were determined to be 2.17×10^{-12} $M\cdot s^{-1}$, 3.73×10^{2} $M\cdot s^{-1}$, and 1.25×10^{5} $M\cdot s^{-1}$, respectively. The observation that the trans isomer slowed down the strand displacement by over 300 times suggests that the trans-cAB could effectively strengthen adjacent base pairs and inhibit the process of branch migration. When increasing the number of cAB modifications, lighter weights, with $W_{(4,9)}$ for two modifications (Fig. 4.12) and $W_{(5,10)}$ for three modifications were required to be kinetically sensitive to the photoisomerization. Slight increase in the rate of isomerization was observed when increasing the number of cAB modification with k_1 equals to 6.40×10^{-12} $M\cdot s^{-1}$ for two modifications and 1.03×10^{-11} $M\cdot s^{-1}$ for three modifications. The rate constants k_2 and k_3 remain to be the same magnitudes comparing to that of a single modification. Collectively, the kinetic analyses quantitatively revealed ~300 times difference between trans- and cis-cAB in terms of their ability to stabilize or destabilize DNA hybridization and a single

modification was sufficient to produce a drastic kinetic difference when deployed to strand displacement.

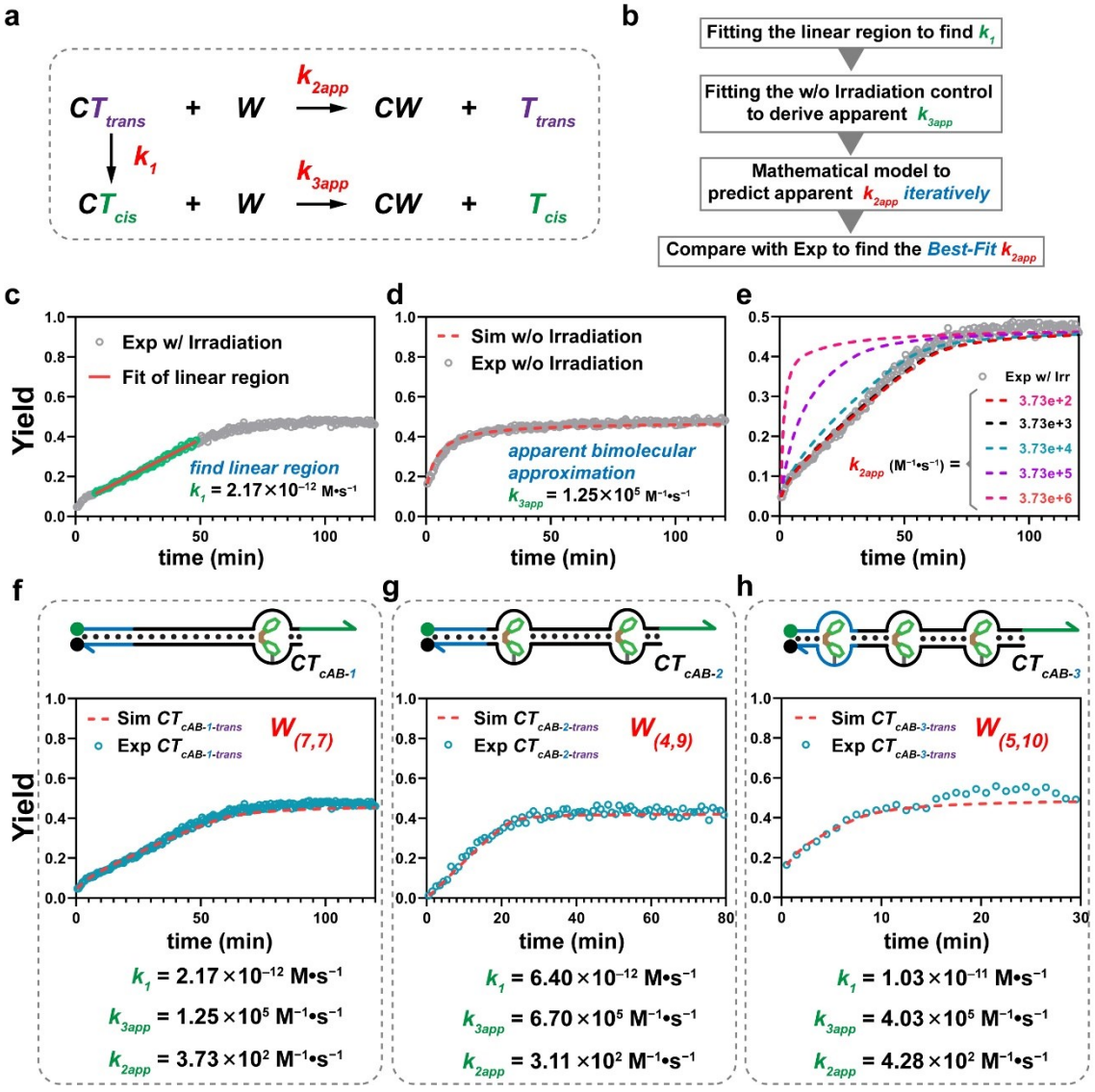


Figure 4.11. Simulation process of the kinetic-sensitive yield curves. (**a**) Proposed reaction model of DNA balance weighing a trans-cAB modification. The apparent isomerization was regarded as zero-order reaction (k_1); DNA balance reactions were simplified as apparent bimolecular reactions (k_{2app}, k_{3app}). (**b**) To fit the rate constants, values of k_1, k_{3app}, and k_{2app} were calculated sequentially. (**c-e**) For the example of single cAB modified DNA balance with weight $W_{(7,7)}$, the isomerization rate k_1 was fitted to be 2.17×10^{-12} M·s^{-1}. The bimolecular rate constant of 'empty' balance was determined to be 1.25×10^{5} M·s^{-1}. With solved k_1 and k_{3app} values, k_{2app} was then fitted to be 3.73×10^{2} M·s^{-1}. (**d-f**) The same simulation approach was applied upon two and three cAB modified DNA balances (TC_{cAB-2} and TC_{cAB-3}) and the derived rate constants were shown. Notably, lighter DNA weights

($W_{(4,9)}$ for $TC_{cAB\text{-}2}$ and $W_{(5,10)}$ for $TC_{cAB\text{-}3}$) were employed to generate the yield curves sensitive to trans-cis isomerization.

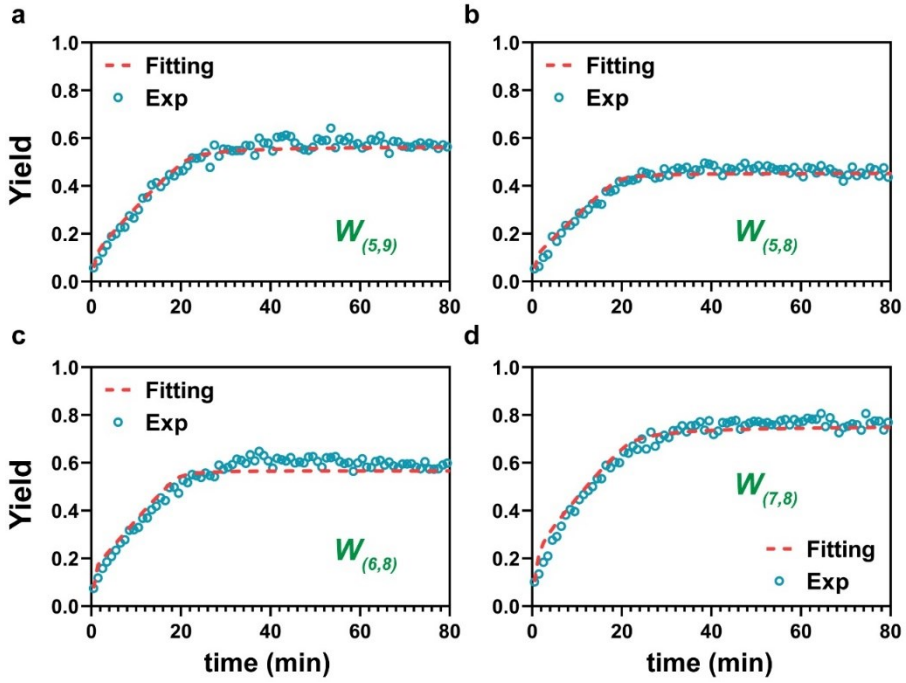


Figure 4.12. Fitting the kinetic-sensitive yield curves for target containing two cAB-modifications ($CT_{cis\text{-}2}$) with DNA weights $W_{(5,9)}$ (**a**), $W_{(5,8)}$ (**b**) $W_{(6,8)}$ (**c**) and $W_{(7,8)}$ (**d**).

Having determined the kinetic parameters, a final goal is to estimate the thermodynamic shift $\Delta\Delta G^{\circ}_{trans}$ of thermally unstable trans-cAB modification compared to non-modification. Because the major difference between the trans- and cis-isomers was in a DNA duplex rather than a single-stranded DNA, we assume the rates of the reverse strand displacement between T and CW were close for T_{trans} and T_{cis} in the DNA balance. As such, the equilibrium constants ratio K_2/K_3 roughly equals to k_2/k_3 and the free energy change of trans-cis isomerization ΔG°_{iso} could thus be calculated using **eq. 4.6**.

$$\Delta G^{\circ}_{iso} = \Delta G^{\circ}_{trans} - \Delta G^{\circ}_{cis} = -\mathrm{RT} \cdot \ln\left(K_2/K_3\right) \approx -\mathrm{RT} \cdot \ln\left(k_2/k_3\right); \textbf{ eq. 4.6}$$

Using the kinetic parameters determined in Figure 4.11f-g, we estimated the trans-cis cAB isomerization in DNA strands possesses an energy change ΔG°_{iso} equals to –5.04. –4.55, and –4.06 kcal·mol^{-1}, for single, double, and triple cAB modified DNA balances, respectively. The reaction free energy of trans-cAB modified DNA balance will then be derived from equations 4.7 and 4.8:

$$\Delta\Delta G^{\circ}_{trans} = \Delta G^{\circ}_{N} - \Delta G^{\circ}_{trans} = \left(\Delta G^{\circ}_{N} - \Delta G^{\circ}_{cis}\right) + \left(\Delta G^{\circ}_{cis} - \Delta G^{\circ}_{trans}\right)$$

$$= \Delta\Delta G^{\circ}_{cis} - \Delta\Delta G^{\circ}_{iso};\ \textbf{eq. 4.7}$$

$$\Delta G^{\circ}_{CT_{trans}} = \Delta G^{\circ}_{CT_N} + \Delta\Delta G^{\circ}_{trans};\ \textbf{(eq. 4.8)}$$

A complete comparison of the free energies of DNA balances ($\Delta G°_{CT}$) bearing trans-/cis-cAB modification(s) is shown in Table 1. The empty balance possesses a natural baseline level of –25.56 kcal·mol^{-1}. Introducing multiple cis-cAB modifications gradually increases the value of $\Delta G°_{CT}$ from –25.56 to –17.66 kcal·mol^{-1}, which is reflected as the easier destabilization of the DNA duplex. On the other hand, a single trans-cAB modification stabilizes the hybridization with an energy shift of –3.44 kcal·mol^{-1}. Increasing the number of trans-cAB modifications, however, destabilizes the CT duplex with $\Delta G°_{CT}$ of –23.81 and –21.72 kcal·mol^{-1} for two and three trans-cAB, respectively. Overall, the trans-cAB modified dsDNA are more stable compared to its cis-configuration, whereas it can both stabilize/destabilize the natural DNA duplex depending on the number of modifications.

Table. 4.2. The sequence design of 'empty' DNA balance and positions of cis-/trans-cAB modifications. Free Gibbs energy of DNA balances (ΔG°_{CT}) were derived from thermodynamic and kinetic analysis of DNA balances.

