

KINETIC CONTROL OF NUCLEIC ACID STRAND DISPLACEMENT REACTIONS

by

Xiaoping Olson

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Materials Science and Engineering

Boise State University

December 2016

© 2016

Xiaoping Olson

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

**DEFENSE COMMITTEE AND FINAL READING APPROVALS**

of the dissertation submitted by

Xiaoping Olson

Dissertation Title: Kinetic Control of Nucleic Acid Strand Displacement Reactions

Date of Final Oral Examination: 23 September 2016

The following individuals read and discussed the dissertation submitted by student Xiaoping Olson, and they evaluated her 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

Elton Graugnard, Ph.D.

Member, Supervisory Committee

Bernard Yurke, 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 thesis was approved by the Graduate College.

## DEDICATION

This dissertation would not have been possible without the loving support of my husband, Jay, and our two sons, Alex and Evan.

Jay is a wonderful life partner and also my mentor in pursuit of science. His passion and persistence in blazing new trails has been an inspiration, and his loving words of encouragement have carried me through the challenges of graduate school life.

Alex and Evan, and their fresh view of the world have constantly humbled and refreshed my mind in a way that will be impossible for them to appreciate until much later. They have brought me into a new world as much as I have brought them. Their laugh and cry refresh my soul daily.

To Jay, Alex, and Evan I dedicated this dissertation.

## ACKNOWLEDGEMENTS

First, I would like to express my sincere appreciation and thanks to my advisor, mentor, and friend, Dr. William L. Hughes. Will recruited me to the Nanoscale Materials and Device Research group at Boise State University four years ago, and in that time I could not have asked for a more perfect advisor. He has provided the ideal mix of opportunities and freedom to come up my own ideas, while helping to manage the many projects we have pursued together in collaboration with my supervisory committee. Will has been extremely helpful and supportive whenever I ran into obstacles during my research. Without his endless support and mentoring, I would not have finished the journey to complete my dissertation.

I also owe an enormous debt of gratitude to Drs. Bernard Yurke, Elton Graugnard and Peter Allen for many insightful discussions on my projects, creating a clear direction for my research. Dr. Yurke was always open and enthusiastic in discussing theoretical challenges, whenever I had questions. Dr. Graugnard continually inspired me to become a good experimentalist, with persistence in my research. Dr. Allen, as an external committee member, provided valuable feedback and a fresh perspective for my research direction.

Many thanks are also due to my lab colleagues, Dr. Natalya Hallstrom, Dr. Jennifer E. Padilla, Shohei Kotani, Mike Tobiason, Sara Goltry, Saodao Takabayashi, Will Klein, and Brittany Cannon. Their contributions and insights were invaluable on

numerous occasions, and they made my time working a challenging set of projects far more enjoyable as well.

Finally, my deepest appreciation goes to my family. My husband, Dr. Jay Olson has provided me with unfailing support and encouragement throughout my time in graduate school. My two sons, Alex and Evan have inspired me and uplifted my spirit each day. My parents and my parents-in-law who have provided far more help and generosity than could be expected of anyone in their position. This dissertation would not have been possible without their love and support.

## ABSTRACT

Nucleic acids are information-dense, programmable polymers that can be engineered into primers, probes, molecular motors, and signal amplification circuits for computation, diagnostic, and therapeutic purposes. Signal amplification circuits increase the signal-to-noise ratio of target nucleic acids in the absence of enzymes and thermal cycling. Amplification is made possible via toehold mediated strand displacement – a process where one nucleic acid strand binds to a nucleation site on a complementary helix, which then displaces one of the two strands in a nucleic acid complex. When compared to polymerase chain reactions (PCR), the sensitivity and stability of toehold-mediated strand displacement reactions suffer from circuit leakage – reactions of the system in the absence of an initiator. Presented here, from a materials science and engineering perspective, defect engineering has improved the leakage performance of model strand displacement systems made from DNA. Engineered defects used in this study included mismatched base pairs and alternative nucleic acids – both of which are known to impact the stability of hybridization.

To identify sources of leakage in a model signal amplification circuit, availability was defined as the probability that a DNA base (A.T.C.G) was unpaired at equilibrium. This design metric was calculated using NUPACK, a thermodynamic modeling tool. To further understand the relationship between leakage rates and secondary structures, mutual availability was defined as the sum of all pairwise products of the availabilities of the corresponding bases in solution. This thermodynamic analysis yielded rational design

principles for how to minimize leakage by as much as 4-fold by site-specifically introducing mismatched base pairs into DNA duplex regions. To further reduce leakage, chemically modified locked nucleic acids (LNAs) were site-specifically introduced into a model DNA strand displacement system. Briefly described, LNAs are geometrically restricted RNA analogues with enhanced thermo-mechanical stability towards their complement base. When compared to a DNA control with identical sequences, the leakage exhibited by a hybrid DNA/LNA system was reduced from  $1.48 \text{ M}^{-1}\text{s}^{-1}$  (for the DNA system) to  $0.03 \text{ M}^{-1}\text{s}^{-1}$ . In addition, the signal-to-noise ratio increased ~50-fold for a similar hybrid system.

This research provides insight into the sources of leakage in DNA strand-displacement systems, as well as how to maximize strand-displacement performance via the selective introduction of hybridization defects. Rational design of future nucleic acid signal amplification circuits will lead to broader applications in a variety of fields that range from DNA computation to point-of-care diagnostics and therapeutics.

## TABLE OF CONTENTS

DEDICATION .....	iv
ACKNOWLEDGEMENTS .....	v
ABSTRACT .....	vii
LIST OF TABLES .....	xiii
LIST OF FIGURES .....	xiv
LIST OF ABBREVIATIONS AND SYMBOLS .....	xxi
CHAPTER ONE: INTRODUCTION.....	1
Prelude .....	1
1.1 Strand Displacement .....	2
1.2 Strand Displacement Tools .....	3
1.3 Dissertation Layout.....	4
REFERENCES .....	6
CHAPTER TWO: AVAILABILITY – A METRIC FOR NUCLEIC ACID STAND DISPLACEMENT SYSTEMS .....	11
Abstract .....	11
2.1 Introduction.....	12
2.1.1 Leakage Problem .....	12
2.1.2 Thermal Fluctuations in DNA .....	15
2.1.3 Insight from Secondary Structure .....	17
2.2 Results and Discussion .....	18

2.2.1 Effect of Mismatch on Leakage Rate Constants.....	18
2.2.2 Availability .....	18
2.2.3 Mutual Availability.....	24
2.2.4 Catalyzed Reactions.....	27
2.2.5 System Performance .....	29
2.2.6 Analysis of Other Networks.....	32
2.3 Conclusion .....	32
2.4 Methods.....	33
2.4.1 Chemicals and DNA Complex Purification.....	33
2.4.2 Spectrofluorimetry .....	34
2.4.3 Reaction Measurements .....	35
2.5 Author Information .....	36
2.5.1 Corresponding Authors .....	36
2.5.2 Author Contributions .....	36
2.6 Acknowledgements.....	36
REFERENCES .....	37
2.7 Supplementary Information .....	43
2.7.1 Strand sequences.....	43
2.7.2 Extinction coefficient calculation .....	45
2.7.3 Reaction kinetics models .....	45
2.7.4 Transient binding of the substrate and fuel.....	48
2.7.5 Minimum free energy structures of the fuel strands .....	49
2.7.6 Availabilities of fuel strand bases .....	50

2.7.7 Analysis in the context of the Intuitive Energy Landscape (IEL) model.....	52
2.7.8 Multiple location fuel modifications.....	56
2.7.9 Analysis of catalytic rates .....	61
2.7.10 Rate constants and mutual availability .....	63
2.7.11 Mutual availability and rate constants in a hairpin system .....	65
REFERENCES .....	66
<b>CHAPTER THREE: KINETICS OF DNA STRAND DISPLACEMENT SYSTEMS WITH LOCKED NUCLEIC ACIDS.....</b>	
Abstract .....	67
3.1 Introduction.....	68
3.2 Results and Discussion .....	71
3.2.1 Leakage reactions between a DNA invader and hybrid DNA/LNA Substrates .....	71
3.2.2 Invasion reactions between a DNA invader and hybrid DNA/LNA Substrates .....	76
3.2.3 Leakage reactions between hybrid DNA/LNA invaders and hybrid DNA/LNA Substrates .....	78
3.2.4 Invasion reactions between hybrid DNA/LNA invaders and hybrid DNA/LNA Substrates .....	81
3.2.5 Optimize the performance enhancement ratio of hybrid DNA/LNA systems .....	83
3.3 Conclusions.....	85
3.4 Methods.....	86
3.4.1 Materials .....	86
3.4.2 Purification.....	87
3.4.3 Spectrofluorimetry .....	87

3.5 Author Information .....	88
3.5.1 Corresponding Authors .....	88
3.5.2 Author Contributions .....	88
3.6 Acknowledgements.....	88
REFERENCES .....	89
3.7.1 Strand sequences.....	94
3.7.2 Reaction models and rate constants .....	96
3.7.3 Kinetics of Reporters .....	101
3.7.4 Control experiments.....	102
3.7.5 Validating the source of leakage.....	103
3.7.6 The effect of toehold length on reaction kinetics.....	105
3.7.7 Hybrid systems.....	106
REFERENCES .....	108
CHAPTER FOUR: FINAL CONCLUSIONS.....	109

## LIST OF TABLES

Table 2.1	Substrate and catalyst sequences and schematic.....	43
Table 2.2	Reporter sequences and schematic.....	43
Table 2.3	Fuel sequences .....	43
Table 2.4	Rate constants and total mutual availability .....	64
Table 3.1	Substrate sequences .....	94
Table 3.2	Invader sequences .....	95
Table 3.3	Reporter sequences .....	96
Table 3.4	Reaction concentrations and rate constants .....	99

## LIST OF FIGURES

- Figure 2.1. Domain representation of the catalytic DNA strand displacement system from Zhang et al.<sup>24</sup> and four example leakage pathways. (a) In the catalyzed strand displacement pathway of the reaction network, a catalyst strand initiates a reaction cycle driven forward thermodynamically by a net gain in entropy. The strand displacement exchanges the catalyst for the signal strand and exposes a sequestered toehold on the substrate backbone for the fuel, which reacts with the intermediate to complete the cycle and form a waste duplex. Sequences and domains are listed in Section. 2.8.1 (b) In the four leakage pathways, the fuel reacts with the substrate backbone in the absence of a catalyst by exploiting fraying at the 5', nick, and 3' locations of the substrate. .... 14
- Figure 2.2. (a) Sequence and domain representation of the substrate with fraying locations highlighted. (b) Sequence and domain representation of the original fuel strand. Corresponding to the fraying locations of the substrate, the locations of fuel base mismatches are numbered, highlighted, and shown in bold font. They are 5' end (bases 1 and 2), nick (bases 24 and 25), and 3' end (bases 43 and 44). (c) Leakage rate constants for fuel modifications. The concentrations for leakage reactions are fuel (1300 nM), substrate (14 nM), and reporter (20 nM). The black bar represents the leakage rate with the original fuel strand. Pink, orange and blue bars represent leakage rates for fuels with 1 and 2 nt modifications at 5', nick, and 3' locations, respectively. The rates are labeled by the identity of the modified base and its location on the fuel (see panel (b) for locations and original base identities). For example, G<sub>1</sub>T<sub>2</sub> indicates that base 1 was changed from C to G and base 2 was changed from C to T. Error bars show the standard deviation from the mean for select samples in triplicate to estimate experimental error. .... 16
- Figure 2.3. Minimum free energy (MFE) structures and base availabilities for select fuel strands and the substrate backbone. (a) MFE structures of the original fuel strand and fuel modifications A<sub>1</sub>A<sub>2</sub>, G<sub>1</sub>T<sub>2</sub>, G<sub>1</sub>G<sub>2</sub>, and C<sub>24</sub> calculated by NUPACK. The Gibbs free energy of each structure is provided in units of kcal/mol. (b) Base availabilities for the original fuel and fuel modifications A<sub>1</sub>A<sub>2</sub>, G<sub>1</sub>T<sub>2</sub>, and G<sub>1</sub>G<sub>2</sub> (upper plot) and the difference in base availabilities ( $\Delta A$ ) for each modification relative to the original fuel (lower plot). (c) Base availabilities for the original fuel and fuel modification C<sub>24</sub> (upper plot) and the difference in

availability relative to the original fuel (lower plot). (d) MFE structure of the substrate calculated by NUPACK. (e) Availability of each base in the backbone strand of the substrate. Because the fuel strand hybridizes with the backbone strand on the substrate, the base positions of the backbone strand were plotted on the x axis and labeled to correspond to the complement of the fuel strand..... 23

Figure 2.4. (a) Leakage rate constants for each fuel modification plotted versus total mutual availability between the fuel strand and the backbone strand on the substrate. The leakage rate for the original fuel is shown in black while the 5' end, nick and 3' end fuel modifications are shown in blue, green and orange, respectively. Representative error bars of select samples are shown, indicating that the scatter of the data is greater than the experimental error. Error bars represent the standard deviation of the mean of three samples. (b) Natural log plot of the leakage rate constant versus the total mutual availability. The green, and blue lines are the fits for the nick modifications, 5' end modifications..... 26

Figure 2.5. Rate constants of catalyzed reactions between the catalyst (1 nM), fuel (13 nM), substrate (14 nM), and reporter (20 nM) monitored via fluorescence. The black bar represents the original fuel strand. Pink, orange and blue bars represent 1 nt and 2 nt modifications at 5', nick, and 3' locations, respectively. Error bars show the standard deviation from the mean for select samples in triplicate to estimate experimental error. .... 29

Figure 2.6. (a) Ratio of the catalyzed to leakage reaction rates ( $k_{cat}/k_{leak}$ ) for single location fuel modifications to evaluate overall system performance. Catalyzed reactions were performed with the catalyst (1 nM), fuel (13 nM), substrate (14 nM), and reporter (20 nM), monitored via fluorescence, and uncatalyzed leakage reactions were performed with fuel (1300 nM), substrate (14 nM), and reporter (20 nM). The black bar represents the original fuel strand. Pink, orange, and blue bars represent 1 and 2 nt modifications at 5', nick, and 3' locations, respectively. Error bars show the standard deviation from the mean for select samples in triplicate to estimate experimental error. (b) Representative fluorescence data of catalytic reactions: the original fuel (empty black circles), G<sub>1</sub>T<sub>2</sub> modified fuel (empty red triangles), C<sub>24</sub>T<sub>25</sub> modified fuel (yellow diamonds), and G<sub>43</sub>A<sub>44</sub> modified fuel (empty blue stars). (c) Representative fluorescence data of leakage reactions: the original fuel (solid black circles), G<sub>1</sub>T<sub>2</sub> modified fuel (solid red triangles), C<sub>24</sub>T<sub>25</sub> modified fuel (solid yellow diamonds), and G<sub>43</sub>A<sub>44</sub> modified fuel (solid blue stars). The gray lines are the calculated fits to each curve, and the solid blue and purple lines

	represent reactions between the reporter and the original fuel and the reporter and the substrate, respectively.....	31
Figure 2.7.	Schematic of the substrate complex (a) and the network for reporting the presence of single-stranded Signal strands (b).....	43
Figure 2.8.	The transient binding between the substrate toehold and the fuel. The y domain of the fuel strand and y* domain of the substrate toehold can hybridize as shown in reaction I. The x domain on the 5' end of the fuel will further displace the x domain of the signal strand via 3-way branch migration (reaction II). This interaction might interfere with the 3' end fuel binding with the substrate and thus affect the leakage rate constants.....	48
Figure 2.9.	Minimum free energy (MFE) structures of fuels with single-location modifications as calculated by NUPACK. The Gibbs free energy of each structure is shown in kcal/mol. ....	49
Figure 2.10.	Leakage rate constants in descending order and corresponding base availabilities of fuel strands. Each base of the fuel strand is represented as a solid circle colored according to the NUPACK calculated availability. Mismatched bases are circled in gray. ....	50
Figure 2.11.	Catalyzed reaction rate constants in descending order and corresponding fuel base availabilities. Each base of the fuel strand is represented as a solid circle colored according to NUPACK calculated availability. The four bases complementary to the toehold are highlighted with the gray box. Mismatched bases are circled in gray. ....	51
Figure 2.12.	Intuitive energy landscape (IEL) for two of the four proposed leakage pathways for G <sub>1</sub> T <sub>2</sub> fuel modification. Red states indicate mismatch penalties. (a) Leakage initiated at 3' end of the fuel (base 1), (b) Leakage initiated at nick left of the fuel (base 24). Yellow circles denote fraying locations. States A-G are described in the text. For illustration, the values of $\Delta G_{bp}$ , $\Delta G_p$ , $\Delta G_s$ , and $\Delta G_{init}$ were taken from Srinivas et al. <sup>4</sup> $ \Delta G_{p1} $ is the energy difference from no overhang to one overhang in the nick, $ \Delta G_{p2} $ is the energy difference from no overhang to two overhangs in the nick, and $ \Delta G_{p3} $ is the energy difference from one overhang to two overhangs in the nick. For comparison, point "Orig." shows the final state energy when the original fuel is used.....	54
Figure 2.13.	Intuitive energy landscape (IEL) for two of the four proposed leakage pathways for G <sub>1</sub> T <sub>2</sub> fuel modification. Red states indicate mismatch penalties. (a) Leakage initiated at nick right of the fuel (base 25), (b) Leakage initiated at 3' end of the fuel (base 44). States A-G are described in the text. Yellow circles denote fraying locations. For illustration, the values of $\Delta G_{bp}$ , $\Delta G_p$ , $\Delta G_s$ , and $\Delta G_{init}$ were taken from Srinivas et al. <sup>4</sup>	

$|\Delta G_{p1}|$  is the energy difference from no overhang to one overhang in the nick,  $|\Delta G_{p2}|$  is the energy difference from no overhang to two overhangs in the nick, and  $|\Delta G_{p3}|$  is the energy difference from one overhang to two overhangs in the nick. For comparison, point “Orig.” shows the final state energy when the original fuel is used. .... 55

- Figure 2.14. (a) Leakage rate constants of multiple location fuel modifications. (b) Natural log plot of the leakage rate constant versus the mutual availability between the fuel strand and the backbone on the substrate. Representative error bars of selected samples are shown, indicating that the scatter of the data is greater than the experimental error. The red line is the fit for all modifications combined. (c) Catalytic rate constants of multiple location fuel modifications. Gray bars represent substrate and fuel reacted with original catalyst and red bars represent substrate and fuel reacted with catalyst with 1 nt deletion at 5' end. (d) The ratio  $k_{cat}/k_{leak}$  for multiple location fuel modifications. Gray bars represent substrate and fuel reacted with original catalyst and red bars represent substrate and fuel reacted with catalyst with 1 nt deletion at 5' end. .... 58
- Figure 2.15. Minimum free energy (MFE) structures of fuels with multiple-location modifications as calculated by NUPACK. The Gibbs free energy for each structure is shown in kcal/mol. .... 59
- Figure 2.16. (a) Minimum free energy (MFE) structures of original fuel and  $G_1G_1T_{24}A_{25}$  modification as calculated by NUPACK and the Gibbs free energy of each structure. (b) Base availabilities for the original fuel and fuel modification  $G_1G_1T_{24}A_{25}$ . (c) An alternative reaction pathway to Fig. 1b between the modified fuel ( $G_1G_2T_{24}A_{25}$ ) and substrate. .... 60
- Figure 2.17. (a) Leakage rate constants for each hairpin 2 modification plotted versus total mutual availability between hairpin 1 and hairpin 2. The leakage rate for the original hairpins is shown in black while the 1 nt, 2 nt, and 3 nt hairpin 2 modifications are shown in blue, red and dark yellow, respectively. (b) Natural log plot of the leakage rate constant versus the total mutual availability. The solid gray line is the fit for all modifications combined with an adjusted R-squared 0.82. .... 65
- Figure 3.1 Schematic of a nucleic acid based, toehold mediated strand displacement system. Functional domains are represented by numbers and complementary domains are denoted by numbers with asterisks. The Substrate complex includes signal (s) and backbone (b) strands, while the Reporter complex includes dye (d) and quencher (q) strands. Various LNA substitutions have been made to the signal, backbone, and invader (i) sequences in this study. During invasion, the invader hybridizes with the backbone at domain 5\* and displaces the signal strand via three way branch migration. The invader and backbone form a Waste complex. The

free signal strand then reacts with the Reporter complex and releases a dye strand, which is monitored by a fluorometer. .... 71

- Figure 3.2. (A) The original sequence-level Substrate (sb0). Black and red letters respectively denote DNA and LNA nucleotides. (B) Hybrid DNA/LNA Substrates. Substrates b1-b6 selectively substituted LNA for DNA nucleotides on the backbone strand. Substrates s1-s4 selectively substituted LNA for DNA nucleotides on the signal strand. Substrates sb1-sb8 selectively substituted LNA for DNA nucleotides on both the signal and backbone strands. (C) Zero nucleotide (nt) invaders (**i1**, **i2** and **i3**) were used for the leakage reactions and 6 nt toehold invaders (**i4**, **i5** and **i6**) were used for the invasion reactions..... 73
- Figure 3.3. The effects of LNA substitutions on strand displacement kinetics. (A) Leakage rate constants for multiple Substrates that were exposed to DNA invader **i1** with a zero nt toehold. (B) Leakage kinetics for select Substrates (20 nM) that were exposed to DNA invader **i1** (2  $\mu$ M). The black line is the original Substrate (sb0) and the red line is the background reaction when the Reporter complex (40 nM) was mixed with DNA invader **i1** (2  $\mu$ M) without the Substrate. (C) Invasion rate constants for multiple Substrates that were exposed to a DNA invader **i4** with a 6 nt toehold. (D) The performance enhancement factor of each Substrate was calculated by taking the ratio of the rate constants in (C) and (A). Error bars represent the standard deviation from three reactions with different invader concentrations. .... 75
- Figure 3.4. Leakage rate constants for select Substrates and zero toehold invaders with and without LNA substitutions. Black and red letters respectively denote DNA and LNA nucleotides. Error bars are the standard deviation from three reactions with different invader concentrations. .... 80
- Figure 3.5. (A) Invasion rate constants for Substrates with LNA nucleotides on both signal and backbone strands that were exposed to pure DNA and hybrid DNA/LNA invaders with a 6 nt toeholds. (B) The performance enhancement ratio of each Substrate was calculated by taking the ratio of the invasion and leak rate constants. Error bars represent the standard deviation from three reactions with different invader concentrations. .... 81
- Figure 3.6. Optimized hybrid DNA/LNA systems. (A) Strands and sequences for four unique systems. Black and red letters respectively denote DNA and LNA. Each system consists of a zero toehold invader, a 6 nt toehold invader, and a Substrate. For example, “sy1” denotes system 1, in which **i1** is the zero toehold invader, **i6** is the 6 nt toehold invader and sb0 is the original Substrate. (B) Leakage rate constants of each system. (C) Invasion rate constants of each system. (D) The performance enhancement ratio of each system was calculated by taking the ratio of the rate constants in (C) and

(B). Error bars represent the standard deviation from three reactions with different invader concentrations. .... 84

- Figure 3.7. Schematic of a nucleic acid based, toehold mediated strand displacement system. Functional domains are represented by numbers and complementary domains are denoted by numbers with asterisks. The Substrate complex includes signal (s) and backbone (b) strands, while the Reporter complex includes dye (d) and quencher (q) strands. Various LNA substitutions have been made to the signal, backbone, and invader (i) sequences in this study. The toehold length of domain 5 on the invader (i) varies from zero to 14 nucleotides (nt). During invasion, the invader hybridizes with the backbone at domain 5\* and displaces the signal strand via three way branch migration. The invader and backbone form a Waste complex. The free signal strand then reacts with the Reporter complex and releases a dye strand, which is monitored by a fluorometer. .... 97
- Figure 3.8. The reaction kinetics between Reporters dq1 and dq2 and the signal strand. (A) Reporters and signal strand in sequence level. (B) Reporter dq1 kinetics. The final fluorescence intensity was normalized to one arbitrary units (a.u). The red dotted line is the fit. (C) Reporter dq2 kinetics. The final fluorescence intensity was normalized to 1 arbitrary units (a.u). The red dotted line is the fit. .... 102
- Figure 3.9. (A) Reporter (dq1), invader (i1), Substrates (sb0 and b6) are depicted at the sequence level. (B) Leakage controls. The black line denotes 40 nM Reporter dq1. The red line denotes the reaction of 40 nM Reporter dq1 and 6 μM zero toehold invader i1. The blue line denotes the reaction of 40 nM Reporter dq1 and 20 nM original Substrate (sb0). The green line denotes the reaction of 40 nM Reporter dq1 and 20 nM LNA substituted Substrate b6..... 103
- Figure 3.10. Leakage kinetics. 1 a.u. is corresponding to 20 nM signal strand. (A) The original sequence-level Substrate (sb0) and invaders. (B) Availability of Substrate (sb0) backbone. (C) Leakage kinetics with zero toehold invaders with deletions. Invader i1d5 denotes zero toehold invader with 2 nt deletion at 5' end. Invader i1d3 denotes zero toehold invader with 2 nt deletion at 3' end. Invader i1db denotes zero toehold invader with 2 nt deletion at both 5' and 3' ends. .... 104
- Figure 3.11. Strand displacement rate constants and the toehold length of DNA invaders. sb0 is the original Substrate. b6 represents LNA substituted Substrate with 15-LNA on backbone strand. The rate constants are plotted against different toehold lengths of DNA invaders. Error bars are the standard deviation of three reactions with three different concentrations. .... 106

Figure 3.12. Optimized hybrid DNA/LNA systems. Strands and sequences for sy5 and sy6 systems. Black and red letters respectively denote DNA and LNA. 107

## LIST OF ABBREVIATIONS AND SYMBOLS

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic acid
LNA	Lock Nucleic Acid
DSD	DNA Strand Displacement
IDT	Integrated DNA Technologies
PCR	Polymerase Chain Reactions
°C	Degrees Celsius
nt	Nucleotide
$k$	Kinetics Rate Constant
$k_{leak}$	Kinetic Rate Constant of a Leakage Reaction
$k_{cat}$	Kinetic Rate Constant of a Catalytic Reaction
$k_{bi}$	Kinetic Rate Constant of a Bi-Molecular Reaction
$k_{bi}$	Kinetic Rate Constant of a Uni-Molecular Reaction
$\Delta G_s$	Gibbs Free Energy Change of the Sawtooth Height
$\Delta G_p$	Gibbs Free Energy Change of the Penalty
$\Delta G_{bp}$	Gibbs Free Energy Change during Base Pairing
$\Delta G_{init}$	Gibbs Free Energy Change during the Initiation of Branch Migration
$\Delta G_{p1}$	Gibbs Free Energy Change between No Overhang and One Overhang in the Nick

$\Delta G_{p2}$	Gibbs Free Energy Change between No Overhang and Two Overhangs in the Nick
$\Delta G_{p3}$	Gibbs Free Energy Change between One Overhang and Two Overhangs in the Nick
kcal/mol	Kilocalorie per mole
mM	Millimoles
nM	Nanomoles
n	Number of Moles
nm	Nanometer
R	Universal Gas Constant
T	Temperature
$\mu\text{L}$	Microliter
C	Cytosine
G	Guanine
T	Thymine
A	Adenine
MFE	Minimum Free Energy
$\text{Mg}^{2+}$	Magnesium
$\text{Na}^+$	Sodium
<i>M</i>	Total Mutual Availability
TE	Tris-EDTA
TAE	Tris Base, Acetic Acid and EDTA
EDTA	Ethylene diamine tetra acetic
$\text{Mg}(\text{C}_2\text{H}_3\text{O}_2)_2$	Magnesium Acetate

MgCl <sub>2</sub>	Magnesium Chloride
N-PAGE	Native polyacrylamide gel electrophoresis
IEL	Intuitive Energy Landscape
a.u.	Arbitrary Unit
RFU	Relative Fluorescence Unit

## CHAPTER ONE: INTRODUCTION

### Prelude

Nucleic acids are information-dense, programmable polymers that can be engineered into primers,<sup>1</sup> probes,<sup>2, 3</sup> molecular motors,<sup>4-6</sup> and signal amplification circuits for computation,<sup>7, 8</sup> diagnostic,<sup>9</sup> and therapeutic purposes.<sup>10</sup> Programming with nucleic acids is made possible by Watson-Crick hybridization.<sup>11</sup> During hybridization, adenine (A) base-pairs with thymine (T), and guanine (G) base-pairs with cytosine (C) in DNA. In RNA, thymine is replaced by uracil (U). Once hybridized, two complementary oligonucleotides, with anti-parallel orientation, form a double helix. The stability of the double helix is dominated by base stacking between the aromatic rings. It is further stabilized by hydrogen bonds between complementary bases. For example, G-C base pairs have 3 hydrogen bonds while T-A base pairs have 2 hydrogen bonds.<sup>11</sup> Although a very small structural difference, the type of nucleotide bases and the number of hydrogen bonds between the bases changes the thermo-mechanical performance of a double helix. For example, the stability and temperature-dependent behavior of any DNA duplex, from the knowledge of the base sequence, can be predicted by thermodynamic data.<sup>12, 13</sup>

Because of known structure-property-performance relationships of nucleic acids, DNA can be rationally engineered into static and dynamic systems that include motifs,<sup>5, 7</sup> origami,<sup>14</sup> bricks,<sup>15</sup> and strand-displacement systems.<sup>16</sup> Further understanding of their structure-property relationships is essential for improving their performance. As an aspirational goal, signal amplification circuits for low-cost and early stage diagnosis of

disease were targeted here. Briefly described, signal amplification circuits increase the signal-to-noise ratio of target nucleic acids in the absence of enzymes or thermal cycling. Signal amplification circuits are capable of: (1) linear single layer amplification, (2) quadratic feed-forward amplification, (3) exponential auto-catalytic amplification, and (4) exponential cross-catalytic amplification. When compared to natural polymerase chain reactions (PCR), the sensitivity and stability of toehold-mediated strand displacement reactions suffer from circuit *leakage* – reactions of the system in the absence of an initiator.

