IMPROVING THE SELECTIVITY AND REDUCING THE LEAKAGE OF DNA STRAND DISPLACEMENT SYSTEMS

by

Shohei Kotani

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Materials Science and Engineering

Boise State University

May 2018

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Shohei Kotani

Dissertation Title: Improving the Selectivity and Reducing the Leakage of DNA Strand Displacement Systems

Date of Final Oral Examination: 28 February 2018

The following individuals read and discussed the dissertation submitted by student Shohei Kotani, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

William L. Hughes, Ph.D.	Chair, Supervisory Committee
Bernard Yurke, Ph.D.	Member, Supervisory Committee
Jeunghoon Lee, Ph.D.	Member, Supervisory Committee
Peter B. Allen, Ph.D.	Member, Supervisory Committee

The final reading approval of the dissertation was granted by William L.Hughes, Ph.D., Chair of the Supervisory Committee. The dissertation was approved for the Graduate College.

ACKNOWLEDGEMENTS

First, I would like to acknowledge my research advisor, William L. Hughes. He gave me the opportunity to do research and guided me to go through graduate school. I also would like thank my committee members, Bernard Yurke, Jeunghoon Lee, and Peter Allen for their supports and feedback to my research. In addition, I would like to acknowledge Bill Knowlton, Chad Watson, Elton Graugnard, Jennifer E. Padilla, Natalya Hallstrom, Paul Davis, Reza Zadegan, and Wang Kuang for their discussion about my research and support to make Boise State University a nice place to work.

I feel so fortunate to have been able to work with Sadao Takabayashi and Ping Olson during my time at Boise State University. Sadao helped me literally in all aspects of my life outside school. I would not have been able to do any research without working with Ping Olson. Also, our research group was filled with very nice neighbors, Ann Delaney, Blake Rapp, Brittany Cannon, Brett Ward, Christopher Green, Daniel Kelly, Donald Kellies, Drew Lysen, Elijah Spears, Juan Flores-Estrada, Jon Huff, Lance Patten, Michael Tobiason, Sara Goltry, and William Klein.

Many thanks to Jessica Economy, Dena Ross, and Krystal Holaday at the office of Micron School of Materials Science and Engineering for their supports in all aspects in school, and Ruth Prince at Center of Global Education for her support in dealing with the many issues I encountered as an international student.

Finally, I would like to thank all the supports from my parents.

ABSTRACT

Because of the elegance of Watson-Crick base pairing and the programmability of toehold-mediated strand displacement, DNA is a model material for designing, building, and testing molecular assemblies. DNA assemblies are categorized as structural when they are at thermodynamic equilibrium and dynamic when they are not. Through programmed perturbations, metastable assemblies perform physical, chemical, and computational work. When integrated into a diagnostic package, disease-specific nucleic acid sequences can be identified, amplified, and analyzed via standard DNA nanotechnology rules. In order for these rules to make an impact, two critical challenges in the field have been undertaken in this dissertation. First, the selectivity to distinguish an on-target sequence from off-target sequences, with a resolution of a single-nucleotide mutation, has been explored by site-specifically integrating locked nucleic acids into DNA sequences. Locked nucleic acids are RNA analogues that have higher thermal and hence mechanical stability than RNA and DNA. Second, the initiation of metastable chemical reaction networks, in the absence of on-target sequences, has been explored to suppress network leakage; which is the single greatest problem in dynamic DNA nanotechnology. To address this challenge, original catalytic substrates were designed, built, and tested to increase the energy barrier of the leakage reactions without sacrificing the performance of the favorable catalytic reactions. The experimental results showed that site-specific integration of LNA into DNA sequences improved the sequence selectivity by over 2 orders of magnitude. They also showed that network leakage could

be suppressed by 2 orders of magnitude by decoupling the leakage pathway from the catalytic pathway in the original catalytic substrates. When combined, these results constitute a substantial contribution to the field of dynamic DNA nanotechnology and represent important steps towards the creation of low-cost, early-stage diagnostic tools for difficult to detect diseases such as lung, breast, and pancreatic cancers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
LNA	Lock Nucleic Acid
RNA	Ribonucleic Acid
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
PCR	Polymerase Chain Reaction
HCR	Hybridization Chain Reaction
CHA	Catalytic Hairpin Assembly
NUPACK	Nucleic Acid Package
EGNAS	Exhaustive DNA Sequence Design Algorithm
HPLC	High Performance Liquid Chromatography
Tris	Tris Hydroxymethyl Aminomethane
EDTA	Ethylenediaminetetraacetic Acid
TE buffer	Tris-EDTA buffer
TAE buffer	Tris-acetate-EDTA buffer
TBE buffer	Tris-Borate-EDTA buffer
HCl	Hydrogen Chloride
PAGE	Polyacrylamide Gel Electrophoresis

$MgCl_2$	Magnesium Chloride
$Mg(C_2H_3O_2)_2$	Magnesium Acetate
Mg^{2+}	Magnesium Ion
IDT	Integrated DNA Technologies
k	Reaction Rate Constant
k_{leak}	Reaction Rate Constant of Leakage Reaction
k_{cat}	Reaction Rate Constant of Catalytic Reaction
kcal/mol	Kilocalorie per Mole
IEL	Intuitive Energy Landscape
R	Universal Gas Constant
T	Temperature
mM	Millimolar
nM	Nanomolar
pM	Picomolar
aM	Attomolar
mL	Milliliter
μL	Microliter
a.u.	Arbitrary Unit
ΔGbp	Free Energy Cost of Breaking a Base Pair
$\Delta Ginit$	Free Energy Cost of an Initial Binding
$\Delta Gs3$	Sawtooth Amplitude of Three-Way Branch Migration
$\Delta Gs4$	Sawtooth Amplitude of Four-Way Branch Migration

$\Delta G2ov$	Free Energy Cost of Introducing Two Single-Stranded DNA Overhangs at an Overhang-Free Nick
$\Delta G3aj$	Free Energy Cost of Introducing a Three-Arm Junction
$\Delta G4aj$	Free Energy Cost of Introducing a Four-Arm Junction

CHAPTER ONE: INTRODUCTION

1.1 DNA as a Biological Material

Deoxyribonucleic acid (DNA) is a biological polymer with four monomers that have four unique bases. The bases are adenine (A), thymine (T), guanine (G), and cytosine (C).[1] During Watson-Crick base pairing, adenine binds with thymine (A-T) and guanine binds with cytosine (G-C). Because of these simple rules, single-stranded DNA hybridizes with its complementary strand, in an anti-parallel orientation, to form the well-known double-stranded helix (Fig 1.1A). Hydrogen bonding between the bases and base-stacking among the bases stabilizes the double helix.[1] Three important characteristics of hydrogen bonds are that they are specific, weak, and reversible near standard physiological conditions. As a result, DNA is able to store and transfer vital information.

If DNA is the archival storage material of life,[1] RNA transfers DNA information into functional proteins that perform chemical, mechanical and biological work for life.[1] In addition, non-coding RNAs, whose information is not transferred into proteins, have biological significance because they up- and down-regulate gene expression.[2, 3] Because of the significance of nucleic acids in the health and life sciences, a variety of powerful molecular techniques have been invented. Among the most significant is the polymerase chain reaction (PCR), which uses proteins and thermal cycling to exponentially amplify a target sequence among a background of random sequences.[4]

1.2 DNA as an Engineering Material

Engineering is the application of science during the design, building, testing, and operation of machines, structures, and systems. As an engineering material, DNA has well-defined chemical, structural,[5] thermodynamic,[6] and kinetic[7] properties. By exploiting these properties, it can be *engineered* into machines, structures, and systems for a diversity of biological and non-biological applications. In response to the human-genome project, it can also be chemically manufactured cost-effectively in support of the rapid growth of DNA nanotechnology. Given below is background information about structural and dynamic DNA nanotechnology.

1.2.1 Structural DNA Nanotechnology

DNA nanotechnology was pioneered in the early 1980's by Nadrian Seeman at NYU. One of his many contributions to the field was the creation of structural lattices made entirely from DNA.[8] Rather than working with double-stranded DNA, he created multi-arm junctions made from DNA.[9] Four-arm junctions were originally observed in genetic recombination.[1] During recombination, the branch point of the four-arm junction is mobile and hence moves around due to the sequence symmetry of the opposing arms. This motion of the junction is called branch migration (Figure 1.1B).[10] The junction becomes immobile if the sequence is carefully designed so that branch migration does not occur. Once immobilized, the four-arm junction can then be connected to other four-arm junctions to create a periodic lattice (Figure 1.1C). Both two-dimensional[11] and three-dimensional[12] lattices have since been engineered. What began as building structural lattices with a limited number of DNA oligomers has grown into the programmable self-assembly of hundreds of oligomers employed in DNA origami[13, 14] and in DNA bricks.[15,

[16] The result has been a revolution towards the creation of nanobreadboards that organize organic and inorganic materials with nanometer precision.[17-19]

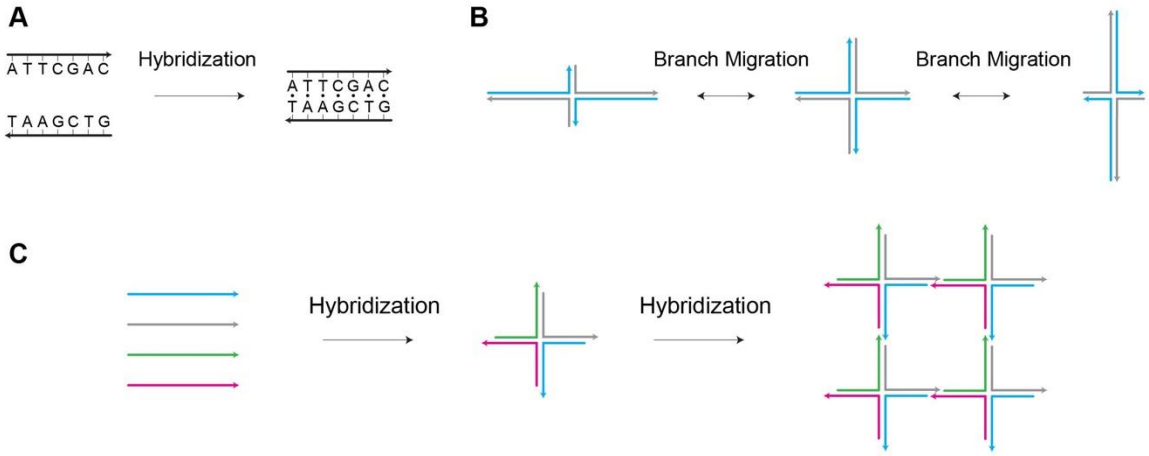

Figure 1.1. DNA as an engineering material. (A) Hybridization of complementary single-stranded DNAs with anti-parallel orientation to form a double-stranded DNA. The arrowheads represent 3' ends. (B) Four-arm junction structure with sequence symmetry undergoes branch migration. This process is called as four-way branch migration. DNA strands are shown as colored lines. (C) Four-arm junction structure without sequence symmetry has an immobile junction. The four-arm junction structure in the middle of the figure has the single-stranded DNA regions at the end of each arm, which are used to form the lattice. All the processes can be executed by simply mixing the four types of single-stranded DNA with unique sequences at an optimum temperature and salt concentration. Therefore, the process is self-assembly

1.2.2 Dynamical DNA Nanotechnology

If the products of structural DNA nanotechnology require thermodynamic equilibrium, dynamic DNA nanotechnology requires assemblies that are metastable until intentionally perturbed by external stimuli. In the case of this dissertation, the stimulus is a target DNA sequence. Common driving forces for DNA reactions to proceed once the target sequence binds to a metastable DNA assembly are an increase in the number of base pairs with enthalpy gain or an increase in the number of components with entropy gain.[20] Regardless of the path, a short segment of single-stranded DNA, called a toehold, at the end of a double-stranded region, accelerates strand displacement by as much as 6

orders of magnitude.[7, 21] The use of a toehold to create dynamic DNA structures was introduced by Bernard Yurke and his colleagues in 2000.[22] Similar to the four-way junction in structural DNA nanotechnology, nature-inspired toeholds catalyzed the field of dynamic DNA nanotechnology. When combined with sequence-specific binding of DNA, a diversity of chemical reaction networks can be engineered (Figure 1.2B).[23] Elegant examples include signal cascade systems,[24] catalytic systems for signal amplification,[25-27] and molecular walkers.[28, 29] In further fundamental work, these systems have been developed to perform molecular computation,[30] chemical oscillation,[31] and sorting of physical cargo at the nanoscale.[32]

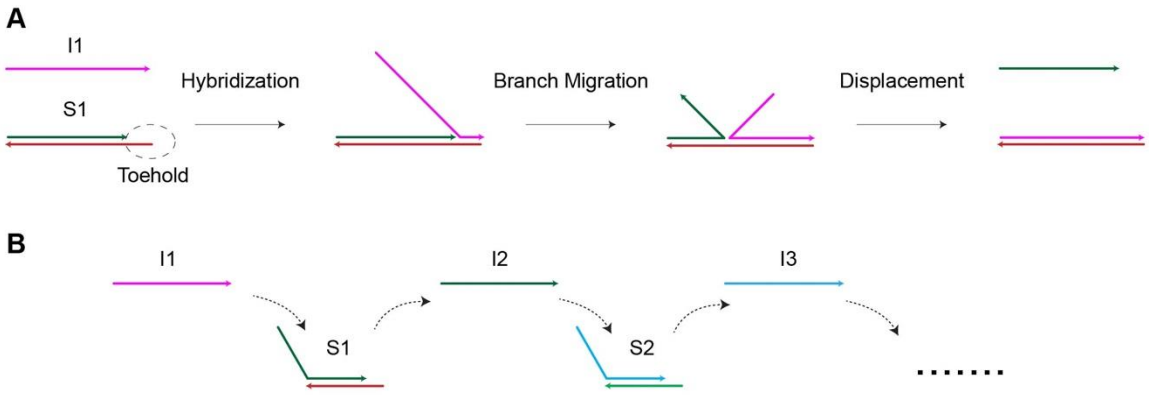

Figure 1.2. Dynamic DNA nanotechnology. (A) Toehold-mediated strand displacement. The invader strand I1 first hybridizes with the toehold, which is the exposed single-stranded region of the double-stranded DNA complex S1. Next, the invader strand and the incumbent strand of the complex (green color) undergoes branch migration. This process is called as three-way branch migration. Finally, the invader strand displaces the incumbent strand. (B) An example of chemical reaction network operated by toehold-mediated strand displacement. First, the invader strand I1 displaces the green color strand from the double-stranded complex S1. Then, the released strand I2 in turn acts as the invader strand to react with the double-stranded DNA complex S2 and releases the invader strand I3 for downstream reactions. This cascaded reaction can be scaled further as long as unique base sequences is available so that the strand displacement reactions do not interfere with each other

1.3 Dissertation Layout

One important application of dynamic DNA nanotechnology is medical diagnostics.[33] In theory, catalytic systems, operated via toehold-mediated strand displacement, are very similar to PCR. For example, they each detect and amplify a small concentration of on-target sequences among a background of off-target sequences. However, unlike PCR, catalytic systems do not require proteins or thermal cycling to perform amplification. In addition, they don't require a computer to perform analysis. Instead, detection, amplification, and analysis can be performed by chemical reaction networks made entirely from DNA. This not only reduces the cost of the tool, it reduces the complexity of the process. However, there are several challenges that prevent the practical application of dynamic DNA nanotechnology.[34, 35] Among them, the first challenge is selectivity for on- versus off-target sequences.[36] To be competitive as a diagnostic tool, single-nucleotide mutations must be resolved because they can have a profound influence on a patient's health.[1] The second challenge is leakage.[37] In catalytic systems, leakage is the initiation of the reaction in the absence of a catalyst. In order to create a catalytic system with elevated signal amplification, it is routine to connect multiple catalytic systems together to create exponential amplification.[27] The practical result is that a minute level of leakage in one catalytic system becomes amplified when connected to other catalytic systems; creating a false positive. Leakage not only prevents the creation of accurate signal amplification systems for medical diagnosis, it also prevents the creation of larger and much more sophisticated chemical reaction networks for molecular computation.[38] This dissertation addresses the challenge of sensitivity by exploring the use of locked nucleic acid (LNA),[39] which is a synthetic RNA analogue

with higher thermo-mechanical stability, and the challenge of leakage by exploring the use of multi-arm junctions as a novel substrate that decouples the leakage reaction pathway from the catalytic reaction pathway. In the spirit of this publication-based dissertation, **Chapter Two** is a manuscript that will be submitted to Nucleic Acid Research and is entitled, "*Improved Selectivity of a Toehold-Mediated Strand Displacement System Using Locked Nucleic Acids*". Building on this work, **Chapter Three** was published in the Journal of the American Chemical Society (JACS) and is entitled, "*Multi-Arm Junctions for Dynamic DNA Nanotechnology*". **Chapter Four** is a conclusion that includes future work.

1.4 References

1. Watson, J. D., Molecular biology of the gene. *Molecular biology of the gene.* **1970,** (2nd edn).
2. Carthew, R. W.; Sontheimer, E. J., Origins and mechanisms of miRNAs and siRNAs. *Cell* **2009,** *136* (4), 642-655.
3. Waters, L. S.; Storz, G., Regulatory RNAs in bacteria. *Cell* **2009,** *136* (4), 615-628.
4. Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **1988,** *239* (4839), 487-491.
5. Seeman, N. C., Nanotechnology and the double helix. *Scientific American* **2004,** *290* (6), 64-75.
6. SantaLucia, J.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-440.
7. Zhang, D. Y.; Winfree, E., Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **2009,** *131* (47), 17303-17314.

8. Seeman, N. C., Nucleic acid junctions and lattices. *Journal of theoretical biology* **1982,** *99* (2), 237-247.

9. Seeman, N. C.; Kallenbach, N. R., DNA branched junctions. *Annu. Rev. Biophys. Biomol. Struct.* **1994,** *23* (1), 53-86.

10. Panyutin, I. G.; Hsieh, P., The kinetics of spontaneous DNA branch migration. *Proc. Natl Acad. Sci. USA* **1994,** *91* (6), 2021-2025.

11. Winfree, E.; Liu, F.; Wenzler, L. A.; Seeman, N. C., Design and self-assembly of two-dimensional DNA crystals. *Nature* **1998,** *394* (6693), 539.

12. Zheng, J.; Birktoft, J. J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P. E.; Ginell, S. L.; Mao, C.; Seeman, N. C., From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **2009,** *461* (7260), 74.

13. Rothemund, P. W. K., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006,** *440* (7082), 297.

14. 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.

15. Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P., Three-dimensional structures self-assembled from DNA bricks. *science* **2012,** *338* (6111), 1177-1183.

16. Wei, B.; Dai, M.; Yin, P., Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **2012,** *485* (7400), 623.

17. Bui, H.; Onodera, C.; Kidwell, C.; Tan, Y.; Graugnard, E.; Kuang, W.; Lee, J.; Knowlton, W. B.; Yurke, B.; Hughes, W. L., Programmable periodicity of quantum dot arrays with DNA origami nanotubes. *Nano letters* **2010,** *10* (9), 3367-3372.

18. Cannon, B. L.; Kellis, D. L.; Davis, P. H.; Lee, J.; Kuang, W.; Hughes, W. L.; Graugnard, E.; Yurke, B.; Knowlton, W. B., Excitonic and logic gates on DNA brick nanobreadboards. *ACS photonics* **2015,** *2* (3), 398-404.

19. Takabayashi, S.; Klein, W. P.; Onodera, C.; Rapp, B.; Flores-Estrada, J.; Lindau, E.; Snowball, L.; Sam, J. T.; Padilla, J. E.; Lee, J., High precision and high yield fabrication of dense nanoparticle arrays onto DNA origami at statistically independent binding sites. *Nanoscale* **2014,** *6* (22), 13928-13938.

20. Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P. K.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **2013,** *41* (22), 10641-10658.

21. Yurke, B.; Mills, A., Jr., Using DNA to power nanostructures. *Genet. Program. Evolvable Mach.* **2003,** *4* (2), 111-122.

22. 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.

23. Soloveichik, D.; Seelig, G.; Winfree, E., DNA as a universal substrate for chemical kinetics. *Proc. Natl Acad. Sci. USA* **2010,** *107* (12), 5393-5398.

24. Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E., Enzyme-free nucleic acid logic circuits. *Science* **2006,** *314* (5805), 1585-1588.

25. Dirks, R. M.; Pierce, N. A., Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **2004,** *101* (43), 15275-15278.

26. Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A., Programming biomolecular self-assembly pathways. *Nature* **2008,** *451* (7176), 318-322.

27. 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.

28. Shin, J.-S.; Pierce, N. A., A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.* **2004,** *126* (35), 10834-10835.

29. Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H., A unidirectional DNA walker that moves autonomously along a track. *Angew. Chem. Int. Ed.* **2004,** *43* (37), 4906-4911.

30. Qian, L.; Winfree, E., Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **2011,** *332* (6034), 1196-1201.

31. Srinivas, N.; Parkin, J.; Seelig, G.; Winfree, E.; Soloveichik, D., Enzyme-free nucleic acid dynamical systems. *Science* **2017,** *358* (6369), eaal2052.

32. Thubagere, A. J.; Li, W.; Johnson, R. F.; Chen, Z.; Doroudi, S.; Lee, Y. L.; Izatt, G.; Wittman, S.; Srinivas, N.; Woods, D., A cargo-sorting DNA robot. *Science* **2017,** *357* (6356), eaan6558.

33. Jung, C.; Ellington, A. D., Diagnostic applications of nucleic acid circuits. *Acc. Chem. Res.* **2014,** *47* (6), 1825-1835.

34. Zhang, D. Y.; Seelig, G., Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **2011,** *3* (2), 103-113.

35. Chen, Y.-J.; Groves, B.; Muscat, R. A.; Seelig, G., DNA nanotechnology from the test tube to the cell. *Nature nanotechnology* **2015,** *10* (9), 748.

36. Chen, S. X.; Seelig, G., An engineered kinetic amplification mechanism for single nucleotide variant discrimination by DNA hybridization probes. *Journal of the American Chemical Society* **2016,** *138* (15), 5076-5086.

37. Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D., Stacking nonenzymatic circuits for high signal gain. *Proc. Natl Acad. Sci. USA* **2013,** *110* (14), 5386-5391.

38. Thachuk, C.; Winfree, E.; Soloveichik, D., Leakless DNA strand displacement systems. *Proceedings of the 21st International Conference on DNA Computing and Molecular Programming 9211* **2015**, 133−153.

39. K. Singh, S.; A. Koshkin, A.; Wengel, J.; Nielsen, P., LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chemical Communications* **1998,** (4), 455-456.

CHAPTER TWO: IMPROVED SELECTIVITY OF A TOEHOLD-MEDIATED STRAND DISPLACEMENT SYSTEM USING LOCKED NUCLEIC ACIDS

Shohei Kotani,[a,] Xiaoping Olson,[a,*] Bernard Yurke,[a,b,] and William L Hughes.[a]*

[a] Micron School of Materials Science & Engineering, Boise State University, 1910 University Drive, Boise, ID, 83725, USA. [b] Department of Electrical & Computer Engineering, Boise State University, 1910 University Drive, Boise, ID, 83725, USA.
[*] These authors contributed equally to this work

Abstract

Toehold-mediated strand displacement is a powerful and programmable method to detect target nucleotide sequences. For diagnostics, single-nucleotide selectivity is vital to distinguish on-target sequences from off-target sequences. Although there are many strategies to improve selectivity, a universally simple method is desired. Here, we site-specifically incorporated various numbers of locked nucleic acids (LNA) into double-stranded DNA complexes and a hairpin system to increase their selectivity during toehold-mediated strand displacement. Hybrid DNA/LNA systems kinetically impeded the mismatched sequences from completing strand displacement, while sustaining the reaction speed of complementary sequences. Single-nucleotide mismatch discrimination improved orders of magnitude when incorporating LNA into DNA strand displacement systems, with discrimination factors exceeding 1,000 in multiple cases and 18,000 in one specific case.

2.1 Introduction

The kinetics of toehold-mediated strand displacement is regulated in part by the toehold length.[1, 2] Short toeholds enable the programmable construction of DNA circuits that detect target nucleic acid sequences,[3] as well as image nucleic acids inside cells.[4, 5] When combined with the polymerase chain reaction, strand displacement can also target nucleic acid sequences.[6-11] For all cases, high selectivity to react with a target sequence is vital because of the occurrence of single-nucleotide polymorphism in human diseases,[12] as well as sequence similarity among a family of micro RNA.[13] One approach to obtain high selectivity, in a strand displacement system, is to adjust the thermodynamic gain to be near zero when an on-target sequence hybridizes with the system. In this situation, the introduction of a mismatch base pair by hybridization of an off-target sequence introduces a small energy penalty that significantly shifts the equilibrium concentration.[6, 14] A second approach is to ensure that the toehold length is short enough to encourage reversible binding. In this case, the energy penalty of a mismatch nucleotide becomes a kinetic barrier for toehold-mediated strand displacement,[15] and hence kinetic selectivity occurs.[16-18] Moreover, when on- and off-target sequences are well known, competitive composition can reduce background signals by consuming off-target sequences.[9, 17, 18]

Although many strand displacement systems have been constructed from DNA or RNA, synthetic nucleic acids are now being adopted to increase hybridization yields[19] and stability against enzymatic digestion.[20] Among them, locked nucleic acids (LNA) are promising because of their thermo-mechanical stability.[21] Briefly described, LNA monomers are conformationally restricted RNA analogs with a covalent bridge between the 2′ oxygen and the 4′ carbon. This bridge locks the ribose ring into an N-type

conformation that dominates A-form DNA.[22] This structural constraint increases the melting temperature of a nucleic acid duplex and enhances the mismatch discrimination of LNA-DNA base pairs when compared to their DNA-DNA controls.[23, 24] As a consequence, LNA have increased the selectivity and stability of molecular beacons.[24, 25]

We have previously shown significant leakage suppression by incorporating LNA into a model DNA strand displacement system.[26] Others have shown that LNA substitutions on the toehold domain are selective to off-target sequences.[27-30] While valuable, this application of LNA is limited to the toehold length, which is relatively short at ~8 nt. In comparison, the substitution of LNA for DNA nucleotides in branch migration domains is attractive because of the increased length of their domains. Here, the number and position of LNA substitutions were studied in detail on the branch migration domains of DNA stand displacement systems. A single-nucleotide mismatch was then introduced into the invader strand to create a single A-A mismatch. The kinetics of the mismatched invaders reacting with hybrid DNA/LNA substrates was investigated using toehold-mediated strand displacement,[1] toehold exchange,[2] and hairpin systems.[31] The discrimination factors, which are defined as the ratios of the reaction rates between on- and off-target sequences that have a single-nucleotide mismatch, were then calculated and compared for every hybrid DNA/LNA substrate.

2.2 Experimental Methods

Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich.

2.2.1 Oligonucleotides and Chemicals

The sequences of the oligomers for the DNA strand displacement systems were designed via NUPACK[32] to have minimal secondary structure, and their sequences are

listed in the Supporting Information 2.8.1. Pure DNA and hybrid DNA/LNA oligonucleotides were purchased from Integrated DNA Technologies (IDT) and Exiqon, respectively, with HPLC purification. Fluorescent reporter strands were labeled by IDT with HPLC purification. Samples were resuspended in 1× TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA) and the concentration was confirmed by a 260 nm absorbance measurement. Substrate and reporter complexes were annealed at 95 °C for 5 minutes and cooled to room temperature over 90 minutes. The resulting complexes were purified via 10% native polyacrylamide gel electrophoresis (acrylamide:bis = 29:1). The bands of interest were cut out of the gel and eluted in 1× TE Mg^{2+} buffer (12.5 mM $MgCl_2$) for 2 days.

2.2.2 Reaction Rate Measurement

Reaction kinetic experiments were performed at 25 °C in glass cuvettes (Starna Cells). The cuvettes were covered with lids to minimize evaporation and had reaction volumes between 120 to 200 µL. The reaction buffer was 1× TE Mg^{2+} buffer. All experiments were conducted using an Agilent or Varian Technologies (Cary Eclipses) flourometer. The calculations of the reaction rate constants are shown in in the Supporting Information 2.8.2 and 2.8.3, which are based on our previous work.[26]

2.3 Results and Discussion

2.3.1 Single-Nucleotide Mismatch Discriminations in a Toehold-Mediated Strand Displacement system

A model toehold-mediated strand displacement system was designed to investigate the kinetics of hybrid DNA/LNA substrates reacting with an invader strand that has a single-mismatched base. The schematic is shown in Figure 2.1. The reaction is

initiated by domain 5 on the invader (**i**) to hybridize with the domain 5* on the substrate (**sb**). Then, the invader displaces the signal (**s**) strand, through a three-way branch migration process, and forms a waste (**W**) complex with the backbone. Finally, the signal strand reacts with the reporter complex (**R**) to release the dye (**d**) strand, whose fluorescence intensity is monitored using a fluorometer.

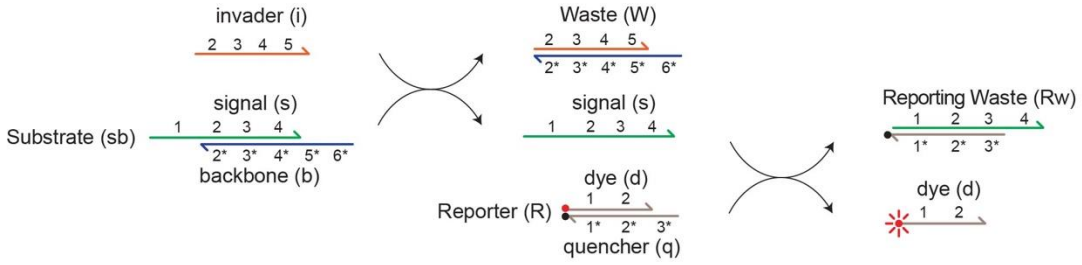

Figure 2.1. Schematic of a model toehold-mediated strand displacement system. Functional domains are represented by numbers and complementary domains are denoted by numbers with asterisks.[2] The substrate complex (sb) includes signal (s) and backbone (b) strands, and the reporter complex (R) includes dye (d) and quencher (q) strands. A single-nucleotide mutation (Thymine to Adenine) was introduced into the invader (i) strand at various positions in order to create a single A-A mismatch, and the detail of the sequences is shown in Figure 2.2

2.3.1.1 The Improvement of Mismatch Discrimination by LNA Substitutions

The sequence level representation of the pure DNA substrate (**sb0**) is shown in Figure 2.2A. In order to investigate the influence of the number and position of LNA substitutions on the kinetics of the strand displacement reaction, three groups of substrates, shown in Figure 2.2B, were utilized: (1) substitutions on the backbone (**b**) strand – substrates **b1** to **b6**, (2) substitutions on the signal (**s**) strand – substrates **s1** to **s5**, and (3) substitutions on both the signal (**s**) and backbone (**b**) strands – substrates **sb1** to **sb5**. Among those groups, the LNA substitutions were distributed along branch migration domains 2, 3, and 4. Here, domain 2 is near the end of the branch migration domain,

domain 3 is the central region of the branch migration domain, and domain 4 is proximal to the toehold domain 5. A single-nucleotide mutation was then introduced to the fully matched invader (**i0**) to generate four single-nucleotide mismatched invaders, as shown in Figure 2.2C. Mismatched invader $\mathbf{m_1}$ has a mutated base on domain 2, invaders $\mathbf{m_2}$ and $\mathbf{m_3}$ have mutated bases on domain 3 as next-nearest neighbors, and invader $\mathbf{m_4}$ has a mutated base on domain 4. To avoid the formation of secondary structures, the thymine (T) of the fully matched invader was changed to adenine (A) at different positions of the mismatched invaders. Therefore, all of the mismatched invaders create a single A-A mismatch in this study.

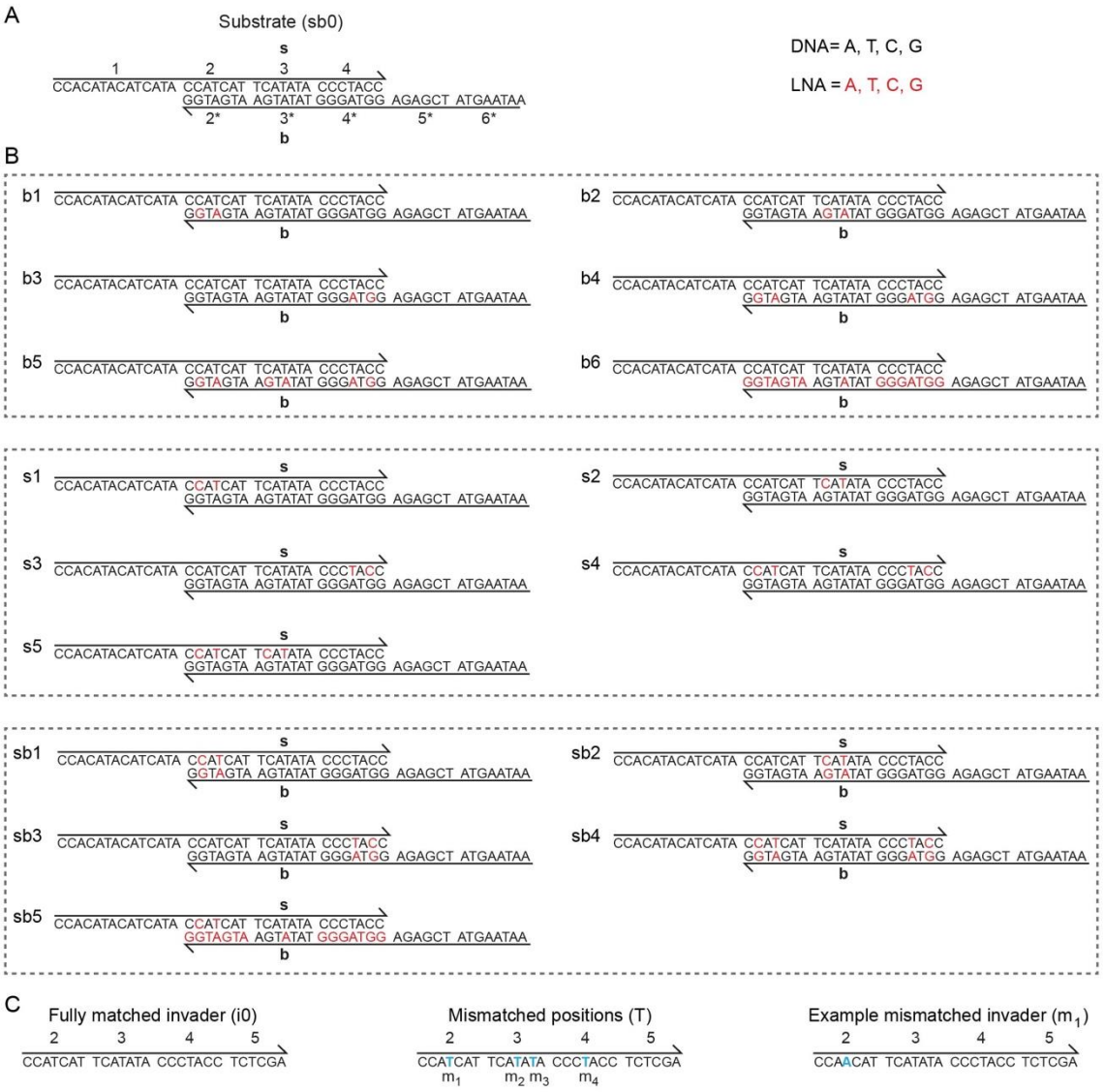

Figure 2.2. (A) The pure DNA substrate (sb0) at the sequence-level. DNA and LNA nucleotides are shown by black and red letters, respectively. (B) Hybrid DNA/LNA substrates. Substrates b1-b6 are LNA substitutions on the backbone (b) strand. Substrates s1-s5 are LNA substitutions on the signal (s) strand. Substrates sb1-sb5 are LNA substitutions on both the signal (s) and backbone (b) strands. (C) Fully matched invader (i0) and single-nucleotide mismatched invaders (m_1, m_2, m_3 and m_4). The positions at which mismatched bases were created by changing a Thymine base, T, to an Adenine base, A, are shown with cyan letters. Also shown is the sequence of the mismatched invader m_1 as an example

The rate constants of hybrid DNA/LNA substrates reacting with single-nucleotide mismatched invaders are shown in Figure 2.3A, C and E. The kinetics data was obtained

by fitting our experimental data with the second-order rate law. For additional details and representative data, see the Supporting Information 2.8.3. As a general trend, we observed that the reaction rate of the strand displacement reaction becomes more sensitive to a mismatched base pair as the mismatch position is closer to the toehold domain 5. This trend held for the strand displacement reaction in Figure 2.3 and Supporting Information 2.8.4, as well as a toehold exchange reaction with a similar design presented in Supporting Information 2.8.5. Since this finding was reported elsewhere,[15] we will not address this point further in our discussion. To quantitatively compare the improvement of selectivity by introducing LNA into the substrates, we also calculated the discrimination factor (DF),[11] which is the ratio between the rate constant for the fully matched invader (**i0**) as an on-target sequence to the rate constant for a mismatched invader as an off-target sequence for each substrate (shown in Figure 2.3B, D and F). For instance, the reaction rate constant of the substrate **b2** in Figure 2.3A is 9.9×10^5 $M^{-1}s^{-1}$ for the fully matched invader **i0** and 1.7×10^4 $M^{-1}s^{-1}$ for the mismatched invader $\mathbf{m_2}$, resulting in the discrimination factor DF = 9.9×10^5 $M^{-1}s^{-1}$ / 1.7×10^4 $M^{-1}s^{-1}$ = 58 in Figure 2.3B.