DNA Balance	Sequence	ΔG°_{CT} $(kcal \cdot mol^{-1})$
CT_N	CTT TAA GAA GGA *GAT ATA CC	–25.56
$CT_{cis\text{-}1N}$	CT Cis T TAA GAA GGA *GAT ATA CC	–23.86
$CT_{cis\text{-}2}$	CT Cis T TAA GAA G Cis GA *GAT ATA CC	–19.26
$CT_{cis\text{-}3}$	CTT TA Cis A GAA G Cis GA *GAT Cis ATA CC	–17.66
$CT_{trans\text{-}1N}$	CT Trans T TAA GAA GGA *GAT ATA CC	–29.00
$CT_{trans\text{-}2}$	CT Trans T TAA GAA G Trans GA *GAT ATA CC	–23.81
$CT_{trans\text{-}3}$	CTT TA Trans A GAA G Trans GA *GAT Trans ATA CC	–21.72

4.2.5 Kinetic model

For DNA balance containing thermally unstable trans-cAB modifications (CT_{trans}), it participates into two parallel reactions: directly with the DNA weight and isomerization to thermally stable CT_{cis} as shown in article Figure 4.11a. The accurate kinetic mathematical model for the reaction process is shown in **equation 4.9 to 4.11**.

$$CT_{trans} + W \rightleftharpoons CW + T_{trans}; \text{ with forward/reverse rate constants } k_{E-f}/k_{E-r}$$

$$CT_{cis} + W \rightleftharpoons CW + T_{cis}; \text{ with forward/reverse rate constants } k_{Z-f}/k_{Z-r}$$

$$CT_{trans} \longrightarrow CT_{cis}; \text{ with isomerization rate constant } k_{iso}$$

$$T_{trans} \longrightarrow T_{cis}; \text{ with isomerization rate constant } k_{iso}$$

As such,

$$\frac{d[CT_{trans}]}{dt} = -k_{E-f}[CT_{trans}][W] + k_{E-r}[CW][T_{trans}] - k_{iso}[CT_{trans}]; \text{ (eq. 4.9)}$$

$$\frac{d[CT_{cis}]}{dt} = -k_{Z-f}[CT_{cis}][W] + k_{E-r}[CW][T_{cis}] + k_{iso}[CT_{trans}]; \text{ (eq. 4.10)}$$

with boundary condition that $[CT]_{total} = [CT_{trans}] + [CT_{trans}] + [CW]$; (**eq. 4.11**)

and the collective fluorescence signal generated from CW is:

$$\frac{dF}{dt} = \frac{d[CW]}{dt} = k_{E-f}[CT_{trans}][W] + k_{Z-f}[CT_{cis}][W] - k_{E-r}[CW][T_{trans}] - k_{E-r}[CW][T_{cis}];$$

Clearly, fitting the kinetic rate constants merely from the fluorescence data is challenging as all the reactions are heavily coupled. Therefore, we simplify the accurate model into a mathematically-solvable one (**equations eq. 4.12-4.14**):

$CT_{trans} + W \longrightarrow CW + T_{trans}$; with apparent bimolecular rate constant k_{E-app}

$CT_{cis} + W \longrightarrow CW + T_{cis}$; with apparent bimolecular rate constant k_{Z-app}

$CT_{trans} \longrightarrow CT_{cis}$; with a constant intramolecular isomerization rate of $k_{iso-app}$

Therefore, the simplified mathematical model is:

$$\frac{d[CT_{trans}]}{dt} = -k_{E-app}[CT_{trans}][W] - k_{iso-app}; \text{ (eq. 4.12)}$$

$$\frac{d[CT_{cis}]}{dt} = -k_{Z-app}[CT_{cis}][W] + k_{iso-app}; \text{ (eq. 4.13)}$$

$$\frac{d[CW]}{dt} = -\frac{d[CT_{trans}]}{dt} - \frac{d[CT_{cis}]}{dt} = -k_{E-app}[CT_{trans}][W] - k_{Z-app}[CT_{cis}][W]; \text{ (eq. 4.14)}$$

Particularly, we regard the commonly accepted first-order isomerization reaction as zero-order because of two reasons: firstly, the cyclic-azobenzene is incorporated into the DNA strand thereby the isomerization is intramolecular; and secondly, the observed kinetic-

sensitive linear domain in fluorescence curves suggest a zero-order rate-limiting step in this reaction process. Even though with assumptions in our model, the overall fitting curves reflect the reaction process very well as shown in article Figure 4.11f-4.11h.

The time-elapsed fluorescence data cannot be fitted directly, and needs to be normalized to dimensionless yield curves at first. To reduce the batch-to-batch experimental variance, we use the self-normalization approach shown below:

$$\eta = \frac{F_{CT}-F_{BG}}{F_{e.q.}-F_{BG}} \times K_{e.q.}; \text{ (eq. 4.15)}$$

where the $F_{e.q.}$ and $K_{e.q.}$ represents the fluorescence level at equilibrated state and the equilibrium constant of corresponding DNA weight, respectively. As the trans-cAB modified DNA balance CT_{trans} will be relaxed to CT_{cis} during the weighing process, a same reaction yield will be reached for both CT_{trans} and CT_{cis}. Therefore, the same $K_{e.q.}$ value is applied for both DNA balances for a given DNA weight.

4.3 Discussion

In this study, we have introduced a rationally designed DNA balance capable of weighing hybridization thermodynamics and strand displacement kinetics of chemically modified DNA through the principle of toehold-exchange. Using the DNA balance, we have successfully determined critical thermodynamics and kinetic parameters of cyclic-azobenzene (cAB), a photoswitchable modification to DNA (Table. 4.2). More importantly, changes in ΔG° for both hybridization and photoisomerization have been quantitatively profiled, which is impossible for classic approaches such as ITC, DSC, and DNA melting

analysis. Critical rate constants for both strand displacement and isomerization have also been successfully determined using the DNA balance by screening kinetically sensitive DNA weights combined with mathematical modeling. As our strategy does not require specialized design and instruction, it can be easily expanded from cAB to other chemical modifications with minimal technical barriers.

Another advantage of DNA balance is that it characterizes chemically modified DNA in native conditions of interest. In previous study,[55] Zhang and colleagues demonstrated that native characterization of nucleic acid motif thermodynamics could lead to measured ΔG° with significant better accuracy. Similarly, the DNA balance measures hybridization thermodynamics of chemically modified DNA at one of the most representative conditions in DNA nanotechnology (room temperature and Tris-EDTA buffer with 12.5 mM Mg^{2+}). As such, critical thermodynamic parameters determined in this work can be used readily to guide the design and uses of cAB modified DNA sequences. Notably, thermodynamic of photoisomerization was estimated by combining experimental measurement with our theoretical modeling. The assumption that photo isomers has minimal impact on reverse strand displacement reaction may introduce error to estimated ΔG° values of the trans- to cis- transition. However, we believed that this error is minimal, as the determined value –4 kcal·mol^{-1} reasonably reflected the overall performance of the DNA balance in the presence of both isomers. Experimental setups, such as performing the measurement upon constant irradiation to maintain the trans-isomer, may help improve the accuracy of the measurement.

The determined rate constants of strand displacement reactions involving chemically modified DNA can also be deployed readily in dynamic DNA nanotechnology.

Quantitative profiling of the kinetics of toehold-mediated strand displacement and toehold exchange reactions has played critical roles in guiding the design of diverse dynamic DNA devices and reaction networks. Despite kinetics of strand displacement has been well-established for unmodified DNA sequences, little was explored for chemically modified DNA. Our work in designing the DNA balance as well as developing the mathematical model offers a powerful tool for characterizing the kinetic behaviors of chemically modified DNA in a quantitative manner. Our work also reveals that cAB with drastic changes in the stability of DNA duplex upon photoisomerization may serve as a unique photoswitch to design DNA probes with high tempo-controllability (*e.g.*, hybridize during photoisomerization and dehybridize as soon as removing light source) or to design photo-regulated dissipative reaction networks.

4.4 Methods and theoretical works

4.4.1 Methods

Materials. Magnesium chloride, 10 × phosphate buffered saline (PBS), TWEEN 20, Parafilm® were purchased from Sigma-Aldrich (Oakville, ON, Canada). Corning™ 96-well plate solid black polystyrene microplates, Axygen™ 0.5 mL microtubes and Fisherbrand™ SureOne™ micropoint pipette tips were purchased from Fisher Scientific (Ottawa, ON, CA). VWR® PCR Tubes purchased from VWR (Mississauga, ON, CA). Cyclic azobenzene was synthesized by the Yan Lab (Brock University, CA)

DNA sequences. All DNA weights strands were obtained from Integrated DNA Technologies (Oakville, IA, USA). Synthesis of the oligonucleotides containing cyclic-azobenzene modification at phosphate backbone was done at the University of Calgary DNA synthesis lab with an ABI 394 DNA synthesizer. The sequence design of DNA balances and weights, along with the modification positions is shown in Table 4.3.

Buffer condition. 1 × PBS (pH 7.4) buffer containing 10 mM $MgCl_2$ and 0.1% (v/v) TWEEN 20 was used as the reaction buffer (PM 10). TWEEN 20 was used to prevent the potential loss of DNA oligonucleotides during dilution and pipetting.

Preparing DNA balance. The strands C and T were firstly diluted from stock concentration 50 µM to 5 µM and 10 µM using buffer PM 10, respectively. 10 µL of C and 10 µL of T were mixed with 30 µL PM10 buffer in a 0.2 mL PCR tube. The DNA balance solution was then annealed in the BIORAD T100 Thermal Cycler. The anneal protocol is set as following: heat to 90°C and hold for 5 minutes, followed by a 5°C decrease in temperature every 2 minutes until a temperature of 20°C was reached.

Typical protocol of DNA balance weighing cAB modification. At room temperature, the DNA balance and weights were diluted to 200 nM using PM 10 for later usage. 10 μL of each DNA weight was added to a 0.5 mL microtube that contained 80 μL of PM 10 and then mixed by vortexing. 80 μL of the DNA weights solutions were transferred to a 96-well plate. Each DNA weight solution was quadrupled. The DNA balance solution that previously stored in dark was then irradiated for 30 minutes in a lab made irradiation box with a 395 nm LED. Once irradiation was complete, 10 μL of non-irradiated DNA balance ($CT_{cis\text{-}cAB}$) and 10 μL of the irradiated balance ($CT_{trans\text{-}cAB}$) were added into two separate wells containing the same DNA weight simultaneously. Each DNA balance was in duplex. The plate was read in the SpectraMax i3 fluorescence plate reader for 2 hours at room temperature with an excitation/emission of 485/535 nm.

Mathematical model building. Free energy of DNA strands and complexes were estimated by NUPACK. For parameters setting in NUPACK, the temperature was set to 25 °C, concentration of Na^+ was 0.1 M, and Mg^{2+} was 0.010 M. Other parameters were used set as default setting. The thermodynamic mathematical model and standard yield-versus-ΔG° landscape were solved and charted through numerical approach in MATLAB (2019a, MathWorks). The real-time kinetic analysis was performed in MATLAB using the corresponding codes.

4.4.2 DNA sequences

Table. 4.3. The DNA sequences of DNA balances and weights, and the cyclic-azobenzene (cAB) modification positions in DNA balances.

DNA Balance	5'-	CTT TAA GAA GGA GAT ATA CC	-3'
	5'-	GGTATAT CT CCT TCT TAA AG TTA CTT ACGA	-3'
		X represents the cAB modification	
$CT_{cis\text{-}1M}$	5'-	CTT TAA GAA G X GA GAT ATA CC	-3'
$CT_{cis\text{-}1N}$	5'-	CT X T TAA GAA GGA GAT ATA CC	-3'
$CT_{cis\text{-}1F}$	5'-	CTT TAA GAA GGA GAT ATA X CC	-3'
$CT_{cis\text{-}2}$	5'-	CT X T TAA GAA G X GA GAT ATA CC	-3'
$CT_{cis\text{-}3}$	5'-	CTT TA X A GAA G X GA GAT X ATA CC	-3'
DNA Weights (W)			
W(4,10)	5'-	GTA ACT TTA AGA AG	-3'
W(5,10)	5'-	A GTA ACT TTA AGA AG	-3'
W(4,9)	5'-	GTA ACT TTA AGA AGG	-3'
W(5,9)	5'-	AGTA ACT TTA AGA AGG	-3'
W(3,8)	5'-	TA ACT TTA AGA AGG A	-3'
W(4,8)	5'-	GTA ACT TTA AGA AGG A	-3'
W(5,8)	5'-	A GTA ACT TTA AGA AGG A	-3'
W(6,8)	5'-	AA GTA ACT TTA AGA AGG A	-3'
W(7,8)	5'-	TAA GTA ACT TTA AGA AGG A	-3'
W(8,8)	5'-	G TAA GTA ACT TTA AGA AGG A	-3'
W(9,8)	5'-	CG TAA GTA ACT TTA AGA AGG A	-3'
W(7,7)	5'-	TAA GTA ACT TTA AGA AGG AG	-3'
W(8,7)	5'-	G TAA GTA ACT TTA AGA AGG AG	-3'
W(9,7)	5'-	CG TAA GTA ACT TTA AGA AGG AG	-3'
W(9,6)	5'-	CG TAA GTA ACT TTA AGA AGG AGA	-3'
W(9,5)	5'-	CG TAA GTA ACT TTA AGA AGG AGAT	-3'

4.5 References

(1) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297–302.