Understanding and then suppressing circuit leakage is of paramount importance for the future development of dynamic DNA systems.<sup>17</sup> Presented here, from a materials science and engineering perspective, *defect engineering* was used to improve the leakage performance of model strand displacement systems made from DNA. Engineered defects used in this study included mismatched base pairs and alternative nucleic acids – both of which are known to impact the stability of hybridization.<sup>18, 19</sup> Listed below are select details about strand displacement systems that are essential for understanding and modeling their performance.

## 1.1 Strand Displacement

Strand displacement is the process by which one oligonucleotide displaces a second oligonucleotide that was originally hybridized to a third strand. The driving force for this reaction is an overall reduction in the Gibbs free energy – which is either dominated by an entropic increase or enthalpic decrease in the system.<sup>5, 16</sup> For example, entropy-driven strand displacement systems can overcome a decrease in the number of base pairs so long as the number of components in the system increases. David Zhang's

entropy-driven strand displacement system was shown to function with a reduction of up to eight base pairs.<sup>16</sup> As an extension to strand displacement, toehold mediated strand displacement is the process where one oligonucleotide binds to a single-stranded nucleation site attached to a complementary double helix, which then displaces one of two strands in the complex. The kinetics of toehold mediated strand displacement is highly dependent on toehold length. For example, the rate constant varies 6 orders of magnitude for toeholds between 0 and 7 nucleotides (nt).<sup>20</sup> To be able to monitor reaction rates at room temperature, toehold lengths often range between 5 and 10 nucleotides during experiments.<sup>4, 21</sup>

## 1.2 Strand Displacement Tools

In support of modeling strand-displacement, a kinetic model was originally proposed by Zhang *et al.* to correlate the hybridization energy to the displacement rates.<sup>20</sup> In support of conceptual understanding at a biophysical level, an intuitive energy landscape (IEL) model was then proposed by Srinivas *et al.*<sup>22</sup>. The IEL model reveals that four distinct steps are involved in toehold-mediated strand displacement: (1) toehold hybridization, (2) blunt end fraying, (3) nucleation of branch migration, and (4) branch migration.<sup>23</sup> More recently, a coarse-grained molecular model for DNA (oxDNA) was developed that accounts for the geometric/steric effects of nucleotides, as well as their nearest neighbors effects.<sup>24-26</sup> Using this model, thermodynamic and kinetic values that are related to strand displacement systems can be captured and visually communicated.<sup>26</sup>

Quantifying secondary structures is essential for designing effective strand-displacement systems such as non-enzymatic signal amplification circuits,<sup>16</sup> catalytic hairpin assembly,<sup>27</sup> and motors.<sup>4</sup> Initially, the secondary structure of DNA was predicted

using Watson-Crick models<sup>28-30</sup> that accounted for base pairing,<sup>31,32</sup> mismatches,<sup>33-36</sup> dangling ends,<sup>37,38</sup> bulges,<sup>39,40</sup> coaxial stacking,<sup>41</sup> and hairpin loops.<sup>42</sup> After the systematic acquisition of thermodynamic parameters, computer algorithms such as NUPACK<sup>43,44</sup> and M-FOLD<sup>45</sup> were then developed to analyze and visualize secondary structures of nucleic acids, as well as predict their minimum free energy structures.

### 1.3 Dissertation Layout

To identify sources of leakage in a model signal amplification circuit, *availability* was defined as the probability that a DNA base (A.T.C.G) was unpaired at equilibrium. This design metric was calculated using NUPACK in Chapter 2. To further understand the relationship between leakage rates and secondary structures, *mutual availability* was also calculated as the sum of all pairwise products of the availabilities of the corresponding bases in solution in Chapter 2. This thermodynamic analysis yielded rational design principles for how to minimize leakage by as much as 4-fold by site-specifically introducing mismatched base pairs into DNA duplex regions. To further reduce leakage, chemically modified *locked nucleic acids* (LNAs) were site-specifically introduced into a model DNA strand displacement system in Chapter 3. Briefly described, LNAs are geometrically restricted RNA analogues with enhanced thermal stability towards their complement base. When compared to a DNA control with identical sequences, the leakage exhibited by a hybrid DNA/LNA system was reduced from 1.48  $M^{-1}s^{-1}$  (for the DNA system) to 0.03  $M^{-1}s^{-1}$ . In addition, the signal-to-noise ratio increased ~50-fold for a similar hybrid system.

Holistically, this research provides insight into the sources of leakage in DNA strand-displacement systems, as well as how to maximize strand-displacement

performance by the selective introduction of hybridization defects. Guidelines for how to rationally design future strand displacement systems made from DNA are summarized in Chapter 4. Applied to nucleic acid signal amplification circuits, these guidelines may lead to broader applications in a variety of fields that range from DNA computation to point-of-care diagnostics and therapeutics.

## REFERENCES

1. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. and Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, 292, 1319-1325.
2. Marras, S.A.E., Kramer, F.R. and Tyagi, S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Research*, 30, 122-122.
3. Tyagi, S. and Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology*, 14, 303-308.
4. Yurke, B., Turberfield, A.J., Mills, A.P., Simmel, F.C. and Neumann, J.L. (2000) A DNA-fuelled molecular machine made of DNA. *Nature*, 406, 605-608.
5. Dirks, R.M. and Pierce, N.A. (2004) Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 15275-15278.
6. Green, S.J., Bath, J. and Turberfield, A.J. (2008) Coordinated chemomechanical cycles: a mechanism for autonomous molecular motion. *Physical Review Letters*, 101, 238101.
7. Qian, L. and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332, 1196-1201.
8. Liu, Q., Wang, L., Frutos, A.G., Condon, A.E., Corn, R.M. and Smith, L.M. (2000) DNA computing on surfaces. *Nature*, 403, 175-179.
9. Jung, C. and Ellington, A.D. (2014) Diagnostic applications of nucleic acid circuits. *Accounts of Chemical Research*, 47, 1825-1835.
10. Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I. and Knights, C. (2005) Targeting the

- DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 434, 917-921.
11. Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature*, 171, 737-738.
  12. Breslauer, K.J., Frank, R., Blöcker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, 83, 3746-3750.
  13. SantaLucia, J., Jr. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annual Review Biophysics Biomolecular Structure*, 33, 415-440.
  14. Maune, H.T., Han, S.-p., Barish, R.D., Bockrath, M., Goddard Iii, W.A., Rothmund, P.W.K. and Winfree, E. (2010) Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nature Nanotechnology*, 5, 61-66.
  15. Cannon, B.L., Kellis, D.L., Davis, P.H., Lee, J., Kuang, W., Hughes, W.L., Graugnard, E., Yurke, B. and Knowlton, W.B. (2015) Excitonic AND logic gates on DNA brick nanobreadboards. *ACS Photonics*, 2, 398-404.
  16. Zhang, D.Y., Turberfield, A.J., Yurke, B. and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318, 1121-1125.
  17. Thachuk, C., Winfree, E. and Soloveichik, D. (2015) In Phillips, A. and Yin, P. (eds.), DNA computing and molecular programming. *Springer International Publishing*, Vol. 9211, pp. 133-153.
  18. You, Y., Moreira, B.G., Behlke, M.A. and Owczarzy, R. (2006) Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Research*, 34, e60-e60.
  19. Natsume, T., Ishikawa, Y., Dedachi, K., Tsukamoto, T. and Kurita, N. (2007) Hybridization energies of double strands composed of DNA, RNA, PNA and LNA. *Chemical Physics Letters*, 434, 133-138.

20. Zhang, D.Y. and Winfree, E. (2009) Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131, 17303-17314.
21. Yurke, B. and Mills, A. (2003) Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines*, 4, 111-122.
22. Srinivas, N., Ouldridge, T.E., Sulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research*, 41, 10641-10658.
23. Srinivas, N., Ouldridge, T.E., Šulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research*, 41, 10641-10658.
24. Ouldridge, T.E., Louis, A.A. and Doye, J.P.K. (2010) DNA nanotweezers studied with a coarse-grained model of DNA. *Physical Review Letters*, 104, 178101.
25. Ouldridge, T.E., Louis, A.A. and Doye, J.P.K. (2011) Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *The Journal of Chemical Physics*, 134, 085101.
26. Doye, J.P.K., Ouldridge, T.E., Louis, A.A., Romano, F., Šulc, P., Matek, C., Snodin, B.E.K., Rovigatti, L., Schreck, J.S. and Harrison, R.M. (2013) Coarse-graining DNA for simulations of DNA nanotechnology. *Physical Chemistry Chemical Physics*, 15, 20395-20414.
27. Yin, P., Choi, H.M., Calvert, C.R. and Pierce, N.A. (2008) Programming biomolecular self-assembly pathways. *Nature*, 451, 318-322.
28. Owczarzy, R., Vallone, P.M., Gallo, F.J., Paner, T.M., Lane, M.J. and Benight, A.S. (1997) Predicting sequence- dependent melting stability of short duplex DNA oligomers. *Biopolymers*, 44, 217-239.
29. SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 1460-1465.

30. SantaLucia, J., Allawi, H.T. and Seneviratne, A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35, 3555-3562.
31. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, 22, 256-263.
32. SantaLucia Jr, J. (2000) The use of spectroscopic techniques in the study of DNA stability. *Spectrophotometry and spectrofluorimetry: A Practical Approach*, 329-356.
33. Allawi, H.T. and SantaLucia, J. (1997) Thermodynamics and NMR of internal G T mismatches in DNA. *Biochemistry*, 36, 10581-10594.
34. Allawi, H.T. and SantaLucia, J. (1998) Nearest-neighbor thermodynamics of internal A C mismatches in DNA: sequence dependence and pH effects. *Biochemistry*, 37, 9435-9444.
35. Allawi, H.T. and SantaLucia, J. (1998) Nearest neighbor thermodynamic parameters for internal G·A mismatches in DNA. *Biochemistry*, 37, 2170-2179.
36. Peyret, N., Seneviratne, P.A., Allawi, H.T. and SantaLucia, J. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches. *Biochemistry*, 38, 3468-3477.
37. Bommarito, S., Peyret, N. and SantaLucia Jr, J. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Research*, 28, 1929-1934.
38. Wang, C., Bae, J.H. and Zhang, D.Y. (2016) Native characterization of nucleic acid motif thermodynamics via non-covalent catalysis. *Nature Communication*, 7.
39. Jacobson, H. and Stockmayer, W.H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. *The Journal of Chemical Physics*, 18, 1600-1606.

40. Ke, S.-H. and Wartell, R.M. (1995) Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragment. *Biochemistry*, 34, 4593-4600.
41. Lane, M.J., Paner, T., Kashin, I., Faldasz, B.D., Li, B., Gallo, F.J. and Benight, A.S. (1997) The thermodynamic advantage of DNA oligonucleotide 'stacking hybridization' reactions: energetics of a DNA nick. *Nucleic Acids Research*, 25, 611-616.
42. Blommers, M.J.J., Walters, J., Haasnoot, C.A.G., Aelen, J.M.A., Van der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1989) Effects of base sequence on the loop folding in DNA hairpins. *Biochemistry*, 28, 7491-7498.
43. Dirks, R.M., Bois, J.S., Schaeffer, J.M., Winfree, E. and Pierce, N.A. (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49, 65-88.
44. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32, 170-173.
45. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31, 3406-3415.

CHAPTER TWO: AVAILABILITY – A METRIC FOR NUCLEIC ACID STAND  
DISPLACEMENT SYSTEMS

*Xiaoping Olson,<sup>a</sup> Shohei Kotani,<sup>a</sup> Jennifer E. Padilla,<sup>a</sup> Natalya Hallstrom,<sup>a</sup> Sara Goltry,<sup>a</sup>  
Jeunghoon Lee,<sup>a,b</sup> Bernard Yurke,<sup>a,c</sup> William L. Hughes,<sup>a,\*</sup> and Elton Graugnard.<sup>a,\*</sup>*

<sup>a</sup> Micron School of Materials Science & Engineering,  
Boise State University, 1910 University Dr., Boise, ID, 83725, USA.

<sup>b</sup> Department Chemistry & Biochemistry,  
Boise State University, 1910 University Dr., Boise, ID, 83725, USA.

<sup>c</sup> Department of Electrical & Computer Engineering,  
Boise State University, 1910 University Dr., Boise, ID, 83725, USA.

*\* To whom correspondence should be addressed.*

This chapter will be published in the American Chemical Society (ACS) Synthetic Biology Journal under the digital object identifier (DOI): 10.1021/acssynbio.5b00231.

### **Abstract**

DNA strand displacement systems have transformative potential in synthetic biology. While powerful examples have been reported in DNA nanotechnology, such systems are plagued by leakage, which limits network stability, sensitivity, and scalability. An approach to mitigate leakage in DNA nanotechnology, which is applicable to synthetic biology, is to introduce mismatches to complementary fuel sequences at key locations. However, this method overlooks nuances in the secondary structure of the fuel and substrate that impact the leakage reaction kinetics in strand displacement systems. In an effort to quantify the impact of secondary structure on leakage, we introduce the concepts of *availability* and *mutual availability* and demonstrate their utility for network

analysis. Our approach exposes vulnerable locations on the substrate and quantifies the secondary structure of fuel strands. Using these concepts, a 4-fold reduction in leakage has been achieved. The result is a rational design process that efficiently suppresses leakage and provides new insight into dynamic nucleic acid networks.

## 2.1 Introduction

Nucleic acids are programmable materials because of their predictable Watson-Crick base pairing<sup>1, 2</sup> and well-documented thermodynamics,<sup>3-7</sup> kinetics,<sup>8-13</sup> and mechanics.<sup>14</sup> In addition to static structures,<sup>15-19</sup> dynamics can be programmed into nucleic acids by toehold-mediated strand displacement<sup>9, 20-22</sup> — whereby kinetic barriers to strand exchange are lowered via short complementary sequences that bring components into proximity. Dynamic nucleic acid technology utilizes toehold-mediated DNA strand displacement (DSD) to construct: (1) nonenzymatic catalytic chemical reaction networks for isothermal signal amplification,<sup>23-26</sup> (2) catalytic hairpin assembly for diagnostics, therapeutics, and theranostics,<sup>27, 28</sup> (3) nanomachines<sup>9, 20, 29</sup> and walkers<sup>30, 31</sup> for work and motility, (4) circuits for energy transport and logic,<sup>32, 33</sup> and (5) networks for computation.<sup>34-37</sup> Although they are compelling, these demonstrations are limited in the scale and complexity necessary for real-world applications by a single fundamental challenge: *network leakage*. Leakage refers to the production of an unwanted output in the absence of an input, and it is the *Achilles' heel* of DSD systems, independent of the DNA/RNA system under consideration. The challenge of leakage must be overcome to achieve the device performance (i.e., speed, sensitivity, selectivity, stability, and scalability) necessary for broader adoption.

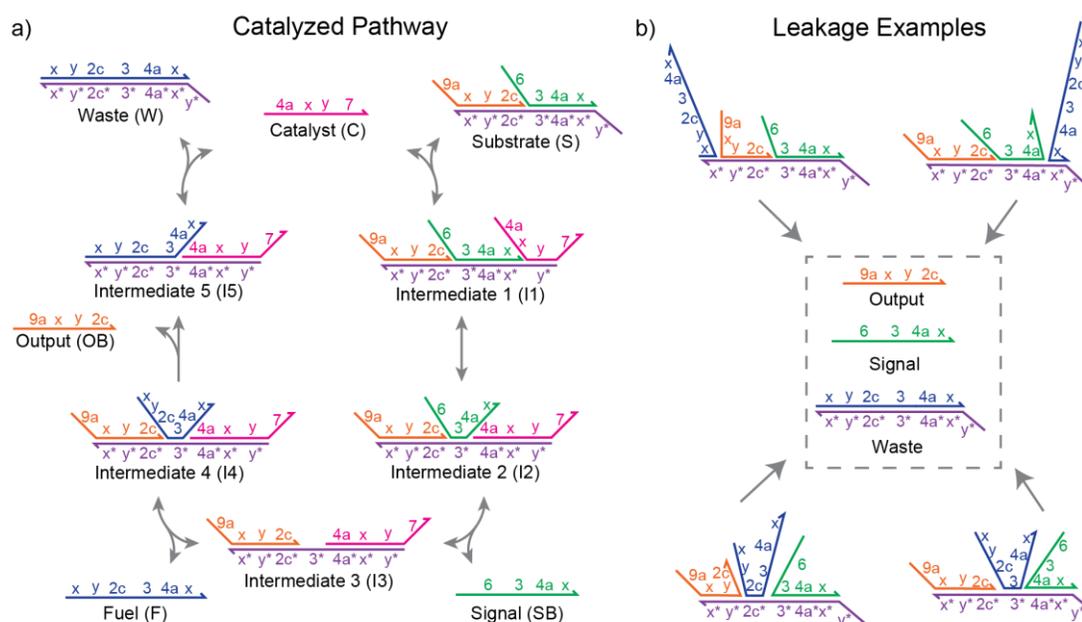
### 2.1.1 Leakage Problem

By design, DSD systems are metastable networks designed to be set into operation by the addition of a specific single-stranded sequence that triggers the reaction. Leakage occurs when system components react in the absence of a trigger, and its effect undermines the performance of catalytic networks,<sup>24, 38</sup> seesaw gates,<sup>36</sup> catalytic hairpin assemblies,<sup>26, 39, 40</sup> and hybridization chain reactions.<sup>25</sup> Extrinsic sources of leakage, including chemical impurities, defective oligonucleotides, and malformed network components, can be minimized with careful processing.<sup>26, 41</sup> In comparison, intrinsic leakage results from the design of the network, even if the components are perfect, and limits the ultimate DSD performance.

Sources of intrinsic leakage may be understood by considering the catalytic reaction network from Zhang *et al.*<sup>24</sup> illustrated in Figure 2.1a. In this representation, unique sequences are represented by labeled domains and complementary domains are denoted with asterisks (domain sequences are provided in Section. 2.8.1). This network consists of a three-strand *substrate* complex in which the “upper” *signal* and *output* strands occupy domains of the lower *backbone* strand. Briefly, network operation is designed to be triggered by a single-stranded *catalyst* strand, that hybridizes with an exposed backbone toehold ( $y^*$ ) and initiates three-way branch migration, to displace the signal strand and expose a sequestered backbone toehold ( $3^*$ ). The catalytic cycle is completed by a similar process with the *fuel* strand reacting with the backbone to displace the output strand, the original catalyst, and form a *waste* product, as illustrated. As the end of a cycle results in no gain or loss of base pairs, this network is driven forward thermodynamically by a net gain in entropy.

In this network, leakage occurs when the substrate and fuel react to produce signal in the absence of catalyst. This leakage reaction depends on successful nucleation of strand invasion by the fuel strand in the absence of an intended toehold. Fuel and substrate must bump into one another favorably, meaning that key bases must have some chance to interact and nucleate. Once nucleated, the leakage reaction proceeds through a branch migration process until strand invasion is complete. In this process, under the conditions reported here, nucleation is the rate-limiting step for the leakage reaction.<sup>8, 11,</sup>

<sup>42</sup> Example leakage reactions are shown in Figure 2.1b.



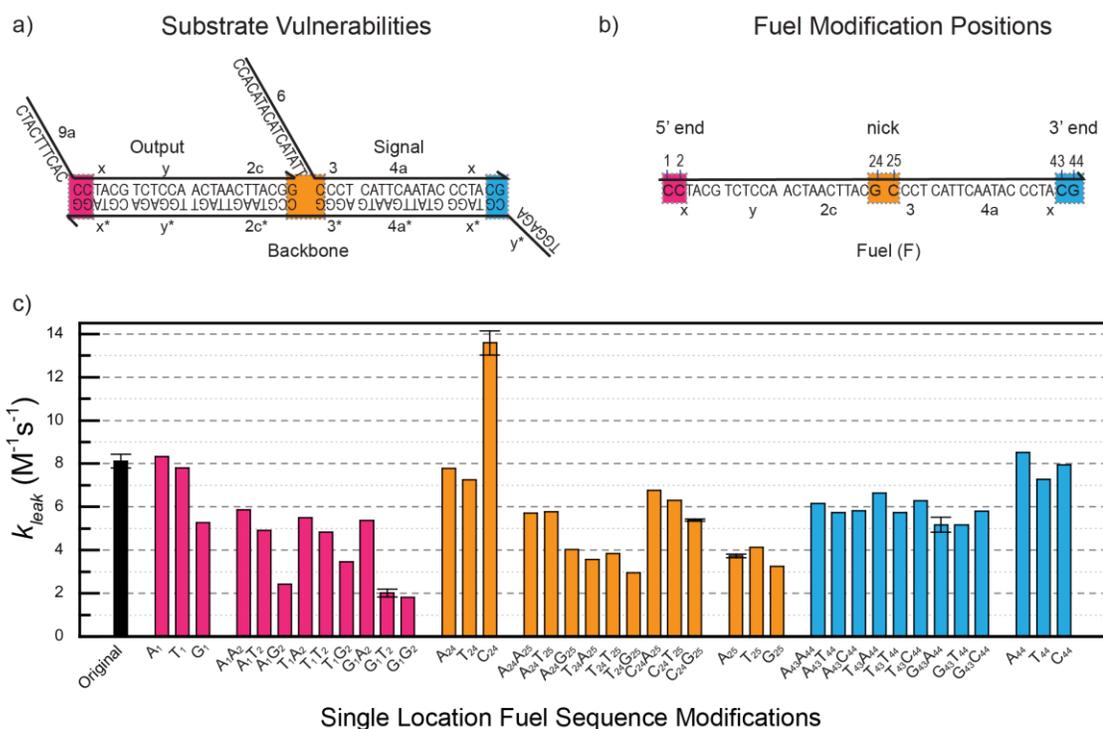
**Figure 2.1.** Domain representation of the catalytic DNA strand displacement system from Zhang et al.<sup>24</sup> and four example leakage pathways. (a) In the catalyzed strand displacement pathway of the reaction network, a catalyst strand initiates a reaction cycle driven forward thermodynamically by a net gain in entropy. The strand displacement exchanges the catalyst for the signal strand and exposes a sequestered toehold on the substrate backbone for the fuel, which reacts with the intermediate to complete the cycle and form a waste duplex. Sequences and domains are listed in Section 2.8.1 (b) In the four leakage pathways, the fuel reacts with the substrate backbone in the absence of a catalyst by exploiting fraying at the 5', nick, and 3' locations of the substrate.

### 2.1.2 Thermal Fluctuations in DNA

Although often considered to be a zero toehold strand displacement event,<sup>43</sup> intrinsic leakage reactions are enabled by transient toeholds created via thermal fluctuations at the ends of the substrate and at the nick between the output and signal strands. *Breathing* refers to the spontaneous dissociation of individual base pairs in the interior of the duplex, and *fraying* is dissociation of the terminal base pairs (at the duplex ends or nicks). Studies of base pair fluctuations indicate that at room temperature the terminal base pairs are 50% open and the penultimate bases (one base pair from the end of the duplex) are 10-20% open, whereas the fraction of open interior base pairs is  $\sim 10^{-6}$  with an open lifetime of  $\sim 0.1 \mu\text{s}$ .<sup>44-46</sup> Additionally, single-stranded DNA overhangs (toeholds or specificity domains, such as domains  $y^*$  and 9a in Figure 2.1a) increase the stability of the neighboring duplex base pairs, but they do not prevent fraying.<sup>44</sup> Thus, fraying of two base pairs at the ends and nick point of the substrate duplex is expected to be the dominant leakage mechanism. These vulnerable regions are highlighted in Figure 2.2a.

Leakage caused by fraying, when compared to toehold invasion, is approximately 4-6 orders of magnitude slower.<sup>11</sup> Even this small leakage drastically limits the scalability of feed-forward, cross-catalytic, and autocatalytic networks, where fuel invasion will unintentionally release the catalyst of the coupled networks. Thermal fluctuations such as fraying have long been suspected as the source of intrinsic leakage, and strategies to suppress it include (1) careful sequence and domain design such as using GC pairs at the fraying locations,<sup>24</sup> (2) use of proper reaction conditions,<sup>47</sup> (3) introduction of buffer or clamp domains that are absent from fuel strands for GC rich sequences,<sup>36, 40</sup> (4)

sequestration of domains in hairpin structures,<sup>48</sup> (5) using extremely pure DNA strands made in bacteria,<sup>26</sup> (6) incorporation of mismatches,<sup>39</sup> and (7) novel domain level redundancy.<sup>49</sup> While each of these approaches has shown some effect, a clear set of design rules have not emerged for consistently and efficiently reducing leakage.



**Figure 2.2.** (a) Sequence and domain representation of the substrate with fraying locations highlighted. (b) Sequence and domain representation of the original fuel strand. Corresponding to the fraying locations of the substrate, the locations of fuel base mismatches are numbered, highlighted, and shown in bold font. They are 5' end (bases 1 and 2), nick (bases 24 and 25), and 3' end (bases 43 and 44). (c) Leakage rate constants for fuel modifications. The concentrations for leakage reactions are fuel (1300 nM), substrate (14 nM), and reporter (20 nM). The black bar represents the leakage rate with the original fuel strand. Pink, orange and blue bars represent leakage rates for fuels with 1 and 2 nt modifications at 5', nick, and 3' locations, respectively. The rates are labeled by the identity of the modified base and its location on the fuel (see panel (b) for locations and original base identities). For example, G<sub>1</sub>T<sub>2</sub> indicates that base 1 was changed from C to G and base 2 was changed from C to T. Error bars show the standard deviation from the mean for select samples in triplicate to estimate experimental error.

### 2.1.3 Insight from Secondary Structure

While previous studies have targeted the location and thermodynamic cost of mismatches in strand displacement systems,<sup>39,50</sup> design principles such as mismatch positions, mismatch numbers, and mismatch identities for suppressing leakage have not emerged. Importantly, base pair mismatch modifications also change the ensemble of DNA secondary structures, which can impact their nucleation and branch migration rates.<sup>51, 52</sup> Here, we report a systematic investigation of the effects of mismatches on intrinsic leakage suppression and network performance using the network shown in Figure 2.1. All one and two base-pair mismatches produced by fuel strand sequence modifications were characterized at the 5', 3', and nick locations (see Figure 2.2a,b) by measuring the reaction rates of uncatalyzed (leakage) and catalyzed reactions using fluorescence photometry. These locations are related to locations on the substrate where fraying is expected to occur and enable nucleation between the backbone and fuel in the absence of catalyst. The results were analyzed on the basis of the mismatch identity, mismatch position, mismatch numbers, and the secondary structure of the fuel strands. To quantify the effects of secondary structure on leakage rates, we calculated the probability that a base is unpaired at equilibrium using NUPACK,<sup>7, 53</sup> as discussed below. We define this probability as the *availability* of a base and introduce availability as a design concept for analyzing and engineering the stability of DNA reaction networks. To further understand the relationship between leakage rates and secondary structures, we define *total mutual availability* as the sum of all pairwise products of the availabilities of corresponding bases between fuels and the backbone. Taking consideration of both mismatches and secondary structure provides a more complete analysis of leakage

suppression, and inclusion of the availability and mutual availability during our analysis provides insight toward rational design principles for minimizing leakage.

## 2.2 Results and Discussion

### 2.2.1 Effect of Mismatch on Leakage Rate Constants

The leakage data for each fuel modification were fit with a second-order kinetics model to extract the leakage rate constant,  $k_{\text{leak}}$  (Section 2.8.3), and the results are shown in Figure 2.2c. The largest leakage suppression was observed for fuel modifications that created two mismatches at the 5' end of the fuel (bases 1 and 2) and one or two mismatches at the nick in the substrate (base 25 and/or bases 24 and 25). While these locations showed consistent leakage suppression, no clear pattern between mismatch base identities and leakage rates emerged. For example, G-A and G-T mismatches show no suppression at base 1 and a factor of 2 suppression at base 25, whereas a G-G mismatch reduces leakage in both locations despite the fact that G-G mismatches have a lower energy penalty than other G or C mismatches when placed within a DNA duplex.<sup>6</sup> While the G-G mismatch consistently reduces leakage at bases 1, 2, and 25, no clear impact from mismatch identity is observed for bases 43 and 44. Although excess fuel in solution could interfere with leakage from the 3' end of the fuel (at the toehold of the substrate backbone; see Section 2.8.4), the data indicates that mismatch identity alone or an associated energy penalty does not ensure leakage suppression.

### 2.2.2 Availability

Beyond mismatch identity, key insight into leakage suppression can be gained by analysis of the secondary structure ensembles of the original and modified fuels. While domain level designs assume the fuel to be purely single-stranded, thermodynamic

analysis using NUPACK reveals a range of secondary structures. The minimum free energy (MFE) structures are shown in Figure 2.3a and have a moderate level of base pairing between six nucleotides of the fuels. Although the MFE structures indicate base pairing between bases 5-23, 6-22, 7-21 for all but one (G<sub>25</sub>) fuel sequence, the probability of pairing is affected by the modifications at bases 1, 2, 24, 25, 43, and 44. The MFE structures for all fuels are provided in Section 2.8.5. G<sub>25</sub> indicates base 25 was changed from C to G. More generally, the letter denotes the base identity and the number denotes the base position from 5' end of the fuel. While the MFE structures are color-coded by the probability for being in the particular MFE structure shown, greater clarity is obtained by plotting the *availability* for each base in the fuel sequences, as shown in Figure 2.3b, c (lower plots). We define availability as the probability that a base is unpaired at equilibrium, and it quantifies the per-base effects of the ensemble of a sequence's secondary structures. Availability is calculated by NUPACK from the predicted secondary structure ensemble lacking pseudoknots and interactions of mismatched base pair.<sup>7, 53</sup> Modifications to the fuel strand alter the availability of the bases since each sequence has a unique ensemble of secondary structures. Figure 2.3b, c (lower plots) shows the changes in base availabilities for modified fuel strands relative to the original fuel sequence. NUPACK calculations were performed using the following parameters: (1) 25 °C operating temperature, (2) 0.05 M Na<sup>+</sup> and 0.0115 M Mg<sup>2+</sup> ion concentrations, (3) 14 nM substrate component concentrations, allowed complex size of 3, (4) 1.3 μM fuel concentration, allowed complex size of 2, and (5) dangles set to “all” in all cases to account for single stranded tails.