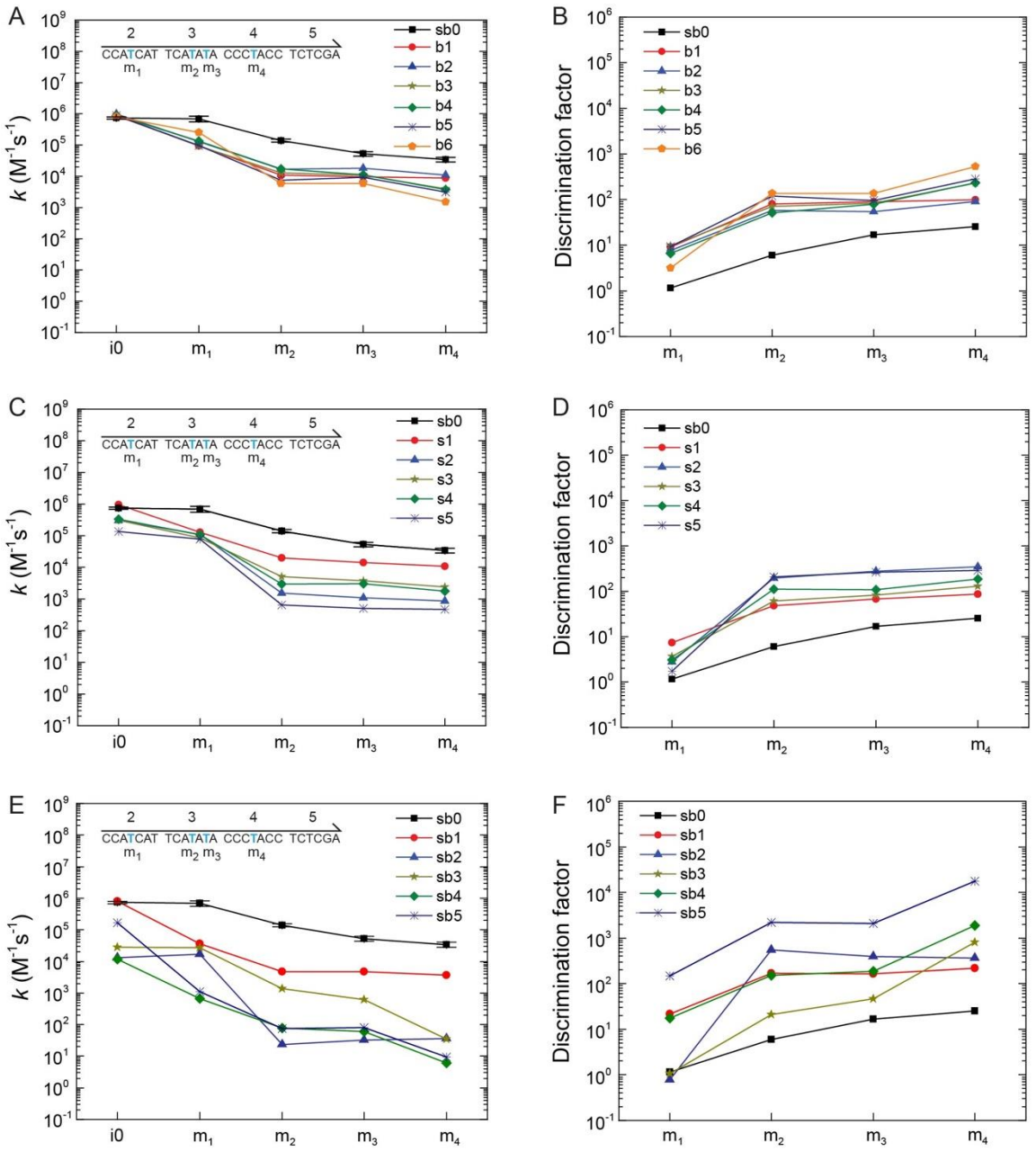

Figure 2.3. The effect of mismatch positions on hybrid DNA/LNA strand displacement rates using double-stranded complexes. The positions at which mismatched bases were created by changing a T base to an A base are shown with cyan letters at the upper left inset of (A), (C), and (E). (A) Rate constants for the invaders reacting with the substrates b1-b6. (B) Discrimination factors for the substrates b1-b6. (C) Rate constants for the invaders reacting with the substrates s1-s5. (D) Discrimination factors for the substrates s1-s5. (E) Rate constants for the invaders reacting with the substrates sb1-sb7. (F) Discrimination factors for the substrates sb1-sb7. Error bars are the standard deviation of three trials

Figure 2.3A-B shows the effect of substituting LNA bases into the backbone (**b**) strand of the substrate on the strand displacement rates for the mismatched invaders. The rate constants for the fully matched invader **i0** remained the same for pure DNA and hybrid DNA/LNA substrates, while the rate constants for the mismatched invaders became significantly slower for hybrid DNA/LNA substrates (Figure 2.3A). As a result, the discrimination factor improved for every hybrid DNA/LNA substrate (Figure 2.3B). The average discrimination factor for the four single-nucleotide mismatches was only 10 for the pure DNA substrate **sb0**, while this value ranged from 53 (substrate **b2**) to 200 (substrate **b6**) for the hybrid DNA/LNA substrates.

Figure 2.3C-D shows the effects of substituting LNA into the signal (**s**) strand of the substrate on the strand displacement rates for mismatched invaders. The LNA substitutions on the signal (**s**) strand affect the strand displacement rates of mismatched invaders more than for the case of LNA substitutions on the backbone (**b**) strand. For example, the discrimination factors of two LNA substitutions in the backbone (**b**) strand for the substrates **b1**, **b2**, **b3** are all similar in Figure 2.3B. In contrast, the substrate **s2** with two LNA substitutions on domain 3 of the signal (**s**) strand showed a higher discrimination factor than the substrates **s1** and **s3** for the mismatched invaders $\mathbf{m_2}$, $\mathbf{m_3}$, and $\mathbf{m_4}$ (Figure 2.3D). The average discrimination factor for the four single-nucleotide mismatches ranged from 52 (substrate **s1**) to 210 (substrate **s2**) for the hybrid DNA/LNA substrates. Because of the cost of LNA bases, it is worth comparing the substrates **b1**, **b2**, **b3**, **s1**, **s2**, and **s3** in Figure 2.2B since all of them contain only two LNA substitutions. Among them, the substrate **s2** outperformed the other substrates even though it had the lowest discrimination factor for the mismatched invader $\mathbf{m_1}$ (DF = 3), because it showed

the highest discrimination factor for the other mismatched invaders **m**$_2$ (DF = 190), **m**$_3$ (DF = 280), and **m**$_4$ (DF = 350), with the average discrimination factor of 210 for the four single-nucleotide mismatches.

Figure 2.3E-F highlights the effect of substituting LNA into both the signal (**s**) and backbone (**b**) strands of the substrate on the strand displacement rate of mismatched invaders. A comparison of **sb1**, **sb2**, and **sb3** in Figure 2.3E shows that the cluster of LNA-LNA base pairs at the center of branch migration domain (substrate **sb2**) or close to the toehold domain (substrate **sb3**) strongly slows down the invasion rate of the fully matched invader **i0**, while clustering at the end of the branch migration domain furthest from the toehold domain (substrate **sb1**) has no effect. The end of the branch migration domain that is furthest from the toehold is called the "*distal end*." In spite of the slow reaction of **i0**, **sb2** showed a better discrimination factor than **sb1** for mismatched invaders **m**$_2$, **m**$_3$, and **m**$_4$ (Figure 2.3F), and a similar trend appears when comparing **s1**, **s2**, and **s3** (Figure 2.3D). Nonetheless, this dataset indicates that when a LNA substitution is introduced into both the top and bottom of the substrate, in a DNA strand displacement system, the best strategy of incorporating LNA-LNA base pairs is to locate them at the distal end of the branch migration domains. In fact, after various types of LNA substitutions were tested, the best design was **sb5**, which has an LNA-LNA base pair at the distal end of the branch migration domain with the majority of the backbone (**b**) strand with LNA substitutions (Figure 2.2B). This substrate significantly slowed down the reaction speed with all four positions of the mismatched invaders while also maintaining a fast reaction speed with the fully matched invader **i0** (Figure 2.3E). As a result, it showed the large discrimination factor for the mismatched invaders **m**$_1$ (DF =

150), **m2** (DF = 2,200), **m3** (DF = 2,100), and **m4** (DF = 18,000), with the average value for the four single-nucleotide mismatches of 5,500 (Figure 2.3F). These improvements indicate that LNA substitutions, on branch migration domains, could aid in the application of nucleic acid circuit systems to diagnostics.[3] Ten additional substrates were tested, and their sequence, reaction rates, and discrimination factors are shown in Supporting Information 2.8.4.

2.3.1.2 The Influence of LNA Substitutions at the Mismatch Site Versus its Next-Nearest Neighbor Site

The effect of LNA substitutions on the melting temperature of fully matched duplexes and a single-nucleotide mismatched duplexes was studied previously.[33] In their work, an LNA substitution at the mismatch sites improved mismatch discrimination by significantly shifting the melting temperature, while an LNA substitution at the next-nearest neighbor of the mismatch sites had almost no effect on the melting temperature. The result indicates that prior knowledge on the exact position of mismatch sites is required for effective usage of LNA substitutions. In contrast, in our case, LNA substitutions are located on the branch migration domain in a strand displacement system, which may result in different rules. Therefore, we compared the effect of an LNA substitution at a mismatch site and its next-nearest neighbor site. For this purpose, the subset of substrates in Figure 2.2B and Figure 2.12B were selected so that mismatched invader **m2** forms a single A-A mismatch at the LNA substitution site, while the mismatched invader **m3** forms a single A-A mismatch at the next-nearest neighbor of the LNA substitution site. An example using substrate **b2** is shown in Figure 2.4A. In order to quantify the improvement of mismatch discrimination for the invaders **m2** and **m3**, by

substituting LNAs on the pure DNA substrate **sb0**, we calculated the fold improvement of the discrimination factor (DF), which is DF (hybrid DNA/LNA substrate) / DF (pure DNA substrate **sb0**) based on the data in Figure 2.3 and Figure 2.13. This quantity is plotted in Figure 2.4B as a function of the substrate type when the invader $\mathbf{m_2}$ (blue) and the invader $\mathbf{m_3}$ (orange) are used. For instance, DF for the invader $\mathbf{m_2}$ is 5 with the substrate **sb0** while 58 with the substrate **b2** in Figure 2.3B, resulting in 11 fold improvement of DF by the LNA substitution at the mismatch site.

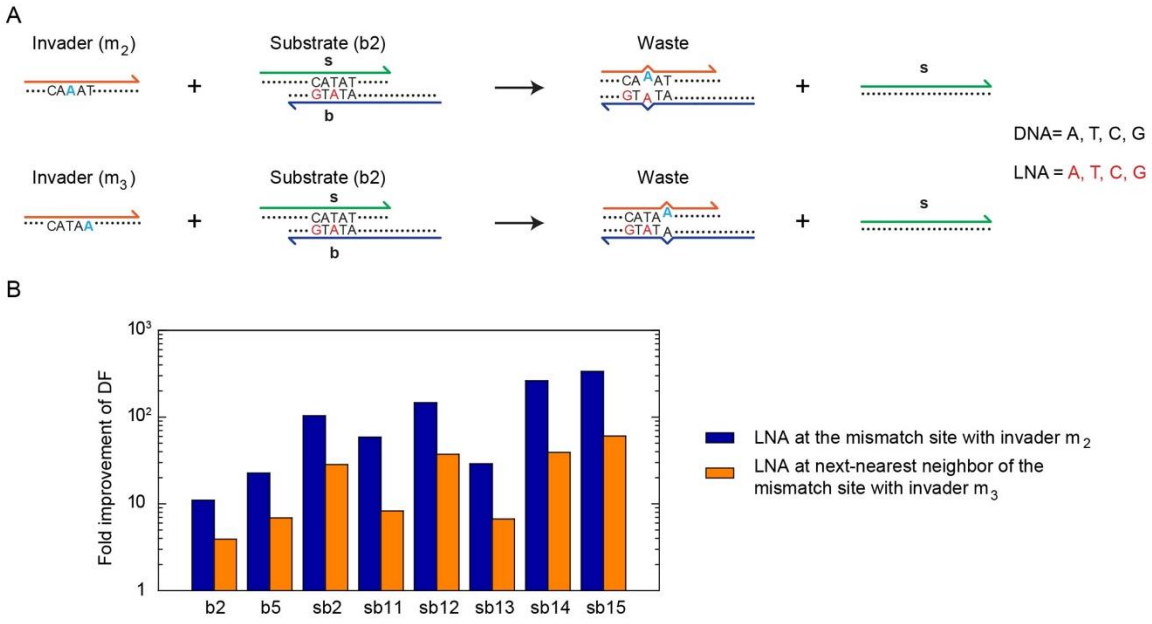

Figure 2.4. The fold improvement of the discrimination factor (DF) for DNA mismatches at LNA substitution sites or next-nearest neighbor sites compared with the pure DNA substrate sb0. (A) An example in which the substrate b2 was reacted with the mismatched invaders $\mathbf{m_2}$ and $\mathbf{m_3}$. DNA and LNA nucleotides are shown by black and red letters, respectively. The positions at which mismatched bases were created by changing a T base to an A base are shown with cyan letters. The mismatched invader $\mathbf{m_2}$ forms a single A-A mismatch at the LNA substitution site, while the mismatched invader $\mathbf{m_3}$ forms a single A-A mismatch at the next-nearest neighbor of the LNA substitution site. (B) Fold improvement of DF, which is DF (hybrid DNA/LNA substrate) / DF (pure DNA substrate sb0) for the invaders $\mathbf{m_2}$ or $\mathbf{m_3}$, for the selected subsets of substrates

As with LNA probes in a hybridization assay,[33] the LNA substitution on the substrate at the mismatch site (**m_2**) is more effective at improving the discrimination factor than the LNA substitution on the substrate at the next-nearest neighbor site (**m_3**), as shown in Figure 2.4B. However, a respectable improvement of the discrimination factor was observed even when the mismatch site was located at the next-nearest neighbor to the LNA substitution site. The data indicates that the mismatch discrimination can be improved even if the LNA substitution is not immediately proximal to the mismatch site. This observation can be supported further if one looks at the substrates **b1**, **s1**, and **sb1**; which have LNA substitutions at the terminal end of the branch migration domain 2 (Figure 2.2B). For example, these substrates showed improvement in the mismatch discrimination for the mismatched invaders **m_2**, **m_3**, and even **m_4** for which the mismatch was 14 nt away from the LNA substituted nucleotides (Figure 2.3B, D, F). This result indicates that a respectable improvement in mismatch discrimination can be obtained even without locating an LNA nucleotide near the mismatch location. This is of advantage in the design of probes for cases where the position at which the mismatch occurs is not known in advance or where the position varies.[11, 13, 18]

2.3.1.3 Detection of Target Sequence at Low Abundances

In addition to nucleic acid circuits, strand displacement systems with LNA substitutions can be used for post-PCR detection similar to molecular beacons.[34] One important requirement for this application is the ability to detect a target sequence with low abundance from the background of off-target sequences with a single-nucleotide mutation.[9, 11] To demonstrate the utility of a hybrid DNA/LNA strand displacement system for this application, we performed the detection of an on-target sequence, which is

the fully matched invader **i0**, with an excess of off-target sequences, which are the mismatched invaders **m_2** and **m_4**. The results are shown in Figure 2.5.

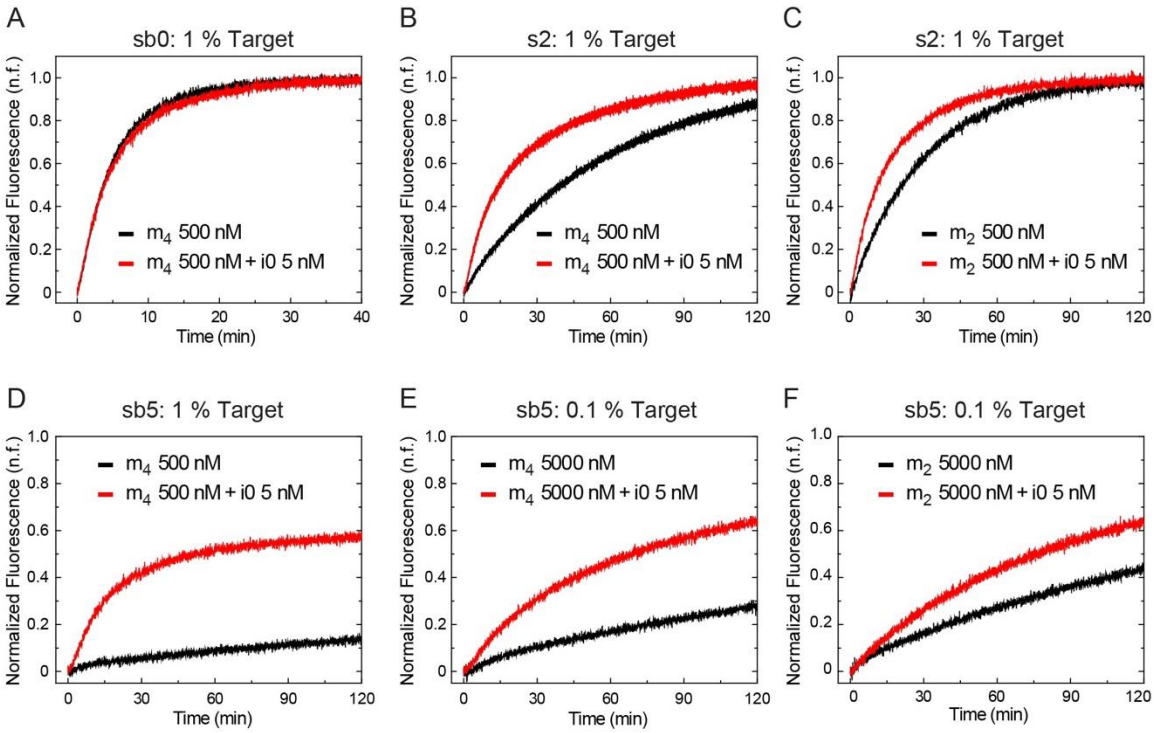

Figure 2.5. Detection of an on-target sequence at low abundances from a background of off-target sequences with a single-nucleotide mutation. For all experiments, fluorescence intensity is normalized so that 0 n.f. (*normalized fluorescence*) is the initial intensity and 1 n.f. is the final intensity after all of the substrates were consumed. In the case of sb5, even excess amount of off-target sequence did not consume all of the substrates. Therefore, excess amount of the target sequence i0 was added to acquire the final intensity. 10 nM of substrates and 20 nM of reporter were used for all of the experiments. (A) The substrate sb0 (pure DNA substrate) and 500 nM of the mismatched invader m_4 with or without 5 nM of the fully matched invader i0. (B) The substrate s2 (hybrid DNA/LNA substrate) and 500 nM of m_4 with or without 5 nM of i0. (C) The substrate s2 and 500 nM of the mismatched invader m_2 with or without 5 nM of i0. (D) The substrate sb5 (hybrid DNA/LNA substrate) and 500 nM of m_4 with or without 5 nM of i0. (E) The substrate sb5 and 5000 nM of m_4 with or without 5 nM of i0. (F) The substrate sb5 and 5000 nM of m_2 with or without 5 nM of i0

Figure 2.5A is the control experiment in which the pure DNA substrate **sb0** was used to detect 1 % of target sequence **i0** with a background of the mismatched invader **m_4**

as the off-target sequence. Little separation is observed between the fluorescence curves for the no target and 1 % target concentration case. In contrast, the hybrid DNA/LNA substrate **s2** could detect 1 % of **i0** within an $\mathbf{m_4}$ background (Figure 2.5B) and the background of the mismatched invader $\mathbf{m_2}$ (Figure 2.5C). This result emphasizes that a single type of LNA substitution can improve the detection of single-nucleotide mutations at various locations. Figure 2.5D shows the detection of 1 % target sequence by the hybrid DNA/LNA substrate **sb5** in the presence of 500 nM $\mathbf{m_4}$. The $\mathbf{m_4}$ only fluorescence curve increases much more slowly than the fluorescence curve with the target present. Figure 2.5E, F shows the successful detection of 0.1 % target by the substrate **sb5**, in the presence of either $\mathbf{m_4}$ and $\mathbf{m_2}$ mismatched invaders. Although those experiments were not conducted on PCR amplified samples, they show the possibility of using a hybrid DNA/LNA strand displacement system for post-PCR detection of single-nucleotide mismatches, even without prior knowledge of the location of the mutations.

We also explored the use of LNA in a toehold exchange system in Supporting Information 2.8.5. In this system, an invader strand does not completely displace the signal strand of a substrate,[2] but the last few base pairs of signal (**s**) strand spontaneously dissociate. We used a system in which both the invasion toehold and the dissociation toehold were 6 nt in length. The nucleotide sequences for this system are shown in Figure 2.14A. Based on the result of a strand displacement system in Figure 2.3, it was expected that LNA substitutions also improve the selectivity of toehold exchange system. However, as shown in Figure 2.14B and C, we did not observe an improvement in the discrimination factor by substituting LNAs into the toehold exchange system.

2.3.2 Single-Nucleotide Mismatch Discriminations in a Hairpin System

In addition to the double-stranded DNA complexes examined above, hairpin structures which employ toehold-mediated strand displacement are also of importance for DNA nanotechnology.[35-37] Therefore, a model hairpin system was designed to investigate the kinetics of hybrid DNA/LNA hairpins reacting with an invader strand with a single-nucleotide mismatch. The schematic is shown in Figure 2.6A. The reaction is initiated by domain 12 on the invader (**hi0**) to hybridize with the domain 12^{*} on the hairpin. Then, the invader opens up the hairpin to form the hairpin complex **HC**. Finally, the complex **HC** reacts with the reporter complex to release the dye (**d**) strand. The sequence level representation of the pure DNA hairpin (**h0**) and LNA/DNA hybrid hairpins (**h1**, **h2**, **h3**) are shown in Figure 2.6B. The hairpin **h1** has LNA substitutions at the side of the duplex to be hybridized with the invader strand, the hairpin **h2** has LNA substitutions at the other side of the duplex to be displaced by the invader strand, and the hairpin **h3** has LNA substations at both sides of the duplex to form LNA-LNA base pairs. Select mismatch modifications were introduced into the branch migration domains (9, 10, and 11) of the fully matched invader **hi0** to create mismatched invaders $\mathbf{m^1}$ to $\mathbf{m^8}$ (Figure 2.6C). Those mismatched invaders form a single A-A mismatch with the duplex region of the hairpin.

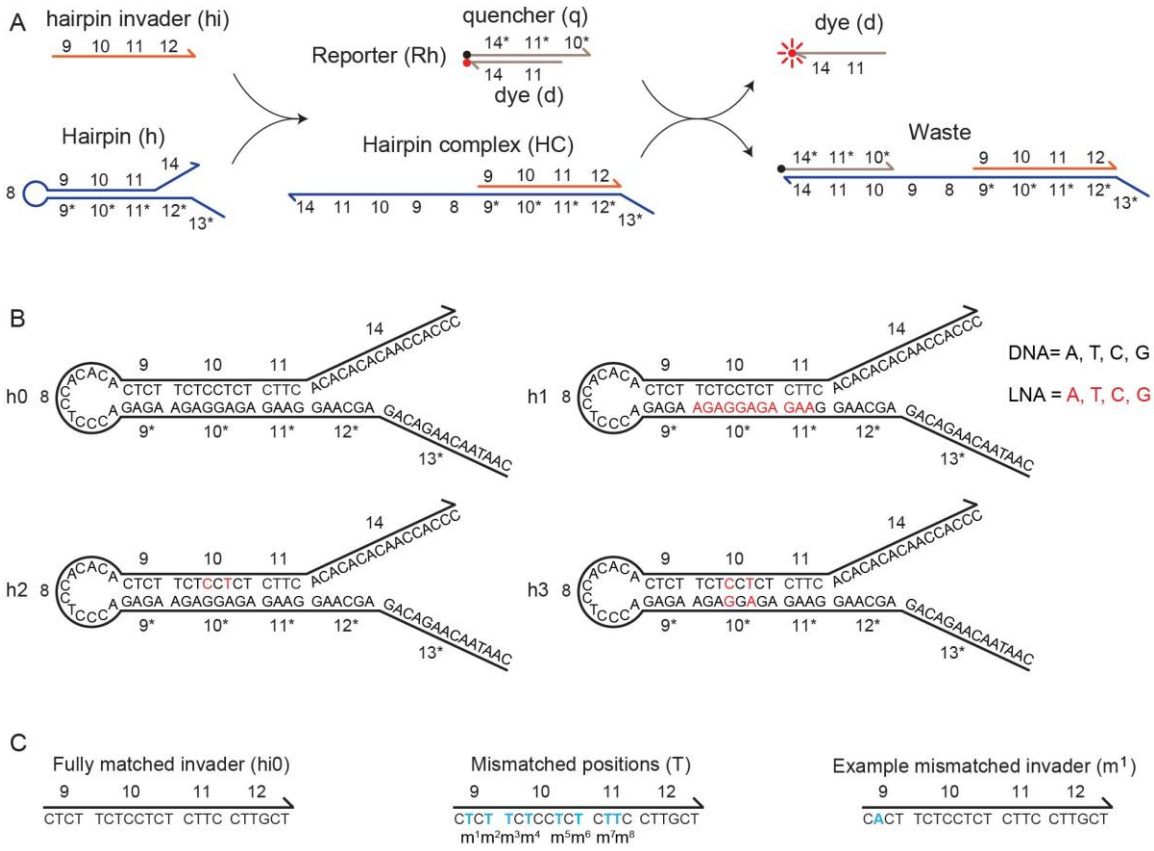

Figure 2.6. (A) Schematic of a hairpin system employing toehold-mediated strand displacement. Functional domains are represented by numbers and complementary domains are denoted by numbers with asterisks. (B) The pure DNA hairpin (h0) and hybrid DNA/LNA hairpins (h1, h2, and h3) displayed at the sequence-level. DNA and LNA nucleotides are presented by black and red letters, respectively. (C) Fully matched invader (hi0) and single-nucleotide mismatched invaders (m^1-m^8). The positions at which mismatched bases were created by changing a T base to an A base are shown with cyan letters. Also shown is the sequence of the mismatched invader m^1 as an example

The rate constants of hybrid DNA/LNA hairpins reacting with single-nucleotide mismatched invaders are shown in Figure 2.7A, and the resulting discrimination factors are shown in Figure 2.7B. Similar to the results of the double-stranded complex systems in Figure 2.3, the rate of hairpin opening becomes more sensitive to a mismatched base pair as the mismatch position comes closer to the toehold domain 12. However, a key difference between the double-stranded complex **sb0** (pure DNA) and the hairpin system

h0 (pure DNA) is the higher discrimination factors exhibited by the **h0** system for the mismatches that occur near the distal end of branch migration domain 9. For the case of the mismatched invader **m^2**, whose mismatch is 4 nt away from the distal end of the branch migration domain, a discrimination factor more than 10 (Figure 2.7B) was observed for the pure DNA hairpin **h0**. In contrast, the double-stranded complex of **sb0** exhibited a discrimination factor similar to invader **m$_1$** (Figure 2.3B), whose mismatch is also 4 nt away from the distal end of the branch migration domain. Even for the invader **m^1** whose mismatch site is only 2 nt away from the distal end of the branch migration domain, a discrimination factor of almost 5 was observed for **h0** (Figure 2.7B).

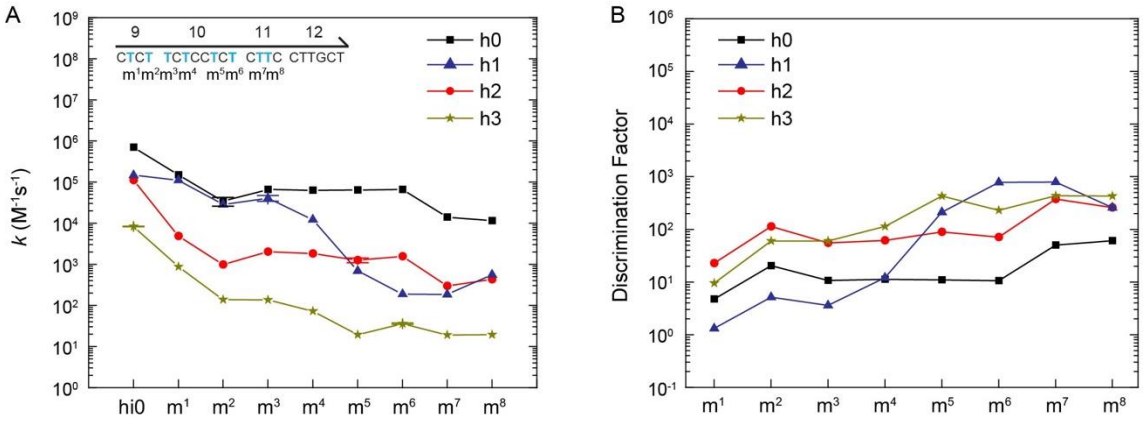

Figure 2.7. The effect of mismatch positions on hybrid DNA/LNA strand displacement rates in a hairpin system. The positions at which mismatched bases were created by changing a T base to an A base are shown with cyan letters at the upper left inset of (A). (A) Rate constants of the invaders reacting with the hairpins h0-h3. Error bars are the standard deviation of the three trials. (B) Discrimination factor for the hairpins h0-h3

For the hybrid DNA/LNA hairpin **h1**, the rate constants of the mismatched invaders **m^1**, **m^2** and **m^3** are about the same as the pure DNA hairpin (Figure 2.7A), while the rate constant for the fully matched invader **hi0** became slower. As a result, the

discrimination factor is worse than that of the pure DNA hairpin at these mismatched positions (Figure 2.7B). This could be due to the expected strong base pair binding between the invaders and the LNA clusters of hairpin **h1** on the domains 10^* and 11^*, which is established before the hairpin encounters the mismatch sites of said invaders. Such strong binding will override the mismatch penalty at the terminal end of the branch migration domain. For the hybrid DNA/LNA hairpin **h2** and **h3**, the reaction rate constants of the mismatched invaders became significantly slower for all of the eight single-nucleotide mismatches (Figure 2.7A) and the discrimination factor improved for all the positions of a single A-A mismatch (Figure 2.7B). However, it appears that LNA substitutions into the hairpin system are less effective than those for the double-stranded complex system in Figure 2.2. For instance, both the double-stranded complex **s2** (Figure 2.2B) and the hairpin **h2** have two LNAs substituted at the center of the branch migration domain at the side of the duplex to be displaced by the invaders. Therefore, one might expect a similar level of improvement in mismatch discrimination for both systems. However, although the substrate **s2** showed a sustained reaction rate constant for the fully matched invader **i0** (Figure 2.3C), the hairpin **h2** showed a decreased reaction rate constant for the fully matched invader **hi0** (Figure 2.7A), which results in the lower DF value. As a result, the average value of DF (hybrid DNA/LNA system) / DF (pure DNA system) for all single-nucleotide mismatched invaders was 19 for substrate **s2**, while it was only 5.8 for hairpin **h2**.

2.4 Conclusion

We site-specifically incorporated LNAs into the branch migration domain of double-stranded DNA complexes and a hairpin system in order to improve their kinetic

single-nucleotide selectivity during toehold-mediated strand displacement. As off-target sequences, single-nucleotide mutations were introduced into on-target sequences to generate a single A-A mismatch at various locations on the invader strands. The double-stranded DNA complex significantly benefited from LNA substitutions when the invaders were designed to completely displace the signal strand on the complexes. The stand displacement reaction rate, for on-target sequences, significantly decreased when LNA-LNA base pairs were introduced at locations other than the distal end of the branch migration domain. Single-nucleotide mismatch discriminations were improved over 2 orders of magnitude when a large number of LNA substitutions were introduced on the bottom strand of the DNA complex, with LNA-LNA base pairs only at the distal end of the branch migration domain. Such a system could detect the on-target sequence of 0.1 % abundance from the background of off-target sequences, even when off-target sequences with a single-nucleotide mismatch at different locations were tested. In the hairpin system, LNA substitutions into the duplex region also improved the mismatch discrimination, although the improvements appeared to be less effective than that of a double-stranded DNA complex. In the toehold exchange system, incorporating LNA substitutions into a double-stranded DNA complex did not improve the single mismatch discrimination. This result indicates that care must be taken when incorporating LNA into toehold exchange systems. For the application of DNA nanotechnology in medical diagnosis, both selectivity and sensitivity are vital.[18, 38] Here we have shown that significant improvement of selectivity can be achieved by incorporating LNA into strand displacement systems. The method is simple, and it is reasonable to expect that even greater selectivity can be achieved once this technique is combined with other techniques,

such as the use of competitive compositions.[9, 17, 18] In addition, our previous work demonstrated the reduction of the leakage reactions in a DNA strand displacement system by LNA substitutions at the branch migration domain,[26] which can be used to improve the sensitivity of catalytic DNA circuits.[35, 36, 39, 40] By combing these techniques, LNA substitutions at branch migration domains should be useful to improve both selectivity and sensitivity of strand displacement systems for future medical diagnostic applications.

2.5 Author Contributions

S.K. and P.O. designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript under with feedback from B. Y. and W. L. H.

2.6 Acknowledgement

This project was supported in part by the following: (1) the W.M. Keck Foundation, (2) NIH Grant No. K25GM093233 from the National Institute of General Medical Sciences, (3) NIH Grant No. P20GM103408 from the INBRE Program of the National Center for Research Resources, and (4) The Micron Foundation. We also thank the Nanoscale Materials and Device Research Group at Boise State University for their continued support.

2.7 Supporting Information

2.7.1 Oligonucleotide Sequences

Table 2.1 Oligonucleotide sequences for double-stranded DNA complexes (Black and red letters represent DNA and LNA nucleotides, respectively)

Substrate	Signal (**s**) strand sequence (5'-3')	Backbone (**b**) strand sequence (5'-3')
sb0	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
b1	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
b2	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
b3	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
b4	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
b5	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
b6	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
s1	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
s2	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
s3	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
s4	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
s5	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb1	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb2	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb3	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG

sb4	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb5	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb6	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb7	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb8	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb9	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb10	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb11	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb12	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb13	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb14	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb15	CCACATACATCATA CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG
sb16	CCACATACATCATA CTCTCT CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG AGAGAG
sb17	CCACATACATCATA CTCTCT CCATCAT TCATATA CCCTACC	AATAAGTA TCGAGA GGTAGGG TATATGA ATGATGG AGAGAG

Table 2.2 Oligonucleotide sequences for hairpin structures (Black and red letters represent DNA and LNA nucleotides, respectively)

Hairpin	Sequence (5'-3')
h0	CAATAACAAGACAG AGCAAG GAAG AGAGGAGA AGAG ACCCTCCACACA CTCT TCTCCTCT CTTC ACACACACAACCACCC
h1	CAATAACAAGACAG AGCAAG GAAG AGAGGAGA AGAG ACCCTCCACACA CTCT TCTCCTCT CTTC ACACACACAACCACCC
h2	CAATAACAAGACAG AGCAAG GAAG AGAGGAGA AGAG ACCCTCCACACA CTCT TCTCCTCT CTTC ACACACACAACCACCC
h3	CAATAACAAGACAG AGCAAG GAAG AGAGGAGA AGAG ACCCTCCACACA CTCT TCTCCTCT CTTC ACACACACAACCACCC

Table 2.3 Oligonucleotide sequences for invaders. The positions at which mismatched bases were created by changing a Thymine base, T, to an Adenine base, A, are shown with cyan letters

Invader		Sequences (5'-3')
Fully matched	**i0**	CCATCAT TCATATA CCCTACC TCTCGA
Mismatch for **i0**	**m_1**	CCAACAT TCATATA CCCTACC TCTCGA
Mismatch for **i0**	**m_2**	CCATCAT TCAAATA CCCTACC TCTCGA
Mismatch for **i0**	**m_3**	CCATCAT TCATAAA CCCTACC TCTCGA
Mismatch for **i0**	**m_4**	CCATCAT TCATATA CCCAACC TCTCGA
Fully matched	**i1**	CTCTCT CCATCAT TCATATA CCCTACC TCTCGA
Mismatch for **i1**	**dm_1**	CTCTCT CCAACAT TCATATA CCCTACC TCTCGA
Mismatch for **i1**	**dm_2**	CTCTCT CCATCAT TCAAATA CCCTACC TCTCGA
Mismatch for **i1**	**dm_3**	CTCTCT CCATCAT TCATAAA CCCTACC TCTCGA
Mismatch for **i1**	**dm_4**	CTCTCT CCATCAT TCATATA CCCAACC TCTCGA
Fully matched	**hi0**	CTCT TCTCCTCT CTTC CTTGCT
Mismatch for **hi0**	**m^1**	CACT TCTCCTCT CTTC CTTGCT
Mismatch for **hi0**	**m^2**	CTCA TCTCCTCT CTTC CTTGCT
Mismatch for **hi0**	**m^3**	CTCT ACTCCTCT CTTC CTTGCT

Mismatch for **hi0**	**m**[4]	CTCT TCACCTCT CTTC CTTGCT
Mismatch for **hi0**	**m**[5]	CTCT TCTCCACT CTTC CTTGCT
Mismatch for **hi0**	**m**[6]	CTCT TCTCCTCA CTTC CTTGCT
Mismatch for **hi0**	**m**[7]	CTCT TCTCCTCT CATC CTTGCT
Mismatch for **hi0**	**m**[8]	CTCT TCTCCTCT CTAC CTTGCT

Table 2.4 Reporter sequences

Reporter complex	Dye strand sequence (5'-3')	Quencher strand sequence (5'-3')
R1	5TET/ CCACATACATCATACCATCAT	TATATGAATGATGGTATGATGTATGTGG/3IABkFQ
R2	5TET/ CCACATACATCATACCAT	TATATGAATGATGGTATGATGTATGTGG/3IABkFQ
Rte	5TET/ CCACATACATCATACTCTCTCCAT	TATATGAATGATGGAGAGAGTATGATGTATGTGG /3IABkFQ
Hr	CT CTTC ACACACAC AACCACCC/3Rox_N	5IAbRQ/GGGTGGTT GTGTGTGT GAAG AGAGGAGA

2.7.2 The Calculation of Reaction Rate Constants for Reporting Reactions

Based on the Figure 2.1 and the second-order rate law, the reactions between the reporter complex (**R**) and the signal strand (**s**) to produce the dye strand (**d**) and the reporting waste product (**Rw**) can be modeled as:

$$\mathbf{R} + \mathbf{s} \xrightarrow{\square\square} \mathbf{d} + \mathbf{Rw} \tag{2.1}$$

Experimentally, dye strand (**d**) is monitored by fluorometer. The rate equation of the dye strand (**d**) can be derived from equation (2.1) as

$$\frac{\square[\mathbf{d}]}{\square\square} = \square_{\square}[\mathbf{R}][\mathbf{s}] \tag{2.2}$$

When the initial concentration of each reactant is $[\mathbf{R}]_0$ and $[\mathbf{s}]_0$ in the absence of the dye strand (**d**), the mass balance equations are:

$$[\mathbf{R}] = [\mathbf{R}]_0 - [\mathbf{d}] \tag{2.3}$$

$$[\mathbf{s}] = [\mathbf{s}]_0 - [\mathbf{d}] \tag{2.4}$$

Insertion of equations (2.3) and (2.4) into (2.2), and integrating yields:

$$\frac{[\mathbf{d}]}{[\mathbf{s}]_0} = \frac{[\mathbf{R}]_0(1 - \square\square\square\{-\square_{\square}([\mathbf{R}]_0 - [\mathbf{s}]_0)\square\})}{[\mathbf{R}]_0 - [\mathbf{s}]_0\square\square\square\{-\square_{\square}([\mathbf{R}]_0 - [\mathbf{s}]_0)\square\}} \tag{2.5}$$

Equation (2.5) is the normalized concentration of the dye strand (**d**), and was fit to the normalized fluorescence intensity in Figure 2.8 in order to calculate the reaction rate constants of the reporting reactions. Here, four types of reporter complexes were used. The reporter complex **R1**, which has 7 nt toehold, was later used to measure the reaction rate constants between the mismatched invaders (**m_1**, **m_2**, **m_3**, and **m_4**) and the substrates in Figure 2.2, Figure 2.12, and Figure 2.5. The reporter complex **R2**, which has 10 nt toehold, was used to measure the reaction rate constants between the fully matched invader (**i0**) and the substrates in Figure 2.2 and Figure 2.12. The reporter complex **Rte** was used for the experiments in Figure 2.14, and the reporter complex **Rh** was used for the experiments in Figure 2.7.