(2) Hong, F.; Zhang, F.; Liu, Y.; Yan, H. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **2017**, *117* (20), 12584–12640.

(3) Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M. Self-Assembly of DNA into Nanoscale Three-Dimensional Shapes. *Nature* **2009**, *459* (7245), 414–418.

(4) Thubagere, A. J.; Li, W.; Johnson, R. F.; Chen, Z.; Doroudi, S.; Lee, Y. L.; Izatt, G.; Wittman, S.; Srinivas, N.; Woods, D.; Winfree, E.; Qian, L. A Cargo-Sorting DNA Robot. *Science* **2017**, *357* (6356).

(5) Chatterjee, G.; Dalchau, N.; Muscat, R. A.; Phillips, A.; Seelig, G. A Spatially Localized Architecture for Fast and Modular DNA Computing. *Nature Nanotechnology* **2017**, *12* (9), 920–927.

(6) Benenson, Y.; Adar, R.; Paz-Elizur, T.; Livneh, Z.; Shapiro, E. DNA Molecule Provides a Computing Machine with Both Data and Fuel. *PNAS* **2003**, *100* (5), 2191–2196.

(7) Woods, D.; Doty, D.; Myhrvold, C.; Hui, J.; Zhou, F.; Yin, P.; Winfree, E. Diverse and Robust Molecular Algorithms Using Reprogrammable DNA Self-Assembly. *Nature* **2019**, *567* (7748), 366–372.

(8) Qian, L.; Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **2011**, *332* (6034), 1196–1201.

(9) Cherry, K. M.; Qian, L. Scaling up Molecular Pattern Recognition with DNA-Based Winner-Take-All Neural Networks. *Nature* **2018**, *559* (7714), 370–376.

(10) Zhang, D. Y.; Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **2009**, *131* (47), 17303–17314.

(11) Wang, J. S.; Zhang, D. Y. Simulation-Guided DNA Probe Design for Consistently Ultraspecific Hybridization. *Nature Chemistry* **2015**, *7* (7), 545–553.

(12) Zhang, D. Y.; Chen, S. X.; Yin, P. Optimizing the Specificity of Nucleic Acid Hybridization. *Nature Chemistry* **2012**, *4* (3), 208–214.

(13) He, L.; Lu, D.-Q.; Liang, H.; Xie, S.; Luo, C.; Hu, M.; Xu, L.; Zhang, X.; Tan, W. Fluorescence Resonance Energy Transfer-Based DNA Tetrahedron Nanotweezer for Highly Reliable Detection of Tumor-Related MRNA in Living Cells. *ACS Nano* **2017**, *11* (4), 4060–4066.

(14) Wan, S.; Zhang, L.; Wang, S.; Liu, Y.; Wu, C.; Cui, C.; Sun, H.; Shi, M.; Jiang, Y.; Li, L.; Qiu, L.; Tan, W. Molecular Recognition-Based DNA Nanoassemblies on the Surfaces of Nanosized Exosomes. *J. Am. Chem. Soc.* **2017**, *139* (15), 5289–5292.

(15) Chang, X.; Zhang, C.; Lv, C.; Sun, Y.; Zhang, M.; Zhao, Y.; Yang, L.; Han, D.; Tan, W. Construction of a Multiple-Aptamer-Based DNA Logic Device on Live Cell

Membranes via Associative Toehold Activation for Accurate Cancer Cell Identification. *J. Am. Chem. Soc.* **2019**, *141* (32), 12738–12743.

(16) Tost, J.; Gut, I. G. Genotyping Single Nucleotide Polymorphisms by MALDI Mass Spectrometry in Clinical Applications. *Clinical Biochemistry* **2005**, *38* (4), 335–350.

(17) Zhang, H.; Chao, J.; Pan, D.; Liu, H.; Qiang, Y.; Liu, K.; Cui, C.; Chen, J.; Huang, Q.; Hu, J.; Wang, L.; Huang, W.; Shi, Y.; Fan, C. DNA Origami-Based Shape IDs for Single-Molecule Nanomechanical Genotyping. *Nature Communications* **2017**, *8* (1), 1–7.

(18) Sadat Mousavi, P.; Smith, S. J.; Chen, J. B.; Karlikow, M.; Tinafar, A.; Robinson, C.; Liu, W.; Ma, D.; Green, A. A.; Kelley, S. O.; Pardee, K. A Multiplexed, Electrochemical Interface for Gene-Circuit-Based Sensors. *Nature Chemistry* **2020**, *12* (1), 48–55.

(19) Wei, F.; Wang, J.; Liao, W.; Zimmermann, B. G.; Wong, D. T.; Ho, C.-M. Electrochemical Detection of Low-Copy Number Salivary RNA Based on Specific Signal Amplification with a Hairpin Probe. *Nucleic Acids Research* **2008**, *36* (11), e65–e65.

(20) Zhang, C.; Zhao, Y.; Xu, X.; Xu, R.; Li, H.; Teng, X.; Du, Y.; Miao, Y.; Lin, H.; Han, D. Cancer Diagnosis with DNA Molecular Computation. *Nature Nanotechnology* **2020**, *15* (8), 709–715.

(21) Wiraja, C.; Zhu, Y.; Lio, D. C. S.; Yeo, D. C.; Xie, M.; Fang, W.; Li, Q.; Zheng, M.; Van Steensel, M.; Wang, L.; Fan, C.; Xu, C. Framework Nucleic Acids as Programmable Carrier for Transdermal Drug Delivery. *Nature Communications* **2019**, *10* (1), 1147.

(22) Chi, Q.; Yang, Z.; Xu, K.; Wang, C.; Liang, H. DNA Nanostructure as an Efficient Drug Delivery Platform for Immunotherapy. *Front. Pharmacol.* **2020**, *10*.

(23) Chen, T.; Ren, L.; Liu, X.; Zhou, M.; Li, L.; Xu, J.; Zhu, X. DNA Nanotechnology for Cancer Diagnosis and Therapy. *International Journal of Molecular Sciences* **2018**, *19* (6), 1671.

(24) Evans, C. H. Gene Therapies for Osteoarthritis. *Curr Rheumatol Rep* **2004**, *6* (1), 31–40.

(25) Li, S.; Jiang, Q.; Liu, S.; Zhang, Y.; Tian, Y.; Song, C.; Wang, J.; Zou, Y.; Anderson, G. J.; Han, J.-Y.; Chang, Y.; Liu, Y.; Zhang, C.; Chen, L.; Zhou, G.; Nie, G.; Yan, H.; Ding, B.; Zhao, Y. A DNA Nanorobot Functions as a Cancer Therapeutic in Response to a Molecular Trigger in Vivo. *Nature Biotechnology* **2018**, *36* (3), 258–264.

(26) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *Journal of Computational Chemistry* **2011**, *32* (1), 170–173.

(27) Chen, X.; Liu, N.; Liu, L.; Chen, W.; Chen, N.; Lin, M.; Xu, J.; Zhou, X.; Wang, H.; Zhao, M.; Xiao, X. Thermodynamics and Kinetics Guided Probe Design for Uniformly Sensitive and Specific DNA Hybridization without Optimization. *Nature Communications* **2019**, *10* (1), 4675.

(28) Wang, G. A.; Xie, X.; Mansour, H.; Chen, F.; Matamoros, G.; Sanchez, A. L.; Fan, C.; Li, F. Expanding Detection Windows for Discriminating Single Nucleotide Variants Using Rationally Designed DNA Equalizer Probes. *Nature Communications* **2020**, *11* (1), 5473.

(29) Lakin, M. R.; Youssef, S.; Polo, F.; Emmott, S.; Phillips, A. Visual DSD: A Design and Analysis Tool for DNA Strand Displacement Systems. *Bioinformatics* **2011**, *27* (22), 3211–3213.

(30) Lennox, K. A.; Behlke, M. A. Chemical Modification and Design of Anti-MiRNA Oligonucleotides. *Gene Therapy* **2011**, *18* (12), 1111–1120.

(31) Gamper, H. B.; Reed, M. W.; Cox, T.; Virosco, J. S.; Adams, A. D.; Gall, A. A.; Scholler, J. K.; Meyer, R. B., Jr. Facile Preparation of Nuclease Resistant 3′ Modified Oligodeoxynucleotides. *Nucleic Acids Research* **1993**, *21* (1), 145–150.

(32) Kawasaki, A. M.; Casper, M. D.; Freier, S. M.; Lesnik, E. A.; Zounes, M. C.; Cummins, L. L.; Gonzalez, C.; Cook, P. D. Uniformly Modified 2'-Deoxy-2'-Fluoro-Phosphorothioate Oligonucleotides as Nuclease-Resistant Antisense Compounds with High Affinity and Specificity for RNA Targets. *J. Med. Chem.* **1993**, *36* (7), 831–841.

(33) Deleavey, G. F.; Damha, M. J. Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing. *Chemistry & Biology* **2012**, *19* (8), 937–954.

(34) Crooke, S. T.; Vickers, T. A.; Liang, X. Phosphorothioate Modified Oligonucleotide–Protein Interactions. *Nucleic Acids Research* **2020**, *48* (10), 5235–5253.

(35) Szymański, W.; Beierle, J. M.; Kistemaker, H. A. V.; Velema, W. A.; Feringa, B. L. Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. *Chem. Rev.* **2013**, *113* (8), 6114–6178.

(36) Liu, X.; Zhang, J.; Fadeev, M.; Li, Z.; Wulf, V.; Tian, H.; Willner, I. Chemical and Photochemical DNA "Gears" Reversibly Control Stiffness, Shape-Memory, Self-Healing and Controlled Release Properties of Polyacrylamide Hydrogels. *Chem. Sci.* **2019**, *10* (4), 1008–1016.

(37) Yan, Y.; Wang, X.; Chen, J. I. L.; Ginger, D. S. Photoisomerization Quantum Yield of Azobenzene-Modified DNA Depends on Local Sequence. *J. Am. Chem. Soc.* **2013**, *135* (22), 8382–8387.

(38) Asanuma, H.; Ito, T.; Yoshida, T.; Liang, X.; Komiyama, M. Photoregulation of the Formation and Dissociation of a DNA Duplex by Using the Cis–Trans Isomerization of Azobenzene. *Angewandte Chemie International Edition* **1999**, *38* (16), 2393–2395.

(39) Asanuma, H.; Liang, X.; Yoshida, T.; Komiyama, M. Photocontrol of DNA Duplex Formation by Using Azobenzene-Bearing Oligonucleotides. *ChemBioChem* **2001**, *2* (1), 39–44.

(40) Zhang, H.; Fu, H.; Shao, X.; Chipot, C.; Monari, A.; Dehez, F.; Cai, W. Conformational Changes of DNA Induced by a Trans-Azobenzene Derivative via Non-Covalent Interactions. *Phys. Chem. Chem. Phys.* **2018**, *20* (35), 22645–22651.

(41) Wang, S.; Yue, L.; Li, Z.-Y.; Zhang, J.; Tian, H.; Willner, I. Light-Induced Reversible Reconfiguration of DNA-Based Constitutional Dynamic Networks: Application to Switchable Catalysis. *Angewandte Chemie International Edition* **2018**, *57* (27), 8105–8109.

(42) Kuzyk, A.; Yang, Y.; Duan, X.; Stoll, S.; Govorov, A. O.; Sugiyama, H.; Endo, M.; Liu, N. A Light-Driven Three-Dimensional Plasmonic Nanosystem That Translates Molecular Motion into Reversible Chiroptical Function. *Nature Communications* **2016**, *7* (1), 10591.

(43) Liao, W.-C.; Willner, I. Synthesis and Applications of Stimuli-Responsive DNA-Based Nano- and Micro-Sized Capsules. *Advanced Functional Materials* **2017**, *27* (41), 1702732.