Consistent with the MFE structures shown in Figure 2.3a, the availabilities of fuel bases 5, 6, 7, 21, 22, and 23 range between 0.1 and 0.6. However, several other bases have availabilities less than 1, which influences the probability of those bases nucleating a leakage reaction. Additionally, availability calculations exhibit subtle changes for modified fuel strands (Figure 2.3b, c) that have a large impact on leakage and are not limited to the modified bases. For example, the availabilities of several bases were considerably different between the original fuel and the G<sub>1</sub>T<sub>2</sub> fuel (Figure 2.3b), especially for bases 1-4 and bases 13-18, which show a drop and bases 21-23, which show a rise. While most modifications decreased availability for certain bases or left them nearly unchanged (for example A<sub>1</sub>A<sub>2</sub>; Figure 2.3b), the C<sub>24</sub> fuel modification increased the availability of several bases when compared to that of the original fuel (Figure 2.3c) and exhibited the highest leakage rate measured (Figure 2.2c). The base availabilities for all fuel modifications are provided in Section 2.8.6 and are ordered in terms of leakage rate in Figure 2.10. This clearly shows the positive correlation between lower fuel base availability and lower leakage rate.

To fully exploit the concept of availability for understanding the source of leakage, the availability of the bases of the substrate backbone must also be considered because both fuel and backbone bases must be available simultaneously for nucleation to occur. Figure 2.3d, e shows the MFE structure of the substrate and availability of the backbone bases. Ideally, the backbone would have zero availability within double-stranded domains (bases 1\* -44\*) and unity availability at the toehold (bases 45\* -50\*). However, the availabilities are ~0.1 at base 1\*, ~0.24 at base 24\*, and ~0.23 at base 25\*, indicating that the substrate is vulnerable to leakage at these locations (i.e., nucleation

with bases 1, 24, and/or 25 of the fuel strands). Thus, in the context of the substrate, availability quantifies the degree of fraying or breathing of the duplex bases.

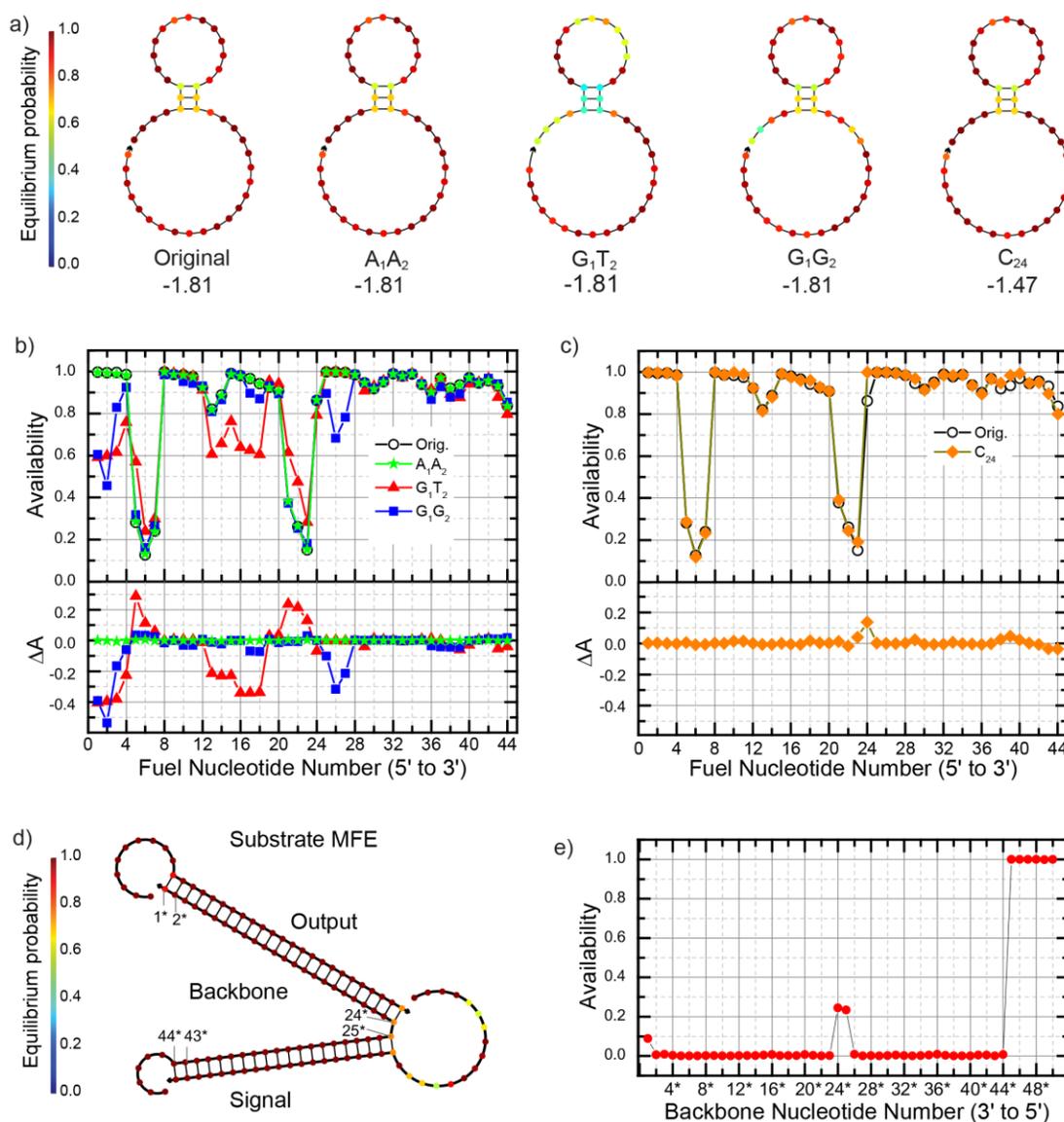
#### 2.2.2.1 Base Modifications at the 5' End of the Fuel

Given that bases 1<sup>\*</sup>, 24<sup>\*</sup>, and 25<sup>\*</sup> of the substrate backbone are most vulnerable to leakage, fuel modifications that reduce availability for fuel bases 1, 24, and 25 can be expected to exhibit the lowest leakage, and this is shown to be the case. For example, leakage was suppressed for the G<sub>1</sub>, G<sub>1</sub>T<sub>2</sub>, and G<sub>1</sub>G<sub>2</sub> fuel modifications. The G<sub>1</sub> leakage drop corresponds to a 5% reduction in the availability of bases 1 and 2 (Section 2.8.6). In addition, the availability of bases 1 and 2 of G<sub>1</sub>T<sub>2</sub> decreased 40%, whereas for G<sub>1</sub>G<sub>2</sub>, the availability of bases 1 and 2 decreased 40 and 54%, respectively (Figure 2.3b and Section 2.8.6). These modified fuels yielded a 4-fold reduction in leakage when compared to that of the original fuel strand. In contrast, the base availabilities in A<sub>1</sub>, T<sub>1</sub>, and A<sub>1</sub>A<sub>1</sub> strands are nearly identical to the original fuel strand (Figure 2.3b and Section 2.8.6), and their leakage suppression was minimal. Here, the changes in availability for single bases on the modified fuel strands provide a compelling explanation for the variation in leakage rates.

#### 2.2.2.2 Base Modifications at the 3' End of the Fuel

The low availabilities at backbone bases 43<sup>\*</sup> and 44<sup>\*</sup> imply a lack of fraying that would be expected to minimize the impact of changes in the availabilities of fuel bases 43 and 44 at the 3' end of the fuels. This hypothesis is consistent with the uniform and relatively minor leakage reductions for fuels with reduced availabilities at bases 43 and 44, such as T<sub>43</sub>T<sub>44</sub> and G<sub>43</sub>T<sub>44</sub> (Section 2.8.6). However, the data for leakage at bases 43<sup>\*</sup> and 44<sup>\*</sup> are confounded by spurious hybridization of the fuel's y domain with the y<sup>\*</sup>

toehold domain of the substrate (bases 45\* to 50\*). This hybridization causes the x domain of both the fuel strand and the signal strand to compete to bind with the x\* domain of the substrate (Section 2.8.4). The competition is expected to be significant since the fuel is at 100x excess concentration. This spurious hybridization is expected to sterically hinder leakage at bases 43\* and 44\* of the backbone and is likely an important factor in the lack of variation in the leakage rate for base modifications at the 3' end of the fuel strand.



**Figure 2.3. Minimum free energy (MFE) structures and base availabilities for select fuel strands and the substrate backbone. (a) MFE structures of the original fuel strand and fuel modifications  $A_1A_2$ ,  $G_1T_2$ ,  $G_1G_2$ , and  $C_{24}$  calculated by NUPACK. The Gibbs free energy of each structure is provided in units of kcal/mol. (b) Base availabilities for the original fuel and fuel modifications  $A_1A_2$ ,  $G_1T_2$ , and  $G_1G_2$  (upper plot) and the difference in base availabilities ( $\Delta A$ ) for each modification relative to the original fuel (lower plot). (c) Base availabilities for the original fuel and fuel modification  $C_{24}$  (upper plot) and the difference in availability relative to the original fuel (lower plot). (d) MFE structure of the substrate calculated by NUPACK. (e) Availability of each base in the backbone strand of the substrate. Because the fuel strand hybridizes with the backbone strand on the substrate, the base positions of the backbone strand were plotted on the x axis and labeled to correspond to the complement of the fuel strand.**

### 2.2.2.3 Base Modifications of the Fuel at the Nick Location

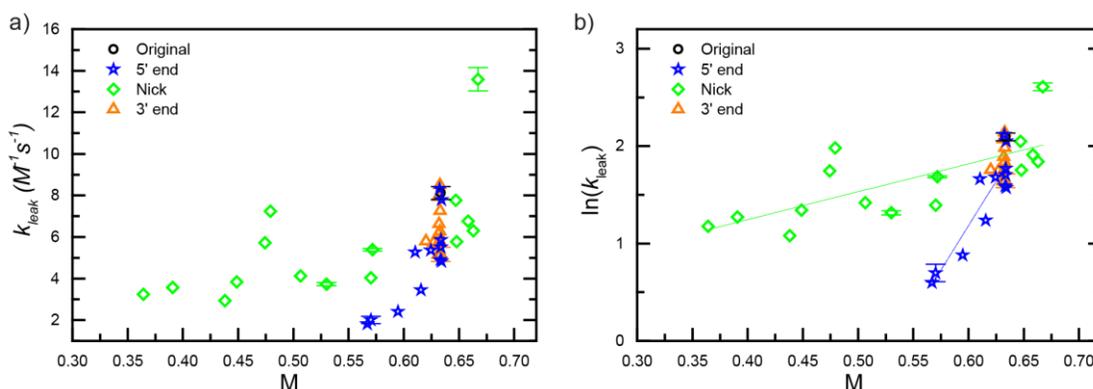
Base 24\* and base 25\* on the substrate backbone have high availabilities, which suggests a greater degree of fraying (Figure 2.3e). Consistent with this expectation, all fuel mismatch modifications at base 25 were observed to suppress the leakage rates. We attribute the reduced leakage for mismatch modifications at fuel base 25 to the lower availabilities at base 25 for the modified fuels compared with the original fuel. For example, availabilities at base 25 for A<sub>25</sub>, T<sub>25</sub> and G<sub>25</sub> were reduced from 21% to 62% and the leakage was reduced from 49% to 68% compared with original fuel. A similar correlation between availability and leakage rate was observed for mismatch modifications at fuel base 24. The single base mismatch at fuel base 24 reduced the leakage for T<sub>24</sub>, for which the availability of base 24 decreased by 72%. The leakage nearly doubled for C<sub>24</sub>, which exhibited a 16% higher availability for base 24. Lastly, no change in leakage rate was observed for A<sub>24</sub>, for which the base 24 availability increased by 9%. An additional factor in the increased leakage observed for C<sub>24</sub> may stem from its increased availability at several bases when compared to the original sequence (Figure 2.3c). An increase in availability corresponds to a decrease in secondary structure, which then lowers the activation energy for nucleation between fuel and substrate. For further consideration, an analysis of base availability in the context of the intuitive energy landscape model of Srinivas *et al.*<sup>11</sup> is provided in Section 2.8.7.

### 2.2.3 Mutual Availability

On the basis of the above observations, base availability is a potentially powerful new design tool with base-specific resolution. In our qualitative explanations, we focused on the separate availabilities of the bases of the fuel or substrate backbone strands.

However, as noted above, leakage reactions require nucleation of these strands with each other. To analyze the combined effects of the availabilities of bases from both strands and to find a quantifiable correlation, we define and analyze a *mutual availability* ( $m_{ij}$ ) and *total mutual availability* ( $M$ ). The mutual availability is simply the product of the availabilities of any two bases, defined as  $m_{ij} = P_{F(i)} \cdot P_{B(j)}$ , where  $P_{F(i)}$  is the availability of base  $i$  of the fuel strand and  $P_{B(j)}$  is the availability of base  $j$  of the backbone strand within the substrate complex. The total mutual availability is defined as  $M = \sum_i(m_{ii^*}) = \sum_i(P_{F(i)} \cdot P_{B(i^*)})$ , where  $i$  indexes the complementary base pairs in the fuel-substrate waste product in correct registration. In other words,  $i^*$  is the base position of backbone strand that matches the complementary position  $i$  of the fuel strand.

For nucleation to occur, key bases of the fuel and backbone must be available to hybridize. Total mutual availability,  $M$ , as defined above, provides a quantitative metric for analyzing fuel and substrate sequence interactions. To assess whether  $M$  could be correlated with leakage rate, Figure 2.4a plots leakage rates versus the calculated values of  $M$  for all fuel sequence modifications. On the basis of the apparent exponential dependence, the natural log of the leakage rate constant is plotted versus  $M$  in Figure 2.4b and is colored coded by 5', 3', and nick modifications of the fuel. Select experiments were repeated in triplicate, and the scatter of the data is greater than the experimental error. Error bars represent the standard deviation of the mean.



**Figure 2.4.** (a) Leakage rate constants for each fuel modification plotted versus total mutual availability between the fuel strand and the backbone strand on the substrate. The leakage rate for the original fuel is shown in black while the 5' end, nick and 3' end fuel modifications are shown in blue, green and orange, respectively. Representative error bars of select samples are shown, indicating that the scatter of the data is greater than the experimental error. Error bars represent the standard deviation of the mean of three samples. (b) Natural log plot of the leakage rate constant versus the total mutual availability. The green, and blue lines are the fits for the nick modifications, 5' end modifications.

Linear fits to the data are provided as guides to the eye. The 5' and nick modifications exhibited linear trends and were fit individually. Given the near zero availability of base 44\* of the backbone, 3' fuel modifications had very little impact on total mutual availability. The nick fuel modifications and their corresponding fit are depicted in green and have a slope of 2.87 with an adjusted  $R^2$  value of 0.50. 5' fuel modifications and their corresponding fit are depicted in blue and have a slope of 18.26 with an adjusted  $R^2$  of 0.81. These results support the correlation of leakage rate with total mutual availability. While the primary discussion here is focused on single location fuel modifications, multiple location modifications (e.g., 5' and nick locations) further reduced the leakage rate to an almost undetectable level (about 100-fold), which are presented and discussed in Section 2.8.8. These data provide further support for total mutual availability as a metric for leakage.

The leakage rate constant appears to be exponentially related to the total mutual availability of the fuel and substrate backbone, suggesting that  $M$  may be related to a nucleation activation energy barrier. However, our data does not distinguish between barriers to nucleation and branch migration nor can they identify the critical nucleus for leakage to proceed. The scatter in Figure 2.4 may result from the incompleteness of our mutual availability model, which does not include branch migration steps, and limitations in total mutual availability as a measure of nucleation barriers. For example, NUPACK does not include pseudoknots, G-quartets, nick overhangs, and the coaxial stacking parameter into its calculations. Additionally, base availability, as defined, does not include tertiary nucleic acid structure. The correlation between the leakage rate constant and total mutual availability also needs careful consideration. For example, as the number of fuel mismatches increases, the leakage rate approaches zero, and the reaction stalls because of a lack of thermodynamic driving force. In comparison, when the total mutual availability is high, an effective toehold is formed, and the nucleation barrier is reduced, which means diffusion is the rate-limiting factor. The relationship between  $M$  and the leakage rate constant is thus constrained by these limits.

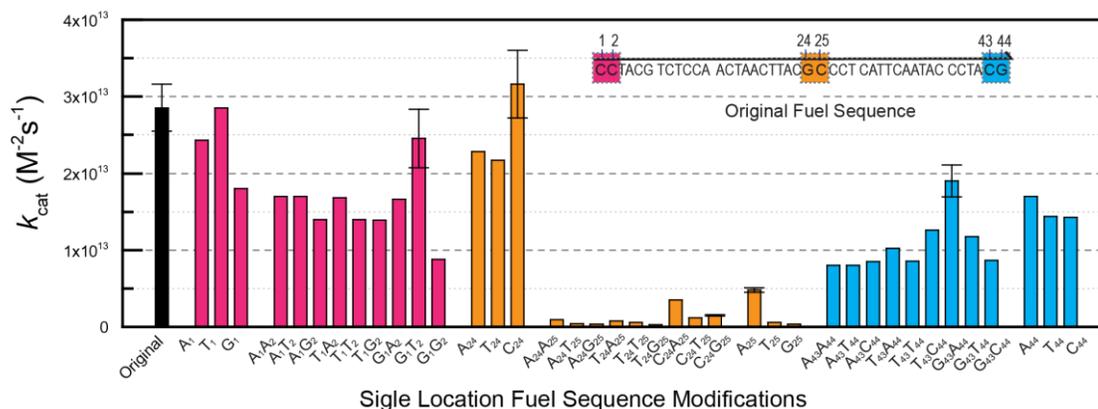
#### 2.2.4 Catalyzed Reactions

It has generally been observed that the rate constants between catalytic reactions and leakage reactions are coupled. It has been shown that when leakage rates were reduced by introducing mismatches, catalytic rates were also decreased or maintained.<sup>39</sup> Likewise, here we also found that some fuel mismatch modifications maintained the original catalytic rate while decreasing the leakage rates. The kinetics data of each fuel modification were fit with a third-order kinetics model with a steady-state approximation

to extract the catalyzed rate constant,  $k_{cat}$  (Section 2.8.3). Catalyzed rate constants ranged 3 orders of magnitude for the fuel modifications and are plotted in Figure 2.5.

The effect of fuel sequence modifications on the catalyzed reactions can be understood via the reaction mechanism. The modification positions play a critical role in the catalyzed reaction as discussed further in Section 2.8.8 and 2.8.9. In Figure 2.5, trends can be observed by grouping the modification positions of the fuel strand at the 5' end (bases 1 and 2) with base 24 of the nick, and at the 3' end (bases 43 and 44) with base 25 of the nick. As expected from the reaction mechanism shown in Figure 2.1, mismatches at base 25 of the nick location have the greatest impact because it affects fuel hybridization with the intermediate (I3), followed by mismatches at the 3' end that impede catalyst release. Fuel modifications at the 5' end and base 24 of the nick locations have minimal impact on the catalytic rate. A strategy to speed up the catalytic reaction is to increase the toehold length by deleting one nucleotide at the 5' end of the catalyst (Section 2.8.8). This strategy has desired effects for the fuels with modification at base 25 whereas it has a counter effect for other fuel modifications.

Since one catalytic reaction cycle in this system has many intermediate steps including toehold exchange, toehold-mediated strand displacement, and spontaneous toehold dissociation, the correlation between the overall catalytic rate constant and the total mutual availability of the catalyzed reaction was not studied in this work.



**Figure 2.5.** Rate constants of catalyzed reactions between the catalyst (1 nM), fuel (13 nM), substrate (14 nM), and reporter (20 nM) monitored via fluorescence. The black bar represents the original fuel strand. Pink, orange and blue bars represent 1 nt and 2 nt modifications at 5', nick, and 3' locations, respectively. Error bars show the standard deviation from the mean for select samples in triplicate to estimate experimental error.

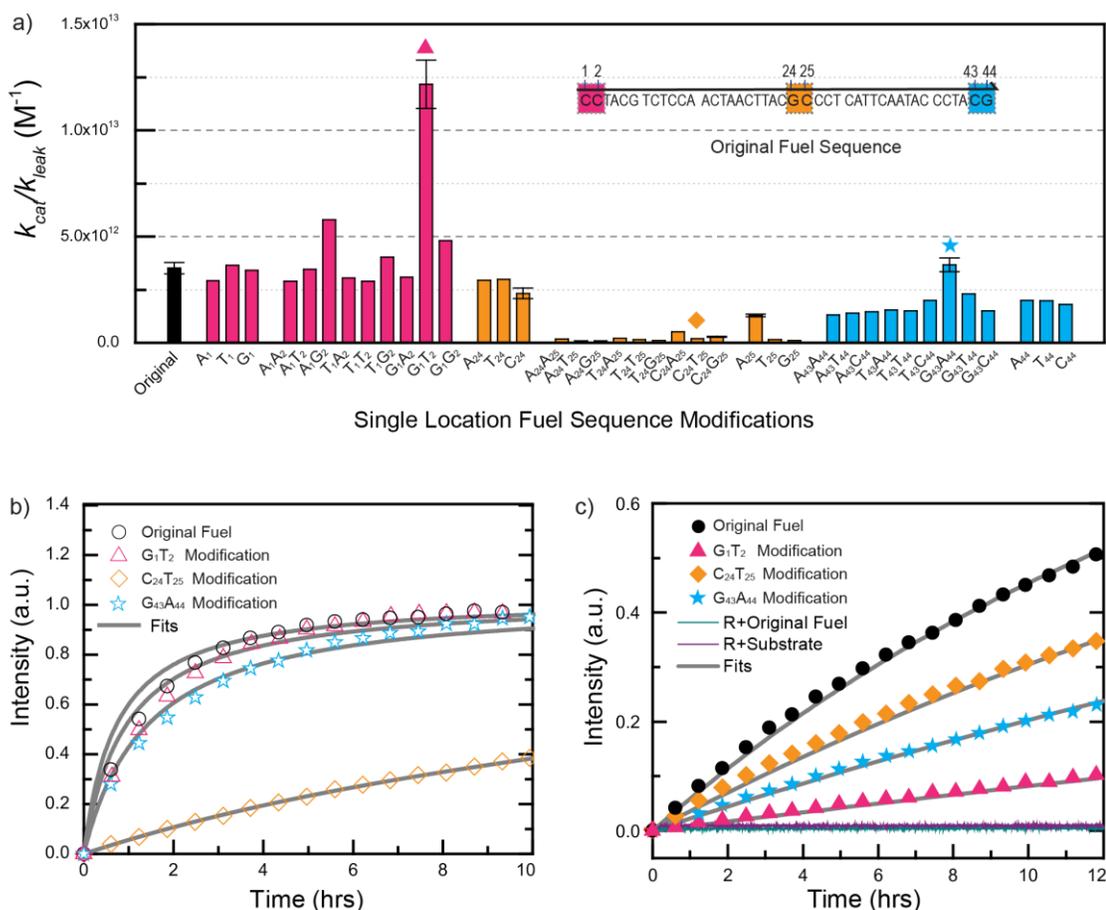
### 2.2.5 System Performance

An ideal DNA strand displacement system would have elevated selectivity to the catalyst, sensitivity to the catalyst, high catalytic turnover (high  $k_{cat}$ ), stability in the absence of the catalyst (low  $k_{leak}$ ), and scalability because of suppressed crosstalk and leakage. Thus, as a practical metric for the performance of the system, we use the ratio,  $k_{cat}/k_{leak}$ . The larger the ratio, the greater will be the capacity to distinguish a response to the catalyst from the background leakage.

Given that the leakage rate is strongly coupled to the catalytic rate for fuel sequence modifications at bases 25, 43, and 44, the suppression of the catalytic reaction reduced performance more than leakage suppression increased it. Locations of strong coupling between catalytic rate and leakage rate can be considered to be limitations of intrinsic leakage suppression; they are a result of the domain design of this system and will be different for other domain level designs. Modifications at base 24 had no net benefit due to the low availability of the substrate at this location. Improvements in

performance came from introducing sequence mismatches at the 5' end of the fuel (bases 1 and 2), where leakage and catalytic reaction rates are decoupled. As measured by the  $k_{cat}/k_{leak}$  ratio, the G<sub>1</sub>T<sub>2</sub> fuel modification has the best performance overall (Section 2.8.10). This modification targeted the vulnerability at base 1\* of the backbone strand due to nonzero availability. It reduced the leakage reaction rate by a factor of 4 but maintained a catalytic rate close to the original fuel strand (Figure 2.6a).

In the literature, mismatch modifications have shown more dramatic improvements to leaky systems and systems using low-quality strands. Mismatches in Jiang *et al.*'s DNA catalytic hairpin design with large leakage showed 25-fold improvements in signal-to-background ratio compared with that of the original hairpins.<sup>39</sup> By contrast, Bhadra *et al.*'s optimized RNA catalytic hairpin system shows only 7-fold leakage reduction, without disturbing the catalytic reaction rate, by introducing mismatch modifications. However, when using unpurified RNA strands in this system, a 13-15-fold reduction in the leakage is observed when compared to that of the control.<sup>54</sup> Zhang *et al.*'s system was optimized and purified, having an intrinsic leakage rate of only  $\sim 8 \text{ M}^{-1}\text{s}^{-1}$ . This work demonstrates that a 4-fold leakage reduction in this system can be achieved while leaving the catalyzed reaction rate nearly unchanged. Mismatches at substrate fraying locations reveal the power of availability to influence circuit performance.



**Figure 2.6.** (a) Ratio of the catalyzed to leakage reaction rates ( $k_{cat}/k_{leak}$ ) for single location fuel modifications to evaluate overall system performance. Catalyzed reactions were performed with the catalyst (1 nM), fuel (13 nM), substrate (14 nM), and reporter (20 nM), monitored via fluorescence, and uncatalyzed leakage reactions were performed with fuel (1300 nM), substrate (14 nM), and reporter (20 nM). The black bar represents the original fuel strand. Pink, orange, and blue bars represent 1 and 2 nt modifications at 5', nick, and 3' locations, respectively. Error bars show the standard deviation from the mean for select samples in triplicate to estimate experimental error. (b) Representative fluorescence data of catalytic reactions: the original fuel (empty black circles), G<sub>1</sub>T<sub>2</sub> modified fuel (empty red triangles), C<sub>24</sub>T<sub>25</sub> modified fuel (yellow diamonds), and G<sub>43</sub>A<sub>44</sub> modified fuel (empty blue stars). (c) Representative fluorescence data of leakage reactions: the original fuel (solid black circles), G<sub>1</sub>T<sub>2</sub> modified fuel (solid red triangles), C<sub>24</sub>T<sub>25</sub> modified fuel (solid yellow diamonds), and G<sub>43</sub>A<sub>44</sub> modified fuel (solid blue stars). The gray lines are the calculated fits to each curve, and the solid blue and purple lines represent reactions between the reporter and the original fuel and the reporter and the substrate, respectively.

### 2.2.6 Analysis of Other Networks

The concept of mutual availability is expected to apply to other network designs as well. In an effort to validate the mutual availability concept with another network design, we analyzed a hairpin design from Jiang *et al.*<sup>39</sup> This study provided sufficient data to apply an analysis of total mutual availability, and we estimated the rate constants for the hairpin design, as described in Section 2.8.11. Even though the total mutual availability values vary relative to the Zhang *et al.* network, the observed trend is the same, even in a different buffer. The results provide compelling support for the validity of mutual availability as a metric for sequence-level network analysis and design.

## **2.3 Conclusion**

The effects of base-pair mismatches on leakage suppression and total network performance were systematically investigated using the well-established catalytic reaction network from Zhang *et al.*<sup>24</sup> Fuel modifications at the 5', 3', and nick locations were chosen because they correspond to vulnerable substrate locations where nucleation often occurs. Qualitatively, availabilities of the substrate and the fuel strand bases were found to correspond well to observed trends in the leakage rate data. Quantitatively, a trend between the total mutual availability and the leakage rates was observed regardless of mismatch identities, mismatch numbers, and mismatch locations. This work suggests availability and mutual availability as design concepts for optimal performance of nucleic acid reaction networks.

Future work can further explore the correlation between the total mutual availability and the activation energy, aiming at a more detailed model of leakage mechanisms. In addition, the correlation between the overall catalytic rate constant ( $k_{cat}$ )

and the total mutual availability of the catalyzed reaction should be studied to allow predictions of the practical metric for the performance of the system ( $k_{cat}/k_{leak}$ ). This study also leaves room for refinement against other interactions that NUPACK does not calculate, such as G-quartets, mismatched interactions, pseudoknots, and geometric constraints. With improved design metrics and refined design tools, nonenzymatic amplification systems can be used as amplifiers for diagnostics, and nucleic acid chemical reaction networks will become more robust tools for theranostics, molecular computation, and synthetic biology.

## 2.4 Methods

### 2.4.1 Chemicals and DNA Complex Purification

Solvents and chemicals were purchased from Sigma-Aldrich unless otherwise noted. DNA oligonucleotides were synthesized and purified with high-performance liquid chromatography by Integrated DNA Technologies (IDT). Reporter strands were labeled with 5' fluorophores (TET) and 3' Iowa Black dark quenchers (IABkFQ) by IDT. Oligonucleotides were prepared in 1x TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, diluted from 100x TE). Final stock concentrations (100  $\mu$ M) were confirmed by measuring the 260 nm absorbance (Eppendorf Biophotometer) using extinction coefficients provided by IDT.