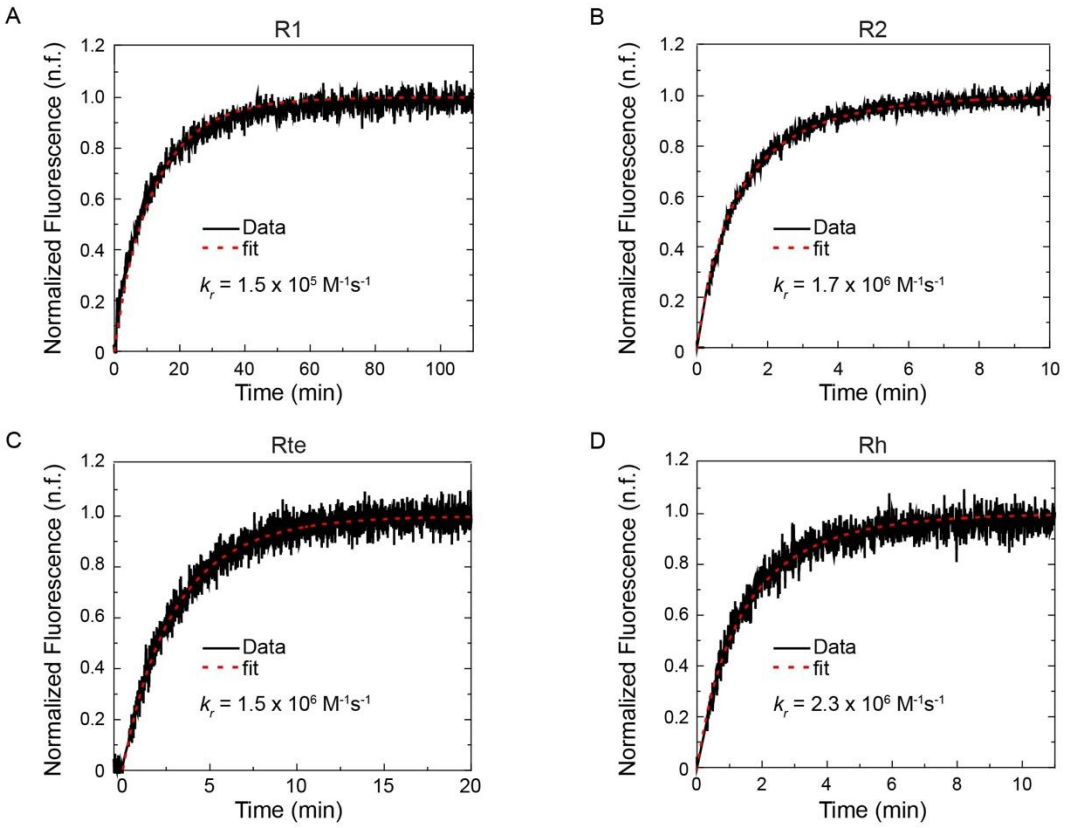

Figure 2.8. The reaction kinetic data for reporter complexes reacting with signal strands. For all experiments, fluorescent intensity was normalized to 1 n.f. (*normalized fluorescence*) at the last time point at which fluorescence data was taken. (A) Kinetics data for 10 nM of the reporter R1 reacting with 2.5 nM of the signal strand (s). The signal strand (s) is the strand in Figure 2.2A without LNA substitutions. (B) Kinetics data for 10 nM of the reporter R2 reacting with 6 nM of the signal strand (s). (C) Kinetics data for 4 nM of the reporter Rte in the toehold exchange system reacting with 1 nM of the signal strand (se). The signal strand (se) is the strand in Figure 2.14A without LNA substitutions. (D) Kinetics data for 6 nM of the reporter Rh reacting with 3 nM of the hairpin complex HC. The hairpin complex HC without LNA modification was purified by native polyacrylamide gel electrophoresis. The schematic of Rh and HC are shown in Figure 2.6A

2.7.3 The Reaction Between DNA/LNA Substrates and Invader Strands

Based on the Figure 2.1 and the second-order rate law, the reactions between the substrate complex (**sb**) and the invader strand (**i**) to produce the signal strand (**s**) and the waste (**W**) can be modeled as:

$$\mathbf{sb} + \mathbf{i} \xrightarrow{\square} \mathbf{s} + \mathrm{W} \tag{2.6}$$

Once the signal strand (**s**) is produced, it reacts with the reporter complex and produces fluorescence signal based on equation (2.1). Under our experimental condition, the reporting reaction in equation (2.1) is much faster than the invading reaction in equation (2.6) due to 2 times extra reporter concentration, quick reaction rate constants for reporting reactions calculated from Figure 2.8, and slow invading reaction when invader strands have a single-nucleotide mutation. Therefore, the fluorescence is considered to be a direct measurement of the concentration of the signal strand (**s**). The rate equation of the signal strand (**s**) can be derived from equation (2.6) as

$$\frac{\square[\mathbf{s}]}{\square\square} = \square[\mathbf{sb}][\mathbf{i}] \tag{2.7}$$

When the initial condition of each species is $[\mathbf{sb}]_0$ and $[\mathbf{i}]_0$ in the absence of the signal strand (**s**), the mass balance equations are:

$$[\mathbf{sb}] = [\mathbf{sb}]_0 - [\mathbf{s}] \tag{2.8}$$

$$[\mathbf{i}] = [\mathbf{i}]_0 - [\mathbf{s}] \tag{2.9}$$

Insertion of equations (2.8) and (2.9) into (2.7), and integrating yields:

$$\frac{[\mathbf{s}]}{[\mathbf{sb}]_0} = \frac{[\mathbf{i}]_0(1 - \square\square\square\{-\square([\mathbf{i}]_0 - [\mathbf{sb}]_0)\square\})}{[\mathbf{i}]_0 - [\mathbf{sb}]_0\square\square\square\{-\square([\mathbf{i}]_0 - [\mathbf{sb}]_0)\square\}} \tag{2.10}$$

Equation (2.10) is the normalized concentration of the signal strand (**s**). Therefore, equation (2.10) can be fit to the experimental data once fluorescence trace is normalized so that 0 n.f. is the initial intensity and 1 n.f. is the final intensity after all of the substrates were consumed. We used equation (2.10) to fit the experimental data when the fluorescence intensity reached completion during the experimental time window.

In the process of deriving equation (2.10), equations (2.7) to (2.9) can be combined to yield the following useful equation:

$$\ln\left(\frac{([\mathbf{sb}]_0 - [\mathbf{s}])[\mathbf{i}]_0}{([\mathbf{i}]_0 - [\mathbf{s}])[\mathbf{sb}]_0}\right) = -\square([\mathbf{i}]_0 - [\mathbf{sb}]_0)\square \tag{2.11}$$

The reaction rate between substrates and mismatched invaders is very slow for the cases where a large number of LNA were substituted into double-stranded DNA complexes and the hairpin system. The slowness of these reactions prevented us from following the reaction to completion. For those cases, the final intensity was obtained by adding an excess amount of fully matched invaders. Then, the fluorescence traces were normalized so that 1 n.f. becomes their final intensity. Equation (2.11) provides a better fitting procedure for fluorescence traces were the reaction did not reach completion. For clarity, equation (2.11) can be rearranged to be:

$$\ln\left(\frac{(1 - [\mathbf{s}]/[\mathbf{sb}]_0)[\mathbf{i}]_0}{([\mathbf{i}]_0/[\mathbf{sb}]_0 - [\mathbf{s}]/[\mathbf{sb}]_0)[\mathbf{sb}]_0}\right) = -\square([\mathbf{i}]_0 - [\mathbf{sb}]_0)\square \tag{2.12}$$

Since [**s**]/[**sb**]$_0$ corresponds to the normalized fluorescence intensity, the data was plotted versus time according the prescription of the left-hand side of equation (2.12). In the resulting plot the data points exhibit a linear trend. Reaction rate constant was calculated by fitting this data to a straight line. Figure 2.9, 2.10, and 2.11 show the representative fluorescence data of the strand displacement reactions together with the fitting traces. For some cases, initial transient reactions with faster reaction rates were observed. An example is the initial portion of fluorescence traces in Figure 2.9C. It can be expected that those initial transient reactions were caused by the sub-population of malformed DNA or hybrid DNA/LNA structures which contain oligonucleotides with synthesis errors. In order to avoid the influence of those structures, the regions of those

initial transients were not used for fitting. Figure 2.9, 2.10, and 2.11 show that second-order rate law fit well to fluorescence traces over the entire range of reaction rates. The same analysis procedure employing the same equations was also used for the toehold exchange reaction and the hairpin system. All the reaction rate constants are shown in Table 2.5.

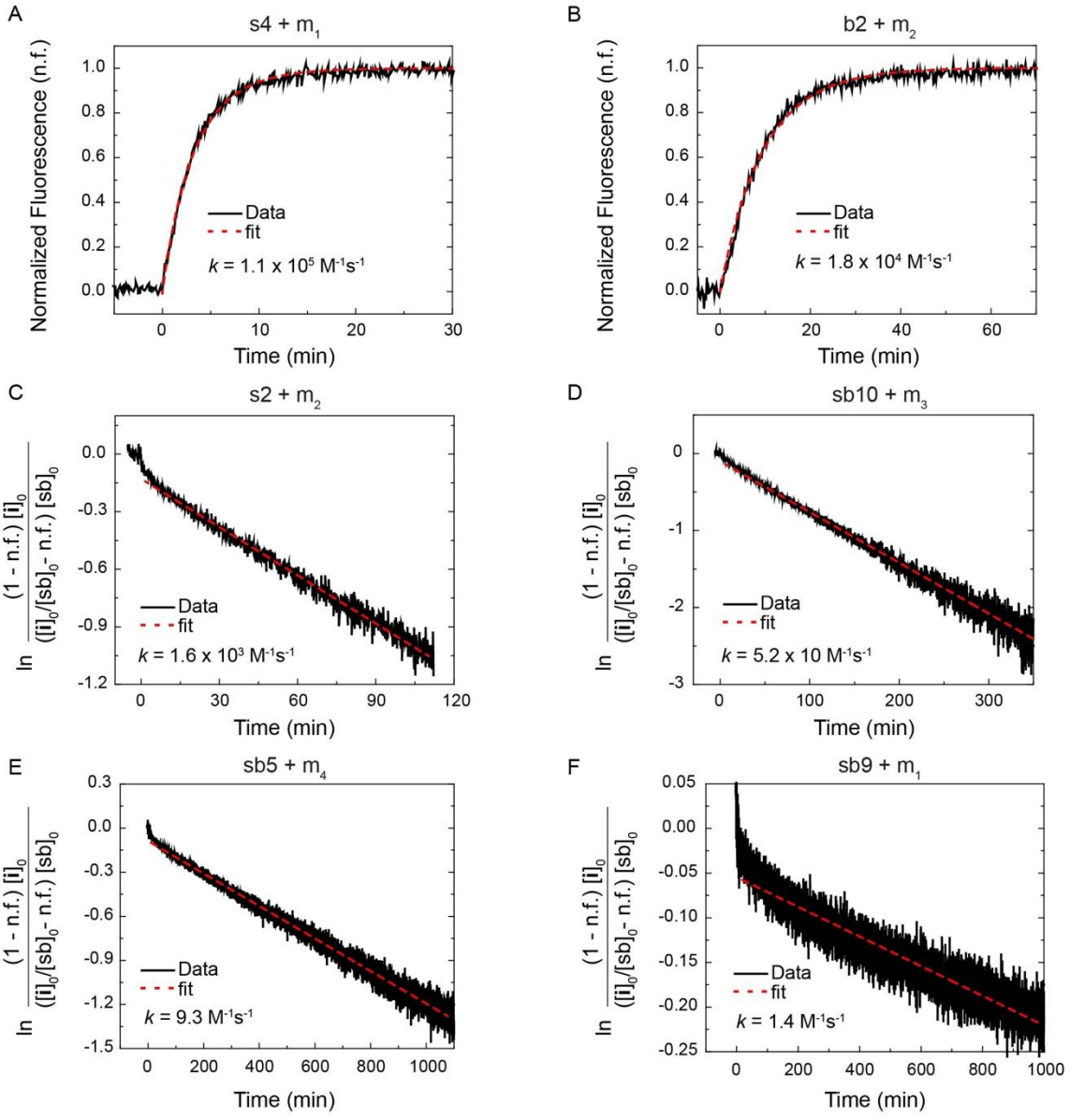

Figure 2.9. Representative kinetics data for strand displacement reactions in which the double-stranded complexes with LNA substitutions and the mismatched invaders are reacted. For all the data shown here, the concentration of substrates was 10 nM, and the concentration of the reporter R1 was 20 nM. (A) The substrate s4 with 50 nM of the mismatched invader m_1. (B) The substrate b2 with 100 nM of the mismatched invader m_2. (C) The substrate s2 with 100 nM of the mismatched invader m_2. (D) The substrate sb10 with 2000 nM of the mismatched invader m_3. (E) The substrate sb5 with 2000 nM of the mismatched invader m_4. (F) The substrate sb9 with 2000 nM of the mismatched invader m_1. Equation (2.10) was used for Figure 2.9A, B, and equation (2.12) was used for Figure 2.9C-F. In Figure 2.9C-F, n.f. stands for *normalized fluorescence* intensity

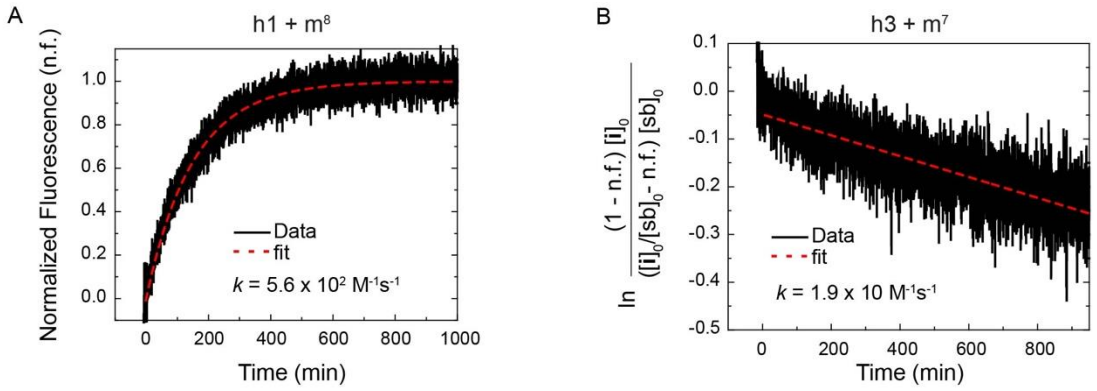

Figure 2.10. Representative reaction kinetics data for strand displacement reactions in which the hairpin system with LNA substitutions and the mismatched invaders were reacted. For all the data shown here, the concentration of hairpins was 10 nM, and the concentration of the reporter Rh was 20 nM. (A) The hairpin h1 with 200 nM of the mismatched invader m^8. (B) The hairpin h3 with 200 nM of the mismatched invader m^7. Equation (2.10) was used for Figure 2.10A, and equation (2.12) was used for Figure 2.10B

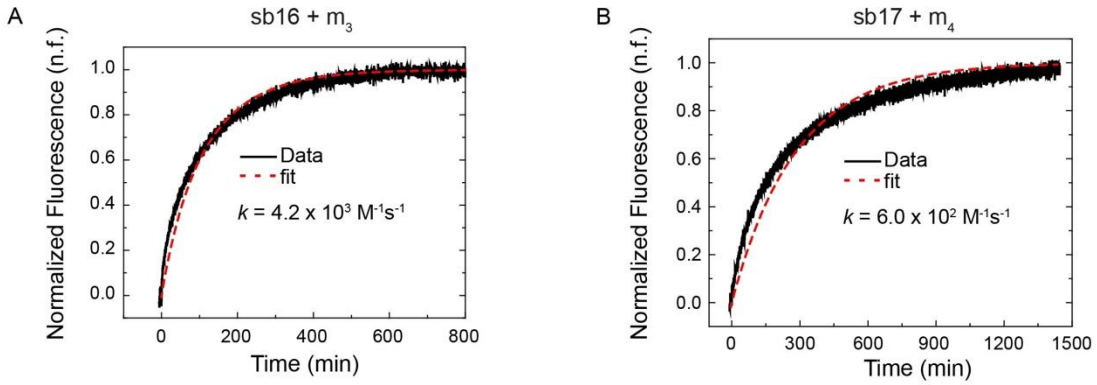

Figure 2.11. Representative reaction kinetics data for toehold exchange reactions. For all the data shown here, the concentration of substrates was 10 nM, and the concentration of the reporter Rte was 20 nM. (A) The substrate sb16 with 40 nM of the mismatched invader m_3. (B) The substrate sb17 with 200 nM of the mismatched invader m_4. Equation (2.10) was used for the fitting function

Table 2.5 Reaction rate constants ($M^{-1}s^{-1}$)

Substrate \Invader	**i0**	**m_1**	**m_2**	**m_3**	**m_4**
sb0	7.4E+05	6.9E+05	1.4E+05	5.3E+04	3.4E+04
b1	8.7E+05	9.5E+04	1.1E+04	9.7E+03	8.8E+03
b2	9.9E+05	1.3E+05	1.7E+04	1.8E+04	1.1E+04
b3	9.1E+05	9.3E+04	1.3E+04	1.1E+04	3.9E+03
b4	8.9E+05	1.4E+05	1.7E+04	1.1E+04	3.8E+03
b5	8.9E+05	9.6E+04	7.5E+03	9.3E+03	3.1E+03
b6	8.1E+05	2.5E+05	5.9E+03	5.9E+03	1.5E+03
s1	9.5E+05	1.3E+05	2.0E+04	1.4E+04	1.1E+04
s2	3.0E+05	1.1E+05	1.6E+03	1.1E+03	8.7E+02
s3	3.1E+05	8.4E+04	5.1E+03	3.8E+03	2.4E+03
s4	3.3E+05	1.1E+05	3.0E+03	3.1E+03	1.8E+03
s5	1.3E+05	7.8E+04	6.5E+02	5.1E+02	4.7E+02
sb1	7.9E+05	3.6E+04	4.7E+03	4.8E+03	3.6E+03
sb2	1.3E+04	1.7E+04	2.4E+01	3.3E+01	3.6E+01
sb3	2.9E+04	2.7E+04	1.4E+03	6.1E+02	3.5E+01
sb4	1.2E+04	6.6E+02	7.7E+01	6.1E+01	6.1E+00
sb5	1.7E+05	1.1E+03	7.5E+01	7.9E+01	9.3E+00
sb6	2.7E+04	5.8E+03	3.1E+01	2.4E+01	6.9E+00
sb7	8.7E+03	3.6E+03	2.8E+01	3.0E+01	5.3E+00
sb8	4.3E+02	2.5E+00	5.5E-01	5.2E-01	3.0E-01
sb9	7.9E+01	1.4E+00	3.1E-01	3.3E-01	4.5E-01
sb10	2.1E+04	6.8E+02	8.7E+01	5.5E+01	9.3E+00
sb11	2.1E+05	2.0E+04	6.8E+02	1.8E+03	4.4E+02
sb12	1.1E+04	6.7E+03	1.5E+01	2.2E+01	3.8E+00

sb13	1.2E+04	1.3E+04	8.2E+01	1.3E+02	2.2E+01
sb14	6.4E+03	2.2E+02	4.6E+00	1.2E+01	2.5E+00
sb15	1.3E+03	2.2E+01	7.5E-01	1.6E+00	6.5E-01
sb16	4.9E+05	1.1E+04	4.7E+03	4.0E+03	7.0E+02
sb17	4.5E+05	9.0E+03	5.3E+03	4.9E+03	3.2E+02
Substrate \Invader	**i1**	**dm_1**	**dm_2**	**dm_3**	**dm_4**
sb16	5.1E+05	2.5E+05	4.1E+04	3.2E+04	1.1E+04
sb17	4.2E+05	1.6E+04	5.6E+03	5.6E+03	6.0E+02

Hairpin\Invader	**hi0**	**m^1**	**m^2**	**m^3**	**m^4**	**m^5**	**m^6**	**m^7**	**m^8**
h0	7.0E+05	1.5E+05	4.4E+04	6.6E+04	6.3E+04	6.4E+04	6.6E+04	1.4E+04	1.1E+04
h1	1.5E+05	6.8E+04	1.7E+04	3.3E+04	7.7E+03	6.9E+02	1.9E+02	1.9E+02	5.6E+02
h2	1.1E+05	4.9E+03	9.9E+02	2.0E+03	1.8E+03	2.0E+03	1.6E+03	3.0E+02	4.3E+02
h3	8.3E+03	8.7E+02	1.4E+02	1.4E+02	7.3E+01	1.9E+01	3.5E+01	1.9E+01	1.9E+01

2.7.4 Single-Nucleotide Mismatch Discriminations in a Toehold-Mediated Strand Displacement system

The pure DNA substrate and the LNA substituted substrates are shown in Figure 2.12A and B, respectively. The substrate **sb5**, already shown in Figure 2.2, is also shown here for convenience. The fully matched invader and the mismatched invaders are shown in Figure 2.12C, which are the same as invaders in Figure 2.2C.

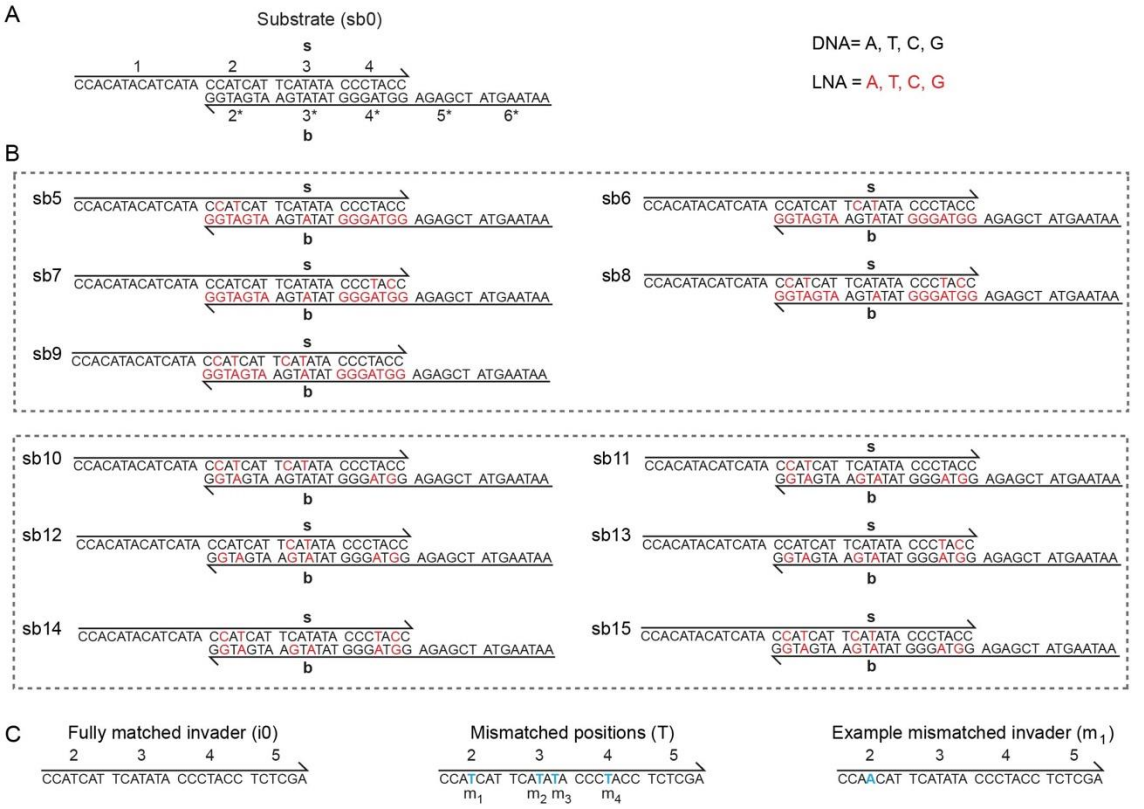

Figure 2.12. (A) The pure DNA substrate (sb0) at the sequence-level. DNA and LNA nucleotides are shown by black and red letters, respectively. (B) Hybrid DNA/LNA substrates with LNA substitutions on both the signal (s) and the backbone (b) strand. Substrates sb5-sb9 have the backbone (b) strand with a large number of LNA substitutions, and substrates s10-s15 have the backbone (b) strand with smaller number of LNA substitutions than substrates sb5-sb9. (C) Fully matched invader (i0) and single-nucleotide mismatched invaders (m_1, m_2, m_3 and m_4). The positions at which mismatched bases were created by changing a Thymine base, T, to an Adenine base, A, are shown with cyan letters. Also shown is the sequence of the mismatched invader m_1 as an example

Figure 2.13 shows the effect of substituting LNA into both the signal (**s**) and backbone (**b**) strands of the substrate on the strand displacement rate of mismatched invaders. For all types of the substrates, the general trend is that the discrimination factor is larger for a single A-A mismatch the closer the mismatch position is to the toehold domain 5, and smaller the closer the mismatch position is to the distal end of branch migration domain.

Based on the behavior of the rate constants of Figure 2.13A, the substrates can be segregated into two distinct groups – the substrates **sb5**, **sb6** and **sb7** with two LNA substitutions on the signal (**s**) strand and the substrates **sb8** and **sb9** with four LNA substitutions on the signal strand. By increasing the number of LNA substitutions on the signal (**s**) strand, the displacement of the signal strand from the substrate is further inhibited. Also, the substrates **sb6** and **sb7** showed a low discrimination factor for the mismatched invader $\mathbf{m_1}$, since those substrates do not have LNA-LNA base pairs at the domain 2, which is the most distal of the branch migration domains.

The average discrimination factor for the four single-nucleotide mismatches is improved for all of the hybrid DNA/LNA substrates (Figure 2.13B), with the value ranged from 182 (substrate **sb9**) to 5549 (substrate **sb5**).

Figure 2.13C and 2.13D show rate constants and discrimination factors for substrates **sb10** to **sb15**. Substrates **sb12** and **sb13**, which do not have LNA-LNA base pairs at the distal end in the domain 2, also did not show improvement of discrimination factor for the mismatched invader ($\mathbf{m_1}$) compared with the pure DNA substrate **sb0** (Figure 2.13D).

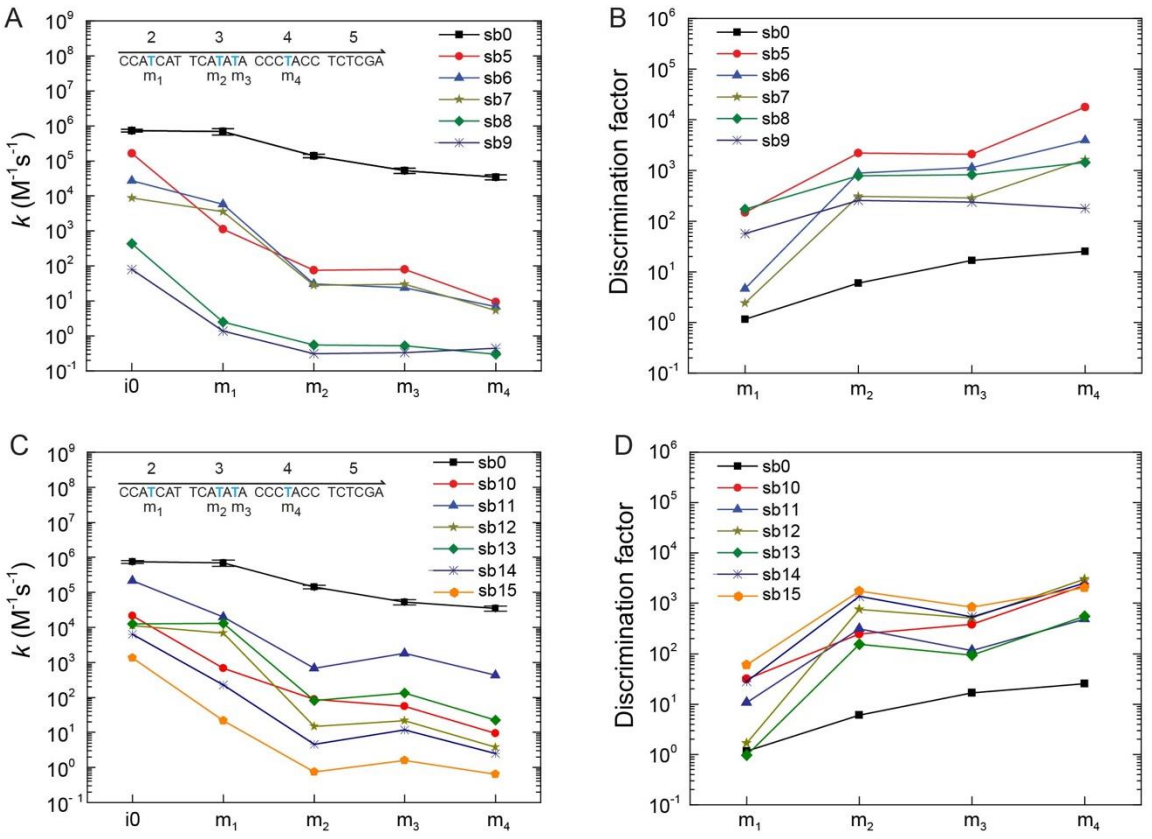

Figure 2.13. The effect of mismatch positions on hybrid DNA/LNA strand displacement rates using double-stranded complexes. The positions at which mismatched bases were created by changing a T base to an A base on the invader are shown with cyan letters at the upper left inset of (A) and (C). (A) Rate constants of the invaders reacting with the substrates sb5-sb9. (B) Discrimination factors for the substrates sb5-sb9. (C) Rate constants of the invaders reacting with the substrates sb10-s15. (D) Discrimination factors for the substrates sb10-s15. Error bars are the standard deviation of three trials

2.7.5 Single-Nucleotide Mismatch Discriminations in a Toehold Exchange System

We explored the kinetics of hybrid DNA/LNA substrate reacting with mismatched invaders in a toehold exchange system together with a strand displacement system as a comparison. The substrates and invaders for these systems are shown in Figure 2.14A. Here, the pure DNA substrate **sb16** was designed by adding 6 bp of domain 7 into the substrate **sb0** in Figure 2.2A. The invader **i1** and corresponding

mismatched invaders **dm$_1$**, **dm$_2$**, **dm$_3$**, and **dm$_4$** are employed in the strand displacement systems, and the invader **i0** and corresponding mismatched invaders **m$_1$**, **m$_2$**, **m$_3$**, and **m$_4$** are employed in the toehold exchange system. Figure 2.14B compares the reaction rate constants of fully matched and single-nucleotide mismatched invaders reacting with the pure DNA substrate **sb16** and hybrid DNA/LNA substrate **sb17**, both in the strand displacement and the toehold exchange systems. For the pure DNA substrate **sb16**, the rate constants of the mismatched invaders **m$_1$**, **m$_2$**, **m$_3$**, and **m$_4$** derived from **i0** in the toehold exchange system (red color data points) are about 1 order of magnitude slower than the rate constants of the mismatched invaders **dm$_1$**, **dm$_2$**, **dm$_3$**, and **dm$_4$** derived from **i1** in the strand displacement system (black color data points). However, the rate constants of fully matched invader are essentially the same for the toehold exchange system with the invader **i0** and the strand displacement system with the invader **i1**. As a result, toehold exchange systems showed higher mismatch discrimination factors (Figure 2.14C).

If we look at the influence of LNA substitutions on the reaction rate, the result in strand displacement system employing the invader **i1** was as expected from the result in Figure 2.3. More specifically, when the pure DNA substrate **sb16** was modified with LNAs to create the hybrid DNA/LNA substrate **sb17**, the reaction rate for single-nucleotide mismatched invaders derived from **i1** (blue color data points in Figure 2.14B) became slower than those for **sb16** (black color data points in Figure 2.14B) and the discrimination factor for those single-nucleotide mismatches improved (Figure 2.14C). However, the result in toehold exchange system employing the invader **i0** was rather surprising, because both the pure DNA substrate **sb16** (red color data points) and the

hybrid DNA/LNA substrate **sb17** (yellow color data points) showed very similar reaction kinetics in Figure 2.14B and discrimination factors (Figure 2.14C) for all types of invaders – no improvement of single-nucleotide selectivity by LNA substitutions. Although we do not have a clear explanation for this result, it does not mean that LNA substitutions for toehold exchange system are not useful. Instead, our results indicate the LNA substitutions must be strategically incorporated based on through understanding of a given system. The influence of LNA substitution on toehold exchange system should be explored in future research.