(44) Zhang, Y.; Zhang, Y.; Song, G.; He, Y.; Zhang, X.; Liu, Y.; Ju, H. A DNA–Azobenzene Nanopump Fueled by Upconversion Luminescence for Controllable Intracellular Drug Release. *Angewandte Chemie International Edition* **2019**, *58* (50), 18207–18211.

(45) Wegner, H. A. Azobenzenes in a New Light—Switching In Vivo. *Angewandte Chemie International Edition* **2012**, *51* (20), 4787–4788.

(46) Kou, B.; Guo, X.; Xiao, S.-J.; Liang, X. Highly Efficient Room-Temperature Photoresponsive DNA Tethering Azobenzene Through Backbone-Inserted Glycerol via Ether Bond. *Small* **2013**, *9* (23), 3939–3943.

(47) Asanuma, H.; Ishikawa, T.; Yamano, Y.; Murayama, K.; Liang, X. Cis-On/Trans-Off of DNA Hybridization with Alkylthio-Azobenzene on L-Threoninol Responding to Visible Light. *ChemPhotoChem* **2019**, *3* (6), 418–424.

(48) Siewertsen, R.; Neumann, H.; Buchheim-Stehn, B.; Herges, R.; Näther, C.; Renth, F.; Temps, F. Highly Efficient Reversible Z−E Photoisomerization of a Bridged Azobenzene with Visible Light through Resolved S1(Nπ*) Absorption Bands. *J. Am. Chem. Soc.* **2009**, *131* (43), 15594–15595.

(49) M, B.; Nl, D.; D, M. Unraveling a Chemically Enhanced Photoswitch: Bridged Azobenzene. *Angew Chem Int Ed Engl* **2010**, *49* (19), 3382–3384.

(50) Carstensen, O.; Sielk, J.; Schönborn, J. B.; Granucci, G.; Hartke, B. Unusual Photochemical Dynamics of a Bridged Azobenzene Derivative. *J. Chem. Phys.* **2010**, *133* (12), 124305.

(51) Böckmann, M.; Doltsinis, N. L.; Marx, D. Enhanced Photoswitching of Bridged Azobenzene Studied by Nonadiabatic Ab Initio Simulation. *J. Chem. Phys.* **2012**, *137* (22), 22A505.

(52) Cao, J.; Liu, L.-H.; Fang, W.-H.; Xie, Z.-Z.; Zhang, Y. Photo-Induced Isomerization of Ethylene-Bridged Azobenzene Explored by Ab Initio Based Non-Adiabatic Dynamics Simulation: A Comparative Investigation of the Isomerization in the Gas and Solution Phases. *J. Chem. Phys.* **2013**, *138* (13), 134306.

(53) Goulet-Hanssens, A.; Utecht, M.; Mutruc, D.; Titov, E.; Schwarz, J.; Grubert, L.; Bléger, D.; Saalfrank, P.; Hecht, S. Electrocatalytic Z → E Isomerization of Azobenzenes. *J. Am. Chem. Soc.* **2017**, *139* (1), 335–341.

(54) Biswas, M.; Burghardt, I. Azobenzene Photoisomerization-Induced Destabilization of B-DNA. *Biophysical Journal* **2014**, *107* (4), 932–940.

(55) Wang, C.; Bae, J. H.; Zhang, D. Y. Native Characterization of Nucleic Acid Motif Thermodynamics via Non-Covalent Catalysis. *Nature Communications* **2016**, *7* (1), 10319.

Chapter 5

Conclusion and Future works

In summary, two single-nucleotide-variants (SNVs) detection probes/assays and chemical modification measurement strategy have been developed: three-dimensional DNA nanomachine (3DDN), DNA equalizer probes (DEPs), and DNA balance. Both experimental and simulation-guided research approaches were adopted to study the fundamental reaction mechanisms.

In Chapter 2, I and my collaborator incorporated the noncovalent catalytic cycle onto the modified gold nanoparticles (AuNPs). Through a simulation-guided approach, the experimental results not only verified our design but also deepened our understanding of the toehold-exchange reaction at interface instead of bulk solution.

In Chapter 3, we proposed a simple yet powerful probe design termed DNA Equalizer Probes (DEPs) that discriminating SNVs in natural double-stranded DNA targets such as PCR amplicons. The development process is backed by accurate simulation works which help understand every single reaction step.

In Chapter 4, we realized the inherent similarity of existence of SNVs and chemical modifications in nucleic acids sequences. Therefore, we employed a dynamic DNA-based reaction, termed DNA balance, to profile the thermodynamic and kinetic properties of cyclic-azobenzene modified DNA. We anticipate this pioneering work harbors the

potential to fulfill dynamic DNA nanotechnology a new function that as a thermodynamic/kinetic measurement toolkit.

Even through powerful DNA-based technologies like the toehold-exchange reaction and the Sink design have been developed to distinguish SNV from its wild counterparts, several practical drawbacks limits their applications in real-world. One of the most severe limitations is the intrinsic trade-off between sensitivity and specificity for a DNA hybridization probe. That is, improving the discrimination performance of a probe sacrifices the reaction yields. In the Sink design, separate toehold-exchange probes aiming to consume undesired targets are utilized to enhance the selectivity; meanwhile a catalytic cycle design was used to amplify the signals simultaneously. To further enhance the thermodynamic difference between correct and spurious targets, X-probe and 4-way-toehold exchange reaction design were developed to increase the single base-pair mismatch to multiple base-pairs mismatch. By doing so, the selectivity can be increased drastically without minimal cost of sensitivity. In our work, we used the 3D DNA nanomachine as an amplification platform rather than using DNA-based circuits. To enhance the selectivity, an external Fuel strand was employed to compensate the loss of sensitivity. And through the simulation and experimental approaches, we have also discovered that the non-covalent catalysis strategy actually increases the selectivity at low target concentration range by lowering the specificity at higher target concentration range.

Another critical limitation of previous technologies is the robustness to the target concentration. That is, a specific probe or reaction network design can only work in a limited target concentration range. However, heavily used nucleic acids amplification techniques such as PCR, LAMP (Loop mediated isothermal amplification) and RCA

(Rolling circle amplification) produce unknown amount of target amplicons, therefore the SNV discrimination performance may lay outside the optimal concentration range for the probe/reaction-network. This limitation can be alleviated by the Sink or Kinetic trap designs but cannot be circumvented, especially if the target concentration was high. In our DNA Equalizer Gate, on the other hand, this problem was completely solved by using a self-competing mechanism. In this design, the complementary strand of a single-stranded target was utilized as a natural inherent competing "Sink" for the DNA Equalizer Probes. The competing activity increases along with increasing concentration of the target. Therefore, this limitation can be solved fundamentally in our DEG approach.

Overall, current SNV discrimination methods can be classified into two categories: the parallel and serial discrimination manners. A typical example for the parallel discrimination pattern is Sink design, in which multiple hybridization events occurring simultaneously. Whereas for the serial manner, multiple discrimination events or reactions occurred sequentially. For example, in our DEG design, the correct/spurious target is selected in the thermal protocol firstly and then discriminated in the following toehold-exchange reaction. Theoretically both strategies can work effectively whereas no systematic study to compare those two has done yet. Therefore, one of the future works is to study the SNV discrimination event systematically guided by those two design strategies.

With specific focus on previous work, more efforts will be done to solve the limitations of the DEG method. That is, exploring the correlating relationship between DEPs, at unbalanced concentration ratios or blocking regions; and reducing the hindrance effect caused by the secondary structure of long single-stranded targets. Besides, with the

high potential of in real-world application potential, more biological/clinical samples will be tested using our DEG approach.

In an open mind point of view, the DNA balance method is applicable to other DNA motifs as well, such as the G-quadraplex and aptamer. The thermodynamic and kinetic properties of the molecular recognition or folding processes can be quantitatively measured in a feasible and low-cost manner. Other than the applications exploration of current DNA balance method, in-depth study of the limits of our methods will also be embarked.

With the drastic development of artificial intelligence in silicon, smart facial, voice, and behavior recognition have been achieved. What's more exciting is that the AI-auxiliary diagnosis show great promises in hospital. In contrast, the in solution DNA-based computation and artificial intelligence is still at early stage. Therefore, in my future works, I would like to start research on the construction of DNA-based artificial intelligence networks and their applications in smart molecular recognition. To achieve this goal, studies will be embarked on the theoretical network theories and the compilation in DNA-based space; in parallel, hierarchical reaction networks will be studied from biological ones occurred within or between cells.

In the end, I would like to end this dissertation with a Chinese poem by Wei Wang (王维) from Tang(唐) dynasty:

老将行

唐・王维

少年十五二十时，步行夺得胡马骑。
射杀山中白额虎，肯数邺下黄须儿！
一身转战三千里，一剑曾当百万师。
汉兵奋迅如霹雳，虏骑奔腾畏蒺藜。
卫青不败由天幸，李广无功缘数奇。
自从弃置便衰朽，世事蹉跎成白首！
昔时飞箭无全目，今日垂杨生左肘。
路旁时卖故侯瓜，门前学种先生柳。
苍茫古木连穷巷，寥落寒山对虚牖。
誓令疏勒出飞泉，不似颍川空使酒。
贺兰山下阵如云，羽檄交驰日夕闻。
节使三河募年少，诏书五道出将军。
试拂铁衣如雪色，聊持宝剑动星文。
愿得燕弓射大将，耻令越甲鸣吾君。
莫嫌旧日云中守，犹堪一战立功勋！

Appendix

ACS SENSORS

Article

Cite This: ACS Sens. 2018, 3, 205−210

pubs.acs.org/acssensors

Paper-Based DNA Reader for Visualized Quantification of Soil-Transmitted Helminth Infections

Alex Guan Wang,[†,#] Tianyu Dong,[†,#] Hayam Mansour,[†,§] Gabriela Matamoros,[∥] Ana L. Sanchez,[⊥,∥] and Feng Li*[,†]

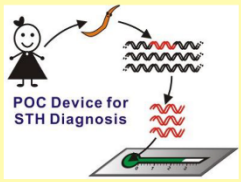
[†]Department of Chemistry, Centre for Biotechnology, and [⊥]Department of Health Sciences, Brock University, St. Catharines, Ontario, Canada, L2S 3A1

[§]Department of Cell Biology, National Research Center, Cairo, Egypt, 12622

[∥]Microbiology Research Institute, National Autonomous University of Honduras, Tegucigalpa, Honduras

Supporting Information

ABSTRACT: Soil-transmitted helminth (STH) infections are a global health issue affecting nearly one-third of the world's population. As most endemic areas of STH are impoverished countries or regions with limited healthcare resources, the accurate diagnosis of STH requires analytical tools that are not only quantitative, but also portable, inexpensive, and with no or minimal demand for external instrument. Herein, we introduce a novel paper-based diagnostic device, termed quantitative paper-based DNA reader (qPDR), capable of quantifying STH at the molecular level by measuring distance as readout, thus eliminating the need for external readers. On the basis of the unique interfacial interaction of a DNA intercalating dye, SYBR Green I, with native cellulose on a chromatographic paper, qPDR allows the distance-based quantification of minute amounts of double-stranded DNA as short as 6 min. By integrating qPDR with polymerase chain reactions that were performed using a smartphone-controlled portable thermal cycler, we were able to quantify minute amount of genetic markers from adult worms of an STH (*Trichuris trichiura*) that were expelled post-treatment by infected children living in the rural areas of Honduras.



KEYWORDS: *point-of-care, nucleic acid testing, paper-based analytical device, distance-based detection, equipment-free quantification, soil-transmitted helminth*

Soil-transmitted helminth (STH) infections are a global health issue, affecting nearly one-third of the world's population (∼2 billion), and are one of the most important causes of childhood malnutrition and cognitive impairment.[1] As the morbidity caused by STH is closely related to the level of infections, the accurate diagnosis thus requires quantitative tools to measure parasitic worms in the faecal samples.[1−3] Classic STH diagnosis relying on microscopy for detecting and counting the number of helminth eggs or worms excreted in faeces is quantitative, but is extremely tedious and time-consuming, requiring well-trained personnel with the competence to identify eggs or larvae, and generally lacks the necessary sensitivity for the accurate diagnosis.[3] Molecular diagnosis using quantitative polymerase chain reaction (qPCR) provides sufficient sensitivity, specificity, and throughput, but is too expensive to be implemented in most endemic regions where poverty is widespread and resources are limited.[2,3] To be readily used in such areas, an ideal diagnostic tool for STH shall be not only quantitative, but also inexpensive, easy to mass produce, and with no or minimal instrumental needs. Herein, we describe a low-cost paper-based diagnostic device capable of quantifying STH infections through the detection of nucleic acid markers at point-of-care (POC) settings.