TAE buffer (10x; 40 mM Tris, 40 mM acetate, 1 mM EDTA) was purchased from Hoefer or Fisher Science and then mixed with 125 mM  $\text{Mg}(\text{C}_2\text{H}_3\text{O}_2)_2 \cdot 4\text{H}_2\text{O}$ . DNA components were diluted to 30  $\mu$ M in 1x TAE buffer with 12.5 mM  $\text{Mg}^{2+}$ . DNA components were annealed at 95 °C for 5 min using a

thermocycler (Eppendorf Mastercycler Nexus Gradient) and cooled to room temperature over ~90 min to form Substrates and Reporters.

Substrate and Reporter complexes were purified by native polyacrylamide gel electrophoresis (N-PAGE). To eliminate malformed substrates, fuel and substrate were stoichiometrically incubated at 15  $\mu\text{M}$  for 1 h at room temperature before loading the gel. The loading buffer contained a 1:1 ratio of bromophenol blue dye and ficoll solution (type 400, 20% water). Substrates were purified by N-PAGE in 14% acrylamide gels (made from 30% acrylamide bis solution in a 29:1 ratio), which were run at 150 V for 7 h. Reporters were also purified by N-PAGE in 10% acrylamide gels, which were run at 150 V for 2 h. For both processes, the cooling system (VWR International) was set to 20  $^{\circ}\text{C}$ .

The bands of interest were cut out of the gels and eluted in 1x TE/ $\text{Mg}^{2+}$  buffer for 2 days at 4  $^{\circ}\text{C}$ . The buffer included 1x TE with 12.5 mM  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$  (Acros Organics) added.

Because  $\text{Mg}^{2+}$  binds to EDTA, the effective  $\text{Mg}^{2+}$  concentration was estimated to be 11.5 mM.<sup>24</sup> Substrate and reporter concentrations were quantified via measuring absorbance at 260 nm and calculated using extinction coefficients predicted by nearest-neighbor models (Section 2.8.2).<sup>36, 55</sup> Typical yields were 30% for the substrate and 50% for the reporter.

#### 2.4.2 Spectrofluorimetry

All experiments were carried out in 1x TE/ $\text{Mg}^{2+}$  buffer with a total volume of 1 mL in 4 mL disposable methacrylate cuvettes (Fisher Scientific) at 25  $^{\circ}\text{C}$ . DNA stock solutions were normally diluted to 2  $\mu\text{M}$  before being added to each

sample. A poly-T strand (dT<sub>20</sub>) was added into all dilute stock samples (1  $\mu$ M and lower) to reach a final concentration of 1  $\mu$ M and prevent DNA loss via nonspecific binding to the microfuge tubes and pipet tips.<sup>24</sup> All solutions were gently mixed by pipetting.

Fluorescence intensity versus time was measured via fluorescence spectrophotometers (Agilent Technologies, Cary Eclipse). Sample solutions were excited at 510 nm, and the emission was measured at 538 nm. Slit sizes used were 2.5 nm for excitation and 10 nm for emission. Fluorescence was normalized so that 1 normalized unit (a.u.) of fluorescence corresponded to 14 nM (the substrate concentration) for leakage reactions and 13 nM (the fuel concentration) for catalyzed reactions.

#### 2.4.3 Reaction Measurements

For leakage reactions, the fluorescence intensity was continuously monitored for the first 12 h (shown in Figure 2.6c) with the samples maintained at 25 °C and then periodically measured until the reaction reached completion at room temperature (~21.5 °C). Substrate (14 nM) and reporter (20 nM) were reacted with ~100-fold excess of the fuel strands (1300 nM) to expedite leakage reactions and to extract intrinsic leakage specific to fuel and substrate interaction. Reaction between substrate and reporter was undetectable under this condition (Figure 2.6c). With the assumption that extrinsic leakage dominates at shorter times and intrinsic leakage dominates at longer times,<sup>26</sup> leakage was measured over the long term to extract intrinsic effects. For catalyzed reactions, the catalyst (1 nM), fuel (13 nM), substrate (14 nM), and reporter (20 nM) were reacted for 10 h (Figure 6b). During the experiments, substrate

reactions were inferred by monitoring the production of signal strand through its reaction with the reporter (Section 2.8.1).

## **2.5 Author Information**

### 2.5.1 Corresponding Authors

WLH [willhughes@boisestate.edu] & EG [eltongraugnard@boisestate.edu]

### 2.5.2 Author Contributions

The primary author performed the research and the primary author wrote the manuscript with support from the corresponding authors.

## **2.6 Acknowledgements**

This project was supported in part by (1) The W.M. Keck Foundation, (2) NIH Grant No. K25GM093233 from the National Institute of General Medical Sciences, (3) NIH Grant No. P20 RR016454 from the INBRE Program of the National Center for Research Resources, and (4) The Micron Foundation. Special thanks are extended to Dr. Peter Allen from the University of Idaho for his thoughtful review of this work.

## REFERENCES

1. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid, *Nature* 171, 737-738.
2. Kool, E. T. (2001) Hydrogen bonding, base stacking, and steric effects in dna replication, *Annual Review of Biophysics and Biomolecular Structure* 30, 1-22.
3. Lu, M., Guo, Q., Marky, L. A., Seeman, N. C., and Kallenbach, N. R. (1992) Thermodynamics of DNA branching, *Journal of Molecular Biology* 223, 781-789.
4. SantaLucia, J., Allawi, H. T., and Seneviratne, A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry-Us* 35, 3555-3562.
5. SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proceedings of the National Academy of Science* 95, 1460-1465.
6. SantaLucia, J., Jr., and Hicks, D. (2004) The thermodynamics of DNA structural motifs, *Annual Review Biophysics Biomolecular Structure* 33, 415-440.
7. Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., and Pierce, N. A. (2007) Thermodynamic analysis of interacting nucleic acid strands, *Siam Review* 49, 65-88.
8. Zhang, D. Y., and Winfree, E. (2009) Control of DNA strand displacement kinetics using toehold exchange, *Journal of the American Chemical Society* 131, 17303-17314.
9. Yurke, B., and Mills, A. (2003) Using DNA to power nanostructures, *Genetic Programming and Evolvable Machines* 4, 111-122.
10. Panyutin, I. G., and Hsieh, P. (1994) The kinetics of spontaneous DNA branch migration, *Proceedings of the National Academy of Science* 91, 2021-2025.

11. Srinivas, N., Ouldrige, T. E., Sulc, P., Schaeffer, J. M., Yurke, B., Louis, A. A., Doye, J. P. K., and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement, *Nucleic Acids Research* 41, 10641-10658.
12. Sikora, J. R., Rauzan, B., Stegemann, R., and Deckert, A. (2013) Modeling stopped-flow data for nucleic acid duplex formation reactions: the importance of off-path intermediates, *Journal of Physical Chemistry B* 117, 8966-8976.
13. Graugnard, E., Cox, A., Lee, J., Jorcyk, C., Yurke, B., and Hughes, W. L. (2010) Kinetics of DNA and RNA hybridization in serum and serum-SDS, *IEEE Transactions on Nanotechnology* 9, 603-609.
14. Hagerman, K. R., and Hagerman, P. J. (1996) Helix rigidity of DNA: the meroduplex as an experimental paradigm, *Journal Molecular Biology* 260, 207-223.
15. Seeman, N. C. (2010) Nanomaterials based on DNA, *Annual Review Biochemistry* 79, 65-87.
16. Linko, V., and Dietz, H. (2013) The enabled state of DNA nanotechnology, *Current Opinion in Biotechnology* 24, 555-561.
17. Edwards, A., and Yan, H. (2014) DNA origami, *Nucleic Acids and Molecular Biology* 29, 93-134.
18. Endo, M., Yang, Y., and Sugiyama, H. (2013) DNA origami technology for biomaterials applications, *Biomaterials Science* 1, 347-360.
19. Saccà, B., Sprengel, A., and Feldkamp, U. (2013) De novo design of nucleic acid structures, In *De Novo Molecular Design*, pp 495-517.
20. Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C., and Neumann, J. L. (2000) A DNA-fuelled molecular machine made of DNA, *Nature* 406, 605-608.
21. Wang, F., Lu, C. H., and Willner, I. (2014) From cascaded catalytic nucleic acids to enzyme-DNA nanostructures: controlling reactivity, sensing, logic operations, and assembly of complex structures, *Chemical Reviews* 114, 2881-2941.
22. Zhang, D. Y., and Seelig, G. (2011) Dynamic DNA nanotechnology using strand-displacement reactions, *Nature Chemistry* 3, 103-113.

23. Wittwer, C. T., Herrmann, M. G., Moss, A. A., and Rasmussen, R. P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification, *Biotechniques* 22, 130-131, 134-138.
24. Zhang, D. Y., Turberfield, A. J., Yurke, B., and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA, *Science* 318, 1121-1125.
25. Dirks, R. M., and Pierce, N. A. (2004) Triggered amplification by hybridization chain reaction, *Proceedings of the National Academy of Science* 101, 15275-15278.
26. Chen, X., Briggs, N., McLain, J. R., and Ellington, A. D. (2013) Stacking nonenzymatic circuits for high signal gain, *Proceedings of the National Academy of Science* 110, 5386-5391.
27. Huttanus, H. M., Graugnard, E., Yurke, B., Knowlton, W. B., Kuang, W., Hughes, W. L., and Lee, J. (2013) Enhanced DNA sensing via catalytic aggregation of gold nanoparticles, *Biosensors & Bioelectronics* 50, 382-386.
28. Jung, C., and Ellington, A. D. (2014) Diagnostic applications of nucleic acid Circuits, *Accounts of Chemical Research* 47, 1825-1835.
29. Goltry, S., Hallstrom, N., Clark, T., Kuang, W., Lee, J., Jorcyk, C., Knowlton, W. B., Yurke, B., Hughes, W. L., and Graugnard, E. (2015) DNA topology influences molecular machine lifetime in human serum, *Nanoscale* 7, 10382-10390.
30. Shin, J.-S., and Pierce, N. A. (2004) A synthetic DNA walker for molecular transport, *Journal of the American Chemical Society* 126, 10834-10835.
31. Yin, P., Yan, H., Daniell, X. G., Turberfield, A. J., and Reif, J. H. (2004) A unidirectional DNA walker that moves autonomously along a track, *Angewandte Chemie International Edition* 43, 4906-4911.
32. Cannon, B. L., Kellis, D. L., Davis, P. H., Lee, J., Kuang, W., Hughes, W. L., Graugnard, E., Yurke, B., and Knowlton, W. B. (2015) Excitonic AND logic gates on DNA brick nanobreadboards, *ACS Photonics* 2, 398-404.

33. Graugnard, E., Kellis, D. L., Bui, H., Barnes, S., Kuang, W., Lee, J., Hughes, W. L., Knowlton, W. B., and Yurke, B. (2012) DNA-controlled excitonic switches, *Nano Letters* 12, 2117-2122.
34. Benenson, Y. (2012) Biomolecular computing systems: principles, progress and potential, *Nature Review Genetics* 13, 455-468.
35. Qian, L., Winfree, E., and Bruck, J. (2011) Neural network computation with DNA strand displacement cascades, *Nature* 475, 368-372.
36. Qian, L., and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades, *Science* 332, 1196-1201.
37. Genot, A. J., Bath, J., and Turberfield, A. J. (2013) Combinatorial displacement of DNA strands: application to matrix multiplication and weighted sums, *Angewandte Chemie International Edition* 52, 1189-1192.
38. Zhang, D. Y., and Winfree, E. (2010) Robustness and modularity properties of a non-covalent DNA catalytic reaction, *Nucleic Acids Res* 38, 4182-4197.
39. Jiang, Y. S., Bhadra, S., Li, B., and Ellington, A. D. (2014) Mismatches improve the performance of strand-displacement nucleic acid circuits, *Angewandte Chemie International Edition* 126, 1876-1879.
40. Yin, P., Choi, H. M., Calvert, C. R., and Pierce, N. A. (2008) Programming biomolecular self-assembly pathways, *Nature* 451, 318-322.
41. Ducani, C., Kaul, C., Moche, M., Shih, W. M., and Hogberg, B. (2013) Enzymatic production of 'monoclonal stoichiometric' single-stranded DNA oligonucleotides, *Nature Methods* 10, 647-652.
42. Green, C., and Tibbetts, C. (1981) Reassociation rate limited displacement of DNA strands by branch migration, *Nucleic Acids Research* 9, 1905-1918.
43. Reynaldo, L. P., Vologodskii, A. V., Neri, B. P., and Lyamichev, V. I. (2000) The kinetics of oligonucleotide replacements<sup>1</sup>, *Journal of Molecular Biology* 297, 511-520.

44. Jose, D., Datta, K., Johnson, N. P., and von Hippel, P. H. (2009) Spectroscopic studies of position-specific DNA "breathing" fluctuations at replication forks and primer-template junctions, *Proceedings of the National Academy of Science* 106, 4231-4236.
45. von Hippel, P. H., Johnson, N. P., and Marcus, A. H. (2013) Fifty years of DNA breathing : reflections on old and new approaches, *Biopolymers* 99, 923--954.
46. Frank-Kamenetskii, M. D., and Prakash, S. (2014) Fluctuations in the DNA double helix: a critical review, *Physics of Life Reviews* 11, 153 - 170.
47. Li, B., Jiang, Y., Chen, X., and Ellington, A. D. (2012) Probing spatial organization of DNA strands using enzyme-free hairpin assembly circuits, *Journal of the American Chemical Society* 134, 13918-13921.
48. Tomov, T. E., Tsukanov, R., Liber, M., Masoud, R., Plavner, N., and Nir, E. (2013) Rational design of DNA motors: fuel optimization through single-molecule fluorescence, *Journal of the American Chemistry Society* 135, 11935-11941.
49. Thachuk, C., Winfree, E., and Soloveichik, D. (2015) Leakless DNA strand displacement systems, In *DNA computing and molecular ;programming*, Springer International Publishing, pp 133-153.
50. Machinek, R. R., Ouldrige, T. E., Haley, N. E., Bath, J., and Turberfield, A. J. (2014) Programmable energy landscapes for kinetic control of DNA strand displacement, *Nature Communications* 5.
51. Chen, C. L., Wang, W. J., Wang, Z., Wei, F., and Zhao, X. S. (2007) Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization, *Nucleic Acids Research* 35, 2875-2884.
52. Gao, Y., Wolf, L. K., and Georgiadis, R. M. (2006) Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison, *Nucleic Acids Research* 34, 3370-3377.
53. Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011) NUPACK: analysis and design of nucleic acid systems, *Journal of Computational Chemistry* 32, 170-173.

54. Bhadra, S., and Ellington, A. D. (2014) Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers, *Nucleic Acids Research* 42, e58-e58.
55. Puglisi, J. D., and Tinoco Jr, I. (1989) [22] absorbance melting curves of RNA, In *Methods in Enzymology* (James E. Dahlberg, J. N. A., Ed.), pp 304-325.

## 2.7 Supplementary Information

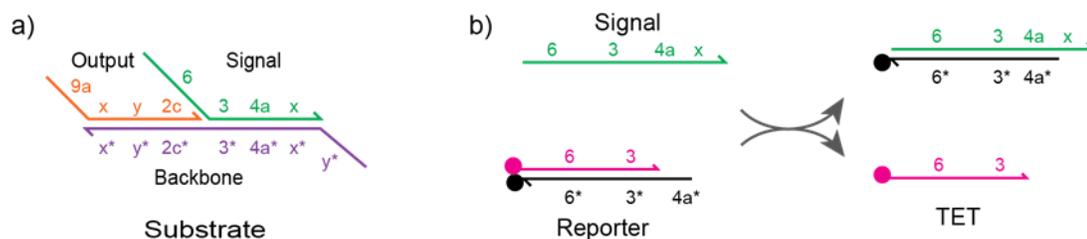
### 2.7.1 Strand sequences

**Table 2.1** Substrate and catalyst sequences and schematic

Name	Sequence (5' to 3')
Signal	CCACATACATCATATTCCCTCATTCAATACCCTACG
Output	CTACTTTACCCCTACGTCTCCAACCTAATTACGG
Backbone	TGGAGACGTAGGGTATTGAATGAGGGCCGTAAGTTAGTTGGAGACGTAGG
Catalyst	CATTCAATACCCTACGTCTCCATACTTATTAGCC
1 nt deletion catalyst	ATTCAATACCCTACGTCTCCATACTTATTAGCC

**Table 2.2** Reporter sequences and schematic

Name	Sequence (5' to 3')
TET	/5TET/CCACATACATCATATTCCCT
Quencher	TTGAATGAGGGAATATGATGTATGTGG/3IABkFQ/



**Figure 2.7.** Schematic of the substrate complex (a) and the network for reporting the presence of single-stranded Signal strands (b).

**Table 2.3** Fuel sequences

	Fuel	Sequence (5' to 3')	MFE(kcal/mol)
Unmodified	Original	CCTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
5' end modifications	A <sub>1</sub>	ACTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
	T <sub>1</sub>	TCTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
	G <sub>1</sub>	GCTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
	A <sub>1</sub> A <sub>2</sub>	AACTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
	A <sub>1</sub> T <sub>2</sub>	ATTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
	A <sub>1</sub> G <sub>2</sub>	AGTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
5' end modifications	T <sub>1</sub> A <sub>2</sub>	TATACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81
	T <sub>1</sub> T <sub>2</sub>	TTTACGTCTCCAACCTAATTACGGCCCTCATTCAATACCCTACG	-1.81

	T <sub>1</sub> G <sub>2</sub>	TGTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTACG	-1.81
	G <sub>1</sub> A <sub>2</sub>	GATACGTCTCCAACCTAACTTACGGCCCTCATTCAATAC CCTACG	-1.81
	G <sub>1</sub> T <sub>2</sub>	GTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTACG	-1.81
	G <sub>1</sub> G <sub>2</sub>	GGTACGTCTCCAACCTAACTTACGGCCCTCATTCAATAC CCTACG	-1.81
Nick location modificatio ns	A <sub>24</sub>	CCTACGTCTCCAACCTAACTTACGACCCTCATTCAATACC CTACG	-2.20
	T <sub>24</sub>	CCTACGTCTCCAACCTAACTTACGTCCTCATTCAATACC CTACG	-2.51
	C <sub>24</sub>	CCTACGTCTCCAACCTAACTTACGCCCTCATTCAATACC CTACG	-1.47
	A <sub>24</sub> A <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGAACCTCATTCAATACC CTACG	-2.20
	A <sub>24</sub> T <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGATCCTCATTCAATACC CTACG	-2.20
	A <sub>24</sub> G <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGAGCCTCATTCAATACC CTACG	-2.20
	T <sub>24</sub> A <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGTACCTCATTCAATACC CTACG	-2.99
	T <sub>24</sub> T <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGTTCTCATTCAATACC CTACG	-2.66
	T <sub>24</sub> G <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGTGCCTCATTCAATACC CTACG	-3.04
Nick location modificatio ns	C <sub>24</sub> A <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGACCTCATTCAATACC CTACG	-1.47
	C <sub>24</sub> T <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGCTCCTCATTCAATACC CTACG	-1.47
	C <sub>24</sub> G <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGCCCTCATTCAATACC CTACG	-1.47
	A <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGACCTCATTCAATACC CTACG	-1.81
	T <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGGTCCTCATTCAATACC CTACG	-1.81
	G <sub>25</sub>	CCTACGTCTCCAACCTAACTTACGGGCCTCATTCAATACC CTACG	-1.89
3' end modificatio ns	A <sub>43</sub> A <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAAA	-1.81
	A <sub>43</sub> T <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAAT	-1.81
	A <sub>43</sub> C <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAAC	-1.81
	T <sub>43</sub> A <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTATA	-1.81
	T <sub>43</sub> T <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTATT	-1.81
	T <sub>43</sub> C <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTATC	-1.81
	G <sub>43</sub> A <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAGA	-1.81
	G <sub>43</sub> T <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAGT	-1.81
	G <sub>43</sub> C <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAGC	-1.81
3' end modificatio ns	A <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTACA	-1.81
	T <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTACT	-1.81

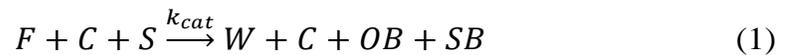
	C <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTACC	-1.81
Multiple- location modificatio ns	G <sub>1</sub> T <sub>2</sub> A <sub>25</sub>	GTTACGTCTCCAACCTAACTTACGGACCTCATTCAATACC CTACG	-1.81
	G <sub>1</sub> G <sub>2</sub> G <sub>43</sub>	GGTACGTCTCCAACCTAACTTACGGCCCTCATTCAATAC CCTAGG	-1.96
	G <sub>1</sub> G <sub>2</sub> T <sub>24</sub> A <sub>25</sub>	GGTACGTCTCCAACCTAACTTACGTACCTCATTCAATACC CTACG	-6.71
	G <sub>1</sub> G <sub>2</sub> G <sub>25</sub> G <sub>43</sub>	GGTACGTCTCCAACCTAACTTACGGGCCTCATTCAATAC CCTAGG	-2.80
	G <sub>1</sub> T <sub>2</sub> G <sub>43</sub> T <sub>44</sub>	GTTACGTCTCCAACCTAACTTACGGCCCTCATTCAATACC CTAGT	-1.81
	G <sub>1</sub> G <sub>2</sub> G <sub>43</sub> T <sub>44</sub>	GGTACGTCTCCAACCTAACTTACGGCCCTCATTCAATAC CCTAGT	-1.81
	T <sub>24</sub> A <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	CCTACGTCTCCAACCTAACTTACGTACCTCATTCAATACC CTAGT	-2.99
	G <sub>1</sub> T <sub>2</sub> A <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	GTTACGTCTCCAACCTAACTTACGGACCTCATTCAATACC CTAGT	-1.81
	G <sub>1</sub> G <sub>2</sub> T <sub>24</sub> G <sub>43</sub> T <sub>44</sub>	GGTACGTCTCCAACCTAACTTACGTCCCTCATTCAATACC CTAGT	-2.51
	G <sub>1</sub> G <sub>2</sub> G <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	GGTACGTCTCCAACCTAACTTACGGGCCTCATTCAATAC CCTAGT	-2.80
	G <sub>1</sub> G <sub>2</sub> T <sub>24</sub> G <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	GGTACGTCTCCAACCTAACTTACGTGCCTCATTCAATACC CTAGT	-4.49

### 2.7.2 Extinction coefficient calculation

The extinction coefficients at 260 nm for single-stranded DNA components were provided by integrated DNA technologies. The extinction coefficients at 260 nm for the reporter and the substrate were calculated by summing up the single and double-stranded regions, as following  $e (total) = e(single) + e(double)$ , where  $e (double) = e(top strand) + e(bottom strand) - 3200 (units)N_{AT} + 2000(units)N_{GC}$ , where  $N_{AT}$  and  $N_{GC}$  are respectively the number of AT pairs and GC pairs in the double-stranded regions.<sup>1, 2</sup>

### 2.7.3 Reaction kinetics models

We consider the following model for the catalytic system:



Equation (1) is the catalyzed reaction for the catalytic system in Figure. 2.1. In this reaction  $F$  is the fuel,  $C$  is the catalyst, and  $S$  is the substrate. In this reaction the fuel  $F$ , catalyst  $C$ , and substrate  $S$  combine to form waste product  $W$  and release catalyst  $C$ , output  $OB$ , and signal  $SB$ . The rate constant is denoted  $k_{cat}$ .

Equation (2) is the leakage reaction for the catalytic system in Figure 2.1. In this reaction the fuel  $F$  and the substrate  $S$  combine to form the waste product  $W$  and release output  $OB$  and signal  $SB$ . This reaction proceeds with the rate constant  $k_{leak}$ .

Equation (3) is the reaction of signal  $SB$  and reporter  $R$  with a rate constant  $k_{TET} = 8 \times 10^5 M^{-1}s^{-1}$ , which is much faster than  $k_{cat}$  and  $k_{leak}$ <sup>3</sup>. As a consequence, we neglect reporter delay and we treat the fluorescence as a direct measure of the  $SB$  concentration.

Equations (1-3) do not include the reverse reactions, that is, the reverse reactions are considered to be negligible.

Leakage and catalyzed reactions have the following rate equation:

$$\frac{d[SB]}{dt} = k_{cat}[F][C][S] + k_{leak}[F][S] \quad (4)$$

Mass balance equations are:

$$[F] = [F]_0 - [SB] \quad (5)$$

$$[S] = [S]_0 - [SB] \quad (6)$$

$$[C] = [C]_0 \quad (7)$$

Equations (4) through (7) yield:

$$\frac{d[SB]}{dt} = k_{cat}([F]_0 - [SB])[C]_0([S]_0 - [SB]) + k_{leak}([F]_0 - [SB])([S]_0 - [SB]) \quad (8)$$

One obtains:

$$[SB] = \frac{[F]_0[S]_0(1-\exp^{[t \cdot (k_{cat}[C]_0 + k_{leak}) \cdot ([S]_0 - [F]_0)])}}{[F]_0 - [S]_0 \exp^{[t \cdot (k_{cat}[C]_0 + k_{leak}) \cdot ([S]_0 - [F]_0)]}} \quad (9)$$

When  $t \rightarrow \infty$  and  $[S]_0 > [F]_0$

$$[SB]_{\infty} = [F]_0 \quad (10)$$

$$[SB]_{cat} = \frac{[SB]}{[SB]_{\infty}} = \frac{[S]_0(1-\exp^{[t \cdot (k_{cat}[C]_0 + k_{leak}) \cdot ([S]_0 - [F]_0)])}}{[F]_0 - [S]_0 \exp^{[t \cdot (k_{cat}[C]_0 + k_{leak}) \cdot ([S]_0 - [F]_0)]}} \quad (11)$$

The fluorescence concentration is approximately equal to the signal strand concentration. Rate constants of catalytic reaction can be obtained by fitting the fluorescence verses time data using the equation (11), as illustrated in Figure 2.6b of chapter 1.

For leakage reaction  $[C]_0 = 0$ , thus

$$[SB] = \frac{[F]_0[S]_0(1-\exp^{[t \cdot k_{leak} \cdot ([S]_0 - [F]_0)])}}{[F]_0 - [S]_0 \exp^{[t \cdot k_{leak} \cdot ([S]_0 - [F]_0)]}} \quad (12)$$

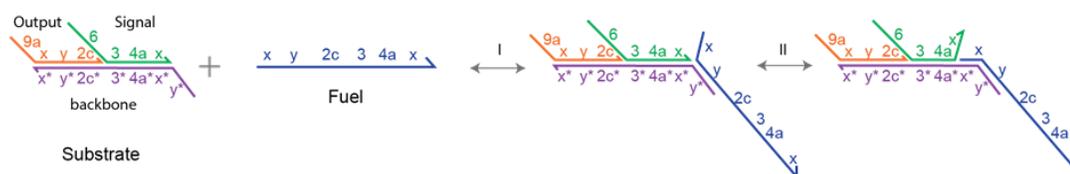
When  $t \rightarrow \infty$  and  $[S]_0 < [F]_0$

$$[SB]_{\infty} = [S]_0 \quad (13)$$

$$[SB]_{leak} = \frac{[SB]}{[SB]_{\infty}} = \frac{[F]_0(1-\exp^{[t \cdot k_{leak} \cdot ([S]_0 - [F]_0)])}}{[F]_0 - [S]_0 \exp^{[t \cdot k_{leak} \cdot ([S]_0 - [F]_0)]}} \quad (14)$$

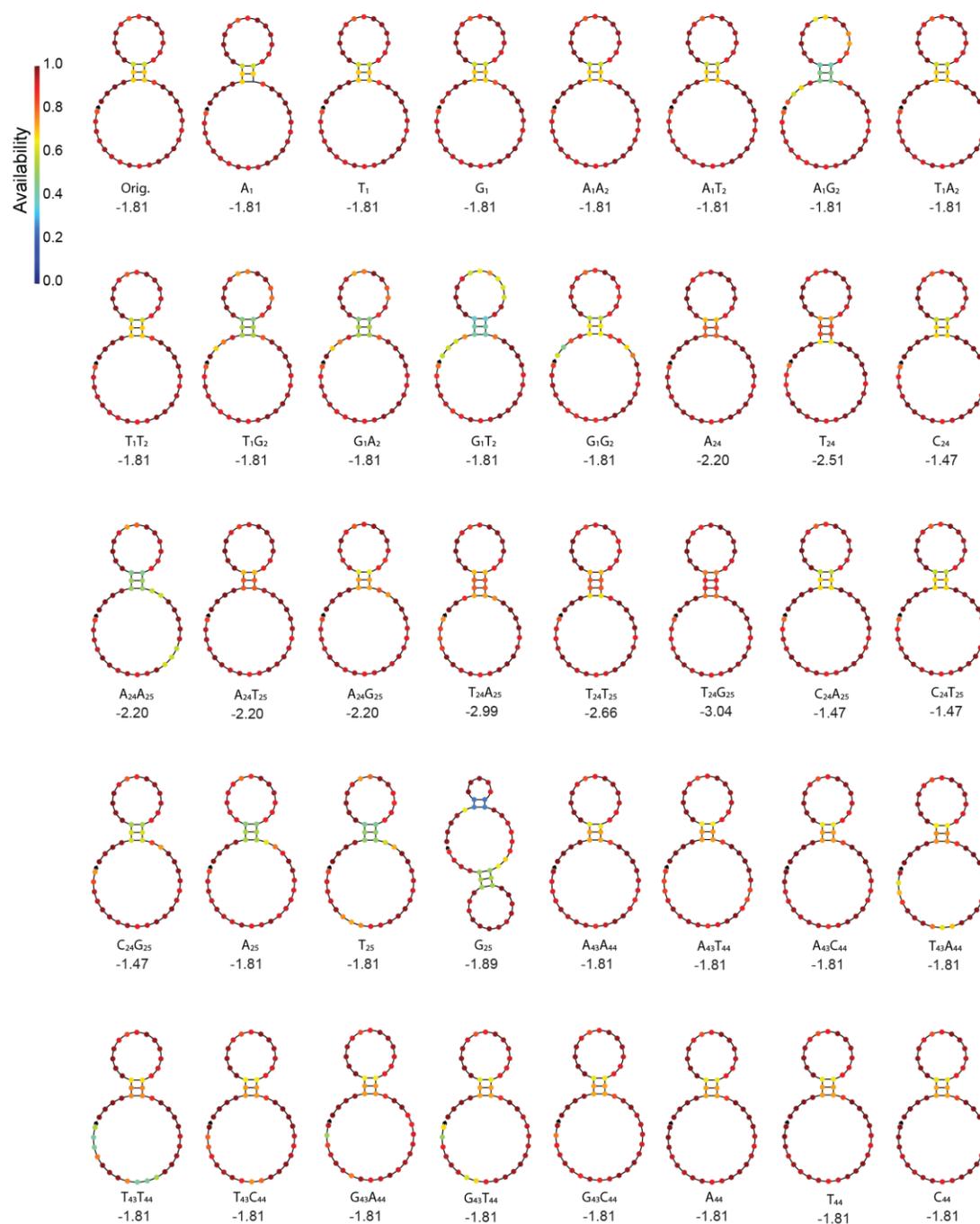
This is the equation to which the fluorescence data is fit in order to extract the leakage rate constant from the fluorescence verses time data of the leakage experiments, as illustrated in Figure 2.6c.