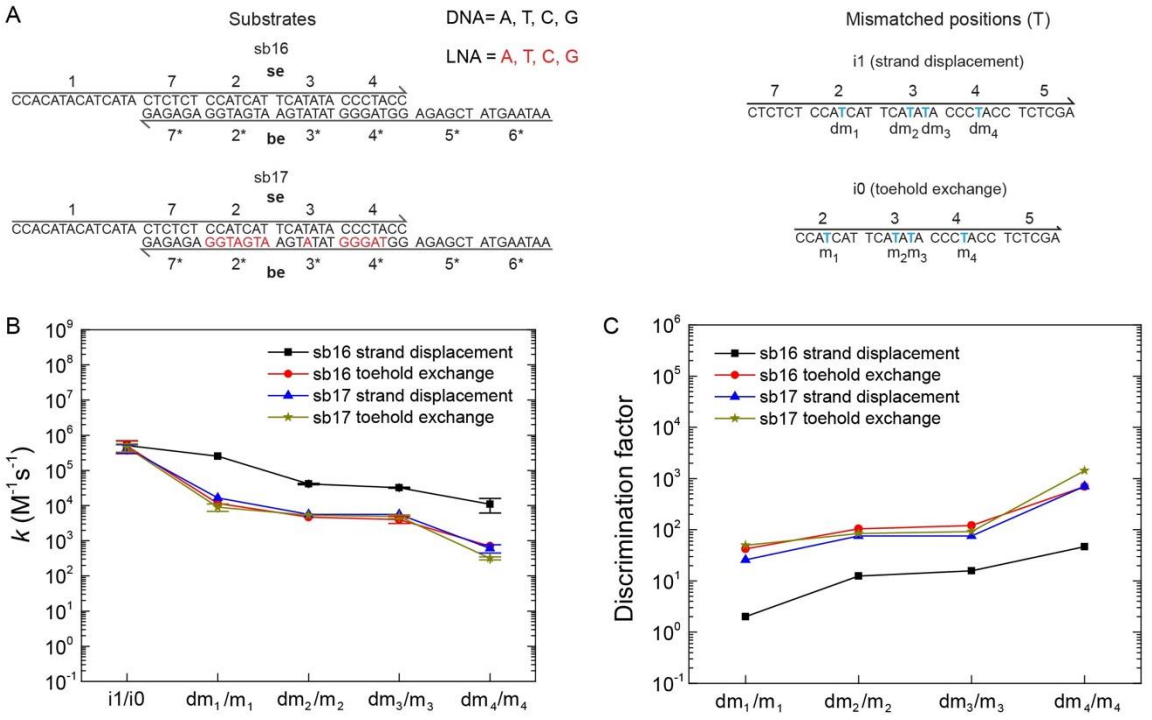

Figure 2.14. The comparison of reaction rates in strand displacement and toehold exchange systems. (A) The pure DNA substrates sb16 and the hybrid DNA/LNA substrate sb17 are shown on the left side. DNA and LNA nucleotides are represented by black and red letters, respectively. When the invader i1 and the corresponding mismatched invaders dm_1, dm_2, dm_3, and dm_4 are employed the systems function as strand displacement systems with 6 nt toehold with which the invader initiates strand displacement. When the invader i0 and the corresponding mismatched invaders m_1, m_2, m_3, and m_4 are employed the systems function as toehold exchange systems for which both the invading toehold (domain 5) and the dissociating toehold (domain 7) are 6 nt in length. Fully matched invaders (i0 and i1) and their corresponding single-nucleotide mismatched invaders are shown on the right side. The positions at which mismatched bases were created by changing a T base to an A base are shown with cyan letters. (B) Rate constants of the invaders reacting with the substrates in the strand displacement and the toehold exchange systems. (C) Discrimination factors for mismatched invaders reacting with the substrates in strand displacement and the toehold exchange systems. Error bars are the standard deviation of three trials

2.8 References

1. Yurke, B.; Mills, A., Jr., Using DNA to power nanostructures. *Genet. Program. Evolvable Mach.* **2003,** *4* (2), 111-122.

2. Zhang, D. Y.; Winfree, E., Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **2009,** *131* (47), 17303-17314.

3. Jung, C.; Ellington, A. D., Diagnostic applications of nucleic acid circuits. *Acc. Chem. Res.* **2014,** *47* (6), 1825-1835.

4. Choi, H. M. T.; Beck, V. A.; Pierce, N. A., Next-Generation in Situ Hybridization Chain Reaction: Higher Gain, Lower Cost, Greater Durability. *ACS Nano* **2014,** *8* (5), 4284-4294.

5. Cheglakov, Z.; Cronin, T. M.; He, C.; Weizmann, Y., Live cell microRNA imaging using cascade hybridization reaction. *Journal of the American Chemical Society* **2015,** *137* (19), 6116-6119.

6. 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.

7. Khodakov, D. A.; Khodakova, A. S.; Linacre, A.; Ellis, A. V., Toehold-mediated nonenzymatic DNA strand displacement as a platform for DNA genotyping. *Journal of the American Chemical Society* **2013,** *135* (15), 5612-5619.

8. Li, Q.; Luan, G.; Guo, Q.; Liang, J., A new class of homogeneous nucleic acid probes based on specific displacement hybridization. *Nucleic Acids Research* **2002,** *30* (2), e5-e5.

9. Wang, J. S.; Zhang, D. Y., Simulation-guided DNA probe design for consistently ultraspecific hybridization. *Nat. Chem.* **2015,** *7* (7), 545-553.

10. Byrom, M.; Bhadra, S.; Jiang, Y. S.; Ellington, A. D., Exquisite allele discrimination by toehold hairpin primers. *Nucleic acids research* **2014,** *42* (15), e120-e120.

11. Xiao, X.; Wu, T.; Xu, L.; Chen, W.; Zhao, M., A branch-migration based fluorescent probe for straightforward, sensitive and specific discrimination of DNA mutations. *Nucleic acids research* **2017,** *45* (10), e90-e90.

12. Kim, S.; Misra, A., SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.* **2007,** *9*, 289-320.

13. Roush, S.; Slack, F. J., The let-7 family of microRNAs. *Trends in cell biology* **2008,** *18* (10), 505-516.

14. Zhang, D. Y.; Chen, S. X.; Yin, P., Optimizing the specificity of nucleic acid hybridization. *Nature chemistry* **2012,** *4* (3), 208-214.

15. Machinek, R. R. F.; Ouldridge, T. E.; Haley, N. E. C.; Bath, J.; Turberfield, A. J., Programmable energy landscapes for kinetic control of DNA strand displacement. *Nat. Commun.* **2014,** *5*, 5324.

16. Zhang, D. Y.; Winfree, E., Robustness and modularity properties of a non-covalent DNA catalytic reaction. *Nucleic Acids Res.* **2010,** *38* (12), 4182-4197.

17. Sternberg, J. B.; Pierce, N. A., Exquisite sequence selectivity with small conditional RNAs. *Nano letters* **2014,** *14* (8), 4568-4572.

18. Chen, S. X.; Seelig, G., An engineered kinetic amplification mechanism for single nucleotide variant discrimination by DNA hybridization probes. *Journal of the American Chemical Society* **2016,** *138* (15), 5076-5086.

19. Vieregg, J. R.; Nelson, H. M.; Stoltz, B. M.; Pierce, N. A., Selective Nucleic Acid Capture with Shielded Covalent Probes. *Journal of the American Chemical Society* **2013,** *135* (26), 9691-9699.

20. Groves, B.; Chen, Y.-J.; Zurla, C.; Pochekailov, S.; Kirschman, J. L.; Santangelo, P. J.; Seelig, G., Computing in mammalian cells with nucleic acid strand exchange. *Nat Nano* **2016,** *11* (3), 287-294.

21. K. Singh, S.; A. Koshkin, A.; Wengel, J.; Nielsen, P., LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chemical Communications* **1998,** (4), 455-456.

22. Koshkin, A. A.; Nielsen, P.; Meldgaard, M.; Rajwanshi, V. K.; Singh, S. K.; Wengel, J., LNA (locked nucleic acid): an RNA mimic forming exceedingly

stable LNA: LNA duplexes. *Journal of the American Chemical Society* **1998,** *120* (50), 13252-13253.

23. Mouritzen, P.; Nielsen, A. T.; Pfundheller, H. M.; Choleva, Y.; Kongsbak, L.; Møller, S., Single nucleotide polymorphism genotyping using locked nucleic acid (LNA™). *Expert review of molecular diagnostics* **2003,** *3* (1), 27-38.

24. Wang, L.; Yang, C. J.; Medley, C. D.; Benner, S. A.; Tan, W., Locked Nucleic Acid Molecular Beacons. *Journal of the American Chemical Society* **2005,** *127* (45), 15664-15665.

25. Yang, C. J.; Wang, L.; Wu, Y.; Kim, Y.; Medley, C. D.; Lin, H.; Tan, W., Synthesis and investigation of deoxyribonucleic acid/locked nucleic acid chimeric molecular beacons. *Nucleic acids research* **2007,** *35* (12), 4030-4041.

26. Olson, X.; Kotani, S.; Yurke, B.; Graugnard, E.; Hughes, W. L., Kinetics of DNA Strand Displacement Systems with Locked Nucleic Acids. *The Journal of Physical Chemistry B* **2017,** *121* (12), 2594-2602.

27. Gao, Z. F.; Ling, Y.; Lu, L.; Chen, N. Y.; Luo, H. Q.; Li, N. B., Detection of Single-Nucleotide Polymorphisms Using an ON–OFF Switching of Regenerated Biosensor Based on a Locked Nucleic Acid-Integrated and Toehold-Mediated Strand Displacement Reaction. *Analytical chemistry* **2014,** *86* (5), 2543-2548.

28. Zhang, X.; Zhang, J.; Wu, D.; Liu, Z.; Cai, S.; Chen, M.; Zhao, Y.; Li, C.; Yang, H.; Chen, J., Ultraselective electrochemiluminescence biosensor based on locked nucleic acid modified toehold-mediated strand displacement reaction and junction-probe. *Analyst* **2014,** *139* (23), 6109-6112.

29. Wu, F.; Chen, M.; Lan, J.; Xia, Y.; Liu, M.; He, W.; Li, C.; Chen, X.; Chen, J., A universal locked nucleic acid-integrated X-shaped DNA probe design for amplified fluorescence detection of single-nucleotide variant. *Sensors and Actuators B: Chemical* **2017,** *241*, 123-128.

30. Zhang, Y.; Wang, L.; Luo, F.; Qiu, B.; Guo, L.; Weng, Z.; Lin, Z.; Chen, G., An electrochemiluminescence biosensor for Kras mutations based on locked nucleic

acid functionalized DNA walkers and hyperbranched rolling circle amplification. *Chemical Communications* **2017,** *53* (20), 2910-2913.

31. Green, S. J.; Lubrich, D.; Turberfield, A. J., DNA hairpins: fuel for autonomous DNA devices. *Biophysical journal* **2006,** *91* (8), 2966-2975.

32. 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. *J. Compu. Chem.* **2011,** *32* (1), 170-173.

33. You, Y.; Moreira, B. G.; Behlke, M. A.; Owczarzy, R., Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Research* **2006,** *34* (8), e60-e60.

34. Tyagi, S.; Kramer, F. R., Molecular beacons: probes that fluoresce upon hybridization. *Nature biotechnology* **1996,** *14* (3), 303-308.

35. Dirks, R. M.; Pierce, N. A., Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **2004,** *101* (43), 15275-15278.

36. Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A., Programming biomolecular self-assembly pathways. *Nature* **2008,** *451* (7176), 318-322.

37. Li, B.; Ellington, A. D.; Chen, X., Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic Acids Res.* **2011,** *39* (16), e110.

38. Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D., Stacking nonenzymatic circuits for high signal gain. *Proc. Natl Acad. Sci. USA* **2013,** *110* (14), 5386-5391.

39. 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.

40. Kotani, S.; Hughes, W. L., Multi-Arm Junctions for Dynamic DNA Nanotechnology. *Journal of the American Chemical Society* **2017,** *139* (18), 6363-6368.

CHAPTER THREE: MULTI-ARM JUNCTIONS FOR DYNAMIC DNA NANOTECHNOLOGY

This chapter was published by American Chemical Society in the journal *Journal of American Chemical Society* and should be referenced appropriately as below.

Shohei Kotani and William L Hughes. "Multi-Arm Junctions for Dynamic DNA Nanotechnology." *J. Am. Chem. Soc.*, 2017, *139* (18), 6363–6368

Boise State University, Micron School of Materials Science and Engineering
1910 University Dr., Boise, Idaho 83725 USA

Abstract

Nonenzymatic catalytic substrates have been engineered using toehold-mediated DNA strand displacement, and their programmable applications range from medical diagnosis to molecular computation. However, the complexity, stability, scalability, and sensitivity of those systems are plagued by network leakage. A novel way to suppress leakage is to increase its energy barrier through four-way branch migration. Presented here, we designed multi-arm junction substrates that simultaneously exploit four-way branch migration, with a high-energy barrier to minimize leakage, and three-way branch migration, with a low-energy barrier to maximize catalysis. Original feed forward, autocatalytic, and cross-catalytic systems have been designed with polynomial and exponential amplification that exhibit the modularity of linear substrates and the stability of hairpin substrates, creating a new phase space for synthetic biologist, biotechnologist,

and DNA nanotechnologists to explore. A key insight is that high-performing circuits can be engineered in the absence of intensive purification and/or extensive rounds of design optimization. Without adopting established leakage suppression techniques, the ratio of the catalytic rate constant to the leakage rate constant is more than 2 orders of magnitude greater than state-of-the-art linear and hairpin substrates. Our results demonstrate that multi-arm junctions have great potential to become central building blocks in dynamic DNA nanotechnology.

3.1 Introduction

Toehold-mediated strand displacement[1] accelerates DNA invasion reactions through branch migration using short single-stranded sequences, called toeholds. Owing to the predictable Watson−Crick binding of DNA,[2] this elegant concept enables the control of reaction kinetics[3,4] and is fundamental to the construction of dynamic DNA systems.[5] One of the key components among them is nonenzymatic catalytic substrates,[6,7] which are used for biomarker sensing,[8] molecular computation,[9-11] and triggered self-assembly.[12,13] Two common catalytic substrates are hairpin substrates[12,14-16] and linear substrates,[10,11,17,18] both of which exist in metastable states prior to being triggered. Once triggered by an external catalyst that reduces the energy barrier, a conformational change proceeds from a metastable state to a lower energy state.

The greatest challenge for catalytic substrates is initiation of the reaction in the absence of a catalyst, known as leakage, which limits the engineering of more complex, scalable, and sensitive tools. Total leakage occurs because of an insufficient energy barrier between the metastable state and equilibrium and is classified as either initial leakage or asymptotic leakage.[19] To address total leakage, select studies have introduced:

(1) highly purified DNA strands,[19] (2) the concept of availability and mutual availability,[20] (3) DNA clamps,[7,9,10,12] (4) DNA mismatches,[20,21] and (5) locked nucleic acids (LNAs).[22] Although these methods are effective at reducing leakage caused by breathing[23] and/or defective DNA,[24] they are repair kits for leaky substrates. It is more desirable to have a substrate whose intrinsic leakage is small in the presence of breathing and defective strands. Hence, a key insight here is the design of substrates that have an intrinsically higher energy barrier for the leakage pathway and a lower energy barrier for the catalytic pathway. Considering that all strand displacement, including leakage, proceeds through branch migration, we engineered energy barriers based on the branch migration process.

There are two classes of branch migration: three-way branch migration[25] and four-way branch migration.[26] Three-way branch migration occurs when a single-stranded oligonucleotide displaces a second strand in a duplex along a complementary region (Figure 3.1A). Four-way branch migration occurs when two duplexes exchange their hybridized strands along a complementary region (Figure 3.1B). An important difference between the two methods is the higher energy barrier for four-way branch migration versus three-way branch migration. This energy difference can be seen in their intuitive energy landscape (IEL)[27] and results in 2–5 orders of magnitude slower reaction rates.[28] The higher energy barrier of four-way branch migration will be due to the larger "sawtooth amplitude"[27] of the branch migration steps, shown as the large difference between ΔG_{s3} (~5 kcal/mol) and ΔG_{s4} (~10 kcal/mol) (Figure 3.1A, B). Therefore, incorporating four-way branch migration, for leakage pathways, and three-way branch migration, for catalytic pathways, offers the potential for extremely low leakage rates and

fast catalytic rates. To implement this strategy, the entropy-driven system by Zhang *et al.* (Figure 3.1C) was considered.[17] This system uses a linear substrate, which is shown as a three-stranded DNA complex. Because of the condition for a single-stranded DNA to be the fuel strand, the leakage of the system occurs through three-way branch migration, resulting in a lower energy barrier for the leakage reaction. However, it is possible to make entropy-driven systems without a single-stranded DNA strand. The solution is the adoption of multi-arm junctions[29] as catalytic substrates (Figure 3.1D). The upper panel uses three-arm junction substrates, and the lower panel uses four-arm junction substrates. In both cases, the catalytic reactions are driven by entropy gains, similar to linear substrates, while the leakage reactions proceed through four-way branch migration.

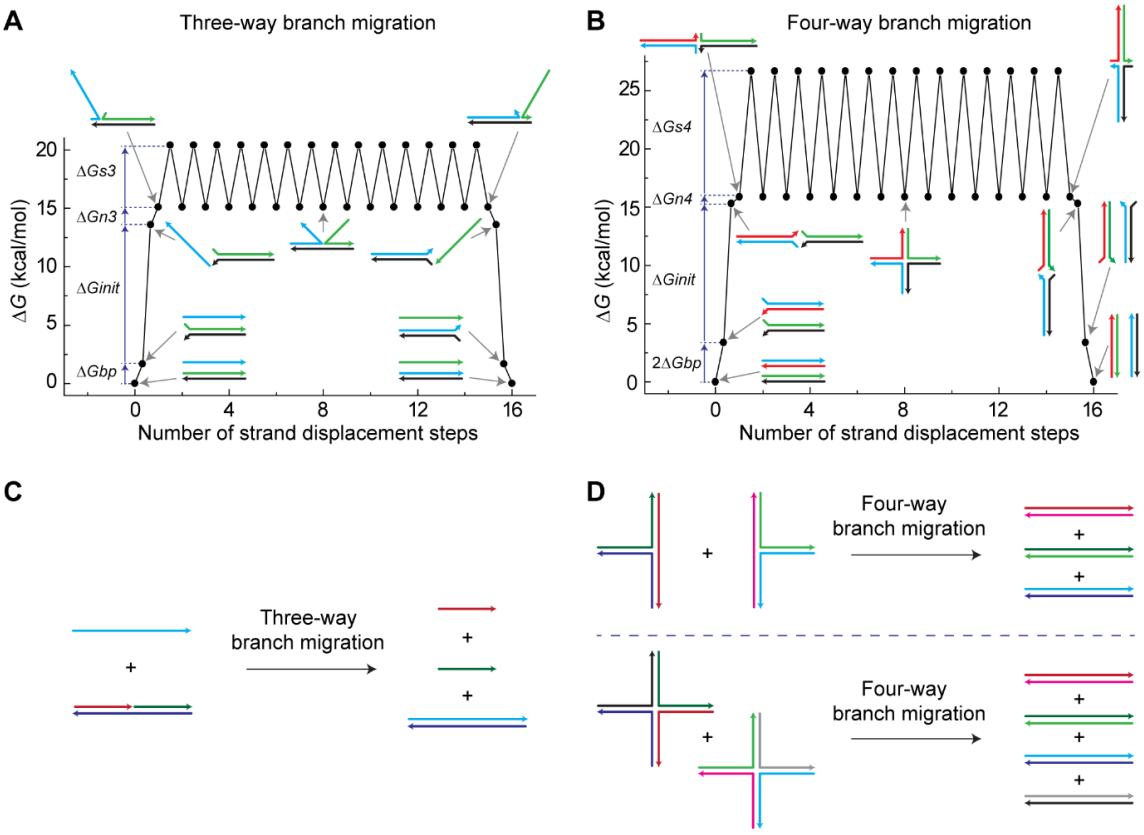

Figure 3.1. Design strategy of multi-arm junction substrates. DNA strands are shown as colored lines with arrowheads representing 3' ends. (A) The IEL of a leakage through three-way branch migration, which was adapted from Srinivas *et al*.[27] The first step is the free energy cost of breaking a base pair, Δ*Gbp* = 1.7 kcal/mol. The second step is the free energy cost of an initial binding, Δ*Ginit* = 11.9 kcal/mol. The third step is the free energy cost of a nucleation, Δ*Gn3* = 1.5 kcal/mol, which is the sum of a base pair gain, –Δ*Gbp* = –1.7 kcal/mol, and the cost of introducing two single-stranded DNA overhangs at an overhang-free nick, Δ*G2ov* = 3.2 kcal/mol. The final step is the "sawtooth amplitude" of three-way branch migration, Δ*Gs3* = 5.3 kcal/mol. Δ*Gbp*, Δ*Ginit*, Δ*G2ov*, and Δ*Gs3* were taken from Srinivas *et al*.[27] (B) The IEL of a leakage through four-way branch migration. The first step requires to break two base pairs (2Δ*Gbp*), and the second step is the initial binding (Δ*Ginit*). The third step is the free energy cost of a nucleation, Δ*Gn4* = 0.6 kcal/mol, which is the sum of two base pairs gain, –2Δ*Gbp* = –3.4 kcal/mol, and the cost of introducing a four-arm junction Δ*G4aj* = 4 kcal/mol.[30] The final step is the sawtooth amplitude of four-way branch migration, Δ*Gs4* = 10.8 kcal/mol, which was calculated based on the step time difference between three-way branch migration[4] and four-way branch migration.[31] (C) The original entropy-driven system based on a linear substrate.[17] Toeholds and single-stranded DNA tails for output formation are removed for simplicity. Leakage reaction occurs through three-way branch migration. (D) The novel entropy-driven system is based on three-arm junction substrates (upper panel) and four-arm junction substrates (lower panel). Leakage reactions occur through four-way branch migration for both cases

3.2 Results and Discussion

3.2.1 Single-Layer Catalytic System with Three-Arm Junction Substrates

We first tested a single-layer catalytic system using a three-arm junction substrate (Figure 3.2A). In this system, the substrate S1 reacts with the catalyst C1 to produce the intermediate I1. Then, I1 reacts with the second substrate S2 to release the catalyst C1. Since both of the reactions proceed through three-way branch migration, the catalytic reaction was expected to be fast. In comparison, the leakage reaction proceeds through four-way branch migration, which was expected to be extremely slow (see Figure 3.5 for the overview, and Figure 3.6-3.12 for details of the catalytic and the leakage pathways). The product P2 reacts with the reporter complex R and displaces the dye strand D from the quencher, which monitors the reaction kinetics via fluorescence emission (Figure

3.2B). Kinetic experiments with 10 nM of substrate showed no detectable leakage in the absence of the catalyst C1 (0 pM), and addition of 500 pM or 1 nM catalyst showed quick catalytic reactions (Figure 3.2C). The catalytic rate constant (k_{cat}) was measured from the 500 pM catalytic reaction and was 2.84×10^{13} $M^{-2}s^{-1}$ (Figure 3.13A), which is similar to a linear substrate ($k_{cat} = 2.86 \times 10^{13}$ $M^{-2}s^{-1}$)[20] and faster than a catalytic hairpin assembly (CHA) system ($k_{cat} = 4.72 \times 10^{11}$ $M^{-2}s^{-1}$).[19] The 50 pM catalytic reaction showed that the reaction nearly stalled at 10 h, yielding catalytic turnover of ~30. This value is less than a linear substrate, which showed catalytic turnover between 80 and 100 at 24 h.[32] Catalytic turnover is known to be very sensitive to the quality of DNA strands[19,32] and can be readily improved by using better purified DNA. We also observed the saturation of the catalytic speed for higher catalyst concentrations (Figure 3.13B) due to the involved unimolecular four-way branch migration where the intermediate I2 is converted into products P2 and P3 (Figure 3.8). The influence of this unimolecular reaction becomes stronger as the catalyst concentration becomes higher and hence the 1 nM catalytic reaction showed worse fitting than the 500 pM catalytic reaction (Figure 3.13C). The numerical integration steps in Section 3.7.3 of the Supporting Information reproduced the saturation behavior at high concentrations of catalyst, as well as lower concentrations of catalyst such as 1 nM and 500 pM (Figure 3.17A).

Next, kinetic experiments were performed with higher concentration of substrates to quantify the leakage and other background reactions (Figure 3.2D). Other than the leakage (R + S1 + S2), the reaction (R + S1) also showed an observable fluorescence increase. In addition, we observed that the maximum fluorescence intensity with 500 nM of substrate was ~30% less than the value expected from 10 nM of substrate, which

would be due to a weak reaction between the reporter waste Rw and the released dye strand D (Figure 3.14A, B). Therefore, the leakage rate constant, k_{leak}, was calculated after subtraction of the background (R + S1), following fluorescence intensity normalization in order to consider the nonlinearity between the fluorescence intensity and the concentration (Figure 3.14C). The leakage rate constant was 2.20×10^{-2} $M^{-1}S^{-1}$, which is in good agreement with the rate constant for four-way branch migration with a zero-toehold (3.4×10^{-2} $M^{-1}s^{-1}$),[28] and is more than 2 orders of magnitude smaller than a linear substrate (8.12 $M^{-1}s^{-1}$).[20] As a metric to compare the performance of different catalytic substrates, the ratio k_{cat} / k_{leak} was summarized in Table 3.1. The data show that the three-arm junction substrate has more than 2 orders of magnitude larger k_{cat} / k_{leak} than other high-performing substrates. Note that the hairpin system compared in Table 3.1 was optimized and operated at higher temperature and lower salinity than the other systems, which will affect the reaction rates. We did not compare other factors, such as maximum catalytic turnover and the initial leakage, because these factors are highly influenced by the quality of DNA strands.[19] In addition, even in the absence of a leakage between S1 and S2, single-stranded tails on those substrates (domains d1s-T2 of S1 and domains d2-3 of S2) may cooperatively displace the D strand from R and produce a fluorescence signal. In order to estimate this background reaction, R was mixed with two single-stranded DNA strands, one with domains d1s-T2 and the other with domains d2-3, and we observed a significant fluorescence signal (Figure 3.14D, E). Therefore, the actual leakage rate is likely smaller than the calculated value above. We also designed the same system with specificity domains shortened from 22 nt to 16 nt, and it showed similar performance (Figure 3.18).

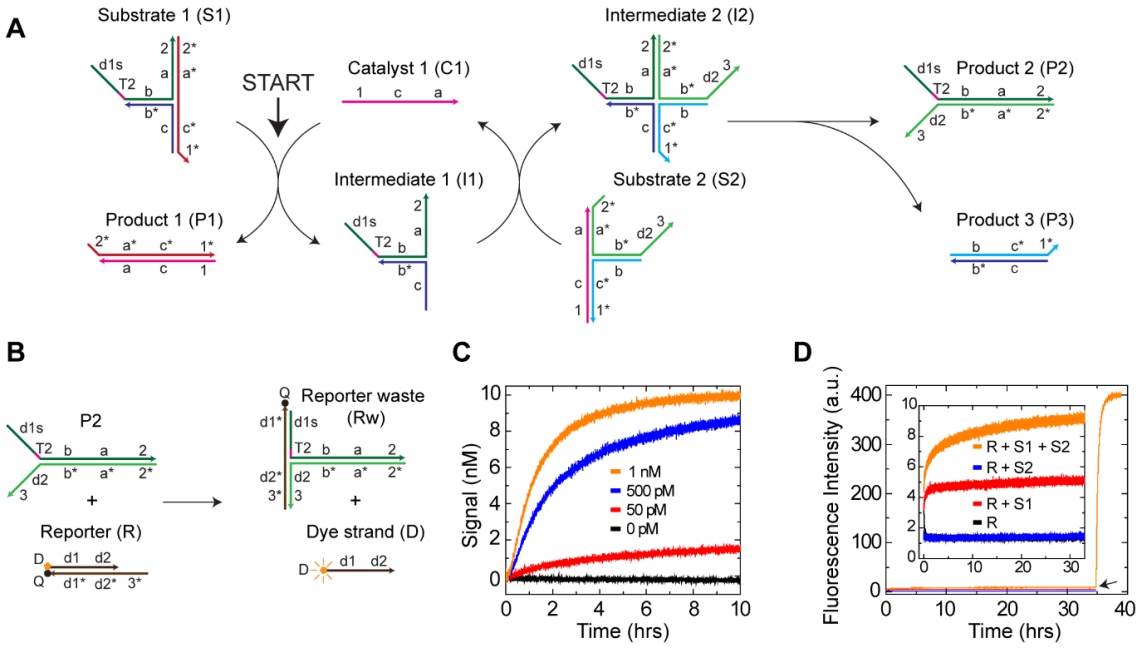

Figure 3.2. Single-layer catalytic system with three-arm junction substrates. (A) A schematic of the catalytic pathway. Functionalities of DNA sequences are represented by domains, which are unique segments of continuous oligonucleotides. Asterisks represent complementary domains, domains with toeholds are represented by numbers, and specificity domains are letters. Domain T2 of S1 represents 2 nt thymidine. (B) Reporting reaction. The reporter complex R has both the dye (D) and the quencher (Q), resulting in quenched fluorescence. The reaction between P2 and R releases the dye strand D and increases fluorescence intensity. (C) Kinetic traces with different concentrations of the catalyst C1. Fluorescence intensity was normalized so that 10 nM corresponds to the maximum fluorescence intensity and 0 nM corresponds to the initial intensity. [S1] = [S2] = 10 nM, [R] = 20 nM. (D) Kinetic traces of the leakage and background reactions. [S1] = [S2] = 500 nM, [R] = 700 nM. The leakage trace is R + S1 + S2, and other traces were performed to measure the background signals. Fluorescence intensity was not normalized. The black arrow shows the addition of catalyst C1 to obtain the maximum fluorescence intensity

Table 3.1 Ratio of the Catalytic to Leakage Rates Constants (k_{cat}/k_{leak}) for Different Catalytic Substrates

Substrate	k_{cat} ($M^{-2}s^{-1}$)	k_{leak} ($M^{-1}s^{-1}$)	k_{cat}/k_{leak} (M^{-1})
Linear [a]	2.86×10^{13}	8.12	3.52×10^{12}
Hairpin [b]	4.72×10^{11}	2.33×10^{-1}	2.02×10^{12}
Three-arm [c]	2.84×10^{13}	2.20×10^{-2}	1.29×10^{15}
Four-arm [c]	2.28×10^{13}	2.11×10^{-2}	1.08×10^{15}

[a] Entropy-driven system at 25 °C with 12.5 mM $MgCl_2$ and 1 mM EDTA.[20] [b] Enzymatically synthesized CHA at 37 °C with 140 mM NaCl, 5 mM KCl, and 0.5 mM EDTA.[19] [c] This work was performed at 25 °C with 12.5 mM $MgCl_2$ and 1 mM EDTA.

3.2.2 Two-Layer Feed Forward Catalytic System with Three-Arm Junction Substrates

Signal cascaded systems for higher signal amplification or molecular computation require smooth connection between multiple layers using outputs of one layer as inputs into other layers. An associative toeholds,[33] which is applied to our design, was originally used for hairpin substrates, and its reaction speed was 2 orders of magnitude slower than a single-stranded invader with a 8 nt toehold. Although a longer toehold was demonstrated to speed up the reaction, it is not suitable for a toehold exchange reaction[4] because a long toehold inhibits its spontaneous dissociation. In contrast with a hairpin substrate, after the investigation of the design parameters (Figure 3.19A-D), a three-arm junction substrate only showed approximately three times the slowdown of a 6 nt toehold (Figure 3.19E, F).

Based on these results, we constructed a two-layer feed forward catalytic system (Figure 3.3A). Kinetic data showed extremely slow leakage, resulting in the large signal-to-noise ratio for the catalyst detection (Figure 3.3B). It will be possible to make more sensitive feed forward systems by using higher quality DNA strands to gain better catalytic turnover,[19,32] or by connecting more layers. In addition, a cross-catalytic system was also constructed by feeding back the signal from the second to the first layer (Figure 3.22). The leakage source was investigated for both catalytic systems by conducting a background check, and the results indicated that the initial leakage was the primary source of leakage (Figure 3.23). We also constructed a feed forward system and an autocatalytic system based on 16 nt specificity domains (Figure 3.24), although they showed larger leakage than the 22 nt designs presented here.

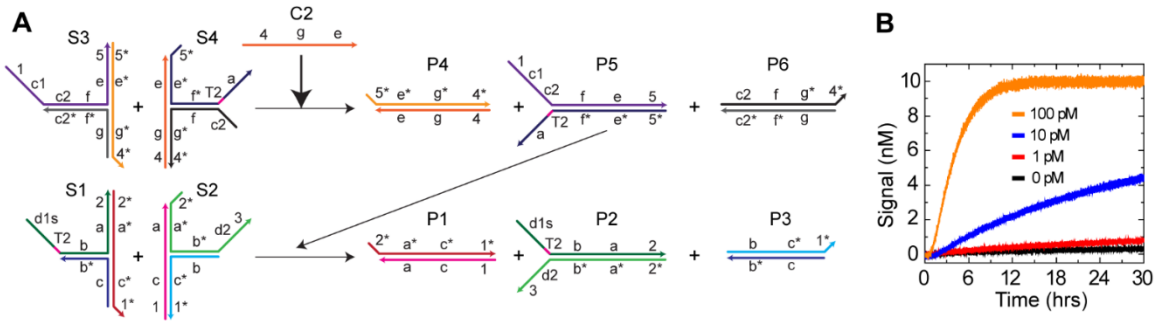

Figure 3.3. Two-layer feed forward catalytic system with three-arm junction substrates. (A) A simplified schematic of the catalytic pathway (see Figure 3.20 for details). The substrates S3 and S4 in the first layer produce the product P5, whose single-stranded sequence acts as the input catalyst for the second layer. The substrates in the second layer (S1, S2) are the same as the single-layer catalytic system (Figure. 3.2) with the same reporting system. Domain c = domains c1 + c2. (B) Kinetic traces with different concentrations of the catalyst C2. [S1] = [S2] = [S3] = [S4] = 10 nM, and [R] = 20 nM

3.2.3 Autocatalytic System with Four-Arm Junction Substrates

The design principle of a three-arm junction substrate can be generalized and extended to other multi-arm junctions. Here, we used a four-arm junction. The reaction converts two substrates S5 and S6 into four products P1, P2, P8, and P9, and the new combinations of domains on each product can trigger downstream reactions (Figure 3.25). In order to characterize the catalytic and leakage reaction, a single-layer catalytic system was initially designed (Figure 3.26). The result showed a very slow catalytic reaction when all of the toeholds were 6 nt long (Figure 3.26B). Therefore, a suitable toehold design was examined (Figure 3.26C-F), together with the connection between multiple layers (Figure 3.19G, H). Based on the results, an autocatalytic system was constructed with an extended toehold for both toehold domains 1 and 2 (Figure 3.4A). The reaction kinetics for the leakage showed good stability without the initiation of clear leakage up to 1.5 h (Figure 3.4B), resulting in a better detection limit. Owing to the

limited catalytic turnover of catalytic systems that are based on DNA strand displacement,[32] autocatalytic or other exponential amplification systems, whose amplification is not limited by maximum catalytic turnover, can be a reasonable candidate for biomarker sensor applications. Because of the very small asymptotic leakage and the quick catalytic reaction (Table 3.1, Figure 3.27A-D), the reduction of the initial leakage is critical for the improvement of the sensitivity. Note that, although three-arm and four-arm junction structures have appeared in dynamic DNA systems, using hairpin substrates[12] or associative toeholds[34] for instance, those multi-arm junction structures exist as products without further structural changes. On the contrary, our work demonstrates that the energy stored in multi-arm junction structures can be catalytically released when coupled with complementary multi-arm junctions. Beyond their application in structural DNA nanotechnology,[35] this research provides a highly novel perspective of multi-arm junctions as a vital tool for dynamic DNA nanotechnology.

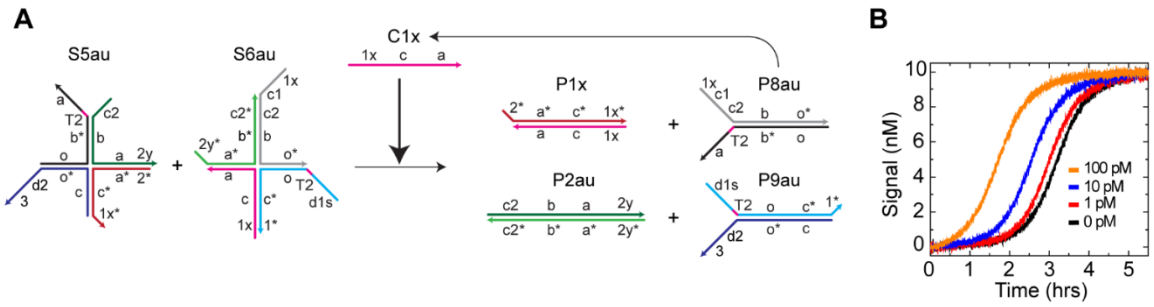

Figure 3.4. Autocatalytic system with four-arm junction substrates. (A) A simplified schematic of the autocatalytic system (see Figure 3.29 for details). P8au has catalytic domains to perform exponential amplification. Two nt were added at the 5' end of domain 1 and the 3' end of domain 2 to generate domains 1x and 2y, respectively. (B) Kinetic traces with different concentrations of the catalyst C1x. [S5au] = [S6au] = 10 nM, [R] = 20 nM

3.3 Conclusions

DNA strand displacement systems are metastable reactions that are triggered by the addition of a specific, single-stranded sequence. Catalytic substrates exploit strand displacement for programmable applications that range from medical diagnosis to molecular computation. However, the complexity, stability, scalability, and sensitivity of said systems are plagued by network leakage – the *Achilles' heel* of dynamic DNA nanotechnology. A novel way to suppress leakage is to increase its energy barrier through four-way branch migration. Multi-arm junction substrates were designed here to simultaneously exploit four-way branch migration (*with a high-energy barrier to minimize leakage*) and three-way branch migration (*with a very low energy barrier to maximize catalysis*). Original feed forward, autocatalytic, and cross-catalytic systems were built with polynomial and exponential amplification that exhibit the modularity of linear substrates and the stability of hairpin substrates. As the stability of hairpin substrates are desired for transcription and used for *in vitro*[36] or *in vivo* applications,[37] and the modularity of linear substrates are desired for circuits,[38] the combined performance of multi-arm-junction substrates creates a new phase space for synthetic biologist, biotechnologists, and DNA nanotechnologist to explore. When compared to state-of-the-art hairpin and linear substrates, our multi-arm junction substrates showed very low leakages without intensive purification of DNA strands,[19] the application of availability and mutual availability,[20] clamps,[7,9,10,12] mismatches,[20,21] or LNAs.[22] They also showed quick catalytic reactions, resulting in more than 2 orders of magnitude larger ratio of the catalytic to leakage reaction rates. It is expected that combinations of our design with other leakage reduction techniques mentioned above, or a proposed novel

domain design,[39] will further improve the performance of our systems. Additionally, the output design of the multi-arm junction substrates showed the modularity to construct a variety of signal cascades. Considering these features, multi-arm junction substrates have great potential to further explore dynamic DNA nanotechnology[5] for the realization of more complex, stable, scalable, and sensitive systems. Future work will focus on optimizing the performance of multi-arm junctions, exploring novel applications for them, and testing our five-arm multi-arm junction substrate (Figure 3.31).