While qualitative paper-based POC devices are available for nucleic acid testing primarily in the form of colorimetric and/or lateral-flow paper devices,[4−6] it remains a grand challenge to develop quantitative devices that are inexpensive and instrument-free. Recent successes in fabricating two-dimensional paper networks and three-dimensional paper origami have advanced the paper-based analytical devices remarkably for nucleic acids in terms of dedicated fluidic controls.[7−11] However, accurate quantification still relies on external electronic readers to collect colorimetric, fluorescent, or electrochemical signals and then convert them into digital readout through electrical domains.[4−11] As suggested by the World Health Organization, an ideal POC diagnostic tool for the developing world should be "reader-free", raising a key technical question on how to directly generate the digital readout without the need for electrical domains.[12,13] Recent efforts toward this challenge have revealed three types of physical readouts that are "digital" and thus are countable, including time,[13−15] distance,[16−22] and the number of colored segments.[23] For example, Philips and co-workers explored "time" as a digital readout on paper-based devices for the reader-free quantification of hydrogen peroxide, metal ions, and active enzymes.[13−15] Henry and co-workers demonstrated the

Received: November 19, 2017
Accepted: December 29, 2017
Published: January 16, 2018

DOI: 10.1021/acssensors.7b00857
ACS Sens. 2018, 3, 205−210

use of "distance" as a quantitative readout for the detection of small molecular targets and metal ions.[17,18] Yang and co-workers also demonstrated a series of distance-based paper devices for small molecules and proteins through the integration of cellulose paper and functional nucleic acids.[19−21] Despite the past success in developing a digital readout for paper-based devices, there is no device or sensing mechanism that is readily available for our need of quantitative STH diagnosis. Particularly, there are very few (only two) "digital" paper-based devices available for nucleic acid testing and neither device is simple, robust, and versatile enough for nucleic acid quantification at clinical settings.[24,25] For example, Hongwarittorrn recently described a distance-based sensing method for DNA amplicons generated by loop mediated isothermal amplification (LAMP) on a paper-based device. However, this device measures changes in Mg^{2+} ions rather than the actual DNA amplicon, the performance of which can be influenced by interferences such as divalent metal ions and complexing agents (e.g., EDTA) in the samples.[24] Kalish et al. described a distance-based paper device that quantifies DNA by inducing the aggregation of DNA-conjugated microspheres.[25] As this strategy can only detect a short ssDNA target, it is thus not feasible for our need of analyzing genomic STH samples. To this end, we introduce a new sensing mechanism that is fully compatible with clinical genomic samples and readily translates the nucleic acid quantification into the measurement of distance on a paper-based diagnostic device. We first describe the development of the distance-based paper-based device, termed quantitative paper-based DNA reader (qPDR), enabled by the unique interfacial interaction between a widely used DNA intercalating dye, SYBR Green I (SG-I), and unmodified cellulose on a chromatographic paper. We then demonstrate the utility of qPDR for the visualized quantification of double-stranded DNA (dsDNA) and nucleic acid amplicons of polymerase chain reactions (PCR). Finally, we demonstrate the applicability of qPDR to clinical samples of STH worms that were expelled post-treatment by infected children living in the rural areas of Honduras.

EXPERIMENTAL SECTION

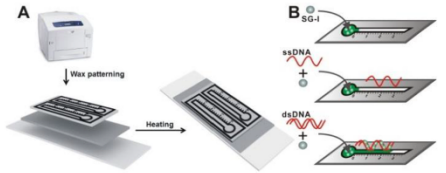
Materials and Reagents. Whatman filter paper (Grade 1), microscope glass slides, paraffin film, sodium citrate ($Na_3C_6H_5O_7$), citrate acid solution ($H_3C_6H_5O_7$, hydrochloric acid (HCl), sodium hydroxide (NaOH), 10× phosphate buffer saline (10 × PBS), TWEEN 20, polyethylene glycol (PEG) 100 000, and SYBR Green I (SG-I) dye were purchased from Sigma (Oakville, ON, Canada). Taq 2 × PCR Master Mix, *N,N,N′,N′*-tetramethylethylenediamine (TEMED), ammonium persulfate (APS), 40% acrylamide/bis-acrylamide solution, and DNA loading buffer were purchased from Bio-Rad Laboratories, Inc. (Mississauga, ON, Canada). NANOpure H_2O (>18.0 MΩ), purified using an Ultrapure Milli-Q water system, was used for all experiments. All synthetic DNA samples (Table S1) were purchased from Integrated DNA Technologies (Coralville, IA) and purified using standard desalting.

Genomic DNA Samples of Soil-Transmitted Helminth (STH). STH worm samples were recovered from eight school-age children infected with *Trichuris trichiura* in the rural region of La Hicaca located in the northwestern area of Honduras. Ethical approval was obtained by the National Autonomous University of Honduras and Brock University. The eight participants received a treatment scheme based on pyrantel pamoate and oxantel pamoate (Conmetel) during the first three days and Albendazole during a fourth day. The adult worms expelled in faeces were washed with saline solution and stored in 70% ethanol. Following the recovery of specimens, DNA was extracted using the Automate Express DNA Extraction System (Thermo Fisher Scientific Inc.) with the commercial kit PrepFiler Express BTA, according to the manufactureer's protocol.

Buffer Conditions. DNA stock solutions were prepared by dissolving oligonucleotides using deionized water and then stored at −20 °C. Concentrations of DNA stock solutions were determined by measurement of absorbance at 260 nm using a Thermo Spectronic Unicam UV−visible spectrophotometer. Unless indicated otherwise, 100 mM citrate buffer (pH of 4) containing 0.5% TWEEN 20 and 1 mg/mL PEG was used as the assay buffer.

Device Design and Fabrication. All qPDR devices were designed to contain a circular sample loading zone (6 mm inner diameter) and a linear test zone (2.0 mm in width ×36 mm in length) using CorelDraw ×8 (Scheme 1A). A ruler with a resolution of 2.0 mm was marked along the test zone to facilitate the measurement of distance as signal readout. Upon fabrication, the patterns were first printed onto cellulose filter paper using a XEROX ColorQube 8580 solid ink printer and then heating on a hot plate at 150 °C for 40 s. The device was then fabricated by stacking the patterned paper and a layer of paraffin film on a microscope glass slide (Scheme 1A). This sandwiched device was then bonded by heating on the hot plate at 110 °C for 30 s. The final width of the testing zone after fabrication was determined to be 1.5 mm.

Scheme 1. Schematic Illustration of the Fabrication Process (A) and the Assay Principle (B) of qPDR



Nucleic Acid Quantification Using qPDR. A typical reaction mixture (10 μL) containing varying concentrations of target nucleic acid and 20 μM SG-I in reaction buffer was first incubated at room temperature for 5 min and then loaded into the sample loading zone of qPDR. After sample loading, qPDR was placed on a pocket-size blue-light box (Mini LED Transilluminator, IO Rodeo Inc.) until the reaction mixture completely wicked through the test zone (typically within 10 min). The migration distance was measured in real-time by naked eye examination. To obtain the kinetic data, videos of qPDR in action were taking by a Nikon D600 digital camera (SI Video). The kinetic curve was then established by converting the video into images at a frequency of 1 snapshot per 20 s. All chromatograms were obtained by first taking photo at 10 min after sample loading using a Nexus 6P smartphone camera and then analyzed using ImageJ.

Quantification of PCR Amplicons Using qPDR. For a typical test, PCR reaction mixtures containing varying concentrations of a 164 bp target DNA, primers at a final concentration of 500 nM each and 1 × Taq master mix were placed in a BioRad T100 thermal cycler. The thermal cycles included an initial incubation at 94 °C for 3 min, followed by 35 cycles (denaturation at 94 °C, annealing at 52 °C and extension at 72 °C for 30 s each) and a final extension at 72 °C for 5 min. After PCR, each amplicon was mixed with SG-I at a final concentration of 20 μM and PEG at a final concentration of 1 mg/mL and then incubated at room temperature for 5 min. The mixture was then loaded and quantified using qPDR using the protocol outline above.

Quantification of STH Samples Using qPDR. A pair of primers was designed to amplify a fragment of 164-bp *β*-tubulin genomic sequence from whipworm *Trichuris trichiura* (TT). For a typical test, the PCR mixture contains 2 ng of TT genomic DNA, a pair of primers (250 nM each), and 1 × Taq mater mix. To meet the needs of future uses at rural areas of Honduras, this mixture was amplified using a low-cost portable thermal cycler (MiniPCR). The PCR protocol and

DOI: 10.1021/acssensors.7b00857
ACS Sens. 2018, 3, 205−210

subsequent detection using qPDR were identical as outlined above. Two round worm *Ascaris lumbricoides* (AL) genomic samples were also included as negative controls to determine the specificity of the assay. After PCR, each reaction was separated into two equal aliquots, with one aliquot loaded and quantified using qPDR and the other analyzed using polyacrylamide gel electrophoresis (PAGE).

■ RESULTS AND DISCUSSION

Assay Principle and Device Fabrication. As illustrated in Scheme 1A, qPDR containing a sample loading zone and a linear test zone is fabricated using a well-established wax printing technique.[26,27] The mechanism of qPDR for nucleic acid analysis is on the basis of the unique interfacial interactions of SG-I with unmodified cellulose on the chromatographic paper (Scheme 1B). In our preliminary tests, we found that a strong interaction exists between SG-I and cellulose on qPDR, evidenced by the strong retention of SG-I at the loading zone upon sample addition (Figure S1). Moreover, we also found two unique features of this interaction, which can potentially be harnessed to design sensing devices for nucleic acid testing. First, the binding affinity between SG-I and cellulose is stronger than that between SG-I and single-stranded DNA (ssDNA) but much weaker than that between SG-I and dsDNA. As a result, SG-I can be efficiently eluted into the test zone by dsDNA, whereas ssDNA has almost no elution capacity to SG-I (Figure 1). Furthermore, the eluent strength of dsDNA is concentration dependent and thus the migration distance of SG-I in the test zone can be quantitatively determined by the concentrations of the dsDNA, making it possible to quantify dsDNA by simply reading the migration distance of SG-I in qPDR. Second, fluorescence of SG-I was greatly enhanced by cellulose upon binding. The magnitude of the fluorescence enhancement was found to be equivalent to that of maximal concentrations of dsDNA (Figure 1 and S1). This feature is highly useful for distance-based quantification of dsDNA, as the fluorescence intensity of SG-I is essentially independent of dsDNA. Therefore, it is possible to visualize the distribution of SG-I within qPDR in the presence of minute amount of dsDNA or in the absence of dsDNA.

Distance-Based Nucleic Acid Quantification Using qPDR. Figure 1 shows a typical test of dsDNA and ssDNA using qPDR. The migration distances (d_R) of SG-I for dsDNA or ssDNA can be clearly visualized upon irradiating qPDR with a pocket-size blue-light box (Figure 1A). A d_R of 22.0 mm was developed for 500 nM 44-bp dsDNA within 6 min, whereas the d_R for equal concentration of a 44-nt ssDNA was only 4.0 mm (Figure 1B), suggesting that qPDR can effectively differentiate dsDNA over ssDNA. To quantitatively study the distribution of SG-I throughout the qPDR device, it is also possible to convert the distribution of SG-I in the test zone into a chromatogram by taking a photo, splitting color channels to enhance the sensitivity, and then digitalizing the test zone (1.5 mm × 30.0 mm) into 18 000 pixels with a resolution of 50 μm per pixel (Figure S2). When the chromatogram was used to facilitate the measurement of d_R, a threshold of 15% normalized fluorescence was used. We chose this threshold value, because it was determined to be close to the sensitivity of the naked eye (details in Supporting Information S2).