### 2.7.4 Transient binding of the substrate and fuel



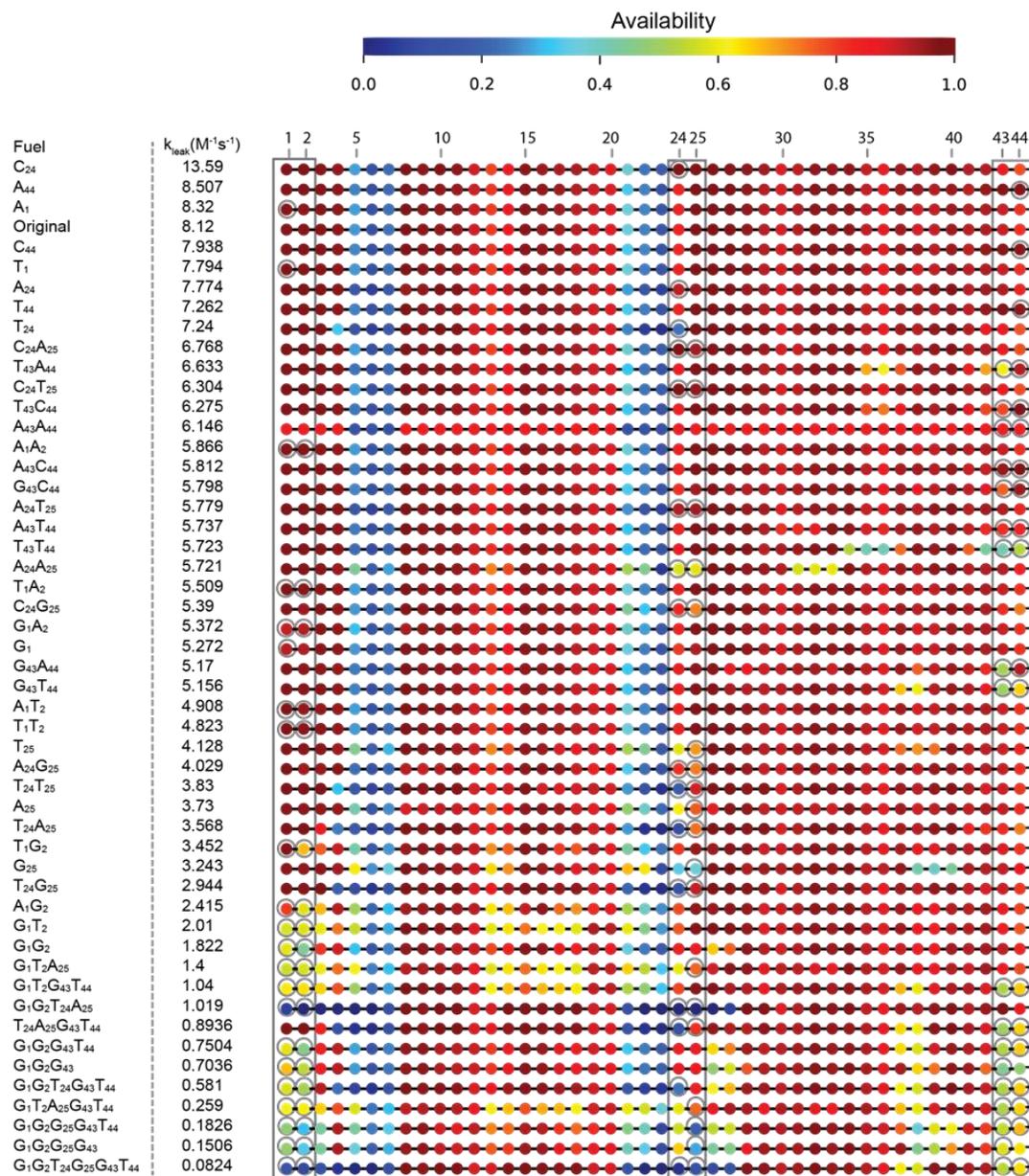
**Figure 2.8.** The transient binding between the substrate toehold and the fuel. The  $y$  domain of the fuel strand and  $y^*$  domain of the substrate toehold can hybridize as shown in reaction I. The  $x$  domain on the 5' end of the fuel will further displace the  $x$  domain of the signal strand via 3-way branch migration (reaction II). This interaction might interfere with the 3' end fuel binding with the substrate and thus affect the leakage rate constants.

### 2.7.5 Minimum free energy structures of the fuel strands

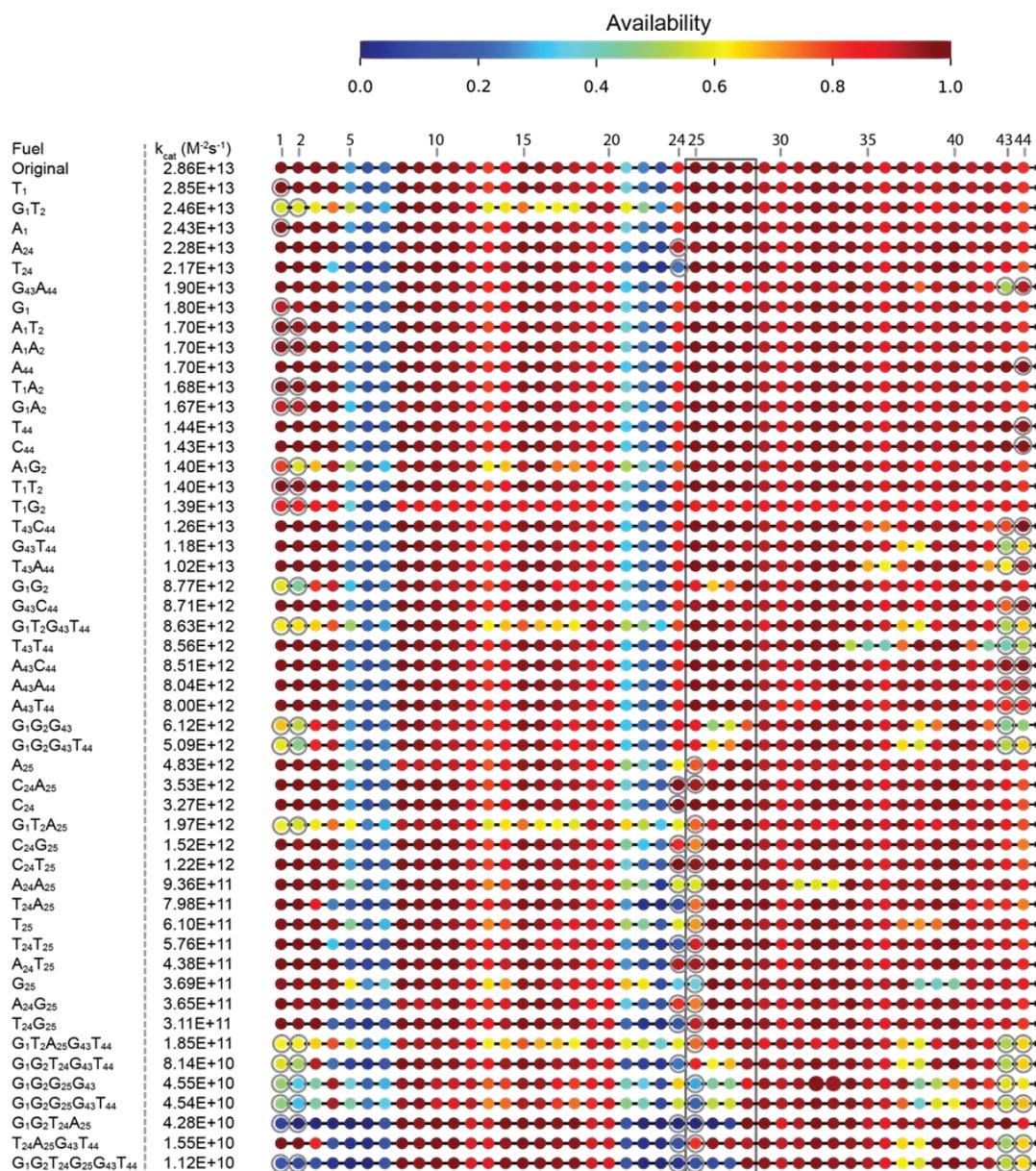


**Figure 2.9.** Minimum free energy (MFE) structures of fuels with single-location modifications as calculated by NUPACK. The Gibbs free energy of each structure is shown in kcal/mol.

## 2.7.6 Availabilities of fuel strand bases



**Figure 2.10.** Leakage rate constants in descending order and corresponding base availabilities of fuel strands. Each base of the fuel strand is represented as a solid circle colored according to the NUPACK calculated availability. Mismatched bases are circled in gray.



**Figure 2.11.** Catalyzed reaction rate constants in descending order and corresponding fuel base availabilities. Each base of the fuel strand is represented as a solid circle colored according to NUPACK calculated availability. The four bases complementary to the toehold are highlighted with the gray box. Mismatched bases are circled in gray.

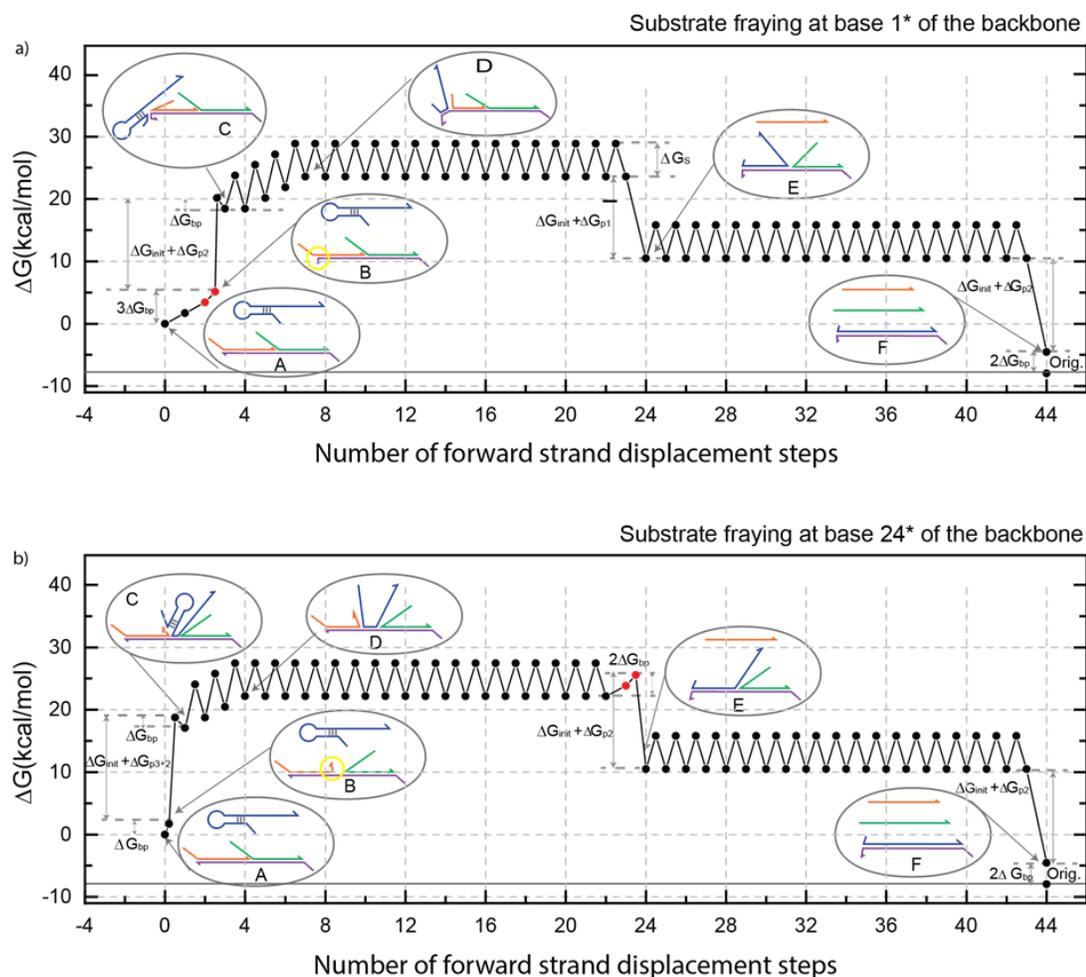
### 2.7.7 Analysis in the context of the Intuitive Energy Landscape (IEL) model

Recently, Srinivas *et al.* published a comprehensive analysis of branch migration and strand displacement and proposed an intuitive energy landscape (IEL) model to describe the experimental rates of strand displacement reactions.<sup>4</sup> In an effort to gain additional insight into leakage, we analyzed the leakage reaction pathways in the framework of the IEL by incorporating the effects of fuel sequence mismatch energy penalties. From the IEL model, we can predict the leakage pathways with the lowest thermodynamic energy barriers (Figure 2.13a,b), yet the most significant fuel modifications were those that affected the high-energy pathways (Figure 2.12a) where the backbone base availabilities were highest. In IEL model, five essential rates and energy parameters  $k_{bi}$ ,  $k_{uni}$ ,  $\Delta G_s$ ,  $\Delta G_p$ , and  $\Delta G_{bp}$  were used to describe strand displacement reaction kinetics, which includes rates of hybridization, fraying, branch migration and branch migration initiation. These parameters were derived or fitted for the case of minimal secondary structures for the single-stranded components and stable duplexes. However, to accurately predict reaction rates, including leakage, base availability from the secondary structures of single strands and unstable duplexes, especially the blunt ends and nick locations, may be an essential part of the IEL model. Future work will include incorporating availability and mutual availability into the IEL model to gain a greater understanding, and thus greater control, of leakage reaction mechanisms.

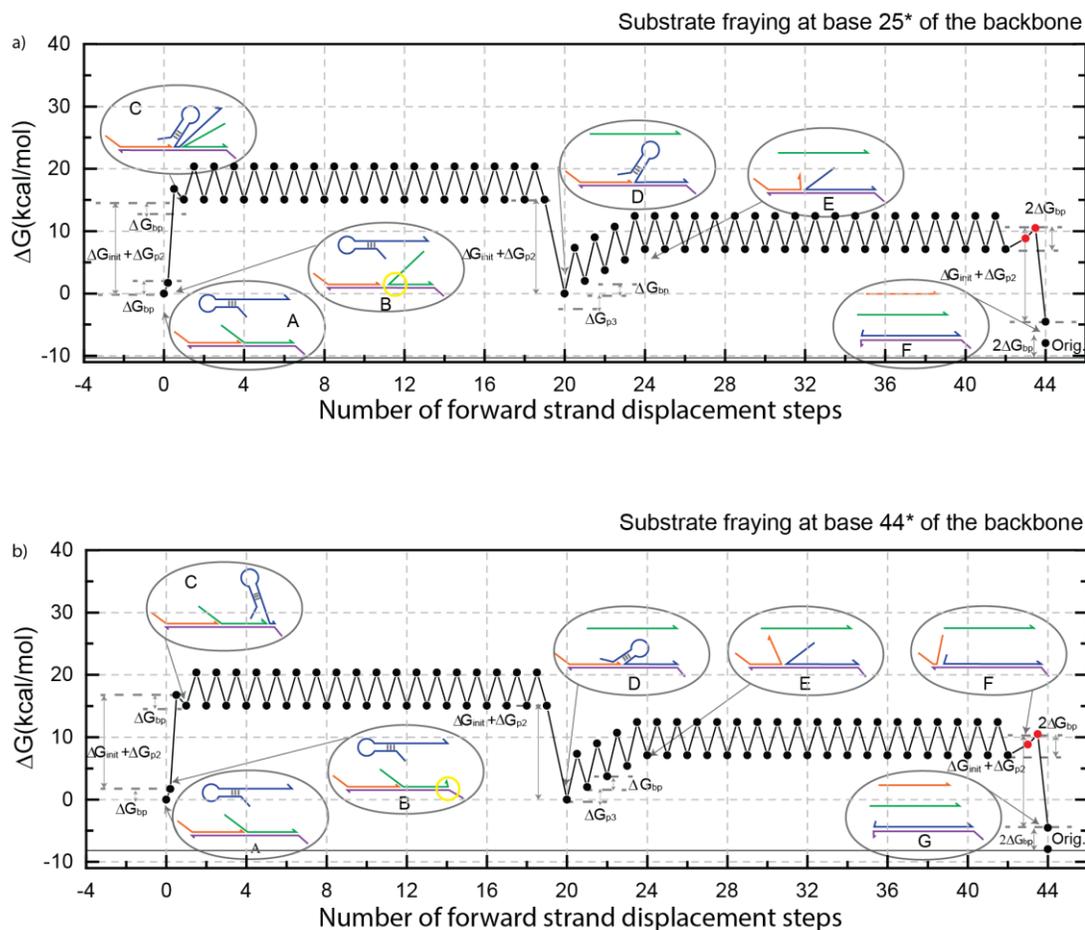
#### 2.7.7.1 The intuitive energy landscape model for leakage pathways

Figures 2.12 and 2.13 shows four possible leakage pathways analyzed in the context of the intuitive energy landscape (IEL) of Srinivas *et al.*<sup>4</sup> In Figure 2.12 and 2.13, state A is the initial state in which the fuel ( $G_1T_2$ ) and substrate are separated, and the Gibbs free energy is taken as zero. For simplicity, all energy parameters were taken from Reference 4. From state A to B, one end of the substrate frays at the cost of one base pair

stacking  $|\Delta G_{bp}|$ . Three base pairs of the substrate have to fray to form the first base pair with the  $G_1T_2$  due to two nucleotide modifications (indicated in the red dots). This results in a higher energy barrier for the pathway illustrating leakage by the fuel 5' end compared to the other three. There is an intermediate step in which the fuel and substrate come close together in correct registry at the cost of  $|\Delta G_{init}|$ . State C is the formation of the first base pair between the fuel and substrate with the energy gain of  $|\Delta G_{bp}|$ , and in this process two overhangs protrude from both sides of the nick with the energy penalty of  $|\Delta G_p|$ . In branch migration process, three base pairs of the fuel have to fray due to the hairpin structure and the energy cost for each base is  $|\Delta G_{bp}|$ . State D is when the fuel hairpin opened and form new base pairs with substrate. The sawtooth pattern represents branch migration steps and the top of the sawtooth is the intermediate transition state with energy penalty of  $|\Delta G_s|$ . In state E, the fuel has completely displaced the output strand, decreasing the system energy by  $|\Delta G_{init}|$  and  $|\Delta G_{p1}|$  (energy difference from no overhang to one overhang in the nick). In the final state F, both the output and signal strands are completely displaced by the fuel, and the gain of the system energy is  $|\Delta G_{init}|$  and  $|\Delta G_{p2}|$  (energy difference from no overhang to two overhangs in the nick). In Figure 2.12a, the final state F is higher than the point Orig. (no mismatches in the fuel strand).



**Figure 2.12.** Intuitive energy landscape (IEL) for two of the four proposed leakage pathways for  $G_1T_2$  fuel modification. Red states indicate mismatch penalties. (a) Leakage initiated at 3' end of the fuel (base 1), (b) Leakage initiated at nick left of the fuel (base 24). Yellow circles denote fraying locations. States A-G are described in the text. For illustration, the values of  $\Delta G_{bp}$ ,  $\Delta G_p$ ,  $\Delta G_s$ , and  $\Delta G_{init}$  were taken from Srinivas et al.<sup>4</sup>  $|\Delta G_{p1}|$  is the energy difference from no overhang to one overhang in the nick,  $|\Delta G_{p2}|$  is the energy difference from no overhang to two overhangs in the nick, and  $|\Delta G_{p3}|$  is the energy difference from one overhang to two overhangs in the nick. For comparison, point "Orig." shows the final state energy when the original fuel is used.



**Figure 2.13. Intuitive energy landscape (IEL) for two of the four proposed leakage pathways for  $G_1T_2$  fuel modification. Red states indicate mismatch penalties. (a) Leakage initiated at nick right of the fuel (base 25), (b) Leakage initiated at 3' end of the fuel (base 44). States A-G are described in the text. Yellow circles denote fraying locations. For illustration, the values of  $\Delta G_{bp}$ ,  $\Delta G_p$ ,  $\Delta G_s$ , and  $\Delta G_{init}$  were taken from Srinivas et al.<sup>4</sup>  $|\Delta G_{p1}|$  is the energy difference from no overhang to one overhang in the nick,  $|\Delta G_{p2}|$  is the energy difference from no overhang to two overhangs in the nick, and  $|\Delta G_{p3}|$  is the energy difference from one overhang to two overhangs in the nick. For comparison, point “Orig.” shows the final state energy when the original fuel is used.**

The leakage rate constant can be expressed by  $k \approx \frac{2e^{-\frac{|\Delta G_{bp}|}{RT}} \rho k_{bi}}{\rho + 2b\gamma e^{(\Delta G_s + \Delta G_p)/RT}}$ , where  $b$  is

the number of bases in the branch migration region, and  $\rho$  is defined as  $\rho = k_{uni}/k_{bi}$ ,  $\gamma$  is defined as  $\gamma = e^{-(|\Delta G_{bp}| - \Delta G_{assoc})/RT} u_0$ , and  $\Delta G_{init} = \Delta G_{volume} + \Delta G_{assoc}$ , where

$\Delta G_{volume} = RT \ln\left(\frac{u_0}{u}\right)$ , and  $\Delta G_{assoc}$  is the free energy cost of association at a standard

concentration of  $u_0 = 1 M$  due to reduction in the entropy caused by lost translational and orientation degrees of freedom.<sup>4</sup> Unimolecular and hybridization rate constants are denoted by  $k_{uni}$  and  $k_{bi}$  respectively, where  $R$  is the universal gas constant and  $T$  is the temperature.

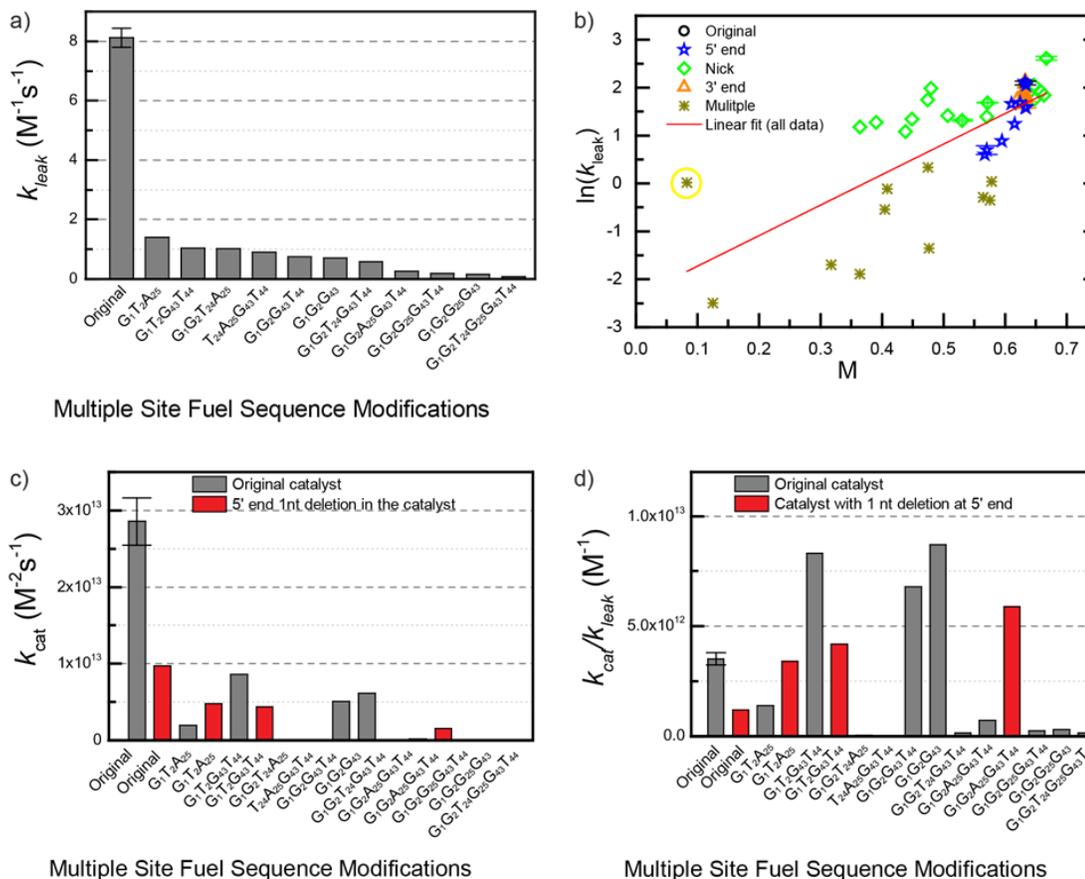
From the above leakage rate constant equation, the availabilities of the fuel and substrate bases can be incorporated into the IEL model. First, the fuel and substrate collide in the nucleation step, which can be described by the bimolecular rate constant  $k_{eff}$ . When the fuel has secondary structures and substrate frays at blunt ends and nick sites, base availabilities of fuel and substrate backbone at nucleation sites must be included in the  $k_{eff}$  to account for the probability of successful nucleation. Then branch migration proceeds after nucleation and this process is a unimolecular reaction, which can be described by  $k_{uni}$ . In the branch migration process, the sawtooth amplitude ( $\Delta G_s$ ) may need to be adjusted to account for secondary structures of the fuel strand. Therefore base availabilities of the fuel could be quantitatively incorporated into local free energy maxima along the sawtooth or through a sequence-dependent base-pairing energy and/or  $k_{uni}$ . By considering base availability of the fuel and substrate backbone, the leakage rate constant can be modified through IEL model to precisely predict reaction rates. The biophysics of nucleation and branch migration steps could be further understood by utility of availability concept in the future.

#### 2.7.8 Multiple location fuel modifications

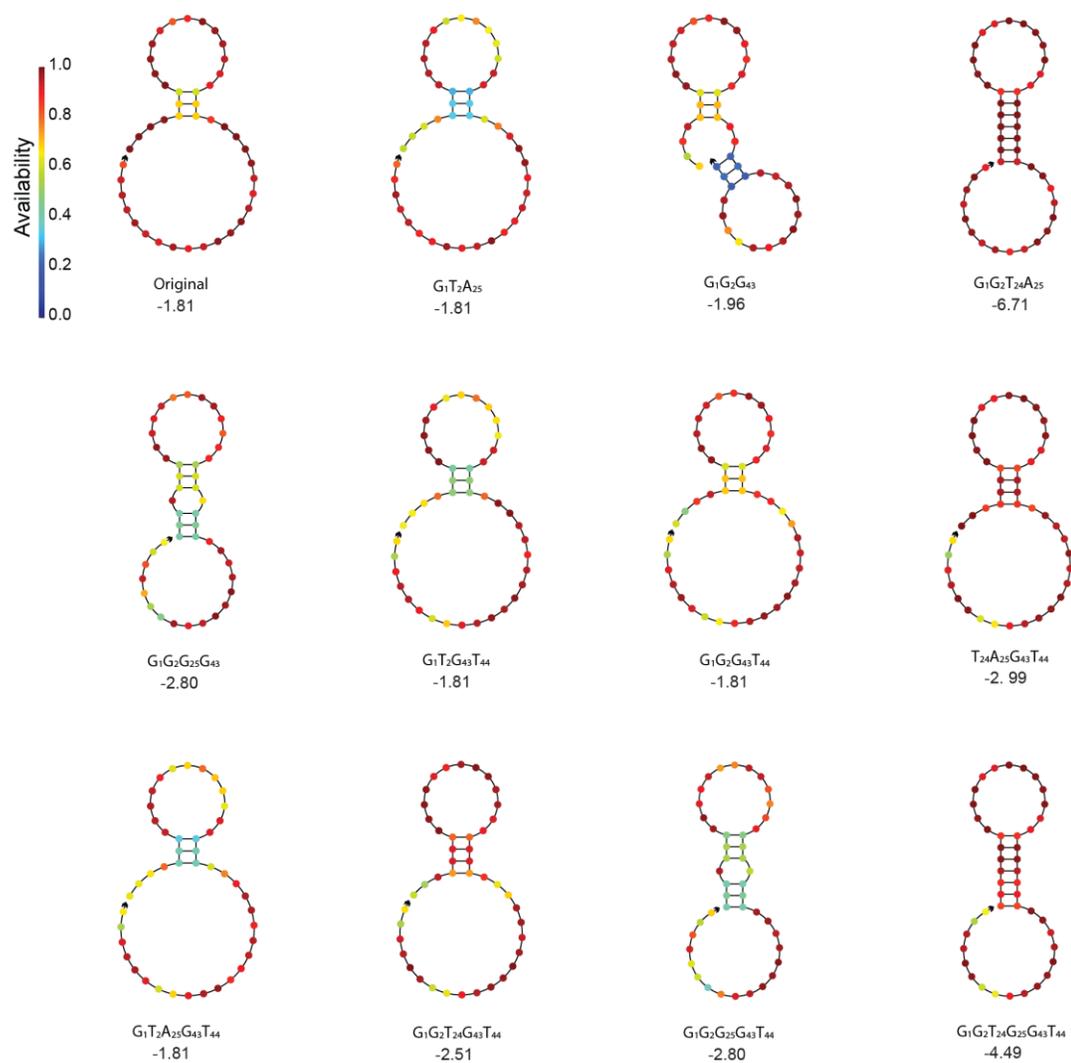
Multiple location fuel modifications have a stronger leakage suppression than the single location modifications, as seen in Figure 2.14. The MFE structures for the fuel strands with sequence modifications at multiple locations are shown in Figure 2.15. For the multiple location modifications, the leakage rate dropped by 100-fold to an almost undetectable level by introducing mismatches at all the four

vulnerable locations. As more mismatches are introduced, leakage reduction can be attributed to an increase in secondary structure of the fuel strand, consistent with the MFE structures seen in Figure 2.15, and a decrease in driving force. As the number of mismatches introduced to the fuel strand increases, the availability of the fuel strand decreases for this system. The changes in availability not only affect the nucleation sites but also appear in regions that can disrupt branch migration. For example, the point in the yellow circle in Figure 2.14(b) with a  $M$  of  $\sim 0.083$  and  $\ln(k_{leak})$  of  $\sim 0$ , is discussed further in Figure 2.16.