3.4 Author Contributions

S.K. conceived the concept, designed the experiments, performed the experiments, analyzed the data, and wrote the manuscript under the supervision and guidance of W.L.H.

3.5 Acknowledgements

This research was supported by: (1) The W.M. Keck Foundation, (2) NIH Grant No. K25GM093233 from the National Institute of General Medical Sciences, (3) NIH Grand No. P20GM103408 from the INBRE Program of the National Center for Research Resources, and (4) The Micron Foundation. We also thank the Nanoscale Materials & Device Research Group at Boise State for their support. We specifically thank B. Yurke and B. Cannon for their editorial comments and X. Olson for her technical insight.

3.6 Supporting Information

3.6.1 Proposed Reaction Mechanism of a Three-Arm Junction Substrate

The overall reaction converts two molecules of substrates S1 and S2 into three molecules of products P1, P2, and P3 (Figure 3.5A), which provides an entropy gain[17] as the driving force. In addition, imperfect base parings at the three-arm junction point[40] of

S1 and S2 disappear by the reaction, which also provides base paring gains. For the construction of signal cascades, the single-stranded tails on each substrate (domains α and β of S1, domains γ and δ of S2) can be used to generate outputs with new combinations of domains (domains α and δ of P1, domains γ and β of P2) by the associative toehold mechanism.[33]

The key concept of the system is the large difference of the energy barrier between the catalytic and leakage pathways. In the catalytic pathway (Figure 3.5B), S1 first consumes the catalyst C1 to produce P1 and the intermediate I1 (step **I**). Then, I1 reacts with S2 to produce C1 and the intermediate I2 (step **II**). Those reactions proceeds through three-way branch migration, and the Intuitive Energy Landscape (IEL) of those reactions show the absence of significant energy barriers after the initiation of the strand displacement, which will allow quick reactions. Finally, I2 is converted into P2 and P3 through four-way branch migration (step **III**) without a reverse reaction, because the long hybridized region (domains c-a-2) in I2 secures the success of the reaction.

In contrast with the catalytic reaction, the leakage reaction is designed to have a high energy barrier. As an example, when a leakage occurs (Figure 3.5C), the initial step is the hybridization of the domain 1* of S1 and 1 of S2 (The left most state to the second state). This reaction occurs via a zero toehold through three-way branch migration.[41] After the hybridization at domain 1, the system needs to overcome an additional energy barrier to shift into four-way branch migration (the second state to the third state) because of the higher sawtooth amplitude of four-way branch migration. This energy barrier, which is shown as $\Delta E1$, is expected to strongly bias the branch migration backward and slow down the reaction speed by orders of magnitude.

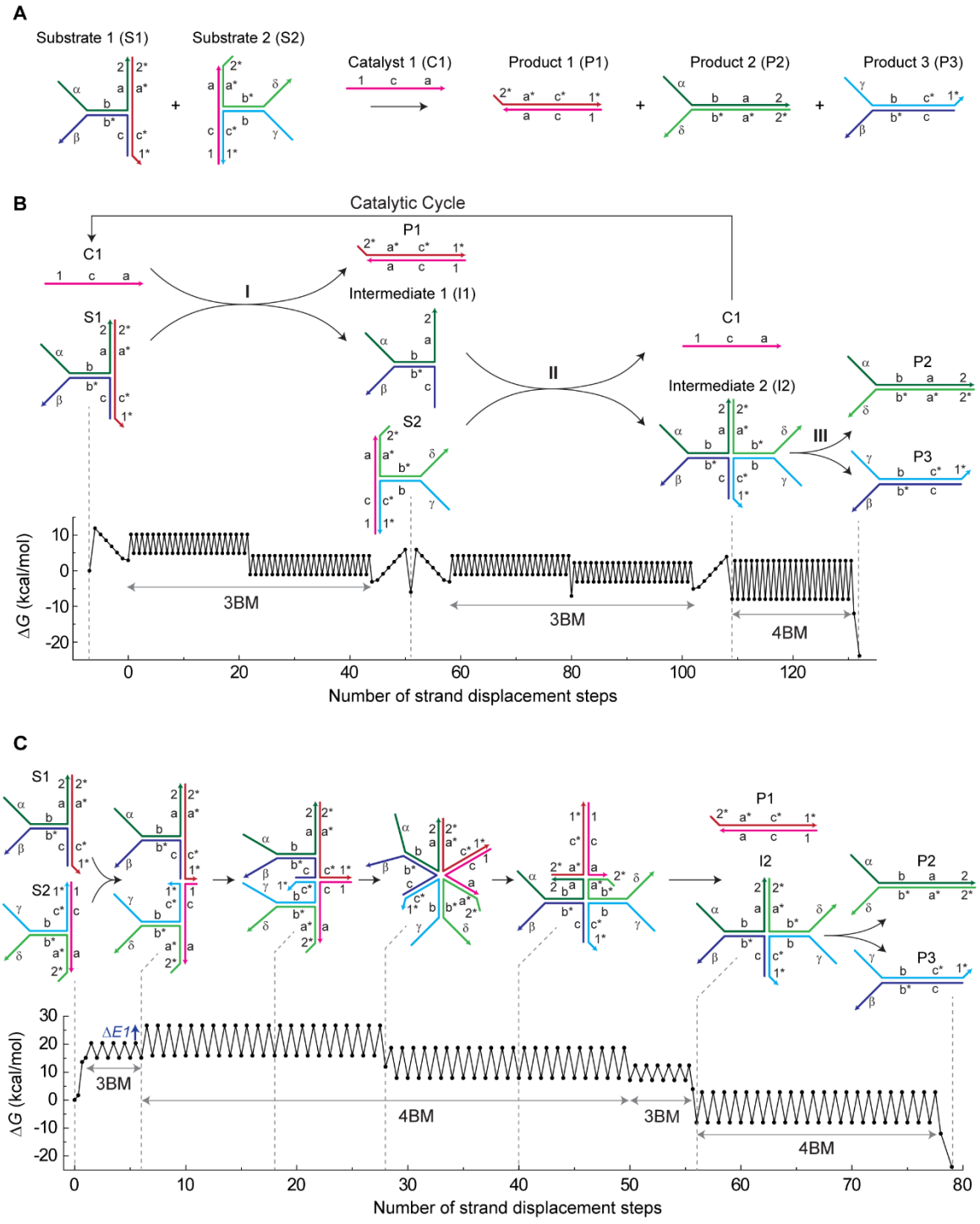

Figure 3.5. Design principles of a three-arm junction substrate for a catalytic system. (A) A general design and its overall reaction of three-arm junction substrates. Functionalities of DNA sequences are represented by domains, which are unique segments of continuous oligonucleotides. Asterisk represents complementary domains. Toehold domains are represented by numbers, specificity domains are letters, and output domains are Greek letters. The reaction converts two substrates

S1 and S2 into three products P1, P2, and P3. Products P2 and P3 have the new combination of output domains α and δ, γ and β, respectively. The sequences of those outputs are independent of the catalyst C1 (domains 1-c-a), therefore can be used for signal cascades. (B) The proposed catalytic reaction pathway. 3BM stands for three-way branch migration, and 4BM stands for four-way branch migration. Step I is the reaction between C1 and S1 through three-way branch migration to produce P1 and the intermediate I1. Step II is the reaction between I1 and S2 through three-way branch migration to produce C1 and the intermediate I2. Step III is the production of P2 and P3 from I2 through four-way branch migration. Also shown is the corresponding Intuitive Energy Landscape (IEL) of each reaction. The details of the IELs are shown in Figure 3.6-3.8. (C) The proposed leakage pathway in the absence of a catalyst. While leakage starting from the hybridization of domain 1 is shown, two other leakage pathways (starting from domain 2 or b) are possible and not shown. Because of the existence of the energy barrier Δ*E1* at the transition from three-way branch migration to four-way branch migration, shown in the IEL, the branch migration will be strongly biased backward. The details of the IEL are shown in Figure 3.9

3.6.2 The Intuitive Energy Landscape (IEL) of Elementary Reactions for Three-Arm Junction Substrates

The values of those energy parameters (Δ*Ginit*, Δ*Gbp*, Δ*G1ov*, Δ*G2ov*, Δ*Gp*, Δ*Gs3*) were from a study on three-way branch migration.[27] Briefly, Δ*Ginit* = 11.9 kcal/mol is the free energy cost to localize two separate molecules together with correct orientation. -Δ*Gbp* = -1.7 kcal/mol is the average value of free energy gain to form a base pair. Δ*G1ov* = 1.2 kcal/mol is the free energy cost of introducing a single-stranded DNA overhang at an overhang-free nick. Δ*G2ov* = 3.2 kcal/mol is the free energy cost of introducing two single-stranded DNA overhangs at an overhang-free nick. Δ*Gp* = 2 kcal/mol, “plateau height”, is the free energy cost of adding a second tail at a junction already possessing one single-stranded DNA tail (thus, Δ*Gp* = Δ*G2ov* - Δ*G1ov*). Δ*Gs3* = 5.3 kcal/mol “sawtooth amplitude”, is the free energy cost for each step of three-way branch migration due to its structural rearrangement.

The values for the structural energy parameters (*ΔG3ajt*, *ΔG3aj*) were calculated as following. *ΔG3ajt* = 5 kcal/mol is the free energy cost of three-arm junction with 2nt thymidine (T2) bulge, assumed the value to be the same as a three-arm junction with 2nt adenine bulge.[42] Δ*G3aj* = 6 kcal/mol is the free energy cost of a three-arm junction without a bulge, deduced from the stabilization effect (-1 kcal/mol) of T2 bulge.[43]

The value of the Δ*Gs4*, sawtooth amplitude of four-way branch migration, was calculated as following. The step time is ~100 μs for three-way branch migration[4] and ~1 s for four-way branch migration.[31] Consequently, their energy barrier difference for each migration step, Δ*G*, can be expected so that $10^4 = \square^{\Delta\square}/\square\square$, resulting in Δ*G* = 5.5 kcal/mol. Therefore, Δ*Gs4* = Δ*G* + Δ*Gs3* = 10.8 kcal/mol. In the study of toehold-mediated strand displacement through four-way branch migration,[28] it was hypothesized that the energy barrier to initiate four-way branch migration originates from the "entropic cost of forming a loop", whose value is ~11 kcal/mol. Although this value was used to explain their kinetic data well enough, it is not clear how to compare this large free energy cost with the small cost of forming a four-arm junction, ~1 kcal/mol, measured at 18 ˚C.[44] The similar value of Δ*Gs4* indicates that the large sawtooth amplitude can be the alternative way to explain their data.

Although the free energy cost of a four-arm junction was studied,[44] the calculated value at 25 °C that was based on the study turned out to be too small. Therefore, we instead used the Δ*G4aj* = 4 kcal/mol,[30] which was calculated using Mfold.[45] The value is in agreement with the study showing more stability of a four-arm junction than a three-arm junction.[46] For a six-arm junction, the predicted free energy cost by Mfold, ~6 kcal/mol, seemed to be too small considering the less stability of a six-

arm junction than a three-arm junction.[47] Therefore, Δ*G6aj* = 12 kcal/mol was used simply as twice of a three-arm junction to be conservative.

For Figure 3.6,3.9-3.12, 0 kcal/mol was set to be the initial free energy of the system. For Figure 3.7, 3.8, the initial free energy was set to be the final value of the previous steps.

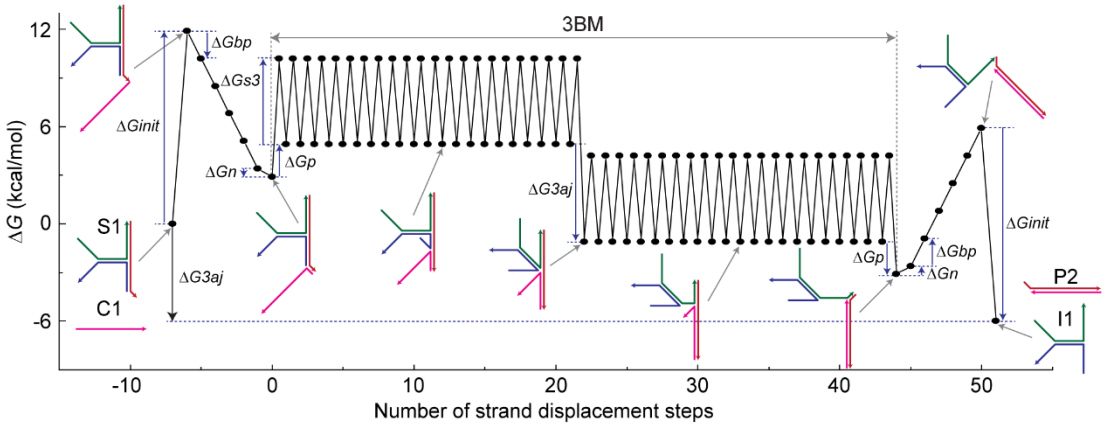

Figure 3.6. The detailed IEL for step I. The thermodynamic driving force of the reaction is the elimination of the three-arm junction point in S1. The process of the toehold dissociation at the end of the reaction is shown as sequential dissociations of base pairs in the toehold domain, 1 bp on each step. However, it was shown that the last several base pairs at the end of a strand displacement can dissociate together spontaneously.[48] Although such alternative pathways are not shown here or for other IELs, it is likely that such an alternative pathway is dominant in toehold exchange reactions. Although a few base pairs at the branch-point of a three-arm junction was shown to be unpaired,[40] it was not shown here and our other IELs. 3BM represents three-way branch migration. Δ*Gn* = 0.5 kcal/mol is the sum of a base pair gain (-Δ*Gbp* = -1.7 kcal/mol) and the cost of introducing a single-stranded DNA overhang at an overhang-free nick (Δ*G1ov* = 1.2 kcal/mol), thus | -Δ*Gbp* + Δ*G1ov* |

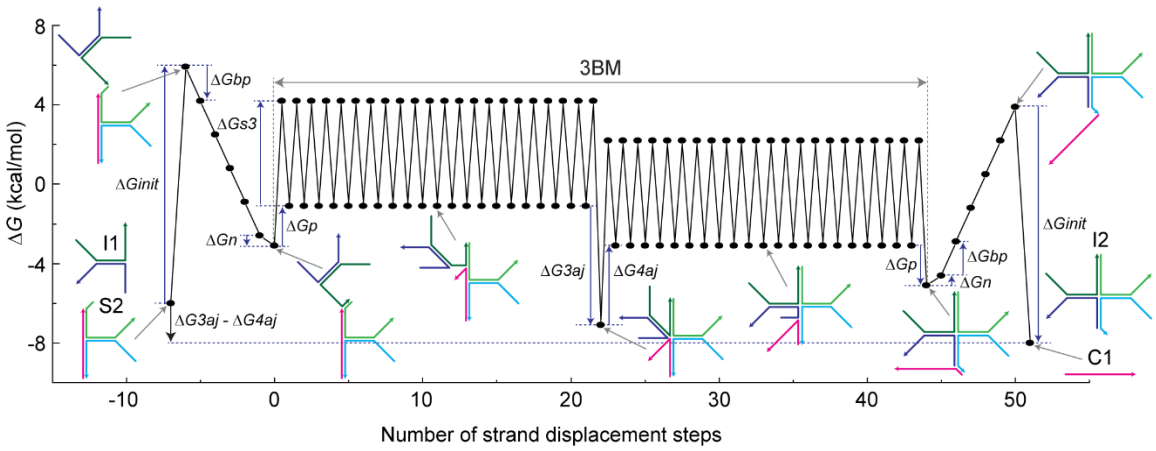

Figure 3.7. The detailed IEL for step II. The thermodynamic driving force for the reaction is the elimination at the three-arm junction point in S2 against the formation of the four-arm junction point in I2

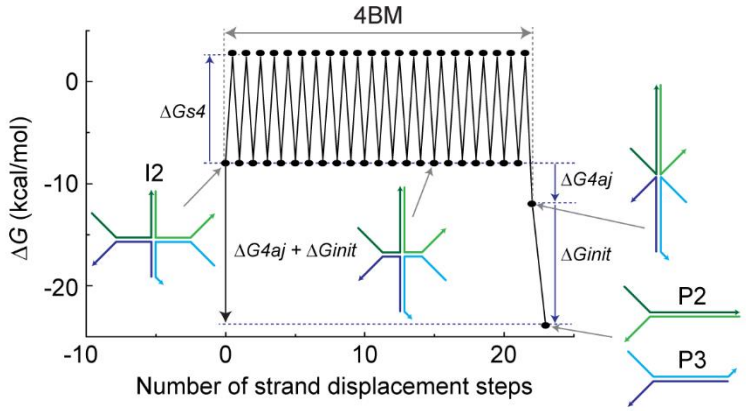

Figure 3.8. The detailed IEL for step III. The thermodynamic driving force is an entropy gain and the elimination of four-arm junction point in I2

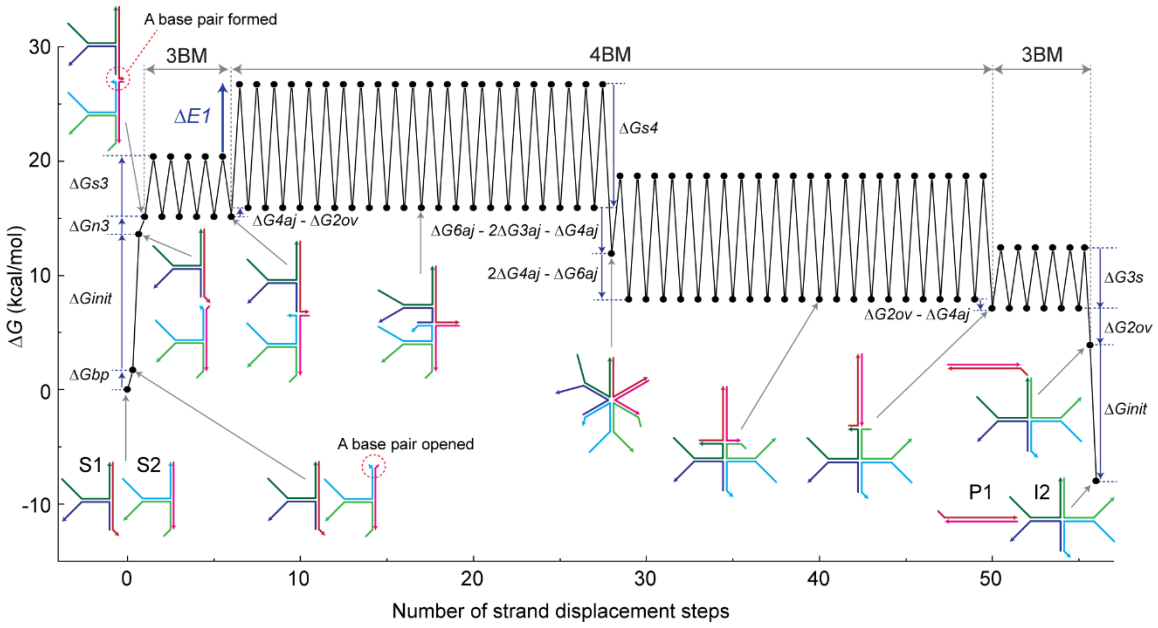

Figure 3.9. The detailed IEL for the leakage reaction. The thermodynamic driving force is the same as the overall catalytic reaction. Because of the existence of the energy barrier Δ*E1*, at the transition between the three-way branch migration and the four-way branch migration, the branch migration is strongly biased backwards. The last step of the leakage reaction, where I2 is converted into P2 and P3 (Figure 3.5C) is exactly the same as the step III (Figure 3.8), thus not shown

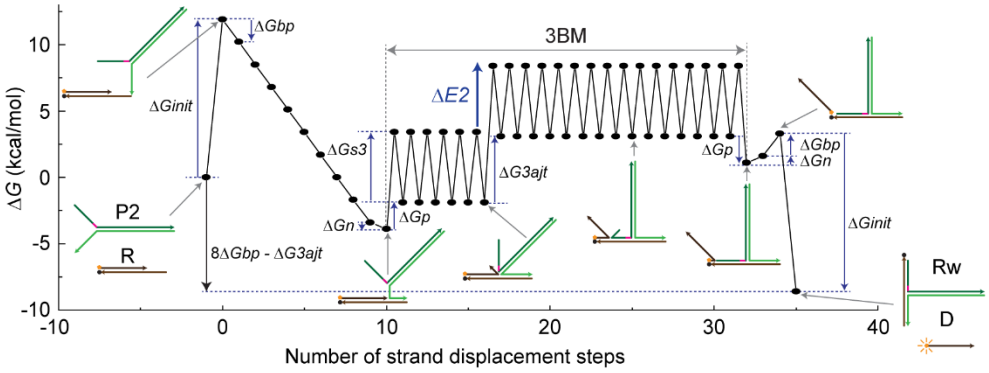

Figure 3.10. The detailed IEL for the reporting reaction. The thermodynamic driving force is the gain of 8 bp against the formation of the three-arm junction point with the T2 bulge in Rw. Because of the existence of the energy barrier Δ*E2* at the transition from no three-arm junction to one three-arm junction, the branch migration will be strongly biased backwards, which requires a longer toehold

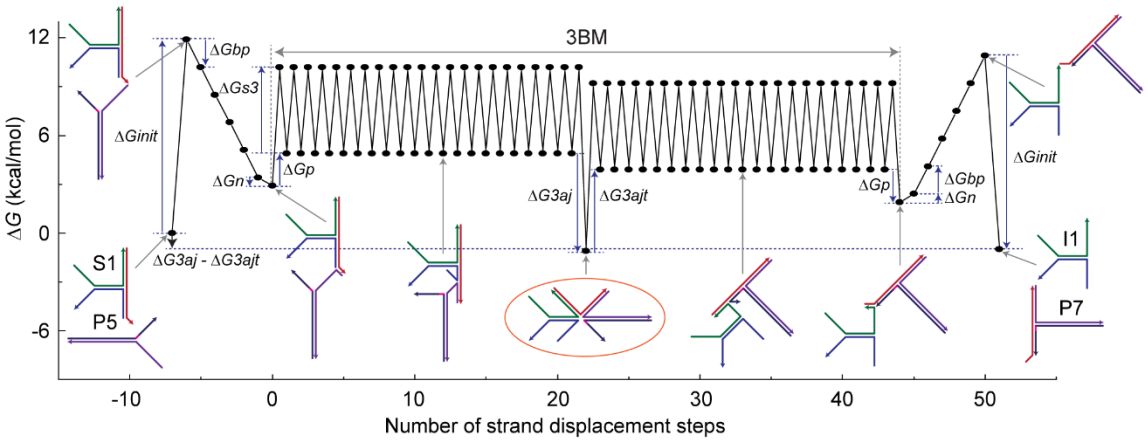

Figure 3.11. The detailed IEL for step I for an output catalyst without a low-energy intermediate. The thermodynamic driving force is the formation of the three-arm junction in P7 with the T2 bulge, by eliminating the three-arm junction point in S1 without a bulge. Note that there exists a low energy reaction intermediate without any junction (circled in orange). Due to the low energy of this intermediate, a deep-well can be seen at the IEL. Thus, it can be expected that the system spends certain time at this intermediate. This might be the reason why the reaction kinetics for P5 was slower than the single-stranded catalyst C1 (Figure 3.19F). Another possibility is the presence of a sawtooth with a higher step height before or after the formation of the stable intermediate

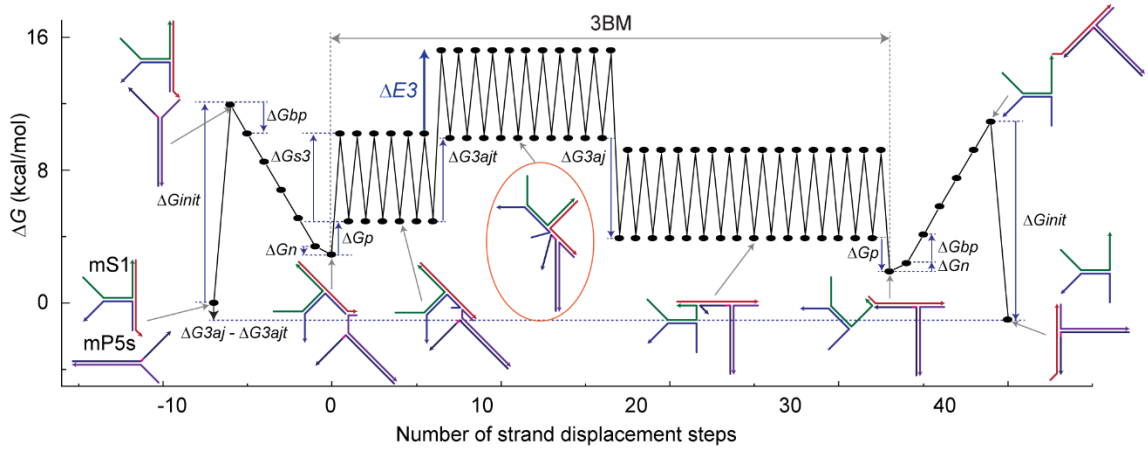

Figure 3.12. The detailed IEL of step I for an output catalyst with a high energy intermediate. The thermodynamic driving force is the same as Figure 3.11. However, in contrast with Figure 3.11, the system will experience the energy barrier Δ*E3* during the shift from one three-arm junction in the system to two three-arm junctions in the system (circled in orange). Because of the existence of this energy barrier, the branch migration will be strongly biased backward. When an associative toehold is applied to a hairpin system, as in the original research,[33] the IEL of the reaction will be similar to this case. In contrast with others IELs, the IEL here was drawn based on the 16 nt specificity domain to agree with the experiment (Figure 3.19B)

3.6.3 Calculation of Rate Constants and Kinetic Simulation for Three-Arm Junction Substrates

The catalytic rate constant k_{cat} – A single-layer catalytic reaction with three-arm junction substrates can be modeled as:

$$\mathrm{S1} + \mathrm{S2} + \mathrm{C1} \xrightarrow{k_{cat}} \mathrm{P1} + \mathrm{P2} + \mathrm{P3} + \mathrm{C1} \tag{3.1}$$

Leakage can initially be ignored because it is orders of magnitude slower than the catalytic reaction. Therefore, the rate equation of P2 can be derived from equation (3.1) as:

$$d[\mathrm{P2}] / dt = k_{cat} [\mathrm{S1}] [\mathrm{S2}] [\mathrm{C1}] \tag{3.2}$$

When the initial condition is $[\mathrm{S1}]_0 = [\mathrm{S2}]_0$, $[\mathrm{C1}]_0$, the mass balance equations are:

$$[\mathrm{S1}] = [\mathrm{S2}] \tag{3.3}$$

$$[\mathrm{P2}] = [\mathrm{S1}]_0 - [\mathrm{S1}] \tag{3.4}$$

$$[\mathrm{C1}] = [\mathrm{C1}]_0 \tag{3.5}$$

Insertion of equations (3.3) through (3.5) into (3.2), and following integration yields:

$$[\mathrm{P2}] = (k_{cat} [\mathrm{S1}]_0^2 [\mathrm{C1}]_0 t) / (1 + k_{cat} [\mathrm{S1}]_0 [\mathrm{C1}]_0 t) \tag{3.6}$$

The delay of the reporting reaction between P2 and R was not considered here, because the reporting reaction is much faster than the catalytic reaction.[49] This allows one to consider the fluorescence signal as a direct measurement of the P2 concentration. Therefore, equation (3.6) was fit to the 500 pM trace (Figure 3.13A), yielding $k_{cat} = 2.84 \cdot 10^{13}$ $M^{-2}s^{-1}$. The inset shows that the initial portion of the reaction does not fit well due to

the initial transient delay. The source of the transient delay occurs during reaction step **III** (Figure 3.5B, 3.8), which is the unimolecular reaction through four-way branch migration. A similar initial transient delay was observed for a kissing loop catalytic substrate,[7] which also has four-way branch migration at the last step of the catalytic reaction. The reaction step **III** eventually becomes the rate-limiting step at higher catalytic concentrations. In fact, the data shows no difference of the reaction speed among 5, 10, and 50 nM catalysts (Figure 3.13B). Also, even at lower catalytic concentrations, the influence of this unimolecular reaction becomes stronger as the catalyst concentration becomes higher and the overall reaction speed approaches its saturated speed. This is a reason for the poor fitting to the 1 nM catalytic reaction (Figure 3.13C). Therefore, k_{cat} = $2.84 \cdot 10^{13}$ $M^{-2}s^{-1}$ should be considered as the rate constant when a bimolecular reaction between a catalyst and a substrate is the rate limiting step of the overall catalytic reaction. The k_{cat} of a linear substrate for the entropy-driven system was adapted from the literature where the same analytical method was used.[49] The k_{cat} of a hairpin substrate was calculated from enzymatically synthesized CHA,[19] where k_{app} was defined as the apparent rate constant including the contribution of the hairpin eA_1. Therefore, $k_{app} = k_{cat}$ $[eA_1]$, where k_{app} = 0.17 $nM^{-1}h^{-1}$ and $[eA_1]$ was assumed to be the initial concentration, 100 nM

.

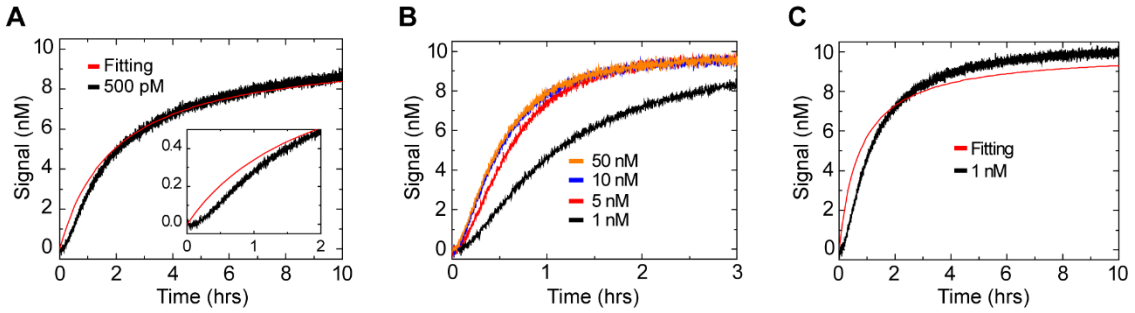

Figure 3.13. Catalytic rate constants for three-arm junction substrates. The fluorescence intensity was normalized so that 10 nM corresponds to the maximum fluorescence intensity and 0 nM corresponds to the initial fluorescence intensity for each kinetic trace. [S1] = [S2] = 10 nM, [R] = 20 nM. (A) The kinetic trace with 500 pM of catalyst C1 was fit to equation (3.6), and $k_{cat} = 2.84 \cdot 10^{13}$ $M^{-2}s^{-1}$ was extracted. The inset shows the deviation of the fitting trace from the kinetic trace due to the initial transient. (B) Kinetic traces with different concentrations of the catalyst C1. (C) The kinetic trace with 1 nM catalyst C1 was fit to equation (3.6). Although $k_{cat} = 3.71 \cdot 10^{13}$ $M^{-2}s^{-1}$ was extracted, the result was not included in Table 3.1 because it yielded a worse fit

The leakage rate constant k_{leak} – A leakage reaction of three-arm junction substrates was modeled as following:

$$S1 + S2 \xrightarrow{k_{leak1}} P1 + P2 + P3 \tag{3.7}$$

The rate equation of P2 can be derived from equation (3.7) as:

$$d[P2] / dt = k_{leak1} [S1] [S2] \tag{3.8}$$

In order to convert a fluorescence intensity into a molecular concentration of the leakage at high concentrations used in Figure 3.2D, it was necessary to consider the influence of the 2 nt deletion at the domain d1s (the 5' ends of the domain d1 was deleted). The 2 nt was initially deleted in order to reduce the background signal from R + S1. However, the deletion leaves 2 nt exposed on the reporting waste Rw, which allows the backward reaction of the reporting reaction by forming a Rw-D complex (Figure 3.14A). The elimination of the three-arm junction in Rw also facilitates the reaction.

NUPACK[50] was used to predict the population of Rw-D complexes, and the population was 0 % at 10 nM (the experimental condition in Figure 3.2C), but increased to 18 % at 500 nM (the experimental condition in Figure 3.2D). Similar to this prediction, although the maximum fluorescence intensity for R + S1 + S2 reaction was 11 a. u. at 10 nM, it was only 399 a. u. at 500 nM (Figure4.14B). The difference between NUPACK prediction and experimental result could be due to the inaccuracy of the energy parameter of the three-arm junction and the presence of the Dye-Quencher pair interaction.[51] Because of this non-linearity between the fluorescence intensity and the concentration, the maximum fluorescence intensity of the leakage at 500 nM substrates concentration was not used to normalize the data. Instead, it can be seen that the fluorescence intensity of the 500 nM leakage reaction before the addition of excess catalysts is the same range of the maximum fluoresce intensity of the 10 nM reaction (Figure 3.14B, inset). Consequently, it is appropriate to convert fluorescence intensities of the leakage reaction into concentrations based on the 10 nM reaction. Therefore, after subtracting the background (R + S1) from the leakage (R + S1 + S2), the leakage at 500 nM substrate concentration was normalized with the final intensity at 10 nM substrate concentration (Figure 3.14C). In order to avoid the influence of the initial leakage, the kinetic trace from 20 to 30 hours was fit to equation (3.8) as the initial slope of the leakage reaction, and $k_{leak1} = 2.20 \cdot 10^{-2}$ $M^{-1}s^{-1}$ was extracted. Note that this time window can be considered as the initial portion of the leakage due to its long time scale. The k_{leak} of a linear substrate for the entropy-driven system was adapted from the literature where the same leakage model was applied.[49] The k_{leak} of a hairpin substrate was calculated from enzymatically synthesized CHA,[19] where k_{Asy} was defined as the apparent rate constant

including the contribution of hairpin eA_1. Therefore, $k_{Asy} = k_{leak}$ [eA_1], where $k_{Asy} = 0.84 \cdot 10^{-4}$ h^{-1} and [eA_1] was assumed to be the initial concentration, 100 nM. In addition to CHA, hairpin chain reaction (HCR)[14] is one of the most widely used catalytic systems based on hairpin substrates. It can be expected that HCR has a smaller leakage rate than CHA considering their leakage pathway differences. Therefore, we cannot conclude that multi-arm junction substrates have better ratio of the catalytic rate constant to the leakage rate constant than HCR.

We also performed an experiment to estimate the reaction between R and the single-stranded output tails on S1 and S2, which does not require the formation of P2 by a leakage in order to increase a fluorescence intensity. For this purpose, two of single-stranded DNAs were designed (Figure 3.14D). ss1 has the same sequence of the output tail on S1 (domains d1s-T2), and ss2 has the same sequence of the output tail on S2 (domains d2-3). As expected from the reaction R + S1 (Figure 3.2D), R + ss1 showed a clear fluorescence increase. In addition, the larger fluorescence increase was observed when R, ss1, and ss2 were mixed together (R + ss1 + ss2), showing that those two single-stranded DNA molecules can cooperatively displace the D strand from R. Importantly, the result indicates that some portion of the observed leakage (Figure 3.2D) might be due to the reaction between R and the single-stranded output tails on the substrates (domains d1s-T2 of S1 and domains d2-3 of S2) with the same mechanism. Thus, the actual leakage modeled in equation (3.7) could be smaller than the calculation above. Consequently, we could not quantify the initial leakage by comparing the fluorescence intensity difference between R + S1 and R + S1 + S2 in Figure 3.2D, because the intensity difference might be due to this background signal. We expect that

this type of leakage will be suppressed easily by using clamps.[7] The stronger fluorescence intensity of R + ss1 + ss2 than R + S1 + S2 would be due to the steric hindrance exerted by the domain b on the substrates S1 and S2.