As we harness the unique binding property of SG-I on cellulose paper as the key sensing mechanism, the concentration of SG-I was thought to play a critical role in maximizing the analytical performance of qPDR. First, the enhanced fluorescence of SG-I on cellulose depends on the concentration

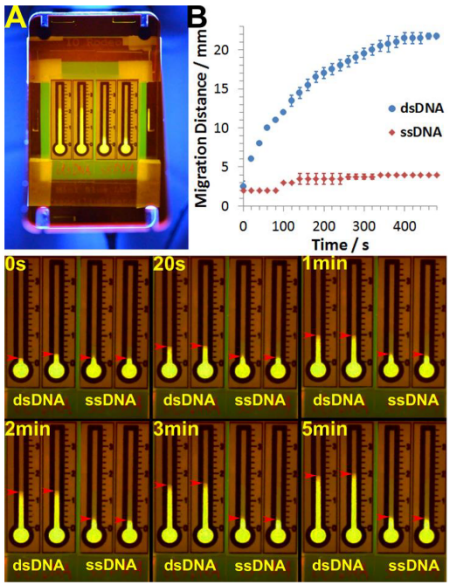


Figure 1. Typical distance-based nucleic acid testing using qPDR. (A) Typical setup of qPDR for visualized measurement of the migration distance of SG-I in the presence of 500 nM dsDNA and 500 nM ssDNA. (B) Kinetic measurements of SG-I migration in the test zone of qPDR in the presence of dsDNA and ssDNA. The migration distance for each data point was visually determined by the naked eye at a frequency of one snapshot per 20 s. The error bar represents one standard deviation from duplicated analyses. The bottom figure shows representative snapshots of qPDR in action at 0 s, 20 s, and 1, 2, 3, and 5 min.

of SG-I. We found that as long as the SG-I concentration is higher than 10 μM, its distribution throughout qPDR can be clearly visualized upon blue light (490 nm) irradiation (Figure S3). Second, the migration distance of SG-I was also found affected by the SG-I concentration. Increasing SG-I concentration from 1.25 μM to 20 μM was also found to effectively enhance the d_R of SG-I from 16.0 mm to 22.0 mm in the presence of 500 nM dsDNA (Figure S3).

As dsDNA serves as an eluting agent to release SG-I from the sample loading zone, the adsorption of dsDNA on cellulose will negatively impact the migration distance of dsDNA and hence SG-I in the test zone. The addition of high molecular weight polyethylene glycol (PEG, MW 100 kDa) was found to effectively address this issue (Figure S4). As shown in Figure S5, increasing amount of PEG from 0 to 1 mg/mL effectively enhances d_R of SG-I in the presence of 500 nM dsDNA by as much as 1.7-fold (Figures S5A,B). The addition of PEG was also found to reduce the d_R of 500 nM ssDNA by a factor of 1.4 (Figure S5C,D). The observed background reduction is likely a result of the increase in the viscosity of the sample solution, which reduces the flow rate and thus allows the additional time for SG-I to be adsorbed by the cellulose at the interface between the sample loading zone and test zone. Collectively,

DOI: 10.1021/acssensors.7b00857
ACS Sens. 2018, 3, 205−210

the overall improvement of signal-to-background ratio was near 2.5 times at the optimal PEG concentration (1 mg/mL). Further increasing the amount of PEG to 10 mg/mL significantly reduced the flow rate due to the dramatic increases in the viscosity and thus resulted in dramatic reduction of d_R for dsDNA.

Having optimized the key assay parameters, we then challenged qPDR with varying concentrations of the 44-bp dsDNA. Figure 2 reveals a concentration dependency between

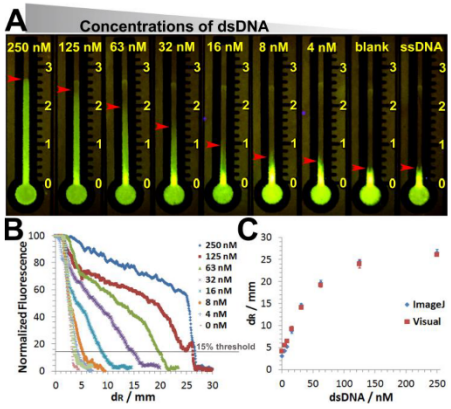


Figure 2. Concentration dependency of the retention distance d_R on the concentration of dsDNA. (A) Representative images of qPDR loaded with 20 μM SG-I, 1 mg/mL PEG, and varying concentrations of dsDNA from 4 nM to 250 nM or 500 nM ssDNA. Red bands indicate retention distances determined using visual examination. (B) Chromatograms of varying concentrations of dsDNA extracted using ImageJ. (C) Retention distance determined either from chromatograms extracted using ImageJ (blue) by direct reading using naked eyes (red) as a function of dsDNA concentrations. Each error bar represents one standard deviation from triplicate analyses.

the observed migration distance d_R and the concentration of dsDNA. In this proof-of-principle assay, the dynamic range is 4−250 nM dsDNA and the detection limit is 4 nM. The assay also revealed remarkably high specificity to dsDNA over ssDNA, which was evidenced by the observation that d_R for 500 nM ssDNA is at the same level of a blank.

Sequence-Specific Nucleic Acid Quantification by Integrating qPDR with PCR. Toward the STH diagnosis, the distance-based nucleic acid quantification using qPDR has to be sequence-specific. One viable solution is to selectively enrich the target nucleic acid with amplification strategies, such as PCR, loop-mediated isothermal amplification (LAMP), or rolling circle amplification (RCA).[28] Here, we chose to use PCR, because this technique has been widely accepted for STH diagnosis.[2,3,29] We also expect that PCR amplicons that are typically much longer than the 44-bp model dsDNA will be favorable targets for qPDR, as longer dsDNA will provide more binding sites for SG-I. The compatibility of qPDR with PCR was then characterized using a 164-bp synthetic dsDNA standard. Figure 3 shows a typical test of the DNA standard with concentrations varying from 1 aM to 1 pM using PCR amplification and then distance-based quantification using

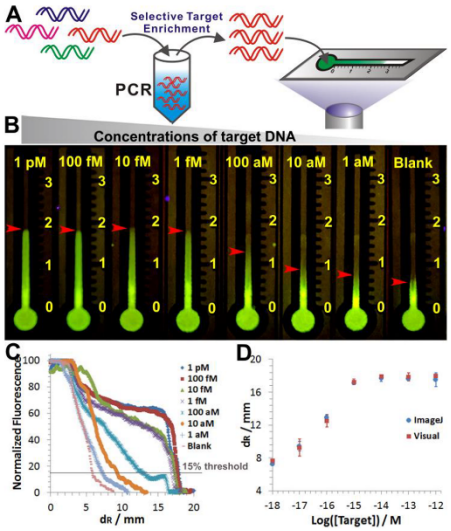
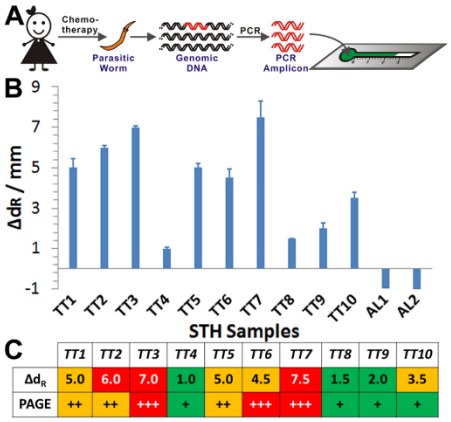


Figure 3. (A) Schematic illustrating the amplification of genetic DNA marker using PCR. DNA standards or genomic DNA samples were amplified using standard PCR protocol for 35 cycles and then mixed with 20 μM SG-I. The reaction mixture was then loaded onto the paper device for quantitative analyses. (B) Representative images of paper-based devices loaded with PCR amplicons of varying concentrations of DNA standards from 1 aM to 1 pM. (C) Chromatograms extracted using ImageJ for the quantitative analysis of PCR amplicons. (D) Migration distance determined either from chromatograms extracted using ImageJ (blue) or by direct reading using naked eyes (red) as a function of concentrations of DNA standards in 2 μL sample. Each error bar represents one standard deviation from triplicate analyses.

qPDR. Excitingly, a quantitative relationship was established when plotting d_R as a function of concentrations of DNA standard (Figure 3D). The dynamic range is from 1 aM to 1 fM with a detection limit at 1 aM.

Quantification of STH Infections Using qPDR. Having confirmed that qPDR is fully compatible with PCR, we then challenged our device with clinical STH samples. All parasitic worm samples, including *Trichuris trichiura* (TT) and *Ascaris lumbricoides* (AL) were collected on-site at the rural areas of La Hicaca, Olanchito, Honduras, where STH prevalence in children is over 50% according to the estimation by the World Health Organization. Genome samples were then isolated from TT and AL worms that were expelled from school age children who had received chemotherapy (Figure 4A). A pair of primers was designed to specifically amplify the 164-bp gene fragment on β-tubulin gene of of TT. This gene fragment containing the codon 200 has been well validated for TT diagnosis and drug resistance tests.[29] To be compatible with our ultimate need for on-site STH diagnosis, the genomic DNA samples were amplified using a low-cost smartphone-controlled portable thermal cycler and qPDR devices post PCR analysis were read by the naked eye (Figure S6). Figure 4B shows the relative migration distances ($\Delta d_R = d_{Rsample} - d_{Rblank}$)

DOI: 10.1021/acssensors.7b00857
ACS Sens. 2018, 3, 205−210



C

	TT1	TT2	TT3	TT4	TT5	TT6	TT7	TT8	TT9	TT10
Δd_R	5.0	6.0	7.0	1.0	5.0	4.5	7.5	1.5	2.0	3.5
PAGE	++	++	+++	+	++	+++	+++	+	+	+

Figure 4. (A) Schematic illustration of the procedure for the quantitative analysis of genomic DNA samples obtained from 10 *Trichuris trichiura* (TT) and 2 *Ascaris lumbricoides* (AL) worms that were collected from infected children in Honduras. (B) Relative migration distances (Δd_R) of genetic markers (β-tubulin gene) determined for each TT or AL sample. Each error bar represents one standard deviation from triplicate analyses. (C) Head-to-head comparison of Δd_R measured using qPDR with band intensities determined by PAGE (Figure S7) for each TT sample. To facilitate the quantitative comparison, levels of β-tubulin gene markers are arbitrarily classified into three groups with each assigned a distinct color (red for high level, orange for intermediate level, and green for low level). Results in qPDR are classified on the basis of Δd_R with Δd_R > 5.0 mm as "high" (red), 2.0 mm < $\Delta d_R \leq$ 5.0 mm as "intermediate" (orange), and $\Delta d_R \leq$ 2.0 mm as "low" (green). Results in PAGE were classified arbitrarily on the basis of the band intensity, with "+" standing for the lowest intensity and "+++" standing for the highest intensity.

for 10 TT worm samples and 2 AL samples (original images in Figure S6). All 10 TT samples were found to be positive with varying levels of β-tubulin marker and the 2 AL samples were both negative. These results were further validated by performing a head-to-head comparison with polyacrylamide gel electrophoresis (PAGE). As shown in Figure 4C, results in both methods are very consistent, with 7 TT samples showing the same concentration levels and the other 3 samples showing minor inconsistencies. Moreover, unlike PAGE where limited quantitative information can be extracted from band intensities (unless imaging software was applied), qPDR can generate digital readings (Δd_R) for all samples by visual examination. Moreover, our device is also much cheaper and faster, and thus more suitable for point-of-care tests at limited resource settings.

CONCLUSION

In conclusion, we have successfully developed a paper-based quantitative device toward the on-site diagnosis and management of STH at resource-limited regions. qPDR is fabricated using unmodified cellulose paper and thus is of low-cost and can be mass produced using high throughput printers. qPDR quantifies nucleic acids by simply reading distance as readout and thus eliminates the need for external electronic readers. qPDR is fully compatible with PCR and thus can be readily integrated into the current STH management. As qPDR is PCR compatible, it is also possible to expand its applications from STH to many other diseases by switching PCR primers. In addition to PCR, qPDR that quantitatively differentiates dsDNA from ssDNA can also be integrated with other isothermal nucleic acid amplification reactions that generate new DNA strands, such as LAMP and RCA. Toward our ultimate goal of on-site STH diagnosis, our next step is to develop a "sample-in, answer-out" diagnostic device through the integration of DNA extraction and isothermal amplification steps with qPDR.

Our device that harnesses the unique binding behaviors of SYBR Green I on unmodified cellulose paper also holds great potential to open up novel sensing mechanisms making use of DNA intercalating dyes and paper chromatography for nucleic acid quantifications. Particularly, the use of migration distance over fluorescence intensity of SG-I for nucleic acid quantification eliminates the need for expensive fluorescence spectrophotometers. The use of unmodified cellulose paper eliminates the need for chemical modification and reagent deposition steps that are commonly found in most distance-based paper devices[17−25] and thus reduces the cost and simplifies the mass productions of devices.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssensors.7b00857.