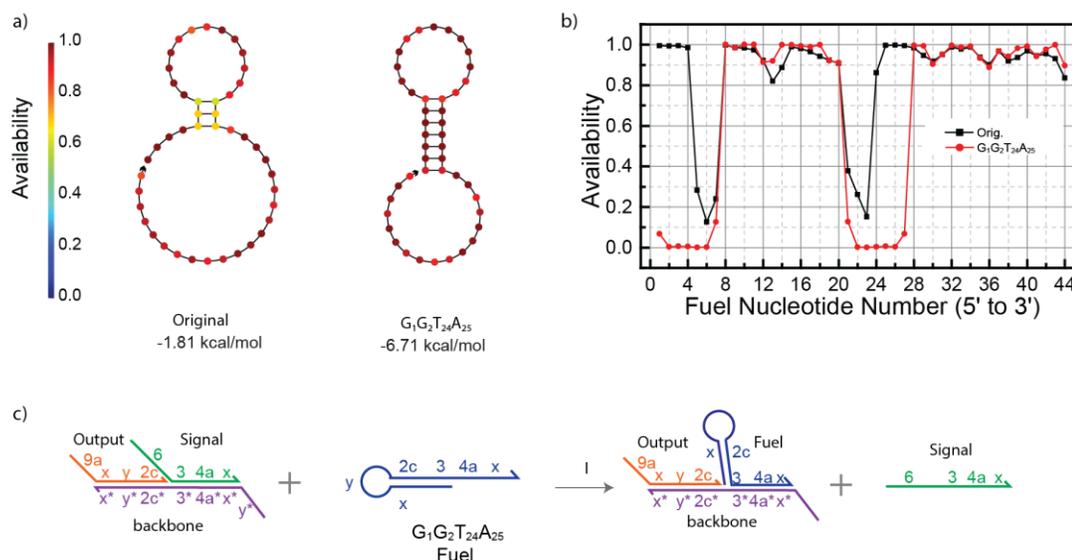
Multiple site fuel modifications also showed lower catalytic reaction rates, as expected from their decreased complementarity to the substrate backbone. As an attempt to recover the rate of the catalytic reaction, we increased the toehold length from 4 nt to 5 nt on domain 3\* of the intermediate 3 (I3) by deleting one nucleotide on the 5' end of the catalyst. This had the desired effect only on the toehold mismatch modifications, while demonstrating the opposite effect on the other modifications (Figure 2.14). There exists a trade-off between slowing the dissociation of the signal strand (SB) and accelerating fuel binding to intermediate 3 (I3) at toehold domain 3\*. This trade-off is advantageous in the case that the fuel strand has a mismatch in the toehold region (base 25), while it is disadvantageous in the case that the fuel is unmodified in the toehold location. Some multiple location modifications achieve a larger  $k_{cat}/k_{leak}$  ratio compared with the original design, however none of these modification exceeded the best of the single location modifications (the ratio of 4, Figure 2.14(d)).



**Figure 2.14.** (a) Leakage rate constants of multiple location fuel modifications. (b) Natural log plot of the leakage rate constant versus the mutual availability between the fuel strand and the backbone on the substrate. Representative error bars of selected samples are shown, indicating that the scatter of the data is greater than the experimental error. The red line is the fit for all modifications combined. (c) Catalytic rate constants of multiple location fuel modifications. Gray bars represent substrate and fuel reacted with original catalyst and red bars represent substrate and fuel reacted with catalyst with 1 nt deletion at 5' end. (d) The ratio  $k_{\text{cat}}/k_{\text{leak}}$  for multiple location fuel modifications. Gray bars represent substrate and fuel reacted with original catalyst and red bars represent substrate and fuel reacted with catalyst with 1 nt deletion at 5' end.



**Figure 2.15. Minimum free energy (MFE) structures of fuels with multiple-location modifications as calculated by NUPACK. The Gibbs free energy for each structure is shown in kcal/mol.**



**Figure 2.16.** (a) Minimum free energy (MFE) structures of original fuel and G<sub>1</sub>G<sub>2</sub>T<sub>24</sub>A<sub>25</sub> modification as calculated by NUPACK and the Gibbs free energy of each structure. (b) Base availabilities for the original fuel and fuel modification G<sub>1</sub>G<sub>2</sub>T<sub>24</sub>A<sub>25</sub>. (c) An alternative reaction pathway to Fig. 1b between the modified fuel (G<sub>1</sub>G<sub>2</sub>T<sub>24</sub>A<sub>25</sub>) and substrate.

Fuel modification G<sub>1</sub>G<sub>2</sub>T<sub>24</sub>A<sub>25</sub> formed a stronger hairpin structure compared with the original fuel (Figure 2.16a) yet yielded a high leakage rate given its low total mutual availability with the substrate backbone (circled in yellow in Figure 2.14b). The fuel base availabilities are shown in Figure 2.16b. However the base availabilities are higher in domain 4a, x and part of domain 3 for the modified fuel, which are corresponding to the double helix region between the signal and backbone of the substrate. An alternative reaction pathway is proposed in Figure 2.16c. Fuel (G<sub>1</sub>G<sub>2</sub>T<sub>24</sub>A<sub>25</sub>) can initiate reaction with the substrate through the 3' end of the fuel, which displaces x domain of the signal due to substrate fraying and then completely displaces signal strand through branch migration. In addition, the toehold domain y\* of the substrate is less likely to stick with y domain of the fuel since it is sequestered in a stable hairpin structure. The x domain of the fuel is also unlikely to further displace x domain of the signal due to a 4-way branch migration. Thus, the right-side of the substrate are more vulnerable for fuel invasion and high leakage rate.

This analysis offers a plausible explanation to the anomalously high leakage rate for a strand with an overall low availability. This also highlights the utility of the concept of base availability and minimum free energy (MFE) structures for analyzing leakage reaction pathways and yielding insight on the leakage reaction rate

### 2.7.9 Analysis of catalytic rates

The effect of mismatches on the catalyzed reaction can be explained via the reaction mechanism. The mismatch positions play a very important role in the catalyzed reaction. Trends can be observed by grouping the mismatch positions of the fuel strand as bases 1, 2 and 24, base 25, and bases 43 and 44. This catalytic cycle can be simplified to four reactions modeled by Zhang *et al.* as shown below.<sup>3</sup>



The reaction between the substrate and the catalyst and the signal reacting with the reporter are not affected by introducing fuel mismatch modifications. Equation (16) is a toehold mediated strand displacement and its rate constant  $k_2$  is mainly determined by the toehold domain 3 of the fuel strand (Figure 1). Thus, this reaction is strongly affected by toehold modifications at base 25. Equation (17) is the release of catalyst from the last intermediate (I5) to produce the waste product. Mismatch modifications at bases 43 and 44 slow down the catalyst release.

Base 25 is in domain 3, which is used by the fuel (F) to bind the intermediate 3 (I3). Single mismatches at base 25 of the fuel strand slow down the

overall catalytic reaction from one order of magnitude to over two orders of magnitude (Figure 2.6). Mismatches at the toehold position are the most detrimental to the catalytic rate.

Bases 1, 2 and 24 modifications had the least impact on the catalyzed rate. To explain this, we look at where they are encountered during strand displacement. According to Machinek *et al.*,<sup>5</sup> the last few base-pairs at the end of the branch migration spontaneously melt during strand displacement. Bases 1 and 2 are involved at the end of the strand displacement in equation (16). The last few bases of the strand OB spontaneously fall off from intermediate 4 before fuel mismatches at base 1 and 2 are involved in the reaction. However, base 24 is right next to toehold domain 3 where it is encountered at the beginning of branch migration according to the pathway depicted in Figure 2.1a. An alternative reaction pathway can proceed as following: the fuel strand releases the catalyst first instead of releasing strand OB due to higher activation energy barrier caused by mismatches at base 24. Mismatches at base 24 is encountered after the fuel establishes a long toehold with the strand backbone. This agrees with Machinek's result that once a sufficiently long toehold has been established, mismatches in the branch migration region do not significantly impede strand displacement.<sup>5</sup>

Equation (17) is a first order reaction, in which 6 base pairs in the catalyst strand spontaneously dissociate from intermediate (15). For the 3' site fuel modifications, the catalyst has to spontaneously detach seven bases for modifications at base 44 and eight bases for modifications at bases 44 and 43. Catalytic reaction rates of fuel modifications at base 44 are slower than the

unmodified fuel strand, which can be attributed to a slower release of the catalyst from intermediate (I5). The effect of delaying catalyst release is even stronger in bases 43 and 44 fuel modifications.

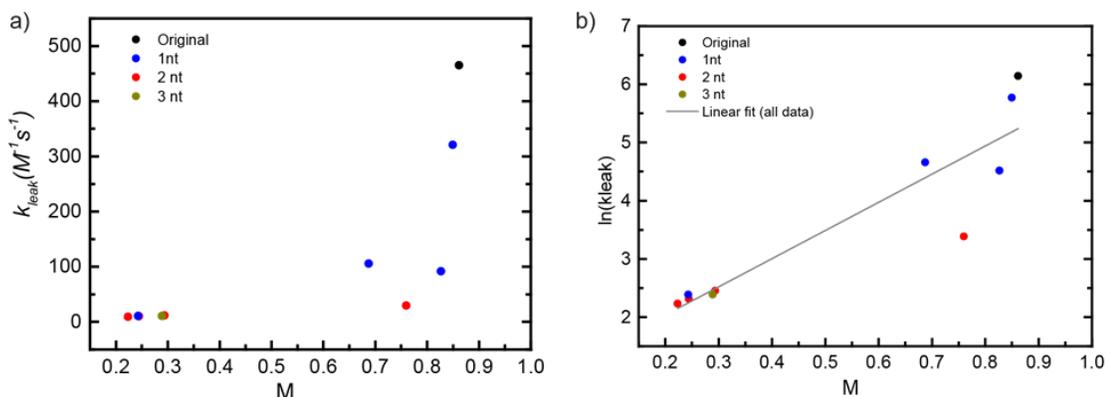
#### 2.7.10 Rate constants and mutual availability

A complete summary of leakage and catalytic rate constants, performance ratios, and total mutual availabilities are provided in Table 2.4 for both single and multiple location fuel modifications.

**Table 2.4 Rate constants and total mutual availability**

	Fuel	$k_{leak}$ ( $M^{-1}s^{-1}$ )	$k_{cat}$ ( $M^{-2}s^{-1}$ )	$k_{cat} / k_{leak}$	$\ln(k_{leak})$	$M$
Unmodified	Original	8.120	2.856E+13	3.512E+12	2.094	0.63336
5' end modifications	A <sub>1</sub>	8.320	2.433E+13	2.924E+12	2.119	0.63252
	T <sub>1</sub>	7.794	2.850E+13	3.656E+12	2.053	0.63431
	G <sub>1</sub>	5.272	1.800E+13	3.414E+12	1.662	0.61041
	A <sub>1</sub> A <sub>2</sub>	5.866	1.700E+13	2.897E+12	1.769	0.63366
	A <sub>1</sub> T <sub>2</sub>	4.908	1.700E+13	3.464E+12	1.591	0.63329
	A <sub>1</sub> G <sub>2</sub>	2.415	1.398E+13	5.790E+12	0.882	0.59486
	T <sub>1</sub> A <sub>2</sub>	5.509	1.683E+13	3.055E+12	1.706	0.63375
	T <sub>1</sub> T <sub>2</sub>	4.823	1.397E+13	2.896E+12	1.573	0.63426
	T <sub>1</sub> G <sub>2</sub>	3.452	1.392E+13	4.032E+12	1.239	0.61562
	G <sub>1</sub> A <sub>2</sub>	5.372	1.665E+13	3.099E+12	1.681	0.62457
	G <sub>1</sub> T <sub>2</sub>	2.010	2.456E+13	1.216E+13	0.698	0.57031
G <sub>1</sub> G <sub>2</sub>	1.822	8.769E+12	4.813E+12	0.600	0.56708	
Nick location modifications	A <sub>24</sub>	7.774	2.283E+13	2.937E+12	2.05	0.64705
	T <sub>24</sub>	7.240	2.167E+13	2.992E+12	1.980	0.47923
	C <sub>24</sub>	13.590	3.16E+13	2.320E+12	2.609	0.66726
	A <sub>24</sub> A <sub>25</sub>	5.721	9.355E+11	1.635E+11	1.744	0.47431
	A <sub>24</sub> T <sub>25</sub>	5.779	4.381E+11	7.580E+10	1.754	0.64776
	A <sub>24</sub> G <sub>25</sub>	4.029	3.650E+11	9.058E+10	1.394	0.57029
	T <sub>24</sub> A <sub>25</sub>	3.568	7.983E+11	2.237E+11	1.272	0.39064
	T <sub>24</sub> T <sub>25</sub>	3.830	5.755E+11	1.502E+11	1.343	0.44876
	T <sub>24</sub> G <sub>25</sub>	2.944	3.109E+11	1.055E+11	1.080	0.4381
	C <sub>24</sub> A <sub>25</sub>	6.768	3.534E+12	5.222E+11	1.912	0.65833
	C <sub>24</sub> T <sub>25</sub>	6.304	1.215E+12	1.926E+11	1.841	0.66309
	C <sub>24</sub> G <sub>25</sub>	5.390	1.518E+12	2.818E+11	1.685	0.5717
	A <sub>25</sub>	3.730	4.829E+12	1.294E+12	1.316	0.5302
	T <sub>25</sub>	4.128	6.103E+11	1.478E+11	1.418	0.50664
G <sub>25</sub>	3.243	3.694E+11	1.138E+11	1.176	0.36412	
3' end modification	A <sub>43</sub> A <sub>44</sub>	6.146	8.039E+12	1.307E+12	1.816	0.6314
	A <sub>43</sub> T <sub>44</sub>	5.737	7.995E+12	1.393E+12	1.747	0.63224
	A <sub>43</sub> C <sub>44</sub>	5.812	8.506E+12	1.463E+12	1.760	0.63252
	T <sub>43</sub> A <sub>44</sub>	6.633	1.020E+13	1.538E+12	1.892	0.63188
	T <sub>43</sub> T <sub>44</sub>	5.723	8.556E+12	1.494E+12	1.745	0.62699
	T <sub>43</sub> C <sub>44</sub>	6.275	1.258E+13	2.004E+12	1.836	0.63264
	G <sub>43</sub> A <sub>44</sub>	5.170	1.900E+13	3.673E+12	1.643	0.63356
3' end modification	G <sub>43</sub> T <sub>44</sub>	5.156	1.180E+13	2.287E+12	1.640	0.6313
	G <sub>43</sub> C <sub>44</sub>	5.798	8.707E+12	1.501E+12	1.757	0.6201
	A <sub>44</sub>	8.507	1.700E+13	1.998E+12	2.141	0.63267
	T <sub>44</sub>	7.262	1.439E+13	1.981E+12	1.983	0.63299
	C <sub>44</sub>	7.938	1.427E+13	1.797E+12	2.072	0.63157
Multiple-location modifications	G <sub>1</sub> T <sub>2</sub> A <sub>25</sub>	1.404	1.97E+12	1.407E+12	0.336	0.47453
	G <sub>1</sub> G <sub>2</sub> G <sub>43</sub>	0.704	6.12E+12	8.698E+12	-0.352	0.57542
	G <sub>1</sub> G <sub>2</sub> T <sub>24</sub> A <sub>25</sub>	1.019	4.28E+10	4.200E+10	0.019	0.08283
	G <sub>1</sub> G <sub>2</sub> G <sub>25</sub> G <sub>43</sub>	0.151	4.55E+10	3.021E+11	-1.893	0.36378
	G <sub>1</sub> T <sub>2</sub> G <sub>43</sub> T <sub>44</sub>	1.049	8.63E+12	8.298E+12	0.039	0.57801
	G <sub>1</sub> G <sub>2</sub> G <sub>43</sub> T <sub>44</sub>	0.750	5.09E+12	6.783E+12	-0.287	0.5646
	T <sub>24</sub> A <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	0.894	1.55E+10	1.735E+10	-0.112	0.4084
	G <sub>1</sub> T <sub>2</sub> A <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	0.259	1.85E+11	7.143E+11	-1.351	0.47614
	G <sub>1</sub> G <sub>2</sub> T <sub>24</sub> G <sub>43</sub> T <sub>44</sub>	0.581	8.14E+10	1.401E+11	-0.543	0.40441
	G <sub>1</sub> G <sub>2</sub> G <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	0.183	4.54E+10	2.486E+11	-1.700	0.31731
	G <sub>1</sub> G <sub>2</sub> T <sub>24</sub> G <sub>25</sub> G <sub>43</sub> T <sub>44</sub>	0.082	1.12E+10	1.359E+11	-2.496	0.12528

### 2.7.11 Mutual availability and rate constants in a hairpin system



**Figure 2.17.** (a) Leakage rate constants for each hairpin 2 modification plotted versus total mutual availability between hairpin 1 and hairpin 2. The leakage rate for the original hairpins is shown in black while the 1 nt, 2 nt, and 3 nt hairpin 2 modifications are shown in blue, red and dark yellow, respectively. (b) Natural log plot of the leakage rate constant versus the total mutual availability. The solid gray line is the fit for all modifications combined with an adjusted R-squared 0.82.

The total mutual availability  $M$  was calculated in NUPACK for Jiang et al.'s hairpin design CircA.<sup>6</sup> In Figures. 3 and 4 of Reference (6), the final intensity for 50 nM H1 is about 4300 relative fluorescence units (RFU) for the original hairpin system, and the original leakage 6 RFU/min is corresponding to 0.0697 nM/min. Thus the original leakage rate constant can be calculated as  $(0.0697 \text{ nM min}^{-1}/50 \text{ nM} * 50 \text{ nM}) = 465 \text{ M}^{-1} \text{s}^{-1}$ . The data of total mutual availability and natural log of leakage rate constants was fit with a linear line. The fit for the dataset has a slope of 4.29 with an adjusted  $R^2$  of 0.82. In NUPACK, the concentration for each hairpin is set to 50 nM, temperature is 25 °C, dangle is set to all and salt concentrations are 0.145 M  $\text{Na}^+$  and 0 M  $\text{Mg}^{2+}$ .

## REFERENCES

1. Qian, L., and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades, *Science* 332, 1196-1201.
2. Puglisi, J. D., and Tinoco Jr, I. (1989) [22] absorbance melting curves of RNA, In *Methods in Enzymology* (James E. Dahlberg, J. N. A., Ed.), pp 304-325.
3. Zhang, D. Y., Turberfield, A. J., Yurke, B., and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA, *Science* 318, 1121-1125.
4. Srinivas, N., Ouldridge, T. E., Sulc, P., Schaeffer, J. M., Yurke, B., Louis, A. A., Doye, J. P. K., and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement, *Nucleic Acids Research* 41, 10641-10658.
5. Machinek, R. R., Ouldridge, T. E., Haley, N. E., Bath, J., and Turberfield, A. J. (2014) Programmable energy landscapes for kinetic control of DNA strand displacement, *Nature Communication* 5.
6. Jiang, Y. S., Bhadra, S., Li, B., and Ellington, A. D. (2014) Mismatches Improve the Performance of Strand-Displacement Nucleic Acid Circuits, *Angewandte Chemie* 126, 1876-1879.

CHAPTER THREE: KINETICS OF DNA STRAND DISPLACEMENT SYSTEMS  
WITH LOCKED NUCLEIC ACIDS

*Xiaoping Olson,<sup>a</sup> Shohei Kotani,<sup>a</sup> Bernard Yurke,<sup>a,b</sup>  
Elton Graugnard,<sup>a,\*</sup> and William L Hughes.<sup>a,\*</sup>*

<sup>a</sup> Micron School of Materials Science & Engineering,  
Boise State University, 1910 University Dr., Boise, ID, 83725, USA.

<sup>b</sup> Department of Electrical & Computer Engineering,  
Boise State University, 1910 University Dr., Boise, ID, 83725, USA.

*\* To whom correspondence should be addressed.*

This chapter will be published in a peer-reviewed journal.

**Abstract**

Locked nucleic acids (LNAs) are conformationally restricted RNA nucleotides. Their increased thermal stability and selectivity towards their complements makes them well-suited for diagnostic and thernostic applications. While the structural and thermodyanmic properties of LNA-LNA, LNA-RNA, and LNA-DNA hybridization are known, the kinetic effects of incorproating LNA nucleotides into DNA strand displacment systems is not. Here we thoroughly studied strand displacment kinetics as a function of the number and position of LNA nucleotides in DNA oligonucleotides. When compared to an all-DNA control, with an identical sequence, the leakage rate constant was reduced from  $1.48 \text{ M}^{-1}\text{s}^{-1}$  to  $0.03 \text{ M}^{-1}\text{s}^{-1}$  and the invasion rate was preserved for a hybrid DNA/LNA system. Total performance enhancement ratio also increased 70 fold

when calculating the ratio of the invading rate to the leakage rate constants for a hybrid system. The rational substitution of LNA nucleotides for DNA nucleotides preserves sequence space while improving the signal to noise ratio of strand displacement systems. Hybrid DNA/LNA systems offer great potential for high performance chemical reactions networks that include catalyzed hairpin assemblies, hairpin chain reactions, motors, walkers, and seesaw gates.

### 3.1 Introduction

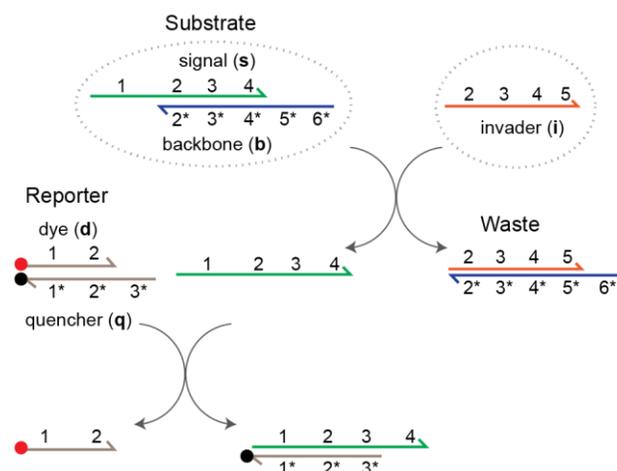
The thermodynamics<sup>1-7</sup> and kinetics<sup>8-12</sup> of Watson-Crick hybridization and strand displacement are well known for DNA and RNA oligonucleotides. As an alternative to naturally occurring nucleic acids, locked nucleic acids (LNAs) are conformationally restricted RNA nucleotides where the 2' oxygen in the ribose bonds to the 4' carbon.<sup>13-16</sup> This covalent bond constrains the sugar in the N-type (C3'-endo) conformation, which in turn preorganizes the phosphate backbone, promotes base stacking, and forces the double helix into its A-form.<sup>16-22,23, 24</sup> These attributes increase LNA's thermal stability and selectivity towards its Watson-Crick complements: including LNA, RNA, and DNA. Naturally occurring nucleotides that neighbor LNA nucleotides also adopt the N-type conformation.<sup>23, 25</sup> When a hybrid DNA/LNA strand binds to an all-DNA oligonucleotide, the structure reflects the number of LNA nucleotides incorporated into the duplex. For example, the A to B-form ratio increases as the number of LNA nucleotides increases.<sup>23</sup>

The stability of LNA containing duplexes can be considered in terms of the Gibbs free energy – which accounts for the entropic and enthalpic contributions of including conformationally restricted nucleotides into a strand. The positive entropic change is

from LNA preorganizing the phosphate backbone.<sup>26</sup> In comparison, the more negative enthalpic change is from an increase in base stacking from LNA when compared to naturally occurring nucleotides such as RNA and DNA.<sup>26</sup> Thermodynamic parameters have been reported for DNA duplexes with single LNA substitutions.<sup>27-30</sup> The results indicate that LNA pyrimidines are more stable than LNA purines and that the overall duplex stability is highly dependent on the DNA nucleotides that neighbor the LNA nucleotides.<sup>27</sup> As a consequence, the melting temperature of a DNA complex ranges from +1 to +8 °C for every LNA nucleotide added.<sup>28, 29</sup> In addition, the thermal stability of a DNA duplex saturates as the number of LNA nucleotides approaches ~50% of the total content.<sup>15, 30</sup> For example, the melting temperature increased on average 5.3 °C per LNA for a 9-nucleotide DNA duplex that had three randomly distributed LNAs on one of its strands. In comparison, the melting temperature increased on average 4.5 °C per LNA when an equivalent duplex was fully saturated with LNA on one of its strands.<sup>19</sup>

In addition to thermodynamic parameters, kinetic parameters have been measured when incorporating LNA nucleotides into DNA systems.<sup>31</sup> The results indicate that LNA-DNA base pairs have an increased binding affinity when compared to DNA-DNA base pairs because they have a slower dissociation rate constant rather than a faster hybridization rate constant.<sup>31, 32</sup> In spite of these structural, thermodynamic, hybridization, and dissociation attributes, the kinetics of incorporating LNA nucleotides into DNA strand displacement systems<sup>8, 33-39</sup> has not been explored. Furthermore, leakage suppression and total system performance (i.e. the signal to noise ratio)<sup>40-43</sup> have not been studied in DNA reactions that include LNA nucleotides.

Presented here for the first time, LNA nucleotides were substituted for DNA nucleotides in a strand displacement system (Figure 3.1). Independent of the number or position of the substitutions, the oligonucleotide sequence was fixed. In our model system, the invader (**i**) hybridizes with the Substrate complex at toehold domain 5 and displaces the signal strand (**s**); creating a Waste complex. The signal strand then reacts with the Reporter complex and releases a dye strand (**d**). Using a fluorometer, intrinsic leakage rates were measured between zero-toehold invaders and Substrates. In contrast, invasion rates were measured between 6 nt toehold invaders and identical Substrates. For both experiments, all-DNA oligonucleotide invader and Substrate controls were compared to hybrid DNA/LNA oligonucleotide invader and Substrate variants – with identical sequences. Leakage was minimized by site-specifically incorporating LNA nucleotides into DNA Substrates. Equally as important, the elevated invasion rates were maintained by incorporating LNA nucleotides into the invader strand. Experimental methods, results, and discussion for how to optimize the kinetic performance of a DNA strand displacement system by site-specifically substituting LNA for DNA nucleotides into the system are described below.