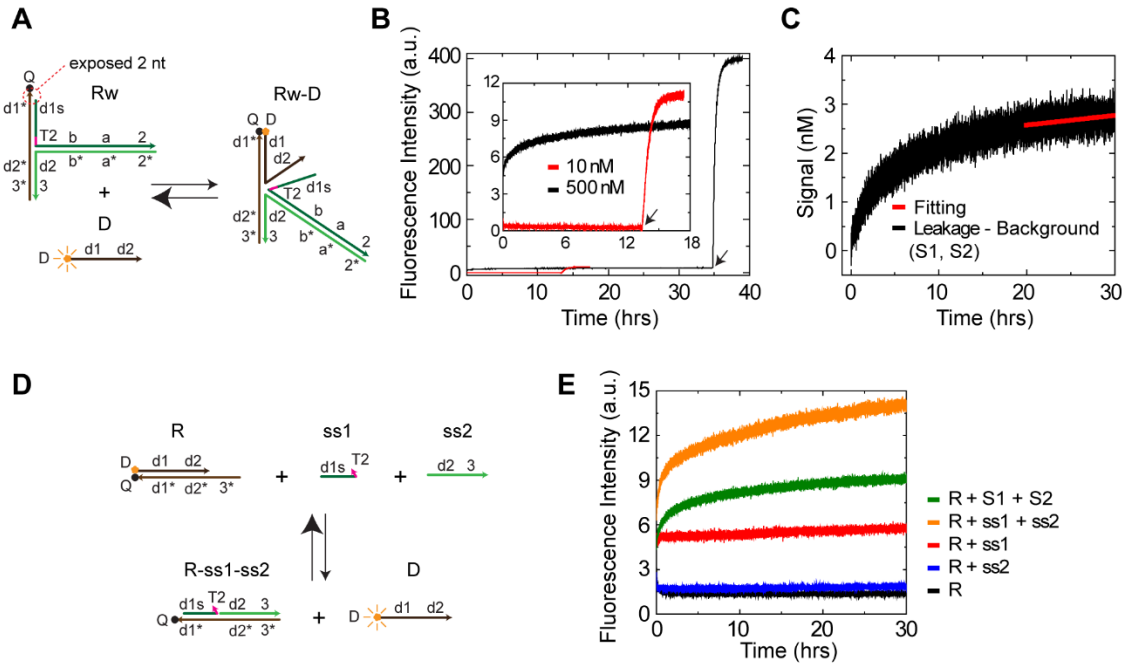

Figure 3.14. The leakage rate constant for three-arm junction substrates. (A) The binding of D with Rw was mediated by the exposed 2 nt at the 3' end of domain d1*. The reaction is negligible under a 10 nM concentration of each species, but was influential at higher concentrations. (B) Non-linearity of the maximum fluorescence intensity when the concentration was high. The maximum fluorescence intensity was 11 a. u. for the10 nM experiment, where [S1] = [S2] = 10 nM, [R] = 20 nM. However, the maximum fluorescence intensity was only 399 a.u. for the 500 nM experiment, where [S1] = [S2] = 500 nM, [R] = 700 nM. Note: the maximum fluorescence intensity would be 550 a.u. if the relationship between the fluorescence intensity and the concentration were linear. The fluorescence data is not normalized. Black arrows show the addition of the catalyst C1 to obtain the maximum fluorescence intensity. (C) The calculation of the leakage rate constant. The background reaction (R + S1) was subtracted from the leakage (R + S1 + S2) from Figure 3.2D at 500 nM, and the initial fluorescence intensity was adjusted to zero. Then, the fluorescence intensity was normalized using the average maximum fluorescence changes of the 10 nM concentration experiments in Figure 3.2C. Considering the slowness of the leakage, the time window before the addition of catalyst can be considered as the initial rate period. Therefore, the slope from 20 to 30 hours was fit to equation (3.8), and k_{leak1} = $2.20 \cdot 10^{-2}$ $M^{-1}s^{-1}$ was extracted. (D) The experimental design to estimate the reaction between R and the single-stranded output tails on S1 and S2. ss1 is the output tail on

S1, and ss2 is the output tail on S2. (E) Kinetic traces of the background reactions. [S1] = [S2] = [ss1] = [ss2] = 500 nM, [R] = 700 nM. The traces of R and R + S1 + S2 are from Figure 3.2D

Kinetic simulation – Equation (3.1) is a simplified catalytic reaction model, which is useful to compare the overall performance of different systems. However, in order to accurately simulate the reaction kinetics, the reaction must be dissected into individual steps based on Figures 3.2A and 3.2B.

$$S1 + S2 \xrightarrow{\square_{\square\square\square\square_1}} P1 + P2 + P3 \tag{3.9}$$

$$C1 + S1 \underset{\square_{-1}}{\overset{\square_1}{\rightleftharpoons}} P1 + I1 \tag{3.10}$$

$$I1 + S2 \underset{\square_{-2}}{\overset{\square_2}{\rightleftharpoons}} I2 + C1 \tag{3.11}$$

$$I2 \xrightarrow{\square_3} P2 + P3 \tag{3.12}$$

$$P2 + R \xrightarrow{\square_4} Rw + D \tag{3.13}$$

Although C1 and P3 can hybridize using toehold domain 1, the reaction was not important in our system if the dissociation rate constant was larger than 10^{-2} s^{-1}, which was used in the literature.[32] The rate constants adopted for our simulations are $k_1 = k_2 = 5 \cdot 10^5$ $M^{-1}s^{-1}$ and also reflect a typical value in the literature for toehold exchange reactions.[4] The backward reaction of equation (3.10) involves the formation of a three-arm junction, whose energy penalty is around 5 kcal/mol. An equivalent energy penalty by a mismatch base pair is known to slow down toehold-mediated strand reactions by 2 to 3 orders of magnitude.[48] Therefore, k_{-1} was assumed to be slow with a value of $1 \cdot 10^3$ $M^{-1}s^{-1}$. During the process of simulation, we realized that the backward reaction of equation (3.11) has to be significantly slower than its forward reaction in

order to reproduce the experimental data. A reason for this is because the strand displacement reaction is difficult between the isoenergetic forms I2 and C1 (Figure 3.15).

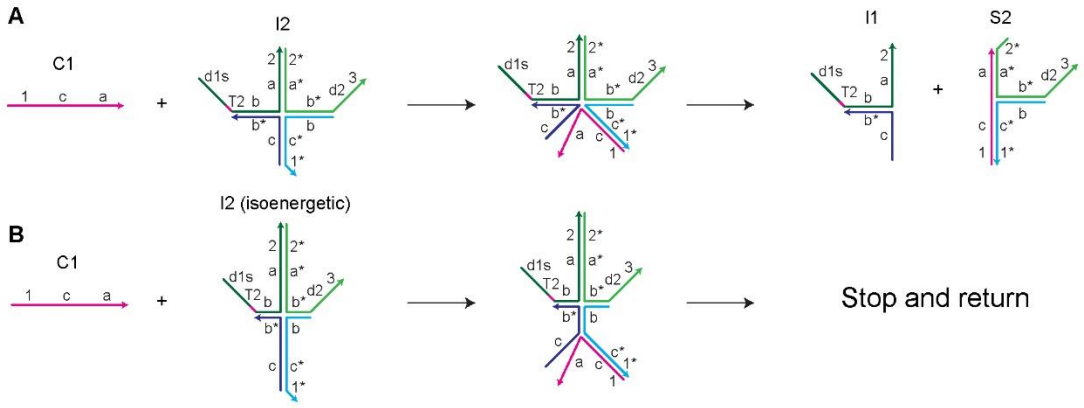

Figure 3.15. C1 cannot complete strand displacement reaction with isoenergetic form of I2. (A) C1 can complete strand displacement with I2 if domains c and a in I2 are connected each other. (B) C1 cannot complete strand displacement with isoenergetic forms of I2 if domain c and a in I2 are separated by domain b. As a result, C1 will continue branch migration backwards and then dissociate from I2

Here, C1 can complete strand displacement with the appropriate form of I2 (Figure 3.15A). In contrast, when C1 tries to perform strand displacement with other isoenergetic forms of I2 (Figure 3.15B), the reaction will be stopped by the domain b of I2, which separates domains a and c. Because four-way branch migration is slow (step time ~ 1s)[31] and three-way branch migration is fast (step time ~ 100 μs)[4], I2 is assumed to be static after C1 initiates strand displacement. Therefore, the available population of I2 for strand displacement (Figure 3.15A) is 1/(length of domain b), which is 1/22. As a result, $k_{-2} = k_2$ /22, which is ~$2 \cdot 10^4$ $M^{-1}s^{-1}$. k_3 in equation (3.12) is the rate constant of four-way branch migration with 22 nt, which is assumed to be $1 \cdot 10^{-3}$ s^{-1} based on the literature from a similar design.[28] k_4 in equation (3.13) was assumed to be $1 \cdot 10^6$ $M^{-1}s^{-1}$, which is a typical value for strand displacement with a long toehold.[17]

In addition to equations (3.9) through (3.13), it was necessary to consider the low catalytic turnover observed. One possible reason for low catalytic turnover is defective DNA strands, which was suggested by Zhang *et al*.[32] The authors reasoned that toehold exchange reactions are sensitive to a few defects in the invader strands. For their entropy-driven system, the Fuel strand was assumed to contain such defects,[32] which corresponds to C1 and I1 for our three-arm junction substrates. Therefore, we hypothesized the existence of defect-substrates here (Figure 3.16).

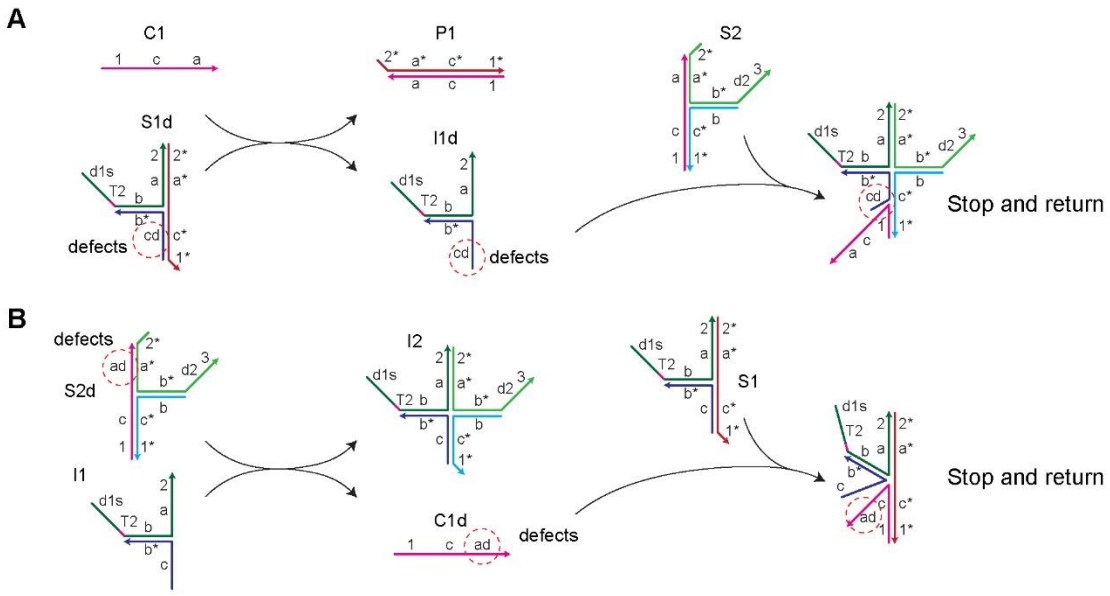

Figure 3.16. Low catalytic turnover caused by defects of DNA strands. (A) Defect-substrate S1d produces defect-intermediate I1d, which is unable to perform strand displacement with S2. (B) Defect-substrate S2d produces defect-catalyst C1d, which is unable to perform strand displacement with S1

For instance, if the defect-substrate S1d has significant sequence defects at domain c (shown as domain cd in Figure 3.16A), the reaction with C1 will produce defect-intermediate I1d. Since I1d also has sequence defects, it is unable to react with substrate S2 to displace a defect-free catalyst C1 (pink strand in S2) for the next round of

reaction. As a result, S1d consumes C1 and poisons the catalytic reaction. Similarly, defect-substrate S2d will consume I1 and produce defect-catalyst C1d, which is unable to perform strand displacement with substrate S1. Although there are multiple scenarios to lower the catalytic turnover, we used this hypothetical model here. Based on Figure 3.16, the following equations can be made.

$$\mathrm{C1} + \mathrm{S1d} \xrightarrow{k_1} \mathrm{P1} + \mathrm{I1d} \tag{3.14}$$

$$\mathrm{I1} + \mathrm{S2d} \xrightarrow{k_2} \mathrm{I2} + \mathrm{C1d} \tag{3.15}$$

In addition to ignoring the reaction between C1 and P3, we also ignored the reactions of I1d with P1 or S2, and the reactions of C1d with I2 or S1, with the assumption that those reactions will only temporary form a complex and then quickly dissociate into their reactants. The forward rate constants for equations (3.14) and (3.15) were assumed to be the same as equations (3.10) and (3.11), and the defect population of S1d and S2d was adjusted to be 2 % each.

Finally, in order to account for the initial leakage with a ~10 hour time scale (Figure 3.14C), we hypothesized the existence of substrates with initial leakage S1i and S2i as below.

$$\mathrm{S1i} + \mathrm{S2i} \xrightarrow{k_i} \mathrm{P1} + \mathrm{P2} + \mathrm{P3} \tag{3.16}$$

The rate constant k_i and population of S1i and S2i were adjusted for experimental data as $k_i = 5 \cdot 10^4\ \mathrm{M^{-1}s^{-1}}$ and 0.46 % each. The simulation result using equations (3.9) through (3.16) is in Figure 3.17A, and agrees well with experimental data and reproduced the observed saturation of reaction speed at high concentration of catalyst. This result supports the proposed reaction model. Figure 3.17B shows the leakage reaction in Figure

3.14C. The agreement is satisfactory, but we would like to note that the observed leakage might be the reaction caused by the reaction between R and the single-stranded output tails on S1 and S2 (Figure 3.14D, E), which is different from the additive of equations (3.9) and (3.16).

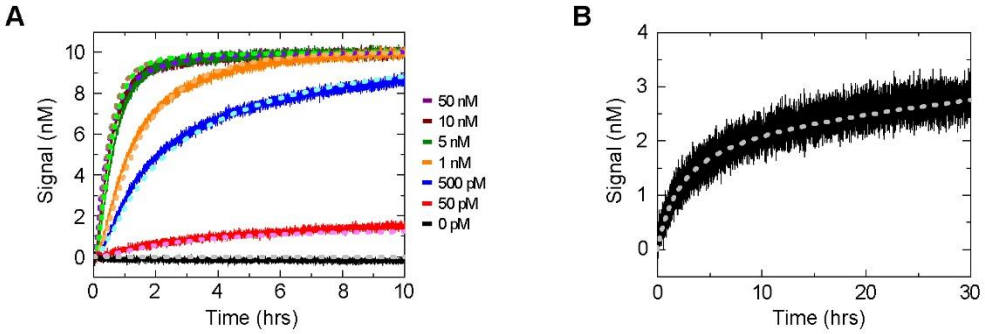

Figure 3.17. Kinetic simulation for single-layer catalytic system with three-arm junction substrate. (A) 10 nM concentration of substrates with different concentrations of catalyst. Kinetic simulations are shown as dots with weak colors. The data is the combination of Figure 3.2C and Figure 3.13B. (B) Leakage reaction at 500 nM concentration of substrates, using data of Figure 3.14C

3.6.4 Single-Layer Catalytic System with Three-Arm Junction Substrates Based on 16nt Specificity Domains

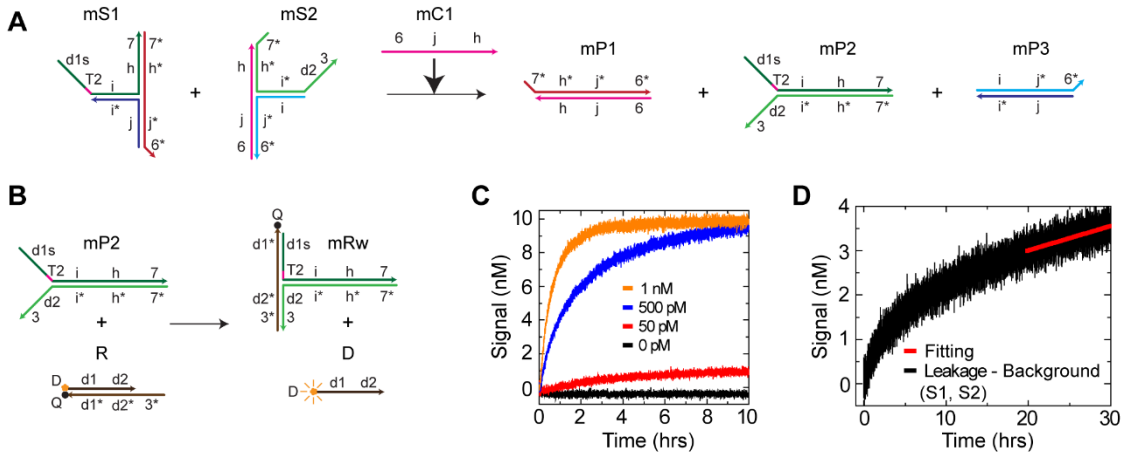

Figure 3.18. Single-layer catalytic system with three-arm junction substrates based on 16 nt specific domains. (A) A simplified schematic of the catalytic pathway. The detailed catalytic pathway is as same as Figure 3.2A with the exception of domains. (B) The reporting reaction. (C) Kinetic traces with different concentrations of the catalyst mC1. The fluorescence intensity was normalized so that 10 nM corresponds to the maximum fluorescence intensity and 0 nM corresponds to the initial intensity. [mS1] = [mS2] = 10 nM, [R] = 20 nM. (D) The calculation of the leakage rate constant. The background reaction (R + mS1) was subtracted from the leakage (R + mS1 + mS2) at 500 nM, and the initial fluorescence intensity was adjusted to zero. Then, the fluorescence intensity was normalized using the average maximum fluorescence changes of the 10 nM concentration experiments in Figure 3.18C. Finally, the slope from 20 to 30 hours was fit to equation (3.8), and $k_{leak} = 6.61 \cdot 10^{-2}$ $M^{-1}s^{-1}$ was extracted

3.6.5 Design Considerations of Output Formation

In order to construct DNA cascade systems, it is necessary for an output from one layer to smoothly trigger downstream layers with a catalyst. In the case of multi-arm junction substrates, outputs are formed via associative toeholds[33] with junction positions being able to be varied, which in turn affects the reaction kinetics. Therefore, two different positions of a junction were tested with three-arm junction substrates based on 16 nt specificity domains (Figure 3.19A). For the first set, the junction position was between domain j1 and j2 (mP5s-t), with a variation having T2 bulge (mP5s). The T2

bulge was used because it was shown to speed up the associate toehold reaction speeds by stabilizing a three-arm junction point.[33] The second set has the junction between domains j2 and h (mP5-t), with a variation having a T2 bulge (mP5). The consequence of the different junction positions can be seen on their reaction kinetics (Figure 3.19B). Comparing with the single-stranded catalyst mC1, the reaction of mP5s-t was significantly slower, even after the addition of the T2 bulge (mP5s). One interesting result is the relative quickness of the reactions when compared to hairpin substrates.[33] This could be due to the difference of the junction position, sequence, or the fundamental difference of the mechanism of strand displacement for hairpins. In comparison, the reaction speed of mP5-t was already similar to mP5s, and the addition of the T2 bulge (mP5) resulted in the same speed as mC1. Because the stoichiometry concentration of the substrates and catalysts (5 nM) were used here, the reaction speeds of mC1 and mP5 were likely limited by the unimolecular reaction step **III**, as shown in Figure 3.13B. Nonetheless, the results showed the best design to be mP5. As such, this design was applied for all other experiments. The reason of the slow reaction speed of mP5s or mP5s-t can be explained using the IEL (Figure 3.12). We have not tested a junction position between domain 6 and j1 for an output catalyst, because it results in the domains j-h to be exposed on a substrate all the time and become the source of leakage.

When the mP5 design is applied, one thing to be considered is that a toehold and its neighboring specificity domains will be exposed together on a substrate initially. As an example, two types of substrates were examined for feed forward catalytic systems (Figure 3.19C). Here, mS3 has the domain j2 covered, but mS3u has the domain j completely exposed (domain j = j1 + j2). NUAPCK[50] predicted that only 12 % of mS1

and mS3 form a complex at 10 nM, because the formation of mS1-mS3 complex needs to form a five-arm junction in order to maintain the maximum number of base pairs. In comparison, 75% of mS1 and mS3u were predicted to form a complex at 10 nM, because the formation of mS1-mS3u complex can eliminate the three-arm junction point in mS1. Those predictions agreed with the PAGE result, which showed that (mS1 + mS3u) has a larger mobility shift than (mS1 + mS3), indicating the formation of a stable mS1-mS3u complex (Figure 3.19D). This type of complex formation can be expected to sequester the toehold domain 6* of the substrate mS1. Therefore, the design of mS3 was applied to all other experiments.

Reaction speed differences between a single-stranded catalyst and a junction catalyst was also tested for three-arm junction substrates with 22 nt specificity domains (Figure 3.19E) and four-arm junction substrates with extended toeholds (Figure 3.19G). Three-arm junctions with 6 nt toehold showed approximately a three times slowness of the overall reaction (estimated from the initial slopes) for the junction catalyst P5 than the single-stranded catalyst C1 (Figure 3.19F). Nonetheless, this reaction is quick enough to construct signal cascade systems based on the toehold exchange reaction.[4] In four-arm junction substrates, the extended 8 nt toehold domain 1 enabled the same overall reaction speed between the single-stranded catalyst C1x and the junction catalyst P8au (Figure 3.19H). Therefore, extended toehold domain 1 can speed up the reaction for three-arm junction substrates. The relative slowness of the reactions in Figure 3.19H, comparing with Figure 3.19F or Figure 3.26E, is likely due to the poor quality of unpurified DNA strands used here.

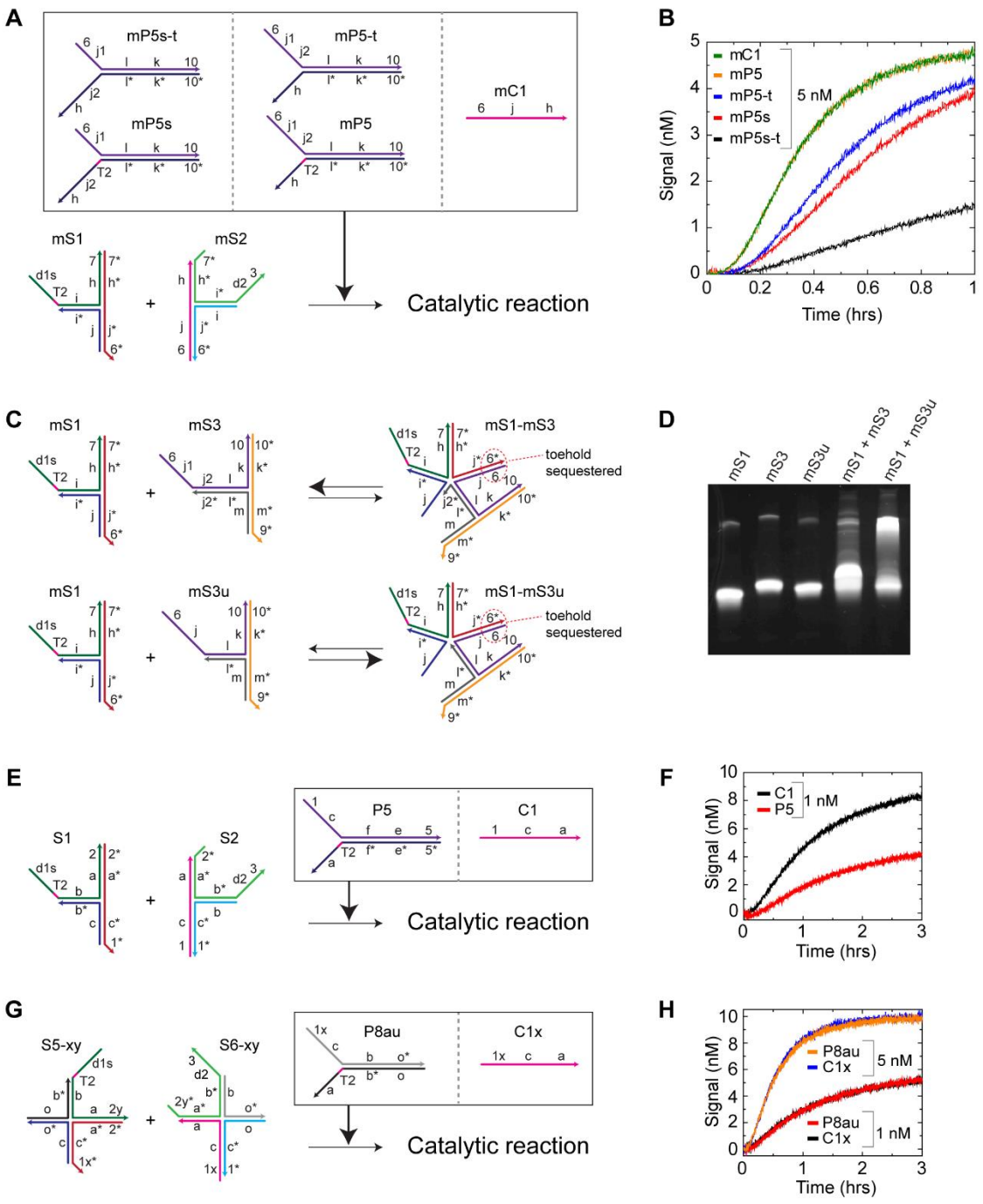

Figure 3.19. Design consideration for output signal formation. (A) Different catalyst designs were tested for single-layer catalytic systems with three-arm junction substrates based on 16 nt specificity domain length. Domain j = domains j1 + j2. The reporting reaction is the same as Figure 3.18B. (B) Kinetic traces with different catalyst species. [mS1] = [mS2] = 5 nM, [R] = 10 nM. Unpurified DNA strands were used here. (C) Comparison of complex formations between mS1-mS3 and mS1-mS3u. While the 6-j domains of mS3u are completely exposed, the j2 of mS3 are covered. (D) 10 % native PAGE analysis of the complex formation. 10 μL of 0.5 μM DNA solution

was used on each lane and stained with dye. mS1 + mS3 showed a small shift during gel migration, indicating that their interaction was weak. In comparison, mS1 + mS3u showed a large shift during gel migration, indicating a stronger interaction. Unpurified DNA strands were used. (E) Different catalyst designs were tested for single-layer catalytic systems with three-arm junction substrates based on 22 nt specificity domain length. (F) Kinetic traces with different catalyst species. [S1] = [S2] = 10 nM, [R] = 20 nM. (G) Different catalyst designs were tested for a single-layer catalytic system with four-arm junction substrates. (H) Kinetic traces with different catalyst species. [S5xy] = [S6xy] = 10 nM, [R] = 20 nM. Unpurified DNA strands were used here

3.6.6 Detailed Reaction Schematic and Kinetic Simulation of Two-Layer Feed Forward Catalytic System with Three-Arm Junction Substrates

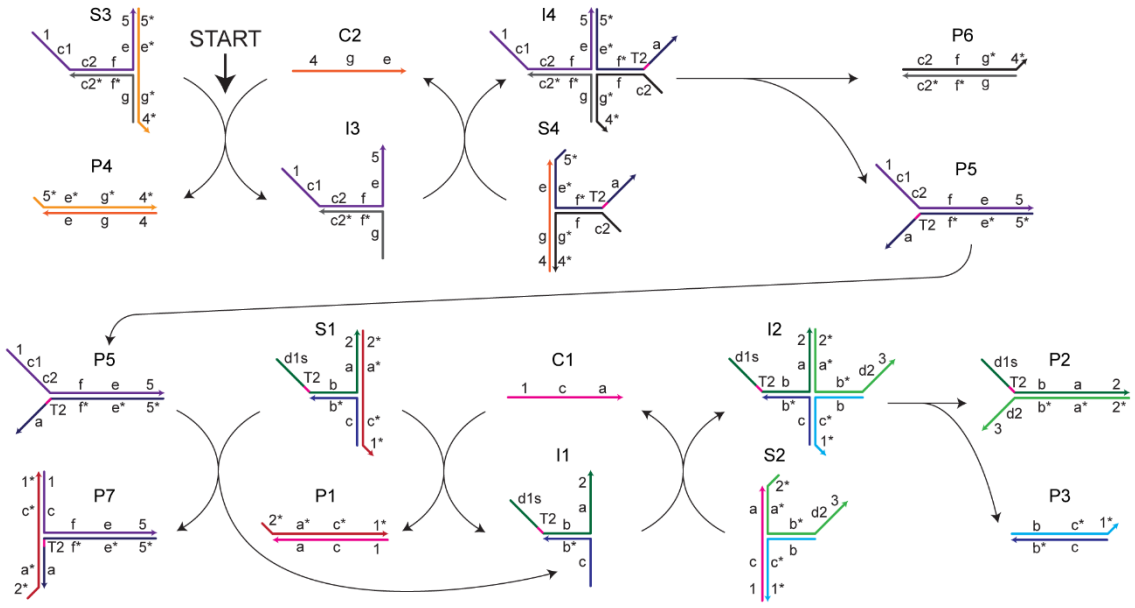

Figure 3.20. Detailed reaction schematic of two-layer feed forward catalytic system with three-arm junction substrates. A schematic of the catalytic pathway. The pathway starts with the reaction between the substrate S3 and the catalyst C2. After one cycle of the catalytic reaction, the first layer produces the product P5, whose single-stranded sequence has the same domains as the catalyst C1 for the second layer. Domain c = domain c1 + c2. The reporting reaction is the same as the single-layer catalytic system (Figure 3.2), thus not shown

Based on Figure 3.20, the following equations can be made.

$$S3 + S4 \xrightarrow{\square_{\square\square\square\square 1}} P4 + P5 + P6 \quad (3.17)$$

$$C2 + S3 \underset{\square_{-5}}{\overset{\square_5}{\rightleftharpoons}} P4 + I3 \quad (3.18)$$

$$I3 + S4 \underset{\square_{-6}}{\overset{\square_6}{\rightleftharpoons}} I4 + C2 \quad (3.19)$$

$$I4 \xrightarrow{\square_7} P5 + P6 \quad (3.20)$$

$$P5 + S1 \underset{\square_{-8}}{\overset{\square_8}{\rightleftharpoons}} P7 + I1 \quad (3.21)$$

$$P5 + S1d \xrightarrow{\square_8} P7 + I1d \quad (3.22)$$

$$C2 + S3d \xrightarrow{\square_5} P4 + I3d \quad (3.23)$$

$$I3 + S4d \xrightarrow{\square_6} I4 + C2d \quad (3.24)$$

$$S3i + S4i \xrightarrow{\square_{\square}} P4 + P5 + P6 \quad (3.25)$$

The leakage rate constant of equation (3.17) was assumed to be the same as equation (3.9). The rate constants used were $k_{-5} = k_{-1}$, $k_{-6} = k_{-2}$, and $k_7 = k_3$. Since we observed a slow feed forward system, the forward rate constants were adjusted to be $k_5 = k_1/2$ and $k_6 = k_2/2$. For equation (3.21), with output catalyst P5, the rate constants were adjusted to be $k_8 = k_{-8} = 5 \cdot 10^4\ M^{-1}s^{-1}$, which is ten times slower than the reaction of the single-stranded catalyst C1 in equation (3.10). The simulation result for the single-layer catalytic system, with the P5 input, is shown in Figure 3.21A, and the simulation agrees well with 5 nM P5. However, at 1 nM P5, the simulation agrees with the data only for the first hour, afterwards the deviation becomes large. One reason for the deviation is that the accumulation of product P7, produced by the reaction between P5 and S1, may inhibit the overall catalytic reaction. A second reason is that there could be side reactions.

Regardless, we are unsure how to accurately model this reaction. Therefore, instead of introducing further hypothetical reactions, we acknowledge that our understanding of the system is not complete. A complete understanding requires measuring the kinetics of the individual reactions to predict the overall kinetics of the system. While this is clearly beneficial and is currently being consideration, it is beyond the scope of this manuscript to perform these experiments. The simulation for the two-layer feed forward system is shown in Figure 3.21B by combining equations (3.9) through (3.25). Here, the concentration of P5, due to initial leakage[17], was adjusted to 1 pM. This value is much smaller than the linear substrate from Zhang *et al.*, where the initial leakage was assumed to be much larger for the 10 nM substrate concetration.[17] A possible explanation is the difference of the leakage mechanism between linear and multi-arm junction substrates. For linear substrates, leakage proceeds though three-way branch migration. As a result, initial leakage is expected to be instantaneous and the majority of malformed substrates will quickly decay into their products. In contrast, leakage proceeds through four-way branch migration for multi-arm junction substrates. As a result, the majority of the malformed substrates will decay slowly, which may cause a slow initial leakage (Figure 3.17B, first ~ 10 hours). S3d and S4d were adjusted to have a 3 % defect rate, while S3i and S4i were adjusted to have a 0.46 % defect rate. Although 100 pM, 1 pM, and 0 pM catalyst reactions agree well to our simulation, the 10 pM catalysis reaction does not agree (Figure 3.21B).

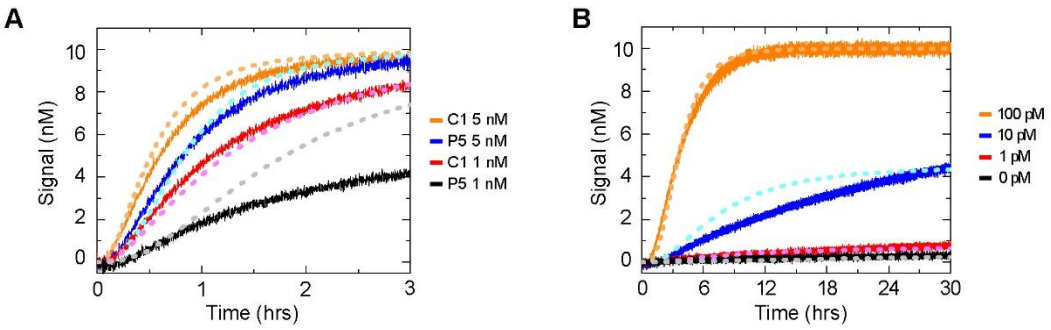

Figure 3.21. Kinetic simulation for two-layer feed forward catalytic system with three-arm junction substrate. (A) Simulation for data in Figure 3.19F, where single-stranded catalyst C1 and junction catalyst P5 were used as input catalyst for single-layer catalytic system with substrates S1 and S2. Kinetic simulations are shown as dots with weak colors. (B) Simulation for the two-layer feed forward system in Figure 3B

3.6.7 Detailed Reaction Schematic and Kinetic Simulation of Cross-Catalytic System with Three-Arm Junction Substrates

In the cross-catalytic system, the first layer is the same as the two-layer feed forward system (Figure 3.20). However, the product of the second layer, P3cr, has the catalytic domains 4-g-e in order to feedback to the first layer (Figure 3.22A). The reporter is the same with our other experiments (Figure 3.22B). Based on Figure 3.22A, B, the following equations can be made.

$$\mathrm{S1cr} + \mathrm{S2cr} \xrightarrow{\square_{\square\square\square\square 1}} \mathrm{P1} + \mathrm{P2cr} + \mathrm{P3cr} \tag{3.26}$$

$$\mathrm{C1} + \mathrm{S1cr} \underset{\square_{-1}}{\overset{\square_1}{\rightleftharpoons}} \mathrm{P1} + \mathrm{I1cr} \tag{3.27}$$

$$\mathrm{I1cr} + \mathrm{S2cr} \underset{\square_{-2}}{\overset{\square_2}{\rightleftharpoons}} \mathrm{I2cr} + \mathrm{C1} \tag{3.28}$$

$$\mathrm{I2cr} \xrightarrow{\square_3} \mathrm{P2cr} + \mathrm{P3cr} \tag{3.29}$$

$$\mathrm{P3cr} + \mathrm{S3} \underset{\square_{-9}}{\overset{\square_9}{\rightleftharpoons}} \mathrm{P8cr} + \mathrm{I3} \tag{3.30}$$

$$\mathrm{P3cr} + \mathrm{S3d} \xrightarrow{\square_9} \mathrm{P8cr} + \mathrm{I3d} \tag{3.31}$$

$$\mathrm{P2cr} + \mathrm{R} \xrightarrow{\square_4} \mathrm{Rwcr} + \mathrm{D} \tag{3.32}$$

$$\mathrm{C1} + \mathrm{S1crd} \xrightarrow{\square_1} \mathrm{P1} + \mathrm{I1crd} \tag{3.33}$$

$$\mathrm{I1cr} + \mathrm{S2crd} \xrightarrow{\square_2} \mathrm{I2cr} + \mathrm{C1d} \tag{3.34}$$

$$\mathrm{S1cri} + \mathrm{S2cri} \xrightarrow{\square\square} \mathrm{P1} + \mathrm{P2cr} + \mathrm{P3cr} \tag{3.35}$$

With the exception of equation (3.30) where the output catalyst P3cr reacts with the substrate S3, the reaction rate constants used were identical to the single-layer three-arm junction substrates in Section S3. Since a quick cross-catalytic system was observed, we used $k_9 = k_{-9} = 5 \cdot 10^5\ \mathrm{M^{-1}s^{-1}}$ in equation (3.30). This is ten times quicker than the reaction between P5 and S1, of which we do not have a reasonable explanation for the difference in the reaction speeds. The simulation and experimental data for the cross-catalytic system is shown in Figure 3.22C by combining equations (3.17) through (3.35). The population of defect-substrates and initial leakage substrates were assumed to be the same as the single-layer catalytic system so that S1crd and S2crd had a 2 % defect rate, and S1cri and S2cri had a 0.46 % defect rate. In order to account for the large leakage at 0 pM, the initial concentration of product P3cr was adjusted to be 70 pM. This large amount of P3cr did not reproduce the clear separation of the leakage and the 10 pM reaction. Again, further characterization of all individual reactions will be necessary to predict the overall kinetics of the system.