Synthetic DNA sequences and modifications; Retention distances of varying concentrations of SG-I; Procedures of converting an image of DNA quantification on qPDR into a chromatogram; Effect of SG-I concentrations on the quantification of dsDNA and the background (ssDNA); Effect of PEG on the retention behavior and migration distances; Quantitative analysis of 12 genomic samples from STH worms; Parallel analyses of PCR amplicons of DNA standards and STH worm samples; Determination of quantification threshold on chromatograms (PDF)

Video of qPDR in action taken by a Nikon D600 digital camera (AVI)

AUTHOR INFORMATION

Corresponding Author

*E-mail: fli@brocku.ca.

ORCID

Feng Li: 0000-0002-2616-5343

Author Contributions

#Alex Guan Wang and Tianyu Dong contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Natural Sciences and Engineering Research Council of Canada and the Brock University Start-Up Fund for financial supports.

REFERENCES

(1) Savioli, L.; Albonico, M. Focus: Soil-Transmitted Helminthiasis. *Nat. Rev. Microbiol.* **2004**, *2*, 618−619.

(2) Bethony, J.; Brooker, S.; Albonico, M.; Geiger, S. M.; Loukas, A.; Diemert, D.; Hotez, P. J. Soil-Transmitted Infections: Ascariasis, Trichuriasis, and Hookworm. *Lancet* **2006**, *367*, 1521−1532.
(3) O'Connell, E. M.; Nutman, T. B. Molecular Diagnostics for Soil-Transmitted Helminths. *Am. J. Trop. Med. Hyg.* **2016**, *95*, 508−513.
(4) Magro, L.; Escadafa, C.; Garneret, P.; Jacquelin, B.; Kwasiborski, A.; Manuguerra, J. C.; Monti, F.; Sakuntabhai, A.; Vanhomwegen, J.; Lafaye, P. Tabeling P. Paper Microfluidics for Nucleic Acid Amplification Testing (NAAT) of Infectious Diseases. *Lab Chip* **2017**, *17*, 2347−2371.
(5) Lafleur, L. K.; Bishop, J. D.; Heiniger, E. K.; Gallagher, R. P.; Wheeler, M. D.; Kauffman, P.; Zhang, X.; Kline, E. C.; Buser, J. R.; Kumar, S. A Rapid, Instrument-Free, Sample-to-Result Nucleic Acid Amplification Test. *Lab Chip* **2016**, *16*, 3777−3787.
(6) Tang, R.; Yang, H.; Gong, Y.; You, M.; Liu, Z.; Choi, J. R.; Wen, T.; Qu, Z.; Mei, Q.; Xu, F. A Fully Disposable and Integrated Paper-Based Device for Nucleic Acid Extraction, Amplification and Detection. *Lab Chip* **2017**, *17*, 1270−1279.
(7) Gong, M. M.; Sinton, D. Turning the Page: Advancing Paper-Based Microfluidics for Broad Diagnostic Applications. *Chem. Rev.* **2017**, *117*, 8447−8480.
(8) Rodriguez, N. M.; Wong, W. S.; Liu, L.; Dewar, R.; Klapperich, C. M. A Fully Integrated Paperfluidic Molecular Diagnostic Chip for the Extraction, Amplification, and Detection of Nucleic Acids from Clinical Samples. *Lab Chip* **2016**, *16*, 753−763.
(9) Connelly, J. T.; Rolland, J. P.; Whitesides, G. M. Paper Machine" for Molecular Diagnostics. *Anal. Chem.* **2015**, *87*, 7595−7601.
(10) Li, X.; Luo, L.; Crooks, R. M. Low-Voltage Paper Isotachophoresis Device for DNA Focusing. *Lab Chip* **2015**, *15*, 4090−4098.
(11) Xu, G.; Nolder, D.; Reboud, J.; Oguike, M. C.; van Schalkwyk, D. A.; Sutherland, C. J.; Cooper, J. M. Paper-Origami-Based Multiplexed Malaria Diagnostics from Whole Blood. *Angew. Chem., Int. Ed.* **2016**, *55*, 15250−15253.
(12) Peeling, R. W.; Holmes, K. K.; Mabey, D.; Ronald, A. Rapid Tests for Sexually Transmitted Infections (STIs): the way forward. *Sex. Transm. Infect.* **2006**, *82*, v1−v6.
(13) Lewis, G. G.; DiTucci, M. J.; Phillips, S. T. Quantifying Analytes in Paper-Based Microfluidic Devices Without Using External Electronic Readers. *Angew. Chem., Int. Ed.* **2012**, *51*, 12707−12710.
(14) Lewis, G. G.; Robbins, J. S.; Phillips, S. T. Point-of-Care Assay Platform for Quantifying Active Enzymes to Femtomolar Levels Using Measurements of Time as the Readout. *Anal. Chem.* **2013**, *85*, 10432−10439.
(15) Lewis, G. G.; Robbins, J. S.; Phillips, S. T. A Prototype Point-of-Use Assay for Measuring Heavy Metal Contamination in Water Using Time as a Quantitative Readout. *Chem. Commun.* **2014**, *50*, 5352−5354.
(16) Tian, T.; Li, J.; Song, Y.; Zhou, L.; Zhu, Z.; Yang, C. Y. Distance-Based Microfluidic Quantitative Detection Methods for Point-of-Care Testing. *Lab Chip* **2016**, *16*, 1139−1151.
(17) Cate, D. M.; Dungchai, W.; Cunningham, J. C.; Volckens, J.; Henry, C. S. Simple, Distance-Based Measurement for Paper Analytical Devices. *Lab Chip* **2013**, *13*, 2397−2404.
(18) Cate, D. M.; Noblitt, S. D.; Volckens, J.; Henry, C. S. Multiplexed paper analytical device for quantification of metals using distance-based detection. *Lab Chip* **2015**, *15*, 2808−2818.
(19) Wei, X.; Tian, T.; Jia, S.; Zhu, Z.; Ma, Y.; Sun, J.; Lin, Z.; Yang, C. J. Target-Responsive DNA hydrogel Mediated "Stop-Flow" Microfluidic Paper-Based Analytic Device for Rapid, Portable and Visual Detection of Multiple Targets. *Anal. Chem.* **2015**, *87*, 4275−4282.
(20) Tian, T.; Wei, X.; Jia, S.; Zhang, R.; Li, J.; Zhu, Z.; Zhang, H.; Ma, Y.; Lin, Z.; Yang, C. J. Integration of Target Responsive Hydrogel with Cascaded Enzymatic Reactions and Microfluidic Paper-Based Analytic Devices (μPADs) for Point-of-Care Testing (POCT). *Biosens. Bioelectron.* **2016**, *77*, 537−542.
(21) Tian, T.; An, Y.; Wu, Y.; Song, Y.; Zhu, Z.; Yang, C. Integrated Distance-Based Origami Paper Analytical Device for One-Step Visualized Analysis. *ACS Appl. Mater. Interfaces* **2017**, *9*, 30480−30487.
(22) Yamada, K.; Henares, T. G.; Suzuki, K.; Citterio, D. Distance-Based Tear Lactoferrin Assay on Microfluidic Paper Device Using Interfacial Interactions on Surface-Modified Cellulose. *ACS Appl. Mater. Interfaces* **2015**, *7*, 24864−24875.
(23) Zhang, Y.; Zhou, C.; Nie, J.; Le, S.; Qin, Q.; Liu, F.; Li, Y.; Li, J. Equipment-Free Quantitative Measurement for Microfluidic Paper-Based Analytical Devices Fabricated Using the Principles of Movable-Type Printing. *Anal. Chem.* **2014**, *86*, 2005−2012.
(24) Hongwarittorrn, I.; Chaichanawongsaroj, N.; Laiwattanapaisal, W. Semi-Quantitative Visual Detection of Loop Mediated Isothermal Amplification (LAMP)-Generated DNA by Distance-Based Measurement on a Paper Device. *Talanta* **2017**, *175*, 135−142.
(25) Kalish, B.; Luong, J.; Roper, J.; Beaudette, C.; Tsutsui, H. Distance-Based Quantitative DNA Detection in a Paper-Based Microfluidic Device. Proceedings of the 12th IEEE International Conference on Nano/Micro Engineered and Molecular Systems, 2017; Los Angeles, USA.
(26) Lu, Y.; Shi, W.; Jiang, L.; Qin, J.; Lin, B. Rapid Prototyping of Paper-Based Microfluidics with Wax for Low-Cost, Portable Bioassays. *Electrophoresis* **2009**, *30*, 1497−1500.
(27) Carrilho, E.; Martinez, A. W.; Whitesides, G. M. Understanding Wax Printing: a Simple Micropatterning Process for Paper-Based Microfluidics. *Anal. Chem.* **2009**, *81*, 7091−7095.
(28) Zhou, Y.; Huang, Q.; Gao, J.; Lu, J.; Shen, X.; Fan, C. A Dumbbell Probe-Mediated Rolling Circle Amplification Strategy for Highly Sensitive MicroRNA Detection. *Nucleic Acids Res.* **2010**, *38*, e156.
(29) Diawara, A.; Drake, L. J.; Suswillo, R. R.; Kihara, J.; Bundy, D. A.; Scott, M. E.; Halpenny, C.; Stothard, J. R.; Prichard, R. K. Assays to Detect Beta-Tubulin Codon 200 Polymorphism in Trichuris Trichuiura and Ascaris Lumbricoides. *PLoS Neglected Trop. Dis.* **2009**, *3*, e397.

DOI: 10.1021/acssensors.7b00857
ACS Sens. 2018, 3, 205−210

Supporting Information

Paper-based DNA reader for visualized quantification of soil-transmitted helminth infections

Alex Guan Wang,[†,#] Tianyu Dong,[†,#] Hayam Mansour,[†,§] Gabriela Matamoros,[‖] Ana L. Sanchez[‡,‖] and Feng Li[†]*

[†]Department of Chemistry, Centre for Biotechnology, [‡]Department of Health Sciences, Brock University, St. Catharines, Canada, L2S 3A1

[§]Department of Cell Biology, National Research Center, Cairo, Egypt, 12622

[‖]Microbiology Research Institute, National Autonomous University of Honduras, Tegucigalpa, Honduras,

[#] A. G. Wang and T. Dong contributed equally to this work

*Corresponding authors.
fli@brocku.ca (F.L.)

S1. Supporting Figures and Tables

Table S1. Synthetic DNA sequences and modifications.

Name		Sequence
44-bp model DNA	Sense	5'-AA ATT CGC AGT CCC CAA CCT CC A ATC ACT CAC CAA CCT CCT GTC-3'
	Rhodamine Green (RG) – labeled Sense	5'-RG-AA ATT CGC AGT CCC CAA CCT CC A ATC ACT CAC CAA CCT CCT GTC-3'
	Antisense	5'-GAC AGG AGG TTG GTG AGT GAT TGG AGG TTG GGG ACT GCG AAT TT-3'
TT-DNA	F-Primer	5'-AGG TTT CAG ATA CAG TTG TAG-3'
	R-Primer	5'-CAA ATG ATT TAA GTC TCC G-3'
	DNA Standard	5'- AGG TTT CAG ATA CAG TTG TAG AAC CAT ATA ATG CAA CTC TGT CAG TCC ACC AGT TGG TAG AGA ACA CGG ACG AAA CAT TCT GCA TAG ATA ATG AAG CGC TTT ACG ATA TTT GTT TCC GAA CTT TGA AGT TAA CAA CAC CAA CTT ACG GAG ACT TAA ATC ATT TG-3'

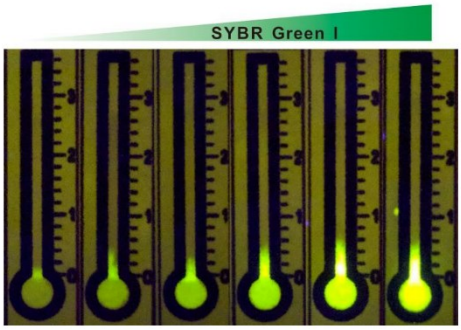


Figure S1. Retention distances of varying concentrations of SG-I from 1.25 μM to 40 μM on qPDR.