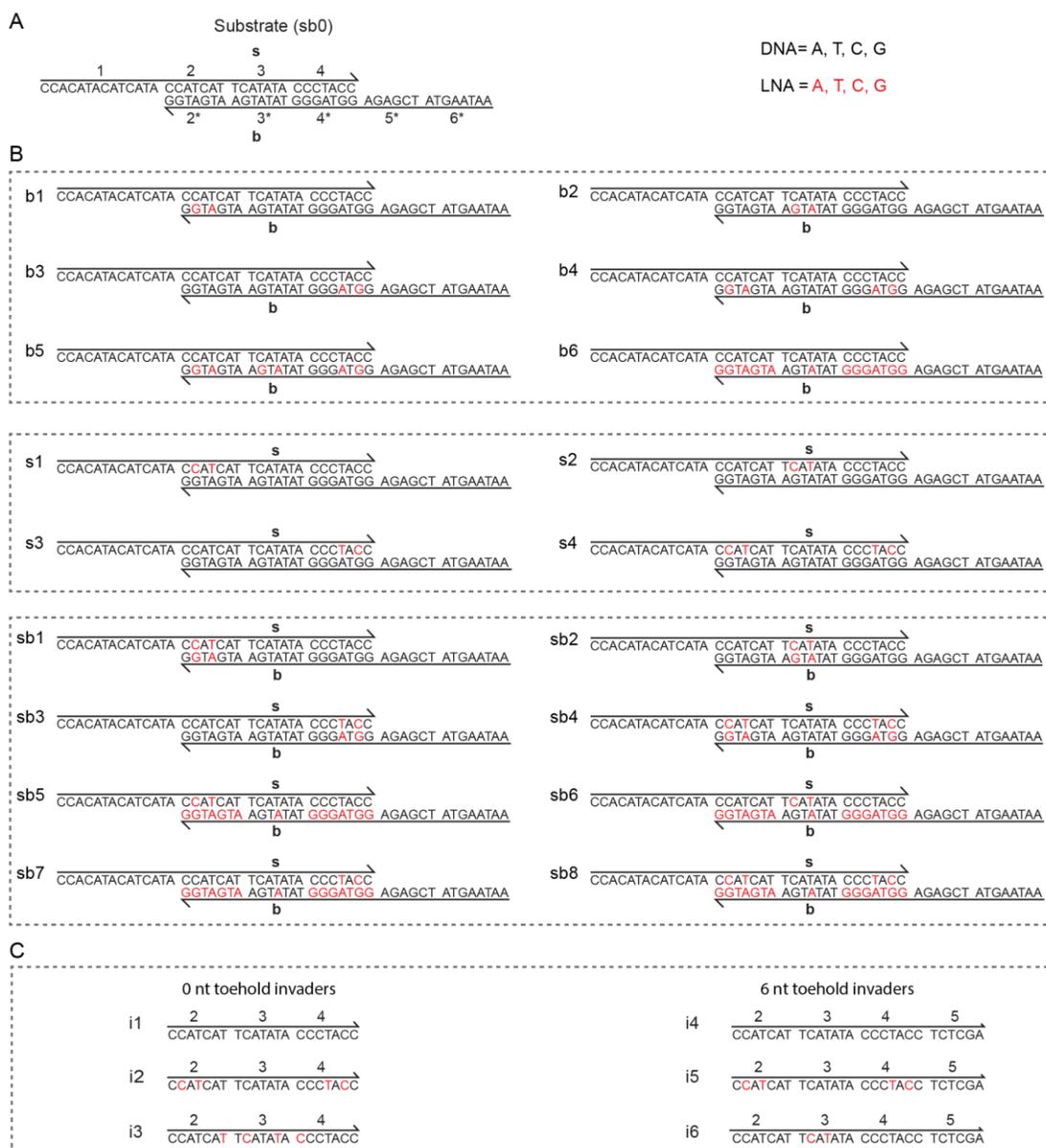
**Figure 3.1 Schematic of a nucleic acid based, toehold mediated strand displacement system. Functional domains are represented by numbers and complementary domains are denoted by numbers with asterisks. The Substrate complex includes signal (s) and backbone (b) strands, while the Reporter complex includes dye (d) and quencher (q) strands. Various LNA substitutions have been made to the signal, backbone, and invader (i) sequences in this study. During invasion, the invader hybridizes with the backbone at domain 5\* and displaces the signal strand via three way branch migration. The invader and backbone form a Waste complex. The free signal strand then reacts with the Reporter complex and releases a dye strand, which is monitored by a fluorometer.**

### 3.2 Results and Discussion

DNA nucleotides were site-specifically substituted by LNA nucleotides in the backbone (b), signal (s), and the backbone and signal of the original sb0 Substrate (Figure. 3.2A,B). LNA substitutions were made in the invaders (i) with 0 and 6 nucleotide (nt) toeholds (Figure 3.2C). All sequences are listed in Section 3.8.1. Experimentally, leakage reactions were measured between invaders with 0 nt toeholds and the Substrates, while invasion reactions were measured between invaders with 6 nt toeholds and the same Substrates. The rate constants were extracted by fitting the data using a 2<sup>nd</sup> order reaction model (Section 3.8.2). Reporter kinetics and control experiments were shown in Section 3.8.3 and 3.8.4 respectively.

#### 3.2.1 Leakage reactions between a DNA invader and hybrid DNA/LNA Substrates

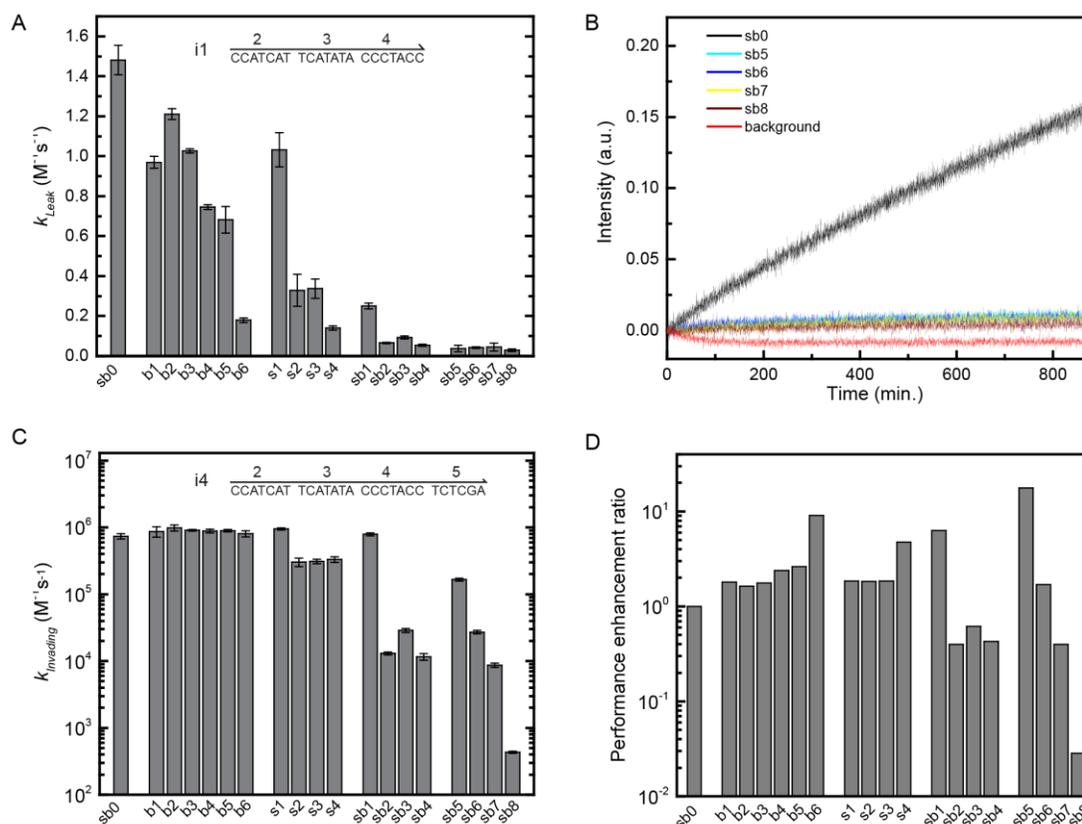
The leakage rate constants between the 0 nt toehold invader **i1** and the Substrates (Figure 3.2B) are shown in Figure 3.3A. Select fluorescence traces for low leak systems are shown in Figure 3.3B. For Substrates with LNA substitutions on the backbone (b1, b2, b3, b4, b5, and b6), there is no enthalpic change in the base pairing during the leakage reaction – in which the DNA invader replaces the signal strand of the Substrate. In general, the more LNAs that are incorporated into the backbone, the greater the leakage suppression. For example, when 15 LNA nucleotides substituted for DNA nucleotides in the b6 backbone, leakage was suppressed by a factor of 7. In addition, Substrates b1 and b3, with two LNAs near the terminal ends of the duplex, had a more pronounced effect on leakage reduction than Substrate b2 with two LNAs in the center of the complex. This is likely attributed to the stronger DNA-LNA base pairing, which reduces fraying frequency and hence reduces the probability for nucleation to occur between the Substrates and the 0 nt toehold invader.



**Figure 3.2.** (A) The original sequence-level Substrate (sb0). Black and red letters respectively denote DNA and LNA nucleotides. (B) Hybrid DNA/LNA Substrates. Substrates b1-b6 selectively substituted LNA for DNA nucleotides on the backbone strand. Substrates s1-s4 selectively substituted LNA for DNA nucleotides on the signal strand. Substrates sb1-sb8 selectively substituted LNA for DNA nucleotides on both the signal and backbone strands. (C) Zero nucleotide (nt) invaders (i1, i2 and i3) were used for the leakage reactions and 6 nt toehold invaders (i4, i5 and i6) were used for the invasion reactions.

For Substrates s2, s3, and s4, with LNA substitutions on the signal strand, the leakage rate constants are smaller than Substrates b2, b3, and b4 with the same number and position of LNAs on the backbone. The reason is because there is an enthalpic penalty in the leakage reaction, where the LNA-DNA base pairs in the Substrates are replaced by DNA-DNA base pairs in the Waste complex.<sup>19</sup> During branch migration, this enthalpic penalty renders a bias for the signal to hybridize to the backbone rather than being replaced by the invader strand.<sup>44</sup> The data also shows that the leakage performance of Substrates s1 and s3 are dramatically different even though they contain the same number and identity of LNAs.

To quantify the effects of secondary structure on the leakage rates of Substrates s1 and s3, the probability that a base was unpaired at equilibrium was calculated for Substrate sb0 using NUPACK (Section 3.8.5).<sup>12</sup> According to our analysis, base availability was higher for the right versus the left side of the sb0 duplex – indicating that the right side of the Substrate is more susceptible to leakage because of fraying and hence favorable nucleation between the Substrate and its zero nt toehold invader. As a consequence, site-specific substitutions of LNA's for DNA nucleotides has a greater leakage suppression effect near the right side (domain 4) versus the left side (domain 2) of the sb0 Substrate. By extension Substrate s3 has greater leakage suppression than Substrate s1. In support of this claim, experimental results in Section 3.8.5 show that leakage is faster from the right versus the left side of the Substrate when the i1 invader was separately truncated by 2 nucleotides at its 5' and then 3' end.



**Figure 3.3.** The effects of LNA substitutions on strand displacement kinetics. (A) Leakage rate constants for multiple Substrates that were exposed to DNA invader i1 with a zero nt toehold. (B) Leakage kinetics for select Substrates (20 nM) that were exposed to DNA invader i1 (2  $\mu$ M). The black line is the original Substrate (sb0) and the red line is the background reaction when the Reporter complex (40 nM) was mixed with DNA invader i1 (2  $\mu$ M) without the Substrate. (C) Invasion rate constants for multiple Substrates that were exposed to a DNA invader i4 with a 6 nt toehold. (D) The performance enhancement factor of each Substrate was calculated by taking the ratio of the rate constants in (C) and (A). Error bars represent the standard deviation from three reactions with different invader concentrations.

Leakage suppression was maximized for Substrates sb1, sb2, sb3, sb4, sb5, sb6, sb7, and sb8 with LNA nucleotides on both their signal and backbone strands. When compared to the the original Substrate (sb0), the leakage reduction for sb8 was  $\sim$ 50-fold. This significant performance increase is because LNA-LNA base pairs are more thermomechanically stable than LNA-DNA or DNA-DNA base pairs.<sup>19</sup> In addition, the energy penalty between LNA-LNA and DNA-LNA base pairs during branch migration

likely contributes to the slower leakage rates. For example, there is a stronger bias to form LNA-LNA base pairs between the signal and the backbone than DNA-LNA base pairs between the invader and the backbone. Surprisingly, Substrate sb2 with two LNA-LNA base pairs near the duplex center has similar leakage performance as Substrate sb4 with two LNA-LNA base pairs near the terminal ends of the duplex. While LNAs near the terminal ends of the Substrate may reduce the fraying frequency and hence lower the probability of invader nucleation, they may not adequately transform the Substrate from the B to the more stable A-form conformation. On the contrary, LNAs in the central region of the Substrate not only change the structural conformation from B-form to the more stable A-form but they also impose a higher energy barrier for branch migration to proceed.<sup>25, 45</sup> This might explain why Substrate sb2, with only two LNA-LNA base pairs in the center of its duplex, reduced the leakage rate from  $1.48 \text{ M}^{-1}\text{s}^{-1}$  to  $0.065 \text{ M}^{-1}\text{s}^{-1}$ .

Overall, Substrates with LNA nucleotides show significant leakage suppression. The leakage rate constant was reduced from  $1.48 \text{ M}^{-1}\text{s}^{-1}$  to  $0.03 \text{ M}^{-1}\text{s}^{-1}$ . LNA-DNA and LNA-LNA base pairs make Substrates less vulnerable to react with the zero toehold invader probably because of their increased thermomechanical stability. LNA nucleotides near the terminal end of the Substrates are more likely to reduce fraying and LNAs in the central region impose a higher energy barrier during branch migration.

### 3.2.2 Invasion reactions between a DNA invader and hybrid DNA/LNA Substrates

The invading rate constants between the 6 nt toehold invader **i4** and the Substrates (Figure 3.2B) are shown in Figure 3.3C. The invading rate constants for Substrates b1-b6 are equivalent to the original Substrate sb0 because: (1) the 6 nt toehold is well established before the branch migration process proceeds, and (2) the enthalpy change

during branch migration is net neutral for Substrates with and without LNA nucleotides. More specifically, stable toeholds provide forward bias for the invader strand to displace the signal strand even though higher energy barriers are confronted during LNA-DNA versus DNA-DNA base pairing. In addition, LNA substitutions at the terminal ends of the duplex for Substrates b1, s1 and sb1 do not affect the invading rates because terminal base pairs spontaneously dissociate during branch migration.<sup>46</sup>

With the exception of s1, Substrates s2, s3, and s4 – with LNA substitutions on the signal strand – reduce the invasion rates because the LNA-DNA base pairs are replaced by DNA-DNA during branch migration. This energy penalty minimizes the forward bias of the random walk process. Forward bias decreases as the number of LNA nucleotides increases on the signal strand. With the exception of sb1, it also decreases for Substrates with LNAs on both the signal and backbone. As a consequence, Substrates sb2-sb8 exhibit substantially lower invasion rates as the number of LNA-LNA base pairs increases. For example, the invading rate constant for Substrate sb8 is reduced by ~3-orders of magnitude when compared to sb0. To further understand the kinetics of strand invasion, the invasion rate constants were measured as a function of toehold length in Section 3.8.6.

During strand invasion, LNA-LNA base pairs in the Substrate behave as barriers to strand invasion. The probability to overcome the barrier is dependent on both the position and number of LNA-LNA base pairs in the Substrate. For example, as equivalent LNA-LNA base pairs move closer to the toehold, the i4 invasion rate generally shrunk – as reflected by sb1 to sb3, and sb5 to sb7. In addition, the invasion rate constants for sb1 and sb5 remained elevated – regardless of the invader used – because the last few bases

at the terminus of branch migration spontaneously melt off.<sup>46</sup> Substrate sb3 with 2 LNA-LNA bases close to the right terminal end had a faster invasion rate than sb2 with 2 LNA-LNA bases in the central region. This may be because LNA substitutions that are close to terminal ends of the duplex are not sufficient to induce structural changes to the duplex.

Overall, LNA substitutions on the Substrate affect the invasion kinetics in the following ways. LNAs at the beginning of branch migration of the Substrate impose a higher penalty for initiating branch migration.<sup>44</sup> LNAs in the center of branch migration of the Substrate slow down the rate of branch migration through a relatively large sawtooth amplitude associated with each step of branch migration.<sup>44</sup> LNAs at the end of branch migration of the Substrate do not affect the strand displacement kinetics.

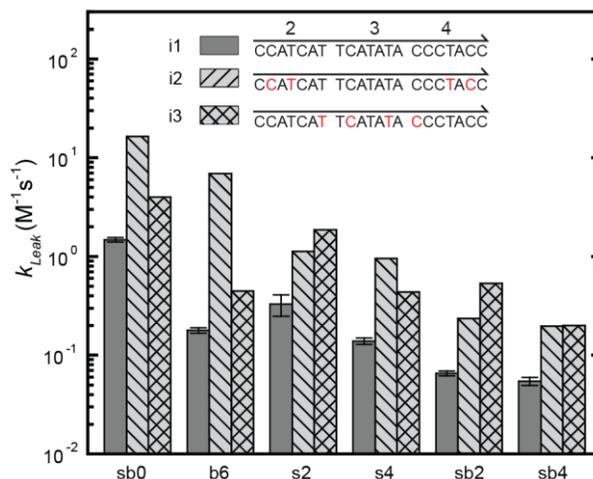
To guide strand invasion design, the performance enhancement ratio is defined as the ratio of the invading rate constant for the 6 nt toehold invader and the leakage rate constant for the 0 nt toehold invader (*performance enhancement ratio* =  $k_{invading}/k_{leak}$ ). The performance enhancement ratio of hybrid DNA/LNA Substrates are shown in Figure 3.3D. The performance enhancement ratio of Substrate sb0 is normalized to one unit and the higher ratios reflect better performance. For example, the performance enhancement ratio of Substrate sb5 is 18 times better than the original sb0 Substrate.

### 3.2.3 Leakage reactions between hybrid DNA/LNA invaders and hybrid DNA/LNA Substrates

The leakage kinetics between zero nt toehold invaders – with LNA substitutions – and select Substrates (Figure 3.2B) are summarized in Figure 3.4. Invaders with LNA substitutions have faster leakage kinetics than the DNA invader for three potential

reasons. First, invader **i2** with LNA substitutions could increase the probability to form stable nuclei with Substrate backbones. Second, the high affinity of LNA-DNA and LNA-LNA base pairs may increase the thermodynamic driving force for strand invasion to proceed. Third, invaders with LNA substitutions likely lower the kinetic barrier and promote forward bias during strand invasion.

In addition to the presence of LNAs in invaders **i2** and **i3**, Substrate **sb0** leakage is highly dependent on LNA location. For example, invader **i2** increased the leakage rate by a factor of 10, while invader **i3** only increased it by a factor of 2.7. In short, LNAs near the terminus of invader **i2** stabilize the nuclei between the invader and the backbone strands – which most likely increases the probability for branch migration to proceed to completion. In contrast, the central location of the LNAs in invader **i3** is expected to support similar nucleation behavior as invader **i1** without any LNA. However, invader **i3** has a higher probability of completing branch migration than **i1** because its LNA-DNA base pairs encourage forward bias to displace the signal strand. Therefore, even though invaders **i2** and **i3** have the same number and identity of LNA substitutions, **i2** exhibits much faster reaction rates than **i3** when invading the original Substrate (**sb0**).



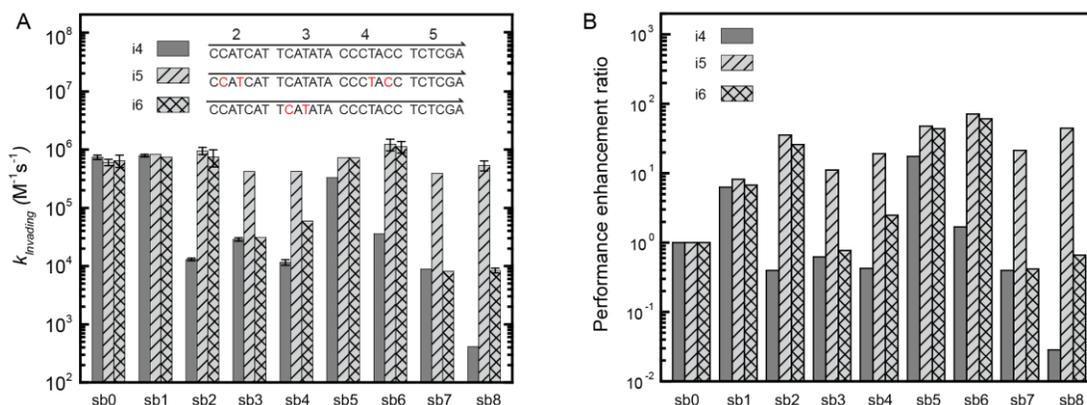
**Figure 3.4. Leakage rate constants for select Substrates and zero toehold invaders with and without LNA substitutions. Black and red letters respectively denote DNA and LNA nucleotides. Error bars are the standard deviation from three reactions with different invader concentrations.**

Invaders with LNA substitutions have faster leakage rates than the all-DNA invader, independent of the Substrates tested. When compared to sb0, b6 includes 15 LNAs on its backbone, most of which are at the terminal ends of domains 2\* and 4\*. When exposed to invaders with LNA substitutions, the leakage rate was ~10 times faster for i2 than i3. Briefly, invader i2 has a faster leakage rate because it stabilizes the nucleation event by replacing 2 DNA-LNA base pairs in the Substrate with 2 LNA-LNA base pairs in the Waste complex. In contrast, Substrates s2 and s4, with LNA substitutions on the signal strand, consistently have greater leakage rates than Substrates sb2 and sb4, which have LNAs on their signal and backbone strands. Regardless of the invaders chosen, Substrates sb2 and sb4 exhibit greater leakage suppression than Substrates s2 and s4, with LNA substitutions only on their signal strand.

### 3.2.4 Invasion reactions between hybrid DNA/LNA invaders and hybrid DNA/LNA

#### Substrates

When using the 0 nt toehold invader **i1**, sb Substrates with LNA on both the signal and backbone strand yielded the greatest leakage suppression (Figure 3.3A). In comparison, the invasion rates decreased for all sb Substrates, excluding sb1, that were invaded by **i4** – a DNA invader with a 6 nt toehold (Figure 3.5A). To improve the invasion performance of the sb Substrates, LNA substitutions were site-specifically included into invaders **i5** and **i6** – both of which had 6 nt toeholds. Regardless of the strand displacement systems in Figure 3.5, toehold hybridization was identical between all Substrates and all invaders. As a consequence, nucleation between the invader strands and the Substrates was assumed to be identical. In many cases, the invasion rates for **i5** and **i6** had comparable invasion performance to the DNA invader **i4** on the original Substrate (sb0). Primary examples include the invasion of Substrates sb1, sb2, sb5, and sb6 with either invader **i5** or **i6**.



**Figure 3.5.** (A) Invasion rate constants for Substrates with LNA nucleotides on both signal and backbone strands that were exposed to pure DNA and hybrid DNA/LNA invaders with a 6 nt toeholds. (B) The performance enhancement ratio of each Substrate was calculated by taking the ratio of the invasion and leak rate constants. Error bars represent the standard deviation from three reactions with different invader concentrations.

Between Substrates sb1 to sb8, the invasion rates were elevated for **i5**. During the invasion reaction, but after the invader toehold binded to the Substrate, 2 LNA substitutions on domain 4 of invader **i5** formed stable LNA-LNA or LNA-DNA base pair with the backbone. The increased stability of these base-pairs renders a forward bias during branch migration, which caused invader **i5** to have an increased probability to displace the signal strand instead of having the branch point return to the toehold binding domain. In comparison, invader **i6**, with 2 LNA substitutions on domain 3, shows far slower invasion rates for Substrates sb3, sb4, sb7, and sb8. LNA-LNA base pairs that are on the Substrate near the onset of branch migration is expected to render a higher energy barrier after **i6** binds to the toehold. The signal strand thus has a stronger bias to hybridize to the backbone strand, which encourages the invader to stay on the toehold domain. Thus, LNAs substitutions on the invader, that are further away from the toehold domain, do not improve the invading rate constants.

Overall, LNA substitutions on the invader affect the invasion kinetics in the following ways. LNAs on the invader bias the random walk forward at the onset of branch migration and hence increase the invasion rate constant. As the LNA substitutions move away from the toehold domain of the invader, the invasion performance gradually decreases.

The performance enhancement ratio of strand displacement systems between Substrates and invaders, with and without LNA substitutions, are shown in Figure 3.5B. The performance enhancement ratio of Substrate sb0 was normalized to 1 unit and higher ratios reflect increased performance. The performance enhancement ratio of all of the LNA substituted Substrates were improved. The performance enhancement of Substrate

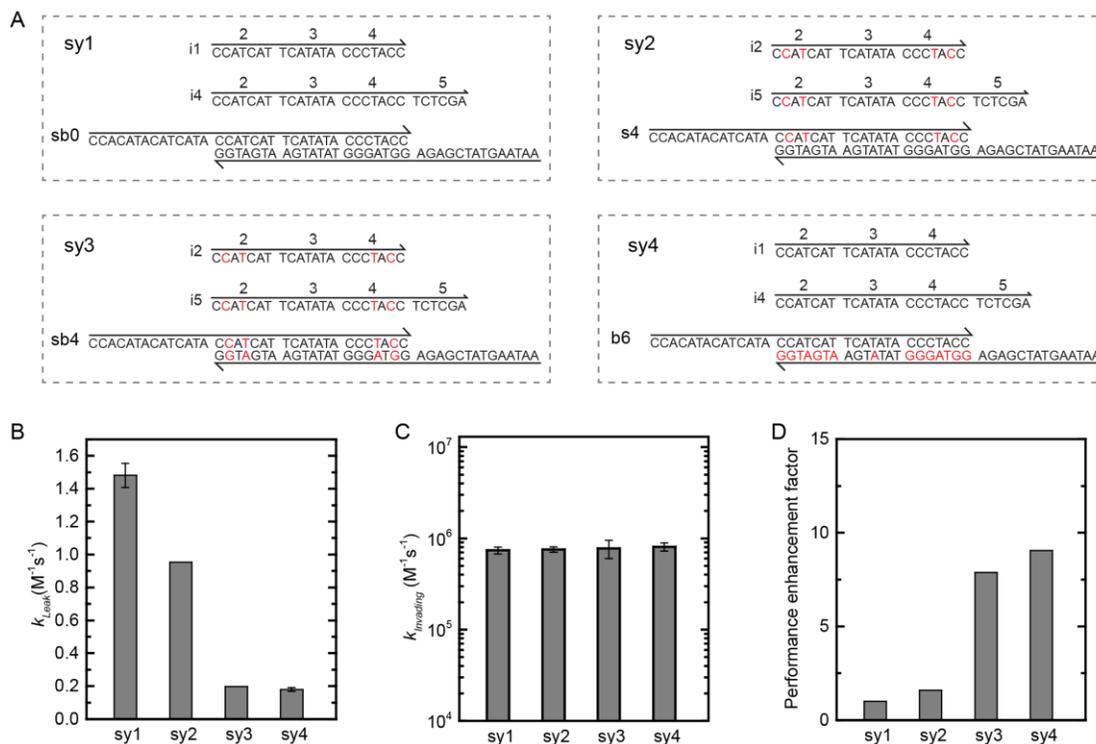
sb6 and invader i5 showed a 70-fold improvement compared with the performance of the original Substrate sb0 and invader i4.

### 3.2.5 Optimize the performance enhancement ratio of hybrid DNA/LNA systems

To explore effective practices for incorporating LNA substitutions into DNA strand displacement systems, four original systems were investigated in Figure 3.6. Identical to Substrate sb0 in Figure 3.3, the control system (sy1) was entirely made from DNA oligonucleotides. For each system, the leakage rate and invasion rate constants were measured and then compared to the DNA control. For all four systems, the performance enhancement ratio was calculated from the ratio of the invasion rate to the leakage rate constants. The performance enhancement ratio of sy1 was normalized to one unit and showed the largest leakage rate (Figure 3.6B). In comparison, the invasion rate constants were equivalent for all four systems (Figure 3.6C). The total system performance enhancement ratio improved from ~2-fold to 9-fold (Figure 3.6D).

In sy1, sy2, and sy3, domains 2 through 4 have identical nucleic acid composition. In sy2 and sy3, LNA substitutions were introduced into the reactants: 0 nt toehold invaders, 6 nt toehold invaders, and the Substrates. Aside from the 6 nt toehold region (domain 5), there is not an enthalpic change in the leakage and invasion reactions. While the thermodynamic driving force was equal for sy1, sy2, and sy3, the performance enhancement ratio of sy2 and sy3 was greater than sy1 by a factor of 1.6 and 8, respectively. The increased performance is attributed to an increase in the number of LNA nucleotides in the backbone of sy3. If the thermodynamic driving force is constant, the more LNA nucleotides that are introduced into the backbone of the Substrate, the

better the performance. As a consequence, it is projected that an all LNA system would outperform an all DNA system.



**Figure 3.6. Optimized hybrid DNA/LNA systems. (A) Strands and sequences for four unique systems. Black and red letters respectively denote DNA and LNA. Each system consists of a zero toehold invader, a 6 nt toehold invader, and a Substrate. For example, “sy1” denotes system 1, in which i1 is the zero toehold invader, i6 is the 6 nt toehold invader and sb0 is the original Substrate. (B) Leakage rate constants of each system. (C) Invasion rate constants of each system. (D) The performance enhancement ratio of each system was calculated by taking the ratio of the rate constants in (C) and (B). Error bars represent the standard deviation from three reactions with different invader concentrations.**

That being said, caution is required when incorporating LNA nucleotides into DNA strand displacement systems because of their strong binding affinity to complementary LNA or DNA nucleotides – especially for single strands.<sup>19</sup> To minimize secondary structure in single stranded reactants, LNA nucleotides can be limited to the duplex backbone. In sy4, the single stranded reactants are DNA oligonucleotides and the Substrate complex has LNA nucleotides in its backbone. The performance enhancement

ratio of sy4 is improved by a factor of 9 when compared to the all-DNA sy1. Unique hybrid DNA/LNA systems are demonstrated in Section 3.8.7 with performance improvement as high as 50 fold.

### 3.3 Conclusions

The kinetics of incorporating LNA nucleotides into a DNA strand displacement system has been studied. LNA substitutions affect the kinetics of strand displacement in three ways. First, LNAs in the Substrates stabilize the duplex probably by reducing the fraying frequency at the terminus of the duplex regions, which lowers the probability of successful nucleation between the Substrates and zero toehold invaders. Second, LNAs in the Substrates induce B-form to A-form structural changes, which may hinder the branch migration process. Third, LNAs in the Substrate or the invaders bias random walks during branch migration – which alters the probability of strand displacement to proceed. When incorporating LNA substitutions into a DNA strand displacement system, the leakage rate was reduced up to 50 fold and the invasion rate was maintained elevated. In comparison, kinetics for hybrid DNA/LNA systems, the performance enhancement ratio can be improved by a factor of 70 – providing insights for how to design future high performance chemical reaction networks made from DNA and LNA. By site-specifically substituting LNA nucleotides for DNA nucleotides, while maintaining the original sequence design, the performance of chemical reaction networks made from nucleic acids can be optimized. For example, LNAs can be strategically incorporated into different systems such as catalyzed hairpin assembly,<sup>36, 47</sup> hairpin chain reaction,<sup>48, 49</sup> DNA walker<sup>50</sup> and seesaw gate systems<sup>51, 52</sup> to minimize unwanted reactions and increase the rate of the desired reactions.