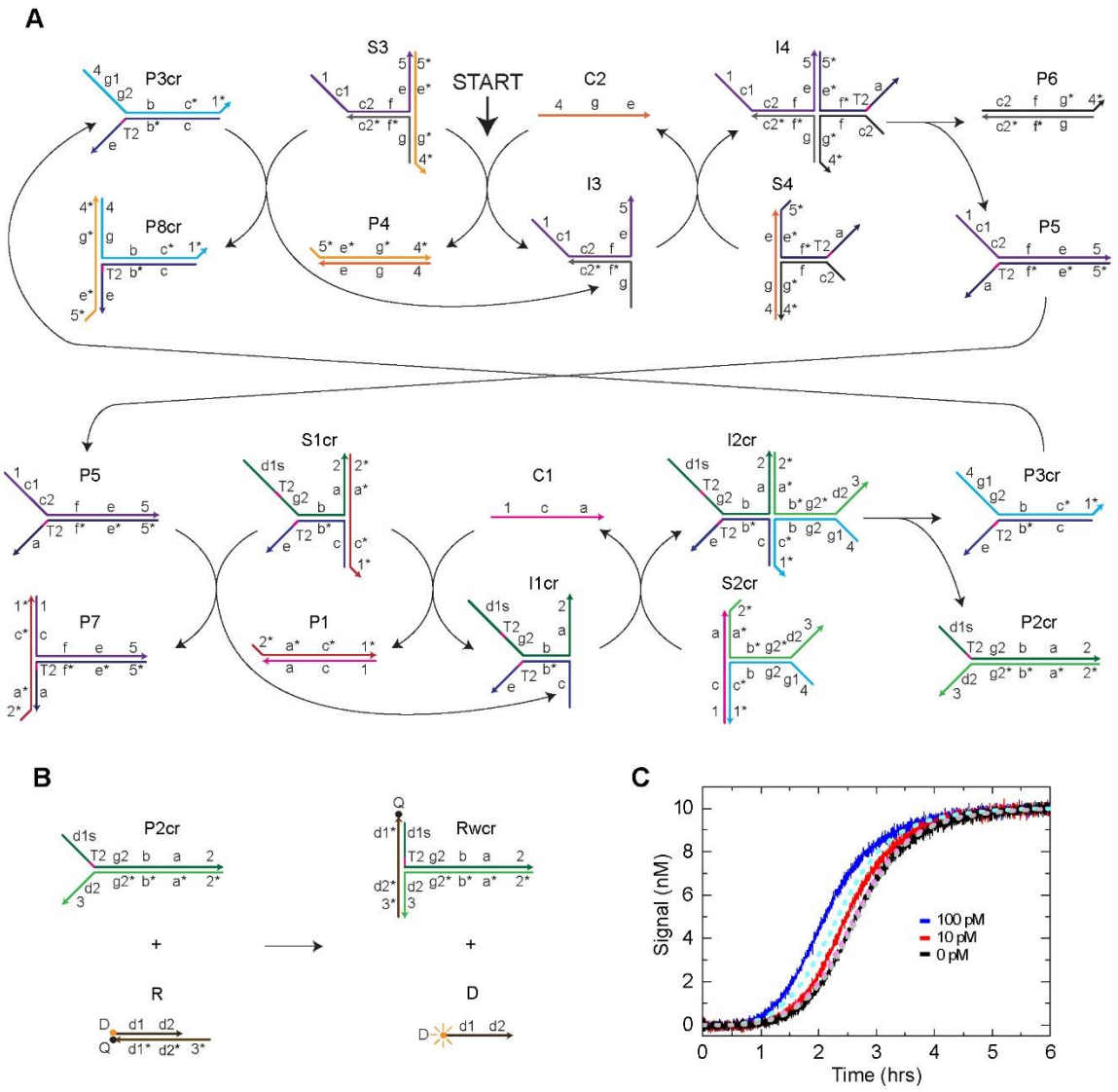

Figure 3.22. Detailed reaction schematic of cross-catalytic system with three-arm junction substrates. (A) A schematic of the catalytic pathway. The catalytic substrates in the first layer (S3, S4) are the same as substrates in two-layer feed forward catalytic system (Figure 3). The pathway starts with the reaction between the substrate S3 and the catalyst C2. After one cycle of the catalytic reaction, the first layer produces the product P5, whose single-stranded sequences have the same domain as the catalyst C1 for the second layer. The second layer then produces the product P3cr, whose single-stranded sequence have the same domain as the catalyst C2 for the first layer. Domain c = domains c1 + c2, and domain g = domains g1 + g2. (B) The reporting reaction. The reporter complex R is the same as all other experiments. (C) Kinetic traces with different concentrations of catalyst C2. [S1cr] = [S2cr] = [S3] = [S4] = 10 nM, [R] = 20 nM. Kinetic simulations are shown as dots with weak colors

3.6.8 Background Check of Two-Layer Feed Forward and Cross-Catalytic System with Three-Arm Junction Substrates

The leakage model in equation (3.7) considers only the reaction between S1 and S2, or other corresponding combinations of substrates. However, other leakage sources can appear if there exist design issues. To check this possibility, reaction components of feed forward and cross-catalytic systems were added step by step and their reaction kinetics were measured. For the feed forward catalytic system, an increase of the leakage was not observed by comparing the single-layer leakage (R + S1 + S2) with traces where other substrates were added, R + S1 + S2 + S3 and R + S1 + S2 + S4 (Figure 3.23A). Similarly, the cross-catalytic system showed no observable difference by comparing R + S1cr + S2cr with R + S1cr + S2cr + S3 and R + S1cr + S2cr + S4 (Figure 3.23B). Considering a very small value of k_{leak} and the fact that we have not performed any method to reduce initial leakage caused by malformed substrates,[17, 19] those results indicate that a main source of leakage for feed forward and cross-catalytic systems is the initial leakage, but not from the single-stranded tails of substrates in upstream layers. Therefore, if rigorous purifications of substrates were performed to reduce the initial leakage, the sensitivity of those systems is expected to significantly improve. The reason why R + S1cr + S2cr (Figure 3.23B) showed a higher fluorescence intensity than R + S1 + S2 (Figure 3.23A) is the higher background of R + S1cr than R + S1. This could be due to the single-stranded domain g2 next to domain d1s of S1cr (Figure 3.22A). This domain will reduce the steric hindrance for domain d1s to be hybridized with R, compared with S1 in which double-stranded domain b exists next to domain d1s (Figure 3.20).

Although we observed the reaction caused by cooperative binding of two single-stranded DNA strands (Figure 3.14D, E), the reaction uses a long toehold (10 nt). Therefore, output tails of substrates with 6 or 8 nt toehold length are less likely to cause the reaction in the same mechanism. Also, even if such reactions become significant once the initial leakage is eliminated, it will be possible to minimize such reactions using clamps.[7]

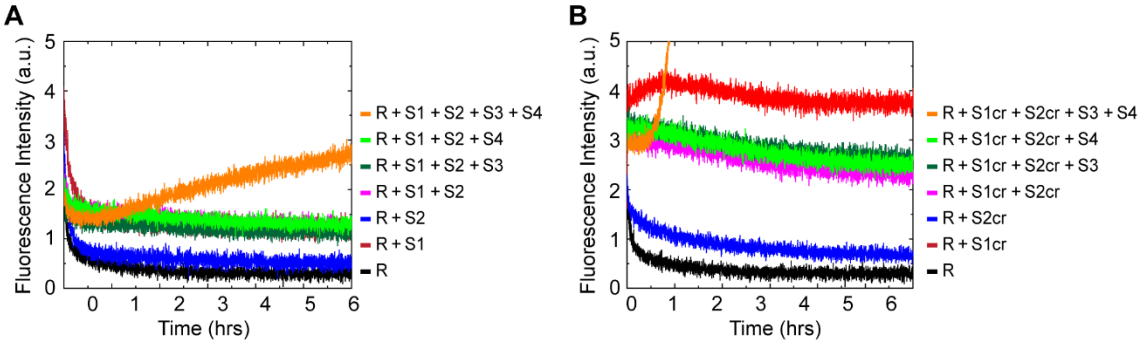

Figure 3.23. Background analysis of the reaction networks with the three-arm junction substrates. For all the data, fluorescence intensity was not normalized. (A) Two-layer feed forward catalytic system. [S1] = [S2] = [S3] = [S4] = 50 nM, [R] = 100 nM. (B) Cross-catalytic system. [Scr1] = [S2cr] = [S3] = [S4] = 50 nM, [R] = 100 nM

3.6.9 Feed Forward and Autocatalytic System with Three-Arm Junction Substrates Based on 16 nt Specificity Domains

In addition to the 22 nt specificity domains, three-arm junctions with 16 nt specificity domains were used to construct a feed forward catalytic system (Figure 3.24A). The leakage (0 pM) was significantly larger than that of the 22 nt design (Figure 3.24B), which decreased the signal to noise ratio. Considering the similar leakage rate constants for the 16 nt and 22 nt specificity domain designs (Figure 3.14C, S8D), we expect that the initial leakage of the 16nt design could be higher than the 22 nt design, resulting in the larger leakage of the feed forward system. An autocatalytic network was

also constructed (Figure 3.24C), and the kinetic trace of the system again showed large leakage (Figure 3.24D). Moreover, the kinetic trace does not appear to have the ideal sigmoid shape. This can be seen because the 0 pM reaction slows down after 2 hours, comparing with the autocatalytic system with four-arm junction substrates (Figure 3.4B). The slow-down of the three-arm junction substrate is attributed to a sequence constraint, which caused one strand (light blue strand, *mB3au*) of mS2au to have a hairpin structure, destabilizing mS2au and mP3au.

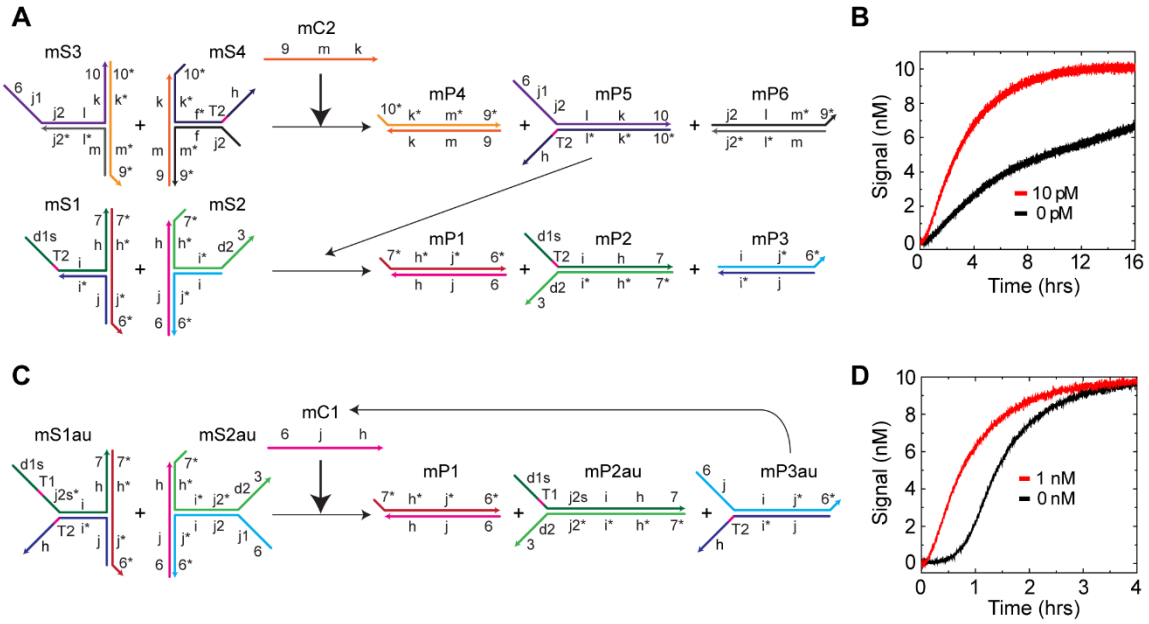

Figure 3.24. Three-arm junction catalytic networks based on 16 nt specific domains. (A) A simplified schematic of the two-layer feed forward catalytic system. Product mP5 has catalytic domains to act as an input catalyst for the second layer. (B) Kinetic traces of the feed forward catalytic system with different concentrations of the catalyst mC2. [mS1] = [mS2] = [mS3] = [mS4] = 10 nM, and [R] = 20 nM. (C) A simplified schematic of the autocatalytic system. The product mP3au has the catalytic domains to cause exponential amplification. Note that the light blue strand of the substrate mS2au has domains 6-j-i-j*-6* (domain j = j1 + j2), resulting in a hairpin structure. mP2au has the output domain for the reporting reaction as mP2 (Figure 3.18B). Domain T1 of mS1au represents 1 nt thymidine. Domain j2s is j2 with 1 nt at 5' end deleted. (D) Kinetic traces of the autocatalytic network with different concentrations of the catalyst mC1. [mS1au] = [mS2au] = 10 nM, and [R] = 20 nM

3.6.10 Design Principle of a Four-Arm Junction Substrate

The design principle of a four-arm junction substrate is identical to a three-arm junction substrate, and the IELs or each reaction is similar. However, the primary difference between them is the existence of an intermediate complex with a six-arm junction, which is shown in the following section.

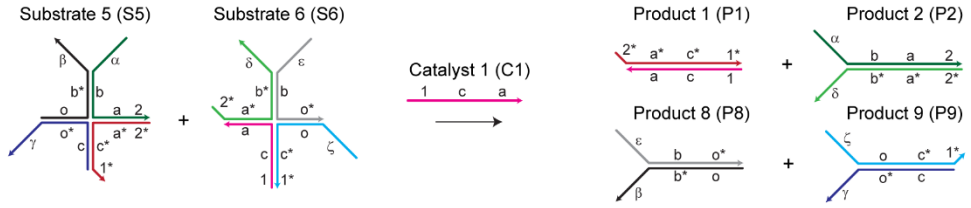

Figure 3.25. Design principles of a four-arm junction substrate for a catalytic system. The reaction converts two substrates S5 and S6 into four products P1, P2, P8, and P9. Products P2, P8 and P9 have a new combination of output domains α and δ, ε and β, ζ and γ, respectively. Sequences of those output are independent of the catalyst C1, therefore can be used for signal cascades

3.6.11 Toehold Length Variations of Four-Arm Junction Substrates

Here, a single-layer catalytic system was used in order to test the operation of four-arm junction substrates (Figure 3.26A), and the initial design with a 6 nt toehold showed a very slow catalytic reaction (Figure 3.26B). Therefore, the toehold length was changed systematically from 6 nt to 8 nt. As a result, the reaction between the intermediate I5 and S6 was found to be the rate limiting step, because the overall catalytic reaction sped up only when domain 2 was extended, which acted as the toehold to drive the reaction between I5 and S6 (Figure 3.26C-E). This performance change is because of the high energy barrier imposed by the six-arm junction structure[47] on I6 (Figure 3.26A). For the above listed substrates, toeholds were extended unevenly so that

toehold hybridization occurred via 8 nt, but dissociation occurred via 6 nt in order to maintain fast dissociation. In fact, the catalytic turnover became lower if both the hybridization and dissociation of the toehold occurred with a 8 nt toehold (Figure 3.26F). The process of toehold dissociation becomes more challenging for longer toeholds, which will be required for five-arm junction substrates because eight-arm junctions will emerge during the catalytic reaction (Figure 3.31). In addition to domain 2, the extended domain 1 allowed the output catalyst to trigger the catalytic reaction as quick as the single-stranded catalyst (Figure 3.19G, H), which is the reason why substrates S5xy and S6xy were adapted for the autocatalytic system in Figure 3.4.

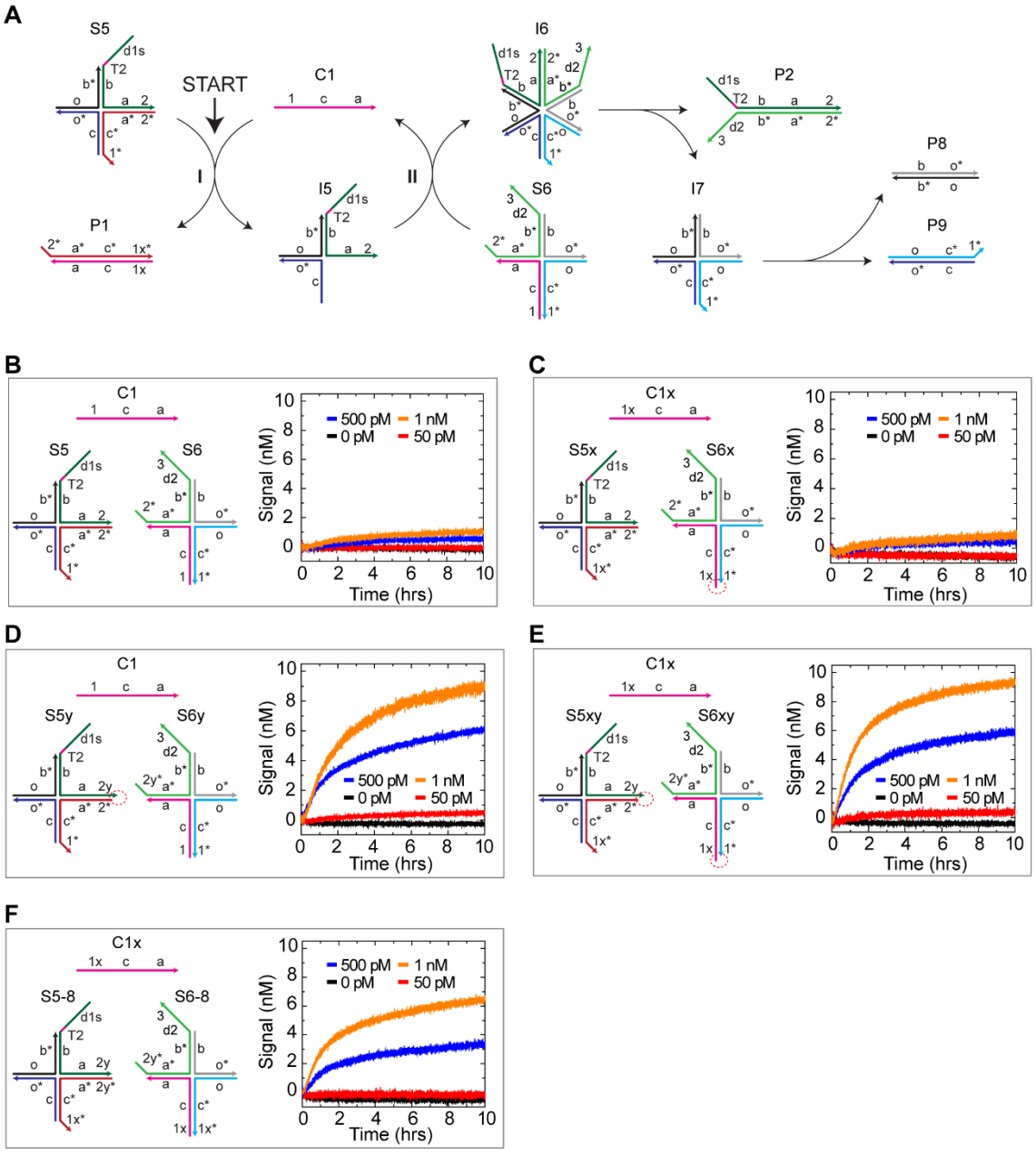

Figure 3.26. Variations of toehold designs for four-arm junction substrates. (A) A schematic of the catalytic pathway for the single-layer catalytic network. (B)-(F) Kinetic traces with different catalytic concentrations. [S5] = [S6] = 10 nM, [R] = 20 nM, and those were the same concentrations for all other variations. 2 nt were added at 5' end of domain 1 and 3' end of domain 2 (both are 6 nt) to generate domains 1x and 2y (both are 8 nt), respectively. The toehold affects the kinetics of step I (the reaction between the substrate S5 and the catalyst C1), and step II (the reaction between the intermediate I5 and the substrate S6). Red dot circles show the exposed

2 nt toeholds. (B) 6 nt design for both toehold domains 1 and 2. (C) The toehold domain 1 was unevenly extended on the substrate S6x. As a result, step I occurs via 8 nt toehold hybridization and 6 nt toehold dissociation, while step II occurs via 6 nt toehold hybridization and 6 nt toehold dissociation. (D) The toehold domain 2 was unevenly extended on the substrate S5y. As a result, step I occurs via 6 nt toehold hybridization and 6 nt toehold dissociation, while step II occurs via 8 nt toehold hybridization and 6 nt toehold dissociation. (E) Both toehold domains 1 and 2 were unevenly extended so that both steps I and II occur via 8 nt toehold hybridization and 6 nt toehold dissociation. (F) Both toehold domain 1 and 2 were evenly extended so that both steps I and II occur via 8 nt toehold hybridization and 8 nt toehold dissociation

3.6.12 Calculation of Rate Constants and Kinetic Simulation for Four-Arm Junction Substrates

The catalytic rate constant and leakage rate constant – The catalytic rate constant for four-arm junction substrates, with extended toeholds at domains 1 and 2, were calculated. Using the same process outlined in Section S3, we obtained $k_{cat} = 2.28 \cdot 10^{13}$ $M^{-2}s^{-1}$ from the 1 nM catalytic reaction (Figure 3.27A), which is almost identical to the three-arm junction substrates (Table 3.1). We did not use the 500 pM catalytic reaction because the reaction nearly stalled at 10 hours (Figure 3.26E). The leakage reactions were also performed for four-arm junction substrates with their corresponding background reactions (Figure 3.27B). Then, after using the same process in Section S3, the leakage rate constants were calculated from 20 to 24 hours (Figure 3.27C, D). The values were $k_{leak2} = 2.11 \cdot 10^{-2}$ $M^{-1}s^{-1}$ for S5xy + S6xy and $k_{leak} = 2.66 \cdot 10^{-2}$ $M^{-1}s^{-1}$ for S5 + S6. While there is no significant difference on the rate constants between the two designs, the S5xy and S6xy have two sets of exposed 2 nt toehold (Figure 3.26E), which can hybridize to each other and initiate strand displacement. This result is in agreement with a prior toehold mediated, four-way branch migration study,[28] which reported the same kinetic rate constants between a zero toehold and an "open" 2 nt toehold. However,

since the background signal caused by the single-stranded tail on the substrates might exist (Figure 3.14D, E), we cannot exclude the possibility that the actual leakage rates of the two designs are significantly different.

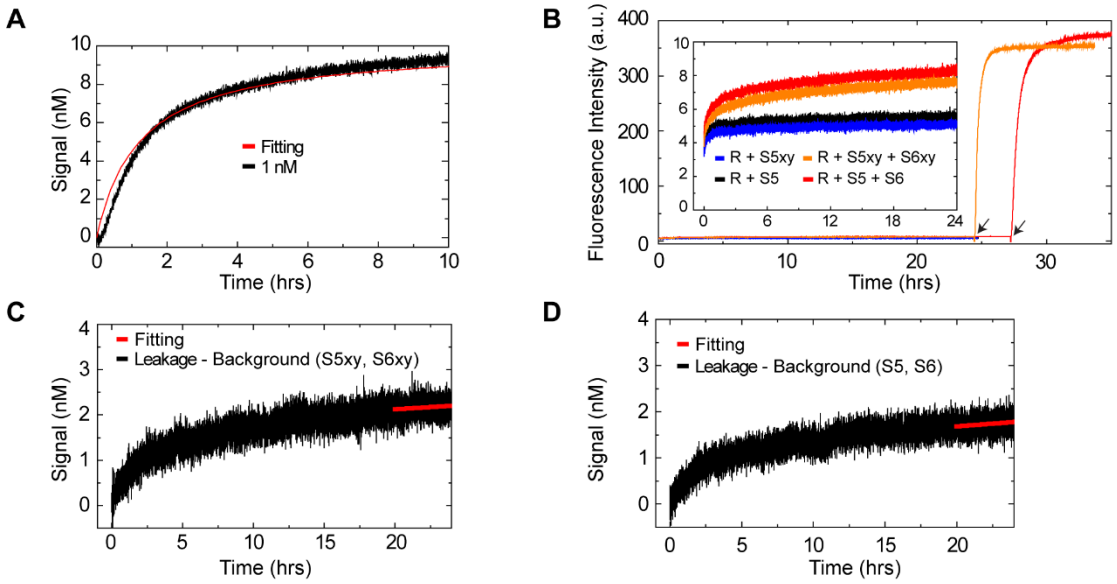

Figure 3.27. The rate constants calculation of single-layer catalytic system with four-arm junction substrates. (A) The kinetic trace of the four-arm junction substrates with 1 nM catalyst C1x was fit to equation (3.6), and $k_{cat} = 2.28 \cdot 10^{13}$ $M^{-2}s^{-1}$ was extracted. [S5xy] = [S6xy] = 10 nM, [R] = 20 nM. (B) Kinetic traces of leakages and the background reactions for the four-arm junction substrates. [S5xy] = [S6xy] = [S5] = [S6] = 500 nM, [R] = 700 nM. Leakage traces are R + S5xy + S6xy and R + S5 + S6, and other traces are performed to measure the background signals. Fluorescence data is not normalized. Black arrows show the addition of catalyst C1. (C) The background reaction (R + S5xy) was subtracted from the leakage (R + S5xy + S6xy), and the initial fluorescence intensity was adjusted to zero. Then, the fluorescence intensity was normalized using the average of maximum fluorescence changes of 10 nM concentration experiments in Figure 3.26E. Finally, the initial slope (20 to 24 hours) was fit to equation (3.8), and $k_{leak2} = 2.11 \cdot 10^{-2}$ $M^{-1}s^{-1}$ was extracted. (D) The background reaction (R + S5) was subtracted from the leakage reaction (R + S5 + S6), and the initial fluorescence intensity was adjusted to zero. Then, the fluorescence intensity was normalized using the average maximum fluorescence changes of the 10 nM concentration experiments in Figure 3.26B. Finally, the initial slope (20 to 24 hours) was fit to equation (3.8), and $k_{leak} = 2.66 \cdot 10^{-2}$ $M^{-1}s^{-1}$ was extracted

Kinetic simulation – Based on Figure 3.26A and E, the following equations can be made for substrates S5xy and S6xy.

$$\mathrm{S5xy} + \mathrm{S6xy} \xrightarrow{\square_{\square\square\square\square 2}} \mathrm{P1} + \mathrm{P2} + \mathrm{P8} + \mathrm{P9} \tag{3.36}$$

$$\mathrm{C1x} + \mathrm{S5xy} \underset{\square_{-1}}{\overset{\square_1}{\rightleftharpoons}} \mathrm{P1} + \mathrm{I5} \tag{3.37}$$

$$\mathrm{I5} + \mathrm{S6xy} \underset{\square_{-2}}{\overset{\square_2}{\rightleftharpoons}} \mathrm{I6} + \mathrm{C1x} \tag{3.38}$$

$$\mathrm{I6} \xrightarrow{\square_3} \mathrm{P2} + \mathrm{P8} + \mathrm{P9} \tag{3.39}$$

$$\mathrm{P2} + \mathrm{R} \xrightarrow{\square_4} \mathrm{Rw} + \mathrm{D} \tag{3.40}$$

$$\mathrm{C1x} + \mathrm{S5xyd} \xrightarrow{\square_1} \mathrm{P1} + \mathrm{I5d} \tag{3.41}$$

$$\mathrm{I5} + \mathrm{S6xyd} \xrightarrow{\square_2} \mathrm{I6} + \mathrm{C1xd} \tag{3.42}$$

$$\mathrm{S5xyi} + \mathrm{S6xyi} \xrightarrow{\square_{\square}} \mathrm{P1} + \mathrm{P2} + \mathrm{P8} + \mathrm{P9} \tag{3.43}$$

Production of P8 and P9 from intermediate I7 was included into equation (3.39) for simplification. Other than the leakage rate constant ($k_{leak2} = 2.11 \cdot 10^{-2}$ $M^{-1}s^{-1}$) we used the same rate constants for the three-arm junction substrates in Figure 3.17. The population of defect-substrates S5xyd and S6xyd were adjusted to be 4 %, and the initial leakage substrates S5xyi and S6xyi were adjusted to be 0.4 %. The simulation result using equations (3.36) through (3.43) is shown in Figure 3.28.

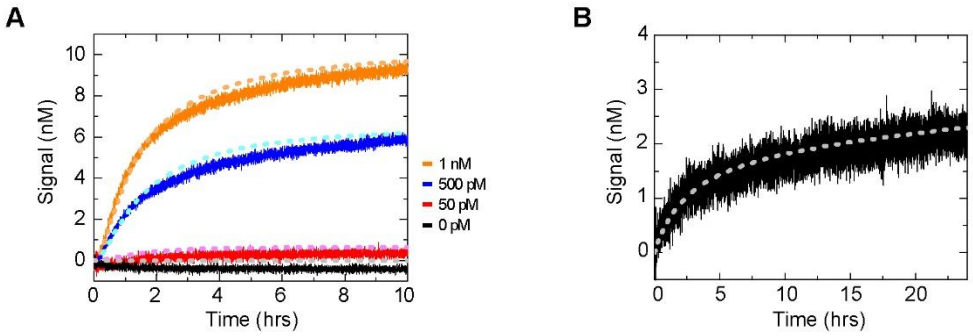

Figure 3.28. Kinetic simulation for single-layer catalytic system with four-arm junction substrate. (A) 10 nM concentration of substrates with different concentration of catalyst for the data in Figure 3.26E. Kinetic simulations are shown as dots with weak colors. (B) Leakage reaction at 500 nM concentration of substrates for the data in Figure 3.27C

3.6.13 Detailed Reaction Schematic and Kinetic Simulation of Autocatalytic System with Four-Arm Junction Substrates

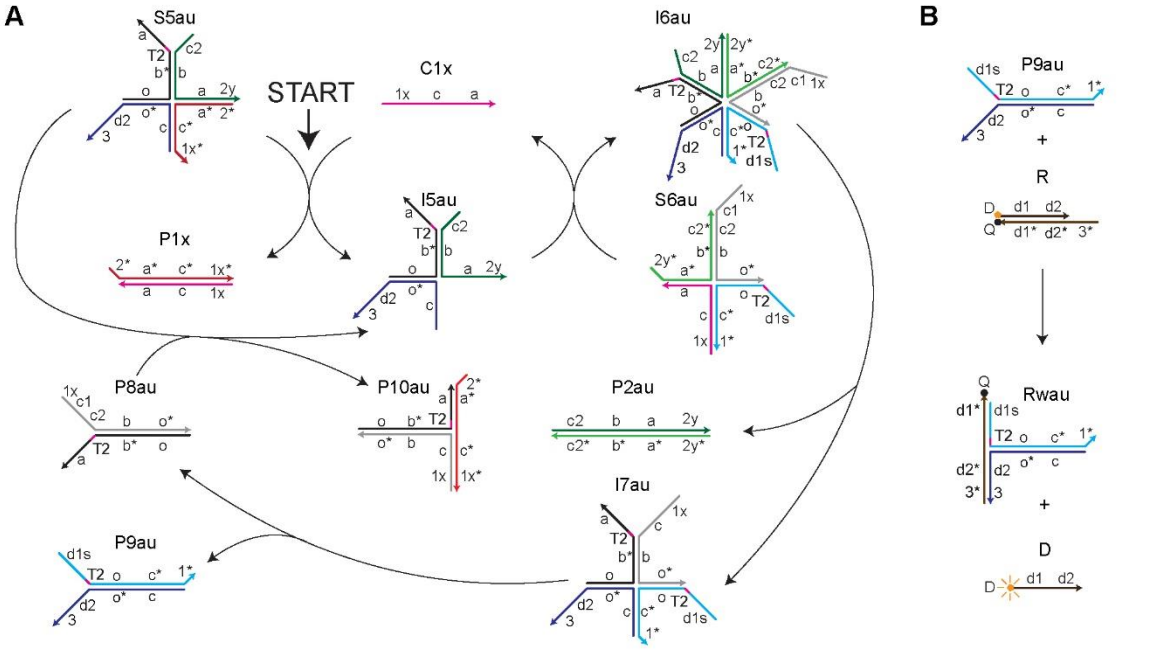

Figure 3.29. The detailed reaction schematic of the autocatalytic system with four-arm junction substrates. (A) A schematic of the catalytic pathway. The pathway starts with the reaction between the substrate S5au and the catalyst C1x. The single-stranded sequence of the product P8au has the same domain as the catalyst C1x. Domain c = domains c1 + c2. (B) The reporting reaction. The reporter complex R is the same as all other experiments

Based on Figure 3.29A and B, the following equations can be made for the autocatalytic system.

$$\mathrm{S5au} + \mathrm{S6au} \xrightarrow{k_2} \mathrm{P1x} + \mathrm{P2au} + \mathrm{P8au} + \mathrm{P9au} \tag{3.44}$$

$$\mathrm{C1x} + \mathrm{S5au} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{P1x} + \mathrm{I5au} \tag{3.45}$$

$$\mathrm{I5au} + \mathrm{S6au} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{I6au} + \mathrm{C1x} \tag{3.46}$$

$$\mathrm{I6au} \xrightarrow{k_3} \mathrm{P2au} + \mathrm{P8au} + \mathrm{P9au} \tag{3.47}$$

$$\mathrm{P8au} + \mathrm{S5au} \underset{k_{-10}}{\overset{k_{10}}{\rightleftharpoons}} \mathrm{P10au} + \mathrm{I5au} \tag{3.48}$$

$$\mathrm{P9au} + \mathrm{R} \xrightarrow{k_4} \mathrm{Rwau} + \mathrm{D} \tag{3.49}$$

$$\mathrm{C1x} + \mathrm{S5aud} \xrightarrow{k_1} \mathrm{P1x} + \mathrm{I5aud} \tag{3.50}$$

$$\mathrm{P8au} + \mathrm{S5aud} \xrightarrow{k_{10}} \mathrm{P10au} + \mathrm{I5aud} \tag{3.51}$$

$$\mathrm{I5au} + \mathrm{S6aud} \xrightarrow{k_2} \mathrm{I6au} + \mathrm{C1xd} \tag{3.52}$$

$$\mathrm{S5aui} + \mathrm{S6aui} \xrightarrow{k_{\square}} \mathrm{P1x} + \mathrm{P2au} + \mathrm{P8au} + \mathrm{P9au} \tag{3.53}$$

The overall speed of the single-layer catalytic reaction was shown to be approximately the same when the output catalyst P8au was used instead of C1x (Figure 3.19H). This trend holds true even if the reaction rate used in equation (3.48), with P8au catalyst, is a few times slower than the reaction rate used in equation (3.45), with the single-stranded catalyst C1x. This trend is because P8au is consumed by S5xy and a new catalyst, C1x, is generated, which causes an initial delay of the overall catalytic reaction that is quickly recovered by releasing C1x from substrate S6xy. Therefore, we used $k_{10} = k_{-10} = k_1/2.5$ and the experimental data was simulated well (Figure 3.30A). The simulation result using equations (3.44) through (3.53) for the autocatalytic system is shown in Figure 3.30B. Here, the population of defect-substrates and initial leakage substrates were assumed to be the same as the single-layer catalytic system so that S5aud and S6aud had a 4 % defect rate, while S5aui and S6aui had a 0.4 % defect rate. The initial concentration of P8au was adjusted to be 1.5 pM. The agreement of the data and the simulation is better than the two-layer feed forward (Figure 3.21B) and the cross-catalytic

system (Figure 3.22C) because the P8au catalyst (Figure 3.30A) is well-simulated when compared to the P5 catalyst (Figure 3.21A).

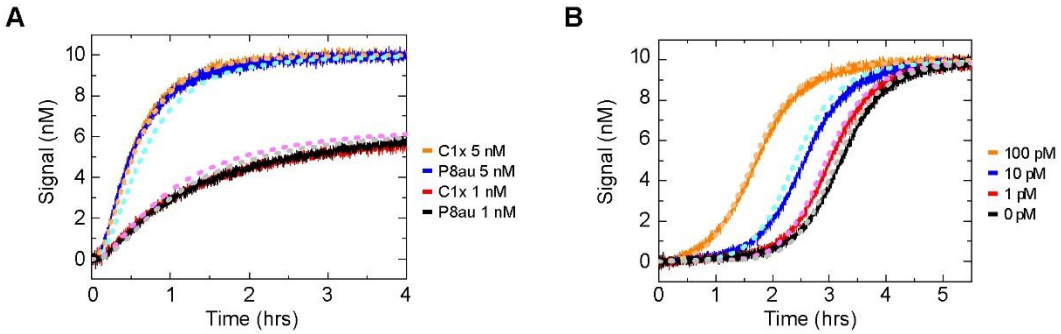

Figure 3.30. Kinetic simulation for autocatalytic system with four-arm junction substrate. (A) Simulation for data in Figure 3.19H, where single-stranded catalyst C1x and junction catalyst P8au were used as input catalyst for single-layer catalytic system with substrate S5xy and S6xy. Since unpurified DNA strands were used for this set of experiment, the population of defect-substrates S5xyd and S6xyd were adjusted to be 8 %. (B) Simulation for the autocatalytic system in Figure 3.4B

3.6.14 Proposed Design of a Five-Arm Junction Substrate

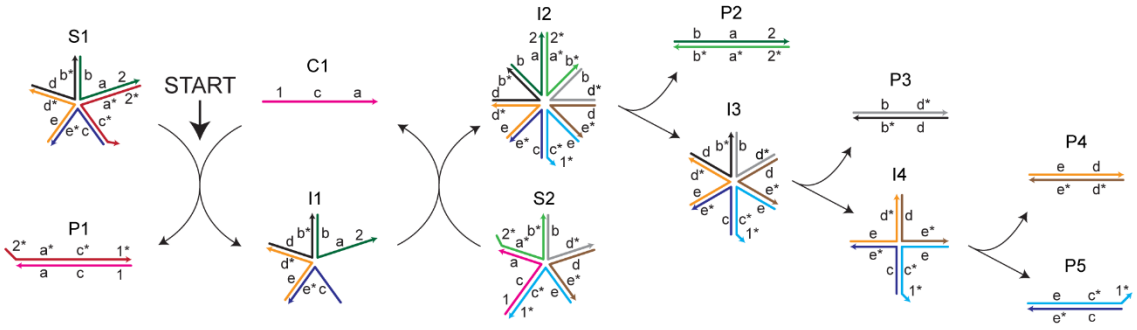

Figure 3.31. Single-layer catalytic system with five-arm junction substrates. A proposed design and a schematic of the catalytic pathway. Naming of domains, a catalyst, and other DNA complexes are unrelated to those in other schematics. For simplicity, single-stranded tails on each substrate, which can be used for the formation of output signals, are not shown.