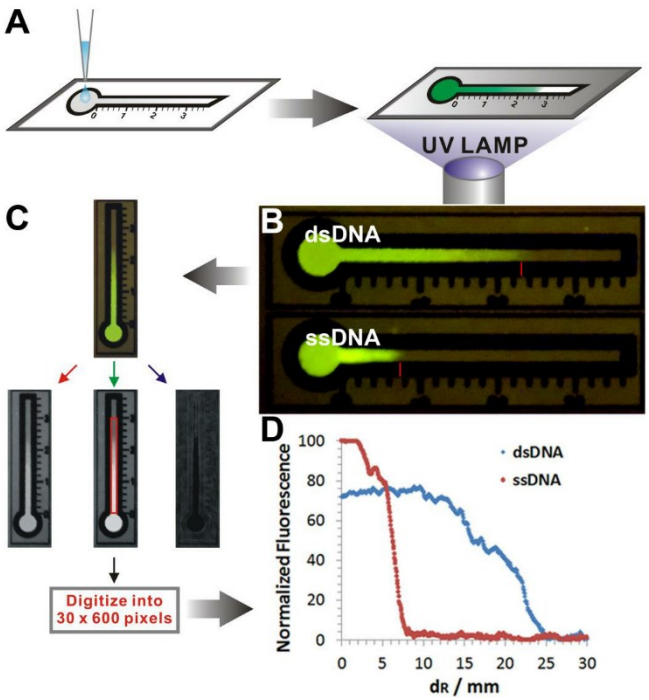


Figure S2. Procedures of converting an image of DNA quantification on qPDR into a chromatogram. (A) Schematic illustration of the procedure for sample loading and data collection using qPDR. (B) Typical fluorescent images of dsDNA/ssDNA-SG-I binding complexes captured using a smartphone camera. (C) Procedure for splitting colour channels and digitizing the image using ImageJ. The red bar indicates the 15% threshold on the chromatogram. (D) Typical chromatograms of DNA-SG-I complexes extracted from the fluorescence images.

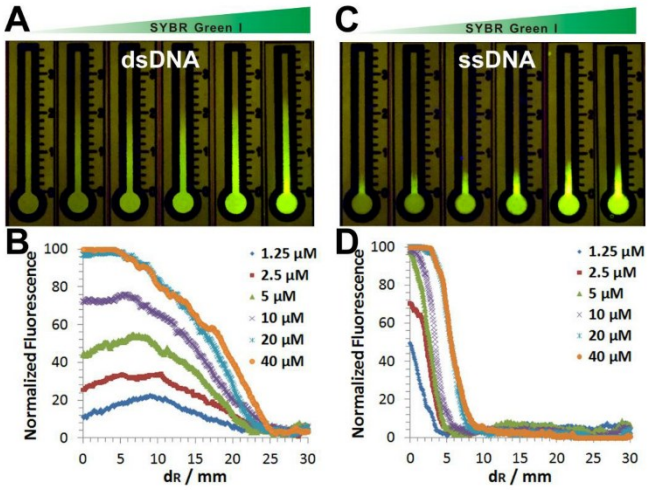


Figure S3. Effect of SG-I concentrations on the quantification of dsDNA and the background (ssDNA). Varying concentrations of SG-I from 1.25 µM to 40 µM were mixed with 500 nM dsDNA (A, B) or 500 nM ssDNA (C, D) and then loaded onto qPDR. Chromatograms (B, D) were extracted using ImageJ with the protocol outlined in Figure S2.

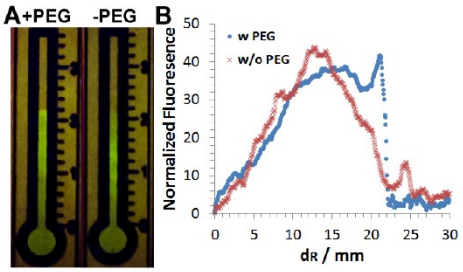


Figure S4. Effect of PEG 100,000 (1 mg/mL) on the retention behavior of 500 nM dsDNA on qPDR. To visualize the distribution of dsDNA on cellulose paper, a fluorescent dye Rhodamine Green (RG) was modified on one of the strand of the dsDNA.

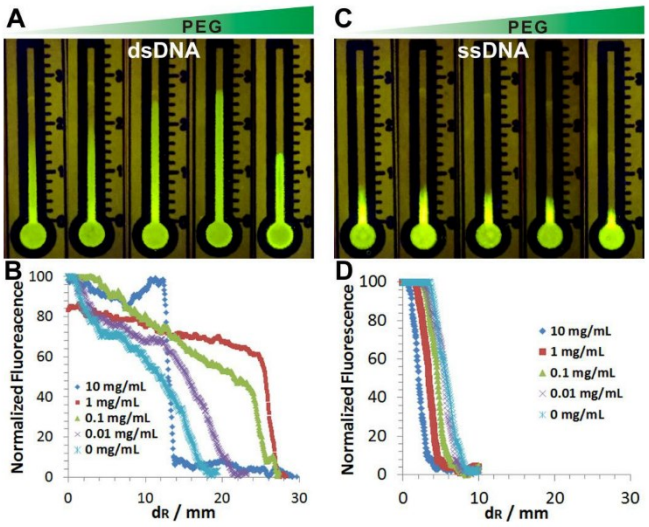


Figure S5. Effect of PEG 100,000 on the migration distances of SG-I in the presence of 500 nM dsDNA and 500 nM ssDNA. Varying concentrations of PEG from 0 to 10 mg/mL were mixed with 20 μM SG-I and 500 nM dsDNA (A, B) or 500 nM ssDNA (C, D), and then loaded into the paper device. Chromatograms (B, D) were extracted using ImageJ with the protocol outlined in Figure S2.

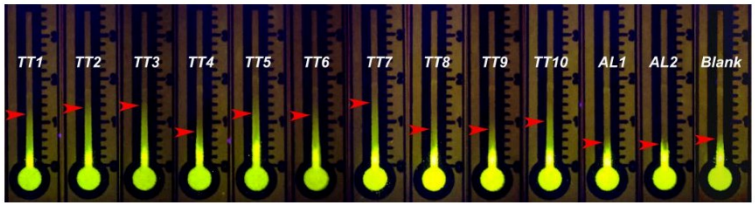


Figure S6. Quantitative analysis of 12 genomic samples from STH worms, include 10 TT worms and 2 AL worms (negative controls), using qPDR.

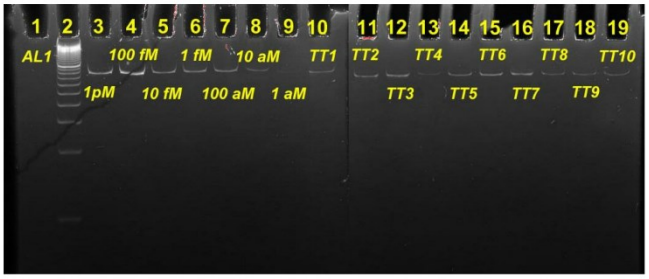
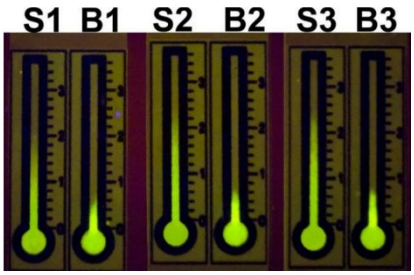


Figure S7. Parallel analyses of PCR amplicons of DNA standards and STH worm samples using gel electrophoresis and qPDR (Figure 4 and S6). **Lane 1:** PCR amplicons from AL1; **Lane 2:** 20-bp DNA ladders; **Lane 3-9:** PCR amplicons from varying concentrations of DNA standards; **Lane 10-19:** PCR amplicons from 10 TT samples.

S2. Determination of quantification threshold on chromatograms

Retention distances of 3 dsDNA samples (S1-3) and 3 ssDNA samples (B1-3) on qPDR read by 59 student volunteers.



Observer	dsDNA-1 d_R (mm)	dsDNA-2 d_R (mm)	dsDNA-3 d_R (mm)	ssDNA-1 d_R (mm)	ssDNA-2 d_R (mm)	ssDNA-3 d_R (mm)
1	16.2	22.0	22.8	7.5	6.4	7.4
2	19.5	22.0	22.5	5.5	5.7	6.4
3	20.0	22.0	24.0	6.0	6.0	8.0
4	20.0	24.0	24.0	6.0	6.0	6.0
5	18.0	22.1	23.6	5.5	6.2	7.8
6	20.0	22.0	23.0	6.0	6.0	8.0
7	20.0	22.0	24.0	7.0	6.0	8.0
8	20.0	22.0	23.5	6.0	6.5	7.0
9	20.0	24.0	26.0	6.0	6.0	8.0
10	20.0	22.0	24.0	6.0	6.0	8.0
11	16.0	20.0	22.0	5.0	6.0	7.0
12	20.0	23.0	24.0	7.0	6.0	7.0
13	21.2	21.9	22.5	5.3	5.9	6.4
14	20.0	21.0	23.0	5.0	5.0	7.0
15	18.0	20.0	22.0	6.0	6.0	6.0
16	12.0	17.6	22.3	4.0	4.2	6.4
17	20.0	22.1	24.5	7.0	6.0	8.0
18	20.0	22.0	23.0	5.0	6.0	7.0
19	18.9	21.7	22.5	6.5	5.9	7.4
20	20.1	20.4	22.1	5.7	6.1	6.3
21	20.0	22.0	22.0	6.0	6.0	8.0
22	20.0	22.0	23.0	6.0	6.0	8.0
23	19.0	22.0	24.5	7.0	6.0	8.0
24	19.0	22.0	22.0	5.0	5.0	6.5
25	20.0	22.0	21.0	7.0	6.0	7.0

26	19.0	21.0	23.0	6.0	6.0	7.0
27	20.0	22.0	23.0	6.0	6.0	8.0
28	20.0	22.0	22.0	5.0	6.0	7.0
29	20.0	22.0	22.0	6.0	6.0	8.0
30	18.0	22.0	22.0	6.0	6.0	6.0
31	18.0	20.0	22.0	6.0	6.0	8.0
32	20.0	22.0	23.0	6.0	6.0	7.0
33	18.9	21.9	22.5	4.1	4.9	6.2
34	13.0	16.1	18.3	4.5	5.2	6.4
35	20.0	20.1	21.0	5.8	6.0	7.6
36	20.0	22.0	24.0	6.0	6.0	8.0
37	20.0	22.0	24.0	6.0	6.0	8.0
38	18.0	20.0	22.0	4.0	6.0	8.0
39	20.0	24.0	26.0	6.0	6.0	8.0
40	18.5	22.5	23.5	5.9	6.1	7.0
41	20.0	24.0	26.0	7.0	8.0	8.0
42	20.0	23.0	23.0	5.0	6.0	7.0
43	20.0	22.0	24.0	6.0	6.0	7.0
44	20.0	23.0	23.5	6.0	6.0	7.5
45	16.2	22.0	22.8	7.5	6.4	7.4
46	20.1	24.5	23.9	6.2	5.7	8.3
47	18.5	22.0	22.5	5.9	6.3	6.9
48	19.0	22.0	22.0	5.0	5.0	6.5
49	18.9	22.0	22.5	5.0	5.8	6.5
50	18.8	23.0	22.0	6.0	5.9	7.0
51	16.0	20.0	22.0	6.0	6.0	7.0
52	18.3	22.0	23.9	6.1	6.3	8.1
53	16.0	21.5	22.0	4.5	5.0	6.0
54	19.8	21.2	21.5	6.7	7.8	7.5
55	20.0	22.0	22.0	5.0	6.0	8.0
56	18.1	22.3	23.5	5.6	6.1	7.9
57	20.0	22.0	22.0	6.0	6.0	8.0
58	20.0	23.0	24.0	6.0	6.0	7.0
59	20.0	22.0	23.0	6.0	6.0	8.0
Average d_R	**19.0**	**21.8**	**22.9**	**6.0**	**6.1**	**7.5**
STD	**1.7**	**1.4**	**1.3**	**0.8**	**0.5**	**0.7**
RSD	**9.1%**	**6.2%**	**5.5%**	**13.5%**	**9.1%**	**9.5%**
Normalized Fluorescence at d_R	**13.9**	**17.3**	**14.3**	**12.8**	**17.3**	**9.7**
Average	**15.18**			**13.30**		