### 3.4 Methods

#### 3.4.1 Materials

DNA and hybrid DNA/LNA oligonucleotides were synthesized with HPLC purification by Integrated DNA Technologies (IDT) and Exqion, respectively. Reporter complexes were also labeled by IDT with 5' TET fluorophores and 3' Iowa Black FQ quenchers. Once received, the oligonucleotides were suspended in a 1x TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The stock concentrations were measured from their 260 nm absorbance using the extinction coefficients provided by IDT and Exqion. To minimize loss from non-specific binding, poly T oligonucleotides were added to the dilute stock solutions (less than 1  $\mu$ M) to reach a final poly T concentration of 1  $\mu$ M. Unless stated otherwise, chemicals and solvents were purchased from Sigma-Aldrich. The 10x TAE buffer (40 mM Tris, 40 mM Acetate, 1 mM EDTA, pH 8.3 ~ 8.5) was purchased from Hoefer or Fisher Scientific. To reach a final concentration of 125 mM  $Mg^{2+}$ ,  $Mg(C_2H_3O_2)_2 \cdot 4H_2O$  was added to the 10x TAE buffer. The oligonucleotide components were diluted to 30  $\mu$ M in a 1x TAE buffer with 12.5 mM  $Mg^{2+}$  and then annealed at 95 °C for 5 minutes using a Eppendorf Mastercycler Nexus Gradient Thermocycler. Once annealed, samples were cooled from 95 °C to room temperature over ~90 minutes to form the Substrates and Reporters used in the below listed experiments.

### 3.4.2 Purification

Substrate and Reporter complexes were purified by gel electrophoresis using a 10% polyacrylamide gel that was made from a 30% acrylamide bis solution in a 29:1 ratio. The native gels were run for two hours at 16 °C using a 150 V bias and a VWR International chiller. To eliminate malformed complexes at room temperature, the Substrates were stoichiometrically incubated with zero toehold DNA invaders at 15  $\mu\text{M}$  for 5 hours before loading the gel. The loading buffer was made from a 1:1 ratio of bromophenol blue dye and ficoll solution (Type 400, 20% water). The desired bands were cut from the gels and then eluted in a 1x TE/Mg<sup>2+</sup> buffer for 2 days at 4 °C. The buffer included 1x TE with 12.5 mM MgCl<sub>2</sub>·6H<sub>2</sub>O (Acros Organics). Once purified, the Substrate and Reporter concentrations were quantified from their 260 nm absorbance.

### 3.4.3 Spectrofluorimetry

Fluorescence spectrophotometers from Agilent and Varian Technologies (Cary Eclipses) were used to measure the reaction kinetics. The slit sizes were set to 2.5 nm for the excitation (510 nm) and 10 nm for the emission (538 nm) wavelengths. All experiments were performed at 25 °C, in 0.4 mL glass cuvettes made from Starna Cells, containing a 1x TE/Mg<sup>2+</sup> buffer, with a total volume of 0.2 mL. The final fluorescence was normalized to 1 arbitrary unit (a.u.) and corresponded to the lower concentration of the Substrate or Invader used in the experiments.

### **3.5 Author Information**

#### 3.5.1 Corresponding Authors

WLH [willhughes@boisestate.edu]

#### 3.5.2 Author Contributions

The primary author performed the research and the primary author wrote the manuscript with support from the corresponding authors.

### **3.6 Acknowledgements**

This project was supported in part by: (1) The W.M. Keck Foundation, (2) NIH Grant No. K25GM093233 from the National Institute of General Medical Sciences, (3) NIH Grant No. P20 RR016454 from the INBRE Program of the National Center for Research Resources, and (4) The Micron Foundation. Special thanks are extended to Dr. Peter Allen from the University of Idaho for his thoughtful review of this work.

## REFERENCES

1. Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature*, 171, 737-738.
2. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31, 3406-3415.
3. Dirks, R.M., Bois, J.S., Schaeffer, J.M., Winfree, E. and Pierce, N.A. (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49, 65-88.
4. SantaLucia, J., Allawi, H.T. and Seneviratne, A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35, 3555-3562.
5. SantaLucia, J., Jr. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annual Review Biophysics Biomolecular Structure*, 33, 415-440.
6. Cooper, A., Johnson, C.M., Lakey, J.H. and Nöllmann, M. (2001) Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophysical Chemistry*, 93, 215-230.
7. Petruska, J. and Goodman, M.F. (1995) Enthalpy-entropy compensation in DNA melting thermodynamics. *Journal of Biological Chemistry*, 270, 746-750.
8. Zhang, D.Y. and Winfree, E. (2009) Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131, 17303-17314.
9. Panyutin, I.G. and Hsieh, P. (1994) The kinetics of spontaneous DNA branch migration. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 2021-2025.

10. Green, C. and Tibbetts, C. (1981) Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Research*, 9, 1905-1918.
11. Morrison, L.E. and Stols, L.M. (1993) Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution. *Biochemistry*, 32, 3095-3104.
12. Olson, X., Kotani, S., Padilla, J.E., Hallstrom, N., Goltry, S., Lee, J., Yurke, B., Hughes, W.L. and Graugnard, E. (2016) Availability: A Metric for nucleic acid strand displacement systems. *ACS Synthetic Biology*.
13. Koshkin, A.A., Singh, S.K., Nielsen, P., Rajwanshi, V.K., Kumar, R., Meldgaard, M., Olsen, C.E. and Wengel, J. (1998) LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron*, 54, 3607-3630.
14. Koshkin, A.A., Rajwanshi, V.K. and Wengel, J. (1998) Novel convenient syntheses of LNA [2.2. 1] bicyclo nucleosides. *Tetrahedron Letters*, 39, 4381-4384.
15. Obika, S., Nanbu, D., Hari, Y., Morio, K.-i., In, Y., Ishida, T. and Imanishi, T. (1997) Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3, -endo sugar pucker. *Tetrahedron Letters*, 38, 8735-8738.
16. Kværnø, L. and Wengel, J. (1999) Investigation of restricted backbone conformations as an explanation for the exceptional thermal stabilities of duplexes involving LNA (Locked Nucleic Acid):† synthesis and evaluation of abasic LNA. *Chemical Communications*, 657-658.
17. K. Singh, S., A. Koshkin, A., Wengel, J. and Nielsen, P. (1998) LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chemical Communications*, 455-456.
18. Koshkin, A.A., Nielsen, P., Meldgaard, M., Rajwanshi, V.K., Singh, S.K. and Wengel, J. (1998) LNA (locked nucleic acid): an RNA mimic forming

- exceedingly stable LNA: LNA duplexes. *Journal of the American Chemical Society*, 120, 13252-13253.
19. Singh, S. (1998) Universality of LNA-mediated high-affinity nucleic acid recognition. *Chemical Communications*, 1247-1248.
  20. Braasch, D.A. and Corey, D.R. (2001) Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chemistry & Biology*, 8, 1-7.
  21. Wengel, J., Petersen, M., Nielsen, K.E., Jensen, G.A., Håkansson, A.E., Kumar, R., Sørensen, M.D., Rajwanshi, V.K., Bryld, T. and Jacobsen, J.P. (2001) LNA (locked nucleic acid) and the diastereoisomeric  $\alpha$ -L-LNA: conformational tuning and high-affinity recognition of DNA/RNA targets. *Nucleosides, Nucleotides and Nucleic Acids*, 20, 389-396.
  22. Petersen, M., Bondensgaard, K., Wengel, J. and Jacobsen, J.P. (2002) Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids. *Journal of the American Chemical Society*, 124, 5974-5982.
  23. Nielsen, K.E., Singh, S.K., Wengel, J. and Jacobsen, J.P. (2000) Solution structure of an LNA hybridized to DNA: NMR study of the d(CTLGCTLTLCTLGC):d(GCAGAAGCAG) duplex containing four locked nucleotides. *Bioconjugate Chemistry*, 11, 228-238.
  24. Nielsen, K.E., Rasmussen, J., Kumar, R., Wengel, J., Jacobsen, J.P. and Petersen, M. (2004) NMR Studies of fully modified locked nucleic Acid (LNA) hybrids: solution structure of an LNA:RNA hybrid and characterization of an LNA:DNA hybrid. *Bioconjugate Chemistry*, 15, 449-457.
  25. Petersen, M., Nielsen, C.B., Nielsen, K.E., Jensen, G.A., Bondensgaard, K., Singh, S.K., Rajwanshi, V.K., Koshkin, A.A., Dahl, B.M., Wengel, J. *et al.* (2000) The conformations of locked nucleic acids (LNA). *Journal of Molecular Recognition*, 13, 44-53.
  26. Searle, M.S. and Williams, D.H. (1993) On the stability of nucleic acid structures in solution: enthalpy - entropy compensations, internal rotations and reversibility. *Nucleic Acids Research*, 21, 2051-2056.

27. McTigue, P.M., Peterson, R.J. and Kahn, J.D. (2004) Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)–DNA Duplex formation. *Biochemistry*, 43, 5388-5405.
28. Braasch, D.A., Liu, Y. and Corey, D.R. (2002) Antisense inhibition of gene expression in cells by oligonucleotides incorporating locked nucleic acids: effect of mRNA target sequence and chimera design. *Nucleic acids research*, 30, 5160-5167.
29. Kurreck, J., Wyszko, E., Gillen, C. and Erdmann, V.A. (2002) Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic acids research*, 30, 1911-1918.
30. Kværnø, L., Kumar, R., Dahl, B.M., Olsen, C.E. and Wengel, J. (2000) Synthesis of abasic locked nucleic acid and two seco-LNA derivatives and evaluation of their hybridization properties compared with their more flexible DNA counterparts. *The Journal of Organic Chemistry*, 65, 5167-5176.
31. Christensen, U., Jacobsen, N., Rajwanshi, V.K., Wengel, J. and Koch, T. (2001) Stopped-flow kinetics of locked nucleic acid (LNA)-oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions. *Biochemical Journal*, 354, 481-484.
32. Arora, A., Kaur, H., Wengel, J. and Maiti, S. (2008) Effect of locked nucleic acid (LNA) modification on hybridization kinetics of DNA duplex. *Nucleic Acids Symposium Series*, 52, 417-418.
33. Zhang, D.Y., Turberfield, A.J., Yurke, B. and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318, 1121-1125.
34. Zhang, D.Y. and Winfree, E. (2010) Robustness and modularity properties of a non-covalent DNA catalytic reaction. *Nucleic Acids Research*, 38, 4182-4197.
35. Yin, P., Yan, H., Daniell, X.G., Turberfield, A.J. and Reif, J.H. (2004) A unidirectional DNA walker that moves autonomously along a track. *Angewandte Chemie International Edition*, 43, 4906-4911.

36. Yin, P., Choi, H.M., Calvert, C.R. and Pierce, N.A. (2008) Programming biomolecular self-assembly pathways. *Nature*, 451, 318-322.
37. Yurke, B., Turberfield, A.J., Mills, A.P., Simmel, F.C. and Neumann, J.L. (2000) A DNA-fuelled molecular machine made of DNA. *Nature*, 406, 605-608.
38. Yurke, B. and Mills, A. (2003) Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines*, 4, 111-122.
39. Wittwer, C.T., Herrmann, M.G., Moss, A.A. and Rasmussen, R.P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, 22, 130-131, 134-138.
40. Thachuk, C., Winfree, E. and Soloveichik, D. (2015) In Phillips, A. and Yin, P. (eds.), *DNA computing and molecular programming*. Springer International Publishing, Vol. 9211, pp. 133-153.
41. Jiang, Y.S., Bhadra, S., Li, B. and Ellington, A.D. (2014) Mismatches improve the performance of strand-displacement nucleic acid circuits. *Angewandte Chemie*, 126, 1876-1879.
42. Chen, X., Briggs, N., McLain, J.R. and Ellington, A.D. (2013) Stacking nonenzymatic circuits for high signal gain. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 5386-5391.
43. Tomov, T.E., Tsukanov, R., Liber, M., Masoud, R., Plavner, N. and Nir, E. (2013) Rational design of DNA motors: fuel optimization through single-molecule fluorescence. *Journal of the American Chemical Society*, 135, 11935-11941.
44. Srinivas, N., Ouldridge, T.E., Sulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research*, 41, 10641-10658.
45. Srinivas, N., Ouldridge, T.E., Šulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research*, 41, 10641-10658.

46. Machinek, R.R., Ouldridge, T.E., Haley, N.E., Bath, J. and Turberfield, A.J. (2014) Programmable energy landscapes for kinetic control of DNA strand displacement. *Nature Communication*, 5.
47. Li, B., Jiang, Y., Chen, X. and Ellington, A.D. (2012) Probing spatial organization of DNA strands using enzyme-free hairpin assembly circuits. *Journal of the American Chemical Society*, 134, 13918-13921.
48. Dirks, R.M. and Pierce, N.A. (2004) Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 15275-15278.
49. Huss, D., Choi, H.M.T., Readhead, C., Fraser, S.E., Pierce, N.A. and Lansford, R. (2015) Combinatorial analysis of mRNA expression patterns in mouse embryos using hybridization chain reaction. *Cold Spring Harbor Protocols*, 2015, pdb.prot083832.
50. Shin, J.-S. and Pierce, N.A. (2004) A synthetic DNA walker for molecular transport. *Journal of the American Chemical Society*, 126, 10834-10835.
51. Qian, L. and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332, 1196-1201.
52. Qian, L., Winfree, E. and Bruck, J. (2011) Neural network computation with DNA strand displacement cascades. *Nature*, 475, 368-372.

### 3.7 Supplementary Information

#### 3.7.1 Strand sequences

**Table 3.1 Substrate sequences (Black and red letters represent DNA and LNA nucleotides, respectively)**

Substrate	Signal (s) name and sequence (5'-3')		Backbone (b) name and sequence (5'-3')	
	sb0	TB	CCACATACATCATACCATCAT TCATATA CCCTACC	LB
b1	TB	CCACATACATCATA CCATCAT TCATATA CCCTACC	LB	AATAAGTATCGAGA <b>GGTAGGG</b> TATATGA ATGATGG
b2	TB	CCACATACATCATA CCATCAT TCATATA CCCTACC	LBm	AATAAGTATCGAGA GGTAGGG TAT <b>ATGA</b> ATGATGG
b3	TB	CCACATACATCATA CCATCAT TCATATA CCCTACC	LB3	AATAAGTATCGAGA GGTAGGG TATATGA ATG <b>ATGG</b>
b4	TB	CCACATACATCATA CCATCAT TCATATA CCCTACC	LBb	AATAAGTATCGAGA <b>GGTAGGG</b> TATATGA ATG <b>ATGG</b>
b5	TB	CCACATACATCATA CCATCAT TCATATA CCCTACC	LBa	AATAAGTATCGAGA <b>GGTAGGG</b> TAT <b>ATGA</b> ATG <b>ATGG</b>
b6	TB	CCACATACATCATA CCATCAT TCATATA CCCTACC	LBI	AATAAGTATCGAGA <b>GGTAGGG</b> TAT <b>ATGA ATGATGG</b>
s1	TB5	CCACATACATCATA <b>CCATCAT</b> TCATATA CCCTACC	LB	AATAAGTATCGAGA GGTAGGG TATATGA ATGATGG
s2	TB m	CCACATACATCATA CCATCAT <b>TCATATA</b> CCCTACC	LB	AATAAGTATCGAGA GGTAGGG TATATGA ATGATGG
s3	TB3	CCACATACATCATA CCATCAT TCATATA CC <b>TACC</b>	LB	AATAAGTATCGAGA GGTAGGG TATATGA ATGATGG
s4	TBb	CCACATACATCATA <b>CCATCAT</b> TCATATA CC <b>TACC</b>	LB	AATAAGTATCGAGA GGTAGGG TATATGA ATGATGG
sb1	TB5	CCACATACATCATA <b>CCATCAT</b> TCATATA CCCTACC	LB3	AATAAGTATCGAGA GGTAGGG TATATGA ATG <b>ATGG</b>
sb2	TB m	CCACATACATCATA CCATCAT <b>TCATATA</b> CCCTACC	LBm	AATAAGTATCGAGA GGTAGGG TAT <b>ATGA</b> ATGATGG
sb3	TB3	CCACATACATCATA CCATCAT TCATATA CC <b>TACC</b>	LB5	AATAAGTATCGAGA <b>GGTAGGG</b> TATATGA ATGATGG
sb4	TBb	CCACATACATCATA <b>CCATCAT</b> TCATATA CC <b>TACC</b>	LBb	AATAAGTATCGAGA <b>GGTAGGG</b> TATATGA ATG <b>ATGG</b>
sb5	TB5	CCACATACATCATA <b>CCATCAT</b> TCATATA CCCTACC	LBI	AATAAGTATCGAGA <b>GGTAGGG</b> TAT <b>ATGA ATGATGG</b>
sb6	TB m	CCACATACATCATA CCATCAT <b>TCATATA</b> CCCTACC	LBI	AATAAGTATCGAGA <b>GGTAGGG</b> TAT <b>ATGA ATGATGG</b>
sb7	TB3	CCACATACATCATA CCATCAT TCATATA CC <b>TACC</b>	LBI	AATAAGTATCGAGA <b>GGTAGGG</b> TAT <b>ATGA ATGATGG</b>
sb8	TBb	CCACATACATCATA <b>CCATCAT</b> TCATATA CC <b>TACC</b>	LBI	AATAAGTATCGAGA <b>GGTAGGG</b> TAT <b>ATGA ATGATGG</b>

**Table 3.2 Invader sequences (Black and red letters represent DNA and LNA nucleotides, respectively)**

Invader	Sequences(5'-3')
<b>i1</b>	CCATCAT TCATATA CCCTACC
<b>i-1</b>	CCATCAT TCATATA CCCTACC T
<b>i-2</b>	CCATCAT TCATATA CCCTACC TC
<b>i-3</b>	CCATCAT TCATATA CCCTACC TCT
<b>i-4</b>	CCATCAT TCATATA CCCTACC TCTC
<b>i-5</b>	CCATCAT TCATATA CCCTACC TCTCG
<b>i4</b>	CCATCAT TCATATA CCCTACC TCTCGA
<b>i-7</b>	CCATCAT TCATATA CCCTACC TCTCGAT
<b>i-8</b>	CCATCAT TCATATA CCCTACC TCTCGATA
<b>i-9</b>	CCATCAT TCATATA CCCTACC TCTCGATAC
<b>i-10</b>	CCATCAT TCATATA CCCTACC TCTCGATACT

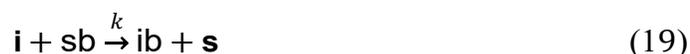
i-11	CCATCAT TCATATA CCCTACC TCTCGATACTT
i-14	CCATCAT TCATATA CCCTACC TCTCGATACTTATT
i5	CCATCAT TCATATA CCCTACC TCTCGA
i6	CCATCAT TCATATA CCCTACC TCTCGA
i2	CCATCAT TCATATA CCCTACC
i3	CCATCAT TCATATA CCCTACC

**Table 3.3 Reporter sequences**

Reporter complex	Dye strand (5'-3')	Quencher strand (5'-3')
dq1	/5TET/CCACATACATCATACCAT CAT	TATATGAATGATGGTATGATGTATGTGG/3IA BkFQ/
dq2	/5TET/ CCACATACATCATACCAT	TATATGAATGATGGTATGATGTATGTGG/3IA BkFQ/

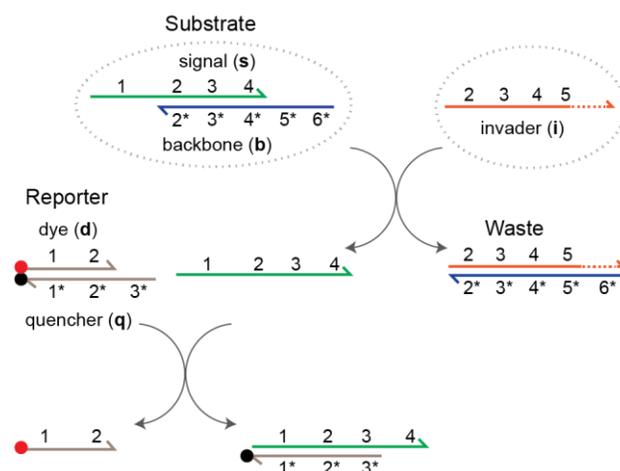
### 3.7.2 Reaction models and rate constants

The schematic model of the nucleic acid based, toehold mediated strand displacement system used in this study is shown in Figure 3.7. Unlike Figure 3.1, the toehold length of domain 5 on the invader (*i*) varies from 0 to 14 nucleotides (nt). Regardless, for both systems the reaction kinetics are assumed to be second order reactions between the invader (*i*) and the Substrate (*sb*), and the signal (*s*) and the Reporter (*dq*).



Equation (19) is the invasion reaction shown in Figures. 3.1 and 3.7. During the reaction, invader *i* and Substrate *sb* combine to form Waste complex *ib* and signal *s*. This reaction proceeds with the rate constant *k*. Equation (20) is the reporting reaction shown in Figures. 3.1 and 3.7. During this reaction, signal *s* and Reporter *dq* combine to release dye *d* and form complex *qs*. This reaction proceeds with the rate constant *k<sub>d</sub>*. The reverse reactions are not accounted in equations (19-

20) and are assumed to be negligible. The Reporter reaction in equation (20) is much faster than the invading reaction in equation (19). The reaction rates were measured in Section 3.8.2. The Reporter concentration was 2 times the Substrate concentration and hence the fluorescence is treated as a direct measurement of the dye (d) concentration.



**Figure 3.7. Schematic of a nucleic acid based, toehold mediated strand displacement system. Functional domains are represented by numbers and complementary domains are denoted by numbers with asterisks. The Substrate complex includes signal (s) and backbone (b) strands, while the Reporter complex includes dye (d) and quencher (q) strands. Various LNA substitutions have been made to the signal, backbone, and invader (i) sequences in this study. The toehold length of domain 5 on the invader (i) varies from zero to 14 nucleotides (nt). During invasion, the invader hybridizes with the backbone at domain 5\* and displaces the signal strand via three way branch migration. The invader and backbone form a Waste complex. The free signal strand then reacts with the Reporter complex and releases a dye strand, which is monitored by a fluorometer.**

The invading reaction is governed by the following rate equation:

$$\frac{d[\mathbf{s}]}{dt} = k[\mathbf{i}][\mathbf{sb}] \quad (21)$$

The two corresponding mass balance equations for the invading reaction are:

$$[\mathbf{i}] = [\mathbf{i}]_0 - [\mathbf{s}] \quad (22)$$

$$[\mathbf{sb}] = [\mathbf{sb}]_0 - [\mathbf{s}] \quad (23)$$

When combined, equations (21) through (23) yielded:

$$\ln \frac{([\mathbf{sb}]_0 - [\mathbf{s}])[\mathbf{i}]_0}{([\mathbf{i}]_0 - [\mathbf{s}])[\mathbf{sb}]_0} = ([\mathbf{sb}]_0 - [\mathbf{i}]_0)kt \quad (24)$$

For the zero-toehold leakage reaction and the short toehold (0-4 nt) invasion reactions, the kinetics are extremely slow. For example, it would take days to months for the reactions to reach completion. As a consequence, the left-hand side of Equation 24 was plotted versus  $t$  for the first 12 hours of each reaction. This approach provided a straight line with a slope of  $([\mathbf{sb}]_0 - [\mathbf{i}]_0)k$ . The leakage rate and short toehold rate constants were then extracted by a linear fit to the line.

From Equation (24), one then obtains:

$$[\mathbf{s}] = \frac{[\mathbf{i}]_0[\mathbf{sb}]_0(1 - \exp[t \cdot k \cdot ([\mathbf{sb}]_0 - [\mathbf{i}]_0)])}{[\mathbf{i}]_0 - [\mathbf{sb}]_0 \exp[t \cdot k \cdot ([\mathbf{sb}]_0 - [\mathbf{i}]_0)]} \quad (25)$$

For the invading reaction, when  $t \rightarrow \infty$  and  $[\mathbf{sb}]_0 < [\mathbf{i}]_0$

$$[\mathbf{s}]_\infty = [\mathbf{sb}]_0 \quad (26)$$

$$[\mathbf{sn}] = \frac{[\mathbf{s}]}{[\mathbf{s}]_\infty} = \frac{[\mathbf{i}]_0(1 - \exp[t \cdot k \cdot ([\mathbf{sb}]_0 - [\mathbf{i}]_0)])}{[\mathbf{i}]_0 - [\mathbf{sb}]_0 \exp[t \cdot k \cdot ([\mathbf{sb}]_0 - [\mathbf{i}]_0)]} \quad (27)$$

In comparison, the invading reactions reached completion for the invaders with toehold lengths greater than 4 nt. To extract the rate constants, Equation (27) was fit to the entire fluorescence curve.

Similar to the invading reaction, the Reporter reaction used the following rate equation:

$$\frac{d[\mathbf{d}]}{dt} = k[\mathbf{s}][\mathbf{dq}] \quad (28)$$

The two corresponding mass balance equations for the Reporter reaction are:

$$[\mathbf{s}] = [\mathbf{s}]_0 - [\mathbf{d}] \quad (29)$$

$$[\mathbf{dq}] = [\mathbf{dq}]_0 - [\mathbf{d}] \quad (30)$$

When combined, Equations (28) through (30) yielded:

$$\frac{d[\mathbf{d}]}{dt} = k_d([\mathbf{dq}]_0 - [\mathbf{d}])([\mathbf{s}]_0 - [\mathbf{d}]) \quad (31)$$

In analogy with Equation (26), one has:

$$[\mathbf{d}] = \frac{[\mathbf{dq}]_0[\mathbf{s}]_0(1 - \exp^{t \cdot k_d \cdot ([\mathbf{s}]_0 - [\mathbf{dq}]_0)})}{[\mathbf{dq}]_0 - [\mathbf{s}]_0 \exp^{t \cdot k_d \cdot ([\mathbf{s}]_0 - [\mathbf{dq}]_0)}} \quad (32)$$

For the Reporter reaction, when  $t \rightarrow \infty$  and  $[\mathbf{s}]_0 < [\mathbf{dq}]_0$

$$[\mathbf{d}]_\infty = [\mathbf{s}]_0 \quad (33)$$

$$\frac{[\mathbf{dn}]}{[\mathbf{d}]_\infty} = \frac{[\mathbf{d}]}{[\mathbf{d}]_\infty} = \frac{[\mathbf{dq}]_0(1 - \exp^{t \cdot k_d \cdot ([\mathbf{s}]_0 - [\mathbf{dq}]_0)})}{[\mathbf{dq}]_0 - [\mathbf{s}]_0 \exp^{t \cdot k_d \cdot ([\mathbf{s}]_0 - [\mathbf{dq}]_0)}} \quad (34)$$

The fluorescence data for the Reporter complex (dq) and the signal strand (s) were fit using Equation (34) to extract the rate constant  $k_d$  which was measured in Figure 3.8. A complete summary of the reaction concentrations and the rate constants of leakage and invasion reactions are provided below in Table 3.4 for various toehold lengths.

**Table 3.4 Reaction concentrations and rate constants**

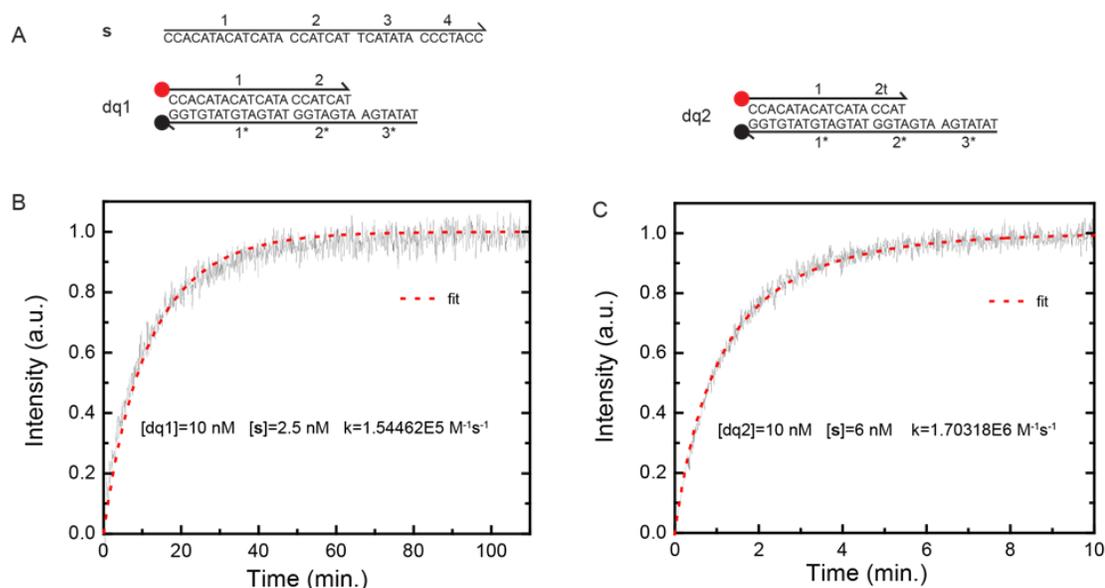
Toehold Length (nt)	Substrate	Invader	Reporter	Rate constant ( $M^{-1}s^{-1}$ )
0	20 nM [sb0]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	1.482
0	20 nM [b1]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.969
0	20 nM [b2]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	1.211
0	20 nM [b3]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	1.027
0	20 nM [