3.6.15 Gel Images

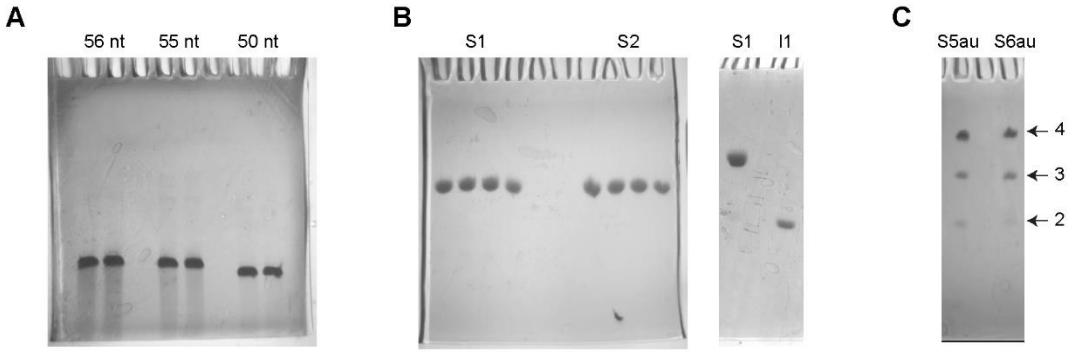

Figure 3.32. Relative mobility of single-stranded DNAs and DNA complexes. (A) A 10 % denature PAGE for unpurified DNA strands purchased from IDT. The first two lanes are the *tC1* strands with 56 nt, the middle two lanes are the *tC3* strands with 55 nt, and the last two lanes are the *tD1* strands with 50 nt. All sequences are shown in Table 3.2. Denature PAGE was performed to remove smear bands from the target strands. Since the separation of 56 nt and 55 nt was very small for our experimental condition, we do not expect purification of full length DNA strands from the n-1 truncated strands that are common during chemical DNA synthesis.[23] (B) 10 % native PAGE. The gel image on the left side shows the purification process of DNA complex S1 and S2, where an equal stoichiometric ratio was used for the complex formation. The running time of the gel was 5 hours. The small gel image on the right side shows the relative mobility of S1, with strands *A1*, *A2*, and *A3*, and intermediate I1, with strands *A2* and *A3*. The running time of the gel was 4 hours. The separation of S1 and I1 are large enough for easy purification at 4 hours. Therefore, even if I1 is formed during the formation of S1, due to concentration error, S1 can be purified under our experimental conditions. (C) 10 % native PAGE for the purification process of DNA complex S5au and S6au. Those complexes were annealed with uneven stoichiometric ratios, as explained in the experimental methods in Section S16. Consequently, three bands appeared on each lane, which included a: (a) four-stranded complex, (b) three-stranded complex, and (c) small amount of two-stranded complex. The number of stands in the gel is labeled and decreases from top to bottom. However, the band separations were large enough so that the target four-stranded DNA complexes could be purified

3.6.16 Experimental Methods

Unless otherwise specified, all of the chemicals were purchased from Thermo Fisher Scientific.

Formation of multi-arm junction substrates – DNA strands were purchased from Integrated DNA Technologies (IDT) and re-suspended to be 100 μM in 1×TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA), diluted from 100×TE (Sigma-Aldrich). The *d* strand was modified with 5' Tetrachlorofluorescein (TET) and the *q* strand was modified with 3' Iowa black dark quenchers (IABkFQ), both of which were purchased with HPLC purification. All other strands were purchased as unpurified, then purified by denature polyacrylamide gel electrophoresis (denature PAGE) in house, otherwise mentioned. For substrate formation, each DNA strand was stoichiometry mixed in 1×TE buffer with 12.5 mM $MgCl_2$ (1×TE/Mg^{2+}). However, a non-stoichiometry concentration was used for the following substrates: (1) S5au (with *A1x* strand 50% shortage), S6au (with *fB4au* strand 50% shortage), and mS1au (with *mA1* strand 30 % shortage), because the excess of those strands can hybridize to the ssDNA tails on the substrates, and might not be separated even after native PAGE purification; (2) mS2 (with *mB3* strand 30% shortage), because the *mB3* strand hybridized with the ssDNA tail on the mS2 substrate due to unintentional complementarity; (3) S4 (with *tD1* strand 50 % excess), because stoichiometry formation showed a small complex population of *tD2-tD3*. This could be due to the inaccuracy of the concentration of either strand; (4) S2 (with *B1* strand 50% excess) only in Figure 3B, because it showed a better catalytic turnover than the stoichiometric formation. Annealing of DNA complexes was performed at 90 °C for 5 minutes and cooled to 20 °C at a rate of 0.1 °C per 10 seconds. This slow annealing rate was necessary to form the mS2au substrate because of the hairpin structure on the *mB3au* strand, but not necessary for the other substrates.

Gel electrophoresis – The gel size used in all the experiments was 1.5 mm thick 10 × 10 cm.

For denature PAGE, 10% gel (acrylamide:bis = 29:1 from Bio-Rad) was made with 1×TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA) and 8 M Urea (ultrapure) from MP Biomedicals. Then, ~15 μL of 100 μM DNA was mixed with equal volume to the loading buffer, which was 8:1:1 volume ratio mixture of formamide (Sigma-Aldrich), bromophenol blue (Sigma-Aldrich), and 10×TBE. Next, the DNA solution was loaded in each lane and run at ~400V for 30 to 40 minutes with circulating water at 60 °C. The bands of interest were cut out, crushed by pipet tips, and eluted in 300 μL of 1×TE buffer overnight. After gel fragments were removed via a centrifuge, 3 M of sodium acetate (Arbor Scientific) (pH 5.2) was added to a final concentration of 0.3 M. Then, 1 mL of 100% Ethanol were added. After cooling at -20 °C for ~1 hour, centrifugation was performed with 15,000 rcf at 4 °C for 20 minutes. The DNA pellet was further rinsed by adding 1 mL of 70% Ethanol, and centrifuged at 15,000 rcf at 4 °C for 20 minutes. Finally, the ethanol solution was removed and the remaining solution was air dried for ~30 minutes, and 30 μL of 1×TE was added.

For native PAGE, 10% gel (acrylamide:bis = 29:1) was made with 1×TAE buffer (40 mM Tris, 40 mM Acetate, 1 mM EDTA) and supplemented with 12.5 mM $Mg(C_2H_3O_2)_2$. Then, DNA complexes were mixed with 1/4 volume of loading buffer, which was 1:1 volume ratio mixture of bromophenol blue dye (Sigma-Aldrich) and a ficoll solution (type 400, 20 % water from Sigma-Aldrich). Next, the DNA complex solution was loaded and run at 150V for ~5 hours with

circulating water at 20 °C. The bands of interest were then cut out of the gels, not crushed, and eluted in 1×TE/Mg^{2+} for 2 days at 4 °C. For Figure 3.19D, the gel was stained with SYBR Gold for 30 minutes.

In all the cases, DNA concentrations were quantified by absorbance at 260 nm and calculated using extinction coefficients provided by IDT for single-stranded and double-stranded DNA.[52]

Kinetics experiments – All kinetic experiments were performed at 25 °C in 1×TE/Mg^{2+} using two fluorescence spectrophotometers (Agilent Technologies, Cary Eclipse) at 521 nm excitation and 538 nm emission with 0.56 mL Special Optical Glass cuvettes (Starna Cells). To prevent solution evaporation, the lid was covered with parafilm. Slit sizes were 2.5 nm for the excitation and 10 nm for the emission. A poly-T solution (dT_{20} or dT_{50}, both did not show any difference) were added to all dilute stock samples (1 μM and lower) with 1 μM in order to minimize the loss of DNA via sticking to the sample tubes and pipet tips.[17] Fluorescence data was then normalized so that the initial fluorescence intensity corresponds to 0 nM and the final intensity corresponded to the initial concentration of the substrates (under the assumption that all substrates were consumed), otherwise mentioned. When the reaction did not reach completion during the experimental time windows, high concentration of catalysts was added to drive the reaction to completion. For the two cases with the four-arm junction substrates (S5 + S6 and S5x + S6x), their maximum intensity was obtained by heating up the sample cuvette in hot water, because of their slow catalytic rates. For fluoresce data where an intensity was not normalized, all kinetic traces were acquired by using the same machine, except the

data in Figure 3.23A. In this data, R + S1 and R + S2 were performed by using a different machine from the other kinetic traces. Therefore, typical fluorescence intensity difference, 5%, was used to correct the intensity for R + S1 and R + S2.

3.6.17 DNA Strand Sequences

The domains and sequences of DNA strands used for each DNA complex are shown in Table 3.2. DNA sequences were designed by NUPACK[50] or EGNAS,[53] then occasionally modified by hand to minimize the secondary structure and hybridization of unrelated domains.

Table 3.2 Oligonucleotide sequences for DNA complexes

Complex	Name	domains	Sequence (5` to 3`)
S1	*A1*	2*-a*-c*-1*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGG
	A2	d1s-T2-b-a-2	CCAAACCTTCATCTTC TT GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCC
	A3	c-b*	CCAGATCAGCAGCCATTCGTTC CCAGGCCTCGTATCGCGAGTGC
S2	*B1*	1-c-a	CCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2	2*-a*-b*-d2-3	GGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC TACTCG CCTCTACTCA
	B3	b-c*-1*	GCACTCGCGATACGAGGCCTGG GAACGAATGGCTGCTGATCTGG AAACGG
S3	*tC1*	5*-e*-g*-4*	GGTGTT CGGGAGTAGGGTAGAGTAAGAG GAAGGTAGAGCGGAGTAACAGG GATAGC
	tC2	1-c1-c2-f-e-5	CCGTTT CCAGATCAGCA GCCATTCGTTC GCGAGTGTCCTGGGTCAAGGCG CTCTTACTCTACCCTACTCCCG AACACC
	tC3	g-f*-c2*	CCTGTTACTCCGCTCTACCTTC CGCCTTGACCCAGGACACTCGC GAACGAATGGC
S4	*tD1*	4-g-e	GCTATC CCTGTTACTCCGCTCTACCTTC CTCTTACTCTACCCTACTCCCG
	tD2	5*-e*-f*-T2-a	GGTGTT CGGGAGTAGGGTAGAGTAAGAG CGCCTTGACCCAGGACACTCGC TT CAGTCCCAAGTCACCACCTAGC
	tD3	c2-f-g*-4*	GCCATTCGTTC GCGAGTGTCCTGGGTCAAGGCG GAAGGTAGAGCGGAGTAACAGG GATAGC
S1cr	*A1*	2*-a*-c*-1*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGG
	A2cr	d1s-T2-g2-b-a-2	CCAAACCTTCATCTTC TT GCTCTACCTTC GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCC
	A3cr	c-b*-T2-e	CCAGATCAGCAGCCATTCGTTC CCAGGCCTCGTATCGCGAGTGC TT CTCTTACTCTACCCTACTCCCG
S2cr	*B1*	1-c-a	CCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2cr	2*-a*-b*-g2*-d2-3	GGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC GAAGGTAGAGC TACTCG CCTCTACTCA
	B3cr	4-g1-g2-b-c*-1*	GCTATC CCTGTTACTCC GCTCTACCTTC GCACTCGCGATACGAGGCCTGG GAACGAATGGCTGCTGATCTGG AAACGG
mS1	*mA1*	7*-h*-j*-6*	GGTGTC GGCGGTGTTGTTAGTG GAATGCGGAGGAGTGG AAACGG
	mA2	d1s-T2-i-h-7	CCAAACCTTCATCTTC TT GCGATGTGGTGAGAGG CACTAACAACACCGCC GACACC
	mA3	j-i*	CCACTCCTCCGCATTC CCTCTCACCACATCGC
mS2	*mB1*	6-j-h	CCGTTT CCACTCCTCCGCATTC CACTAACAACACCGCC

	mB2	7*-h*-i*-d2-3	GGTGTC GGCGGTGTTGTTAGTG CCTCTCACCACATCGC TACTCG CCTCTACTCA
	mB3	i-j*-6*	GCGATGTGGTGAGAGG GAATGCGGAGGAGTGG AAACGG
mS3	*mC1*	10*-k*-m*-9*	GCTTGT GGAGATTTAGGCGTTG GTTCAGGTAGGTTCGG AATAGC
	mC2	6-j1-j2-l-k-10	CCGTTT CCACTCCTCCGCATTC GGTTATGGCTGGTTTC CAACGCCTAAATCTCC ACAAGC
	mC3	m-l*-j2*	CCGAACCTACCTGAAC GAAACCAGCCATAACC GAATGC
mS3u	*mC1*	10*-k*-m*-9*	GCTTGT GGAGATTTAGGCGTTG GTTCAGGTAGGTTCGG AATAGC
	mC2	6-j1-j2-l-k-10	CCGTTT CCACTCCTCC GCATTC GGTTATGGCTGGTTTC CAACGCCTAAATCTCC ACAAGC
	mC3u	m-l*	CCGAACCTACCTGAAC GAAACCAGCCATAACC
mS4	*mD1*	9-m-k	GCTATT CCGAACCTACCTGAAC CAACGCCTAAATCTCC
	mD2	10*-k*-l*-2T-h	GCTTGT GGAGATTTAGGCGTTG GAAACCAGCCATAACC TT CACTAACAACACCGCC
	mD3	j2-l-m*-9*	GCATTC GGTTATGGCTGGTTTC GTTCAGGTAGGTTCGG AATAGC
mP5	*mC2*	6-j1-j2-l-k-10	CCGTTT CCACTCCTCCGCATTC GGTTATGGCTGGTTTC CAACGCCTAAATCTCC ACAAGC
	mD2	10*-k*-l*-2T-h	GCTTGT GGAGATTTAGGCGTTG GAAACCAGCCATAACC TT CACTAACAACACCGCC
mP5-t	*mC2*	6-j1-j2-l-k-10	CCGTTT CCACTCCTCCGCATTC GGTTATGGCTGGTTTC CAACGCCTAAATCTCC ACAAGC
	mD2-T	10*-k*-l*-h	GCTTGT GGAGATTTAGGCGTTG GAAACCAGCCATAACC CACTAACAACACCGCC
mP5s	*mC2s*	6-j1-l-k-10	CCGTTT CCACTCCTCC GGTTATGGCTGGTTTC CAACGCCTAAATCTCC ACAAGC
	mD2s	10*-k*-l*-2T j2-h	GCTTGT GGAGATTTAGGCGTTG GAAACCAGCCATAACC TT GCATTC CACTAACAACACCGCC
mP5-t	*mC2s*	6-j1-l-k-10	CCGTTT CCACTCCTCC GGTTATGGCTGGTTTC CAACGCCTAAATCTCC ACAAGC
	mD2s-T	10*-k*-l*-j2-h	GCTTGT GGAGATTTAGGCGTTG GAAACCAGCCATAACC GCATTC CACTAACAACACCGCC
mS1au	*mA1*	7*-h*-j*-6*	GGTGTC GGCGGTGTTGTTAGTG GAATGCGGAGGAGTGG AAACGG
	mA2au	d1s-T1-j2s-i-h-7	CCAAACCTTCATCTTC T CATTC GCGATGTGGTGAGAGG CACTAACAACACCGCC GACACC
	mA3au	j-i*-T2-h	CCACTCCTCCGCATTC CCTCTCACCACATCGC TT CACTAACAACACCGCC
mS2au	*mB1*	6-j-h	CCGTTT CCACTCCTCCGCATTC CACTAACAACACCGCC
	mB2au	7*-h*-i*-j2*-d2-3	GGTGTC GGCGGTGTTGTTAGTG CCTCTCACCACATCGC GAATGC TACTCG CCTCTACTCA
	mB3au	6-j1-j2-i-j*-6*	CCGTTT CCACTCCTCC GCATTC GCGATGTGGTGAGAGG GAATGCGGAGGAGTGG AAACGG
S5	*A1*	2*-a*-c*-1*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGG
	A2	d1s-T2-b-a-2	CCAAACCTTCATCTTC TT GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCC
	fA3	o-b*	GGCACCTATCGACACCTCACGC CCAGGCCTCGTATCGCGAGTGC
	fA4	c-o*	CCAGATCAGCAGCCATTCGTTC GCGTGAGGTGTCGATAGGTGCC
S5x	*A1x*	2*-a*-c*-1x*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGGTG
	A2	d1s-T2-b-a-2	CCAAACCTTCATCTTC TT GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCC
	fA3	o-b*	GGCACCTATCGACACCTCACGC CCAGGCCTCGTATCGCGAGTGC
	fA4	c-o*	CCAGATCAGCAGCCATTCGTTC GCGTGAGGTGTCGATAGGTGCC
S5y	*A1*	2*-a*-c*-1*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGG
	A2y	d1s-T2-b-a-2y	CCAAACCTTCATCTTC TT GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCCTC
	fA3	o-b*	GGCACCTATCGACACCTCACGC CCAGGCCTCGTATCGCGAGTGC
	fA4	c-o*	CCAGATCAGCAGCCATTCGTTC GCGTGAGGTGTCGATAGGTGCC
S5xy	*A1x*	2*-a*-c*-1x*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGGTG
	A2y	d1s-T2-b-a-2y	CCAAACCTTCATCTTC TT GCACTCGCGATACGAGGCCTGG

			CAGTCCCAAGTCACCACCTAGC ACATCCTC
	fA3	o-b*	GGCACCTATCGACACCTCACGC CCAGGCCTCGTATCGCGAGTGC
	fA4	c-o*	CCAGATCAGCAGCCATTCGTTC GCGTGAGGTGTCGATAGGTGCC
S5-8	*A1xy*	2y*-a*-c*-1x*	GAGGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGGTG
	A2y	d1s-T2-b-a-2y	CCAAACCTTCATCTTC TT GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCCTC
	fA3	o-b*	GGCACCTATCGACACCTCACGC CCAGGCCTCGTATCGCGAGTGC
	fA4	c-o*	CCAGATCAGCAGCCATTCGTTC GCGTGAGGTGTCGATAGGTGCC
S6	*B1*	1-c-a	CCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2	2*-a*-b*-d2-3	GGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC TACTCG CCTCTACTCA
	fB3	b-o*	GCACTCGCGATACGAGGCCTGG GCGTGAGGTGTCGATAGGTGCC
	fB4	o-c*-1*	GGCACCTATCGACACCTCACGC GAACGAATGGCTGCTGATCTGG AAACGG
S6x	*B1x*	1x-c-a	CACCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2	2*-a*-b*-d2-3	GGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC TACTCG CCTCTACTCA
	fB3	b-o*	GCACTCGCGATACGAGGCCTGG GCGTGAGGTGTCGATAGGTGCC
	fB4	o-c*-1*	GGCACCTATCGACACCTCACGC GAACGAATGGCTGCTGATCTGG AAACGG
S6y	*B1*	1-c-a	CCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2y	2y*-a*-b*-d2-3	GAGGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC TACTCG CCTCTACTCA
	fB3	b-o*	GCACTCGCGATACGAGGCCTGG GCGTGAGGTGTCGATAGGTGCC
	fB4	o-c*-1*	GGCACCTATCGACACCTCACGC GAACGAATGGCTGCTGATCTGG AAACGG
S6xy	*B1x*	1x-c-a	CACCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2y	2y*-a*-b*-d2-3	GAGGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC TACTCG CCTCTACTCA
	fB3	b-o*	GCACTCGCGATACGAGGCCTGG GCGTGAGGTGTCGATAGGTGCC
	fB4	o-c*-1*	GGCACCTATCGACACCTCACGC GAACGAATGGCTGCTGATCTGG AAACGG
S6-8	*B1x*	1x-c-a	CACCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	B2y	2y*-a*-b*-d2-3	GAGGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC TACTCG CCTCTACTCA
	fB3	b-o*	GCACTCGCGATACGAGGCCTGG GCGTGAGGTGTCGATAGGTGCC
	fB4x	o-c*-1x*	GGCACCTATCGACACCTCACGC GAACGAATGGCTGCTGATCTGG AAACGGTG
S5au	*A1x*	2*-a*-c*-1x*	GGATGT GCTAGGTGGTGACTTGGGACTG GAACGAATGGCTGCTGATCTGG AAACGGTG
	fA2au	c2-b-a-2y	GCCATTCGTTC GCACTCGCGATACGAGGCCTGG CAGTCCCAAGTCACCACCTAGC ACATCCTC
	fA3au	o-b*-T2-a	GGCACCTATCGACACCTCACGC CCAGGCCTCGTATCGCGAGTGC TT CAGTCCCAAGTCACCACCTAGC
	fA4au	c-o*-d2-3	CCAGATCAGCAGCCATTCGTTC GCGTGAGGTGTCGATAGGTGCC TACTCG CCTCTACTCA
S6au	*B1x*	1x-c-a	CACCGTTT CCAGATCAGCAGCCATTCGTTC CAGTCCCAAGTCACCACCTAGC
	fB2au	2y*-a*-b*-c2*	GAGGATGT GCTAGGTGGTGACTTGGGACTG CCAGGCCTCGTATCGCGAGTGC GAACGAATGGC
	fB3au	1x-c-b-o*	CACCGTTT CCAGATCAGCAGCCATTCGTTC GCACTCGCGATACGAGGCCTGG GCGTGAGGTGTCGATAGGTGCC
	fB4au	d1s-T2-o-c*-1*	CCAAACCTTCATCTTC TT GGCACCTATCGACACCTCACGC GAACGAATGGCTGCTGATCTGG AAACGG
R	*d*	TET-d	/5TET/ CTCCAAACCTTCATCTTCTACTCG
	q	3*-d*-IABkFQ	TGAGTAGAGG CGAGTAGAAGATGAAGGTTTGGAG /3IABkFQ/

Note: domain c = c1 + c2, d = 5'CT + d1s + d2, g = g1 + g2, j = j1 + j2, j2=5'G + j2s, 1x = 5'CA + 1, 2y = 2 + 3'TC.

3.7 References

1. 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.

2. SantaLucia, J.; Hicks, D., The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 415-440.

3. Yurke, B.; Mills, A., Jr., Using DNA to power nanostructures. *Genet. Program. Evolvable Mach.* **2003,** *4* (2), 111-122.

4. Zhang, D. Y.; Winfree, E., Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **2009,** *131* (47), 17303-17314.

5. Zhang, D. Y.; Seelig, G., Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **2011,** *3* (2), 103-113.

6. Turberfield, A. J.; Mitchell, J. C.; Yurke, B.; Mills, A. P.; Blakey, M. I.; Simmel, F. C., DNA fuel for free-running nanomachines. *Phys. Rev. Lett.* **2003,** *90* (11), 118102.

7. Seelig, G.; Yurke, B.; Winfree, E., Catalyzed relaxation of a metastable DNA fuel. *J. Am. Chem. Soc.* **2006,** *128* (37), 12211-12220.

8. Jung, C.; Ellington, A. D., Diagnostic applications of nucleic acid circuits. *Acc. Chem. Res.* **2014,** *47* (6), 1825-1835.

9. Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E., Enzyme-free nucleic acid logic circuits. *Science* **2006,** *314* (5805), 1585-1588.

10. Qian, L.; Winfree, E., Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **2011,** *332* (6034), 1196-1201.

11. Qian, L.; Winfree, E.; Bruck, J., Neural network computation with DNA strand displacement cascades. *Nature* **2011,** *475* (7356), 368-372.

12. Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A., Programming biomolecular self-assembly pathways. *Nature* **2008,** *451* (7176), 318-322.

13. Zhang, D. Y.; Hariadi, R. F.; Choi, H. M. T.; Winfree, E., Integrating DNA strand-displacement circuitry with DNA tile self-assembly. *Nat. Commun.* **2013,** *4*.

14. Dirks, R. M.; Pierce, N. A., Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **2004,** *101* (43), 15275-15278.

15. Bois, J. S.; Venkataraman, S.; Choi, H. M.; Spakowitz, A. J.; Wang, Z. G.; Pierce, N. A., Topological constraints in nucleic acid hybridization kinetics. *Nucleic Acids Res.* **2005,** *33* (13), 4090-4095.

16. Li, B.; Ellington, A. D.; Chen, X., Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic Acids Res.* **2011,** *39* (16), e110.

17. 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.

18. Chen, Y.-J.; Dalchau, N.; Srinivas, N.; Phillips, A.; Cardelli, L.; Soloveichik, D.; Seelig, G., Programmable chemical controllers made from DNA. *Nat. Nanotechnol.* **2013,** *8* (10), 755-762.

19. Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D., Stacking nonenzymatic circuits for high signal gain. *Proc. Natl Acad. Sci. USA* **2013,** *110* (14), 5386-5391.

20. Olson, X.; Kotani, S.; Padilla, J. E.; Hallstrom, N.; Goltry, S.; Lee, J.; Yurke, B.; Hughes, W. L.; Graugnard, E., Availability: a metric for nucleic acid strand displacement systems. *ACS Synth. Biol.* **Article ASAP**.

21. Jiang, Y. S.; Bhadra, S.; Li, B.; Ellington, A. D., Mismatches improve the performance of strand-displacement nucleic acid circuits. *Angew. Chem. Int. Ed.* **2014,** ***53*** (7), 1845–1848.

22. Olson, X.; Kotani, S.; Yurke, B.; Graugnard, E.; Hughes, W. L., Kinetics of DNA strand displacement systems with locked nucleic acids. *Manuscript in Preparation* **2016**.

23. Jose, D.; Datta, K.; Johnson, N. P.; von Hippel, P. H., Spectroscopic studies of position-specific DNA "breathing" fluctuations at replication forks and primer-template junctions. *Proc. Natl Acad. Sci. USA* **2009,** *106* (11), 4231-4236.

24. Temsamani, J.; Kubert, M.; Agrawal, S., Sequence identity of the n-1 product of a synthetic oligonucleotide. *Nucleic Acids Res.* **1995,** *23* (11), 1841-1844.

25. Green, C.; Tibbetts, C., Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Res.* **1981,** *9* (8), 1905-1918.

26. Panyutin, I. G.; Hsieh, P., The kinetics of spontaneous DNA branch migration. *Proc. Natl Acad. Sci. USA* **1994,** *91* (6), 2021-2025.

27. Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P. K.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **2013,** *41* (22), 10641-10658.

28. Dabby, N. L., Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study. PhD thesis, California Institute of Technology: California Institute of Technology, 2013.

29. Seeman, N. C.; Kallenbach, N. R., DNA branched junctions. *Annu. Rev. Biophys. Biomol. Struct.* **1994,** *23* (1), 53-86.

30. Wang, J. S.; Zhang, D. Y., Simulation-guided DNA probe design for consistently ultraspecific hybridization. *Nat. Chem.* **2015,** *7* (7), 545-553.

31. McKinney, S. A.; Freeman, A. D. J.; Lilley, D. M. J.; Ha, T., Observing spontaneous branch migration of Holliday junctions one step at a time. *Proc. Natl Acad. Sci. USA* **2005,** *102* (16), 5715-5720.

32. Zhang, D. Y.; Winfree, E., Robustness and modularity properties of a non-covalent DNA catalytic reaction. *Nucleic Acids Res.* **2010,** *38* (12), 4182-4197.

33. Chen, X., Expanding the rule set of DNA circuitry with associative toehold activation. *J. Am. Chem. Soc.* **2012,** *134* (1), 263-271.

34. Zhu, J.; Zhang, L.; Dong, S.; Wang, E., Four-way junction-driven DNA strand displacement and its application in building majority logic circuit. *ACS Nano* **2013,** *7* (11), 10211-10217.

35. Seeman, N. C., Nanomaterials based on DNA. *Annu. Rev. Biochem.* **2010,** *79* (1), 65-87.

36. Bhadra, S.; Ellington, A. D., Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers. *Nucleic Acids Res.* **2014,** *42* (7), e58.

37. Green, A. A.; Silver, P. A.; Collins, J. J.; Yin, P., Toehold switches: de-novo-designed regulators of gene expression. *Cell* **2014,** *159* (4), 925-939.

38. Phillips, A.; Cardelli, L., A programming language for composable DNA circuits. *J. R. Soc. Interface* **2009,** *6* (Suppl 4), S419-S436.

39. Thachuk, C.; Winfree, E.; Soloveichik, D., Leakless DNA strand displacement systems. *Proceedings of the 21st International Conference on DNA Computing and Molecular Programming 9211* **2015**, 133−153.

40. Sabir, T.; Toulmin, A.; Ma, L.; Jones, A. C.; McGlynn, P.; Schröder, G. F.; Magennis, S. W., Branchpoint expansion in a fully complementary three-way DNA junction. *J. Am. Chem. Soc.* **2012,** *134* (14), 6280-6285.

41. Reynaldo, L. P.; Vologodskii, A. V.; Neri, B. P.; Lyamichev, V. I., The kinetics of oligonucleotide replacements. *J. Mol. Biol.* **2000,** *297* (2), 511-520.

42. Ladbury, J. E.; Sturtevant, J. M.; Leontis, N. B., The thermodynamics of formation of a three-strand, DNA three-way junction complex. *Biochemistry* **1994,** *33* (22), 6828-6833.

43. Zhong, M.; Rashes, M. S.; Leontis, N. B.; Kallenbach, N. R., Effects of unpaired bases on the conformation and stability of three-arm DNA junctions. *Biochemistry* **1994,** *33* (12), 3660-3667.

44. Lu, M.; Guo, Q.; Marky, L. A.; Seeman, N. C.; Kallenbach, N. R., Thermodynamics of DNA branching. *J. Mol. Biol.* **1992,** *223* (3), 781-789.

45. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003,** *31* (13), 3406-3415.

46. Kadrmas, J. L.; Ravin, A. J.; Leontis, N. B., Relative stabilities of DNA three-way, four-way and five-way junctions (multi-helix junction loops): unpaired nucleotides can be stabilizing or destabilizing. *Nucleic Acids Res.* **1995,** *23* (12), 2212-2222.

47. Wang, Y.; Mueller, J. E.; Kemper, B.; Seeman, N. C., Assembly and characterization of five-arm and six-arm DNA branched junctions. *Biochemistry* **1991,** *30* (23), 5667-5674.

48. Machinek, R. R. F.; Ouldridge, T. E.; Haley, N. E. C.; Bath, J.; Turberfield, A. J., Programmable energy landscapes for kinetic control of DNA strand displacement. *Nat. Commun.* **2014,** *5*, 5324.

49. Olson, X.; Kotani, S.; Padilla, J. E.; Hallstrom, N.; Goltry, S.; Lee, J.; Yurke, B.; Hughes, W. L.; Graugnard, E., Availability: a metric for nucleic acid strand displacement systems. *ACS Synth. Biol.* **2017,** *6*, 84–93.

50. 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. *J. Compu. Chem.* **2011,** *32* (1), 170-173.

51. You, Y.; Tataurov, A. V.; Owczarzy, R., Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers* **2011,** *95* (7), 472-486.

52. Cantor, C. R.; Warshaw, M. M.; Shapiro, H., Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides. *Biopolymers* **1970,** *9* (9), 1059-1077.

53. Kick, A.; Boensch, M.; Mertig, M., EGNAS: an exhaustive DNA sequence design algorithm. *BMC Bioinformatics* **2012,** *13*, 138.

CHAPTER FOUR: FINAL CONCLUSIONS

4.1 Conclusion and Future Work

This dissertation addressed two issues of DNA strand displacement systems for its application in medical diagnosis.[1] The first one is selectivity, which was addressed in **Chapter Two**. We explored the use of LNA[2] as a simple method to improve the selectivity for single-nucleotide mutations. For ease of analysis, double-stranded DNA complexes[3] and a hairpin structure[4] were chosen as model systems without the functionality of signal amplification. Our result showed that LNA substitutions can greatly enhance the selectivity of those model systems. Over two orders of magnitude improvement in selectivity appears to be good enough for single-nucleotide mutations for the tested sequence. Therefore, it will be interesting to combine hybrid DNA/LNA strand displacement with available biotechnologies.[5] In terms of design guideline, although more LNA substitutions generally results in higher selectivity, the introduction of LNA-LNA base pairs at locations other than the distal end of the branch migration domain has to be avoided. In addition, we did not observe the improvement of selectivity in the toehold exchange system,[6] which is an important mechanism for dynamic DNA nanotechnology.[7] Since we tested only one type of LNA substitution for one toehold exchange system, further investigation is necessary in order to draw deeper conclusions. Previous work by Ping Olson *et al*. also used the simple strand displacement system outlined here and showed that LNA substitutions can indeed reduce leakage.[8] Thus, our group demonstrated that LNA can enhance selectivity and reduce leakage in simple

systems. Therefore, the next step will be to use LNA substitutions for more complicated systems, such as catalytic substrates (hairpin,[9] linear,[10] and multi-arm junction[11]), and test whether the lessons we learned from simple systems can be easily applied to those systems or not. The large variations of designs tested so far will go through additional selection during this process, and we may be able to obtain more universal guidelines for the use of LNA in strand displacement systems. Although very expensive, LNA substitutions are a simple and powerful method for harnessing strand displacement systems, and widespread use of LNA may reduce its cost in the future.

The second issue we addressed is leakage, which was studied in **Chapter Three**. Catalytic systems operated via DNA strand displacement perform signal amplification in solution in a way that is similar to which transistors perform amplification in electrical circuits. This quality is important not only for medical diagnostics but also for molecular computation. To date, there are two types of catalytic substrates used widely. The first substrate is the hairpin substrate,[9] which was developed in 2004 by Dirks *et al*. The hairpin substrate does not require PAGE purification for use, and even has a potential to be generated *in vivo* and perform signal amplification for applications in synthetic biology.[12] The second substrate is the linear substrate,[10] which was developed in 2007 by Zhang *et al.* Since then, the linear substrates were adapted for molecular computation because they readily connected to form larger circuits.[13] Our contribution to the field was to find out that multi-arm junction structures, which have been central building blocks in structural DNA nanotechnology,[14] have exceedingly low leakage and smooth catalytic reactions when used as substrates for chemical reaction networks. In terms of sensitivity, which is the best metric for diagnostics, our results showed the detection of an on-target

sequence having a 1 pM concentration. While competitive in the field of dynamic DNA nanotechnology, this value is not adequate for practical applications, where sensitivity to aM level concentrations will be required. In general, the structure dictates the function of a molecular machine. In other words, each unique structure will likely find its unique application space. Since multi-arm junction structures appear frequently in structural DNA nanotechnology but not dynamic DNA nanotechnology, applications have yet to be explored. If we aim for the further reduction of leakage and improvement of sensitivity, we likely need a completely different approach for leakage reduction in order to achieve sensitivity which can compete with PCR. I propose the physical separation of substrates in solution.

4.2 References

1. Jung, C.; Ellington, A. D., Diagnostic applications of nucleic acid circuits. *Acc. Chem. Res.* **2014,** *47* (6), 1825-1835.
2. K. Singh, S.; A. Koshkin, A.; Wengel, J.; Nielsen, P., LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chemical Communications* **1998,** (4), 455-456.
3. Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P. K.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **2013,** *41* (22), 10641-10658.
4. Green, S. J.; Lubrich, D.; Turberfield, A. J., DNA hairpins: fuel for autonomous DNA devices. *Biophysical journal* **2006,** *91* (8), 2966-2975.
5. Khodakov, D.; Wang, C.; Zhang, D. Y., Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Advanced drug delivery reviews* **2016,** *105*, 3-19.
6. Zhang, D. Y.; Winfree, E., Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **2009,** *131* (47), 17303-17314.

7. Zhang, D. Y.; Seelig, G., Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **2011,** *3* (2), 103-113.

8. Olson, X.; Kotani, S.; Yurke, B.; Graugnard, E.; Hughes, W. L., Kinetics of DNA Strand Displacement Systems with Locked Nucleic Acids. *The Journal of Physical Chemistry B* **2017,** *121* (12), 2594-2602.

9. Dirks, R. M.; Pierce, N. A., Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **2004,** *101* (43), 15275-15278.

10. 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.

11. Kotani, S.; Hughes, W. L., Multi-Arm Junctions for Dynamic DNA Nanotechnology. *Journal of the American Chemical Society* **2017,** *139* (18), 6363-6368.

12. Bhadra, S.; Ellington, A. D., Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers. *Nucleic Acids Res.* **2014,** *42* (7), e58.

13. Soloveichik, D.; Seelig, G.; Winfree, E., DNA as a universal substrate for chemical kinetics. *Proc. Natl Acad. Sci. USA* **2010,** *107* (12), 5393-5398.

14. Seeman, N. C., Nanotechnology and the double helix. *Scientific American* **2004,** *290* (6), 64-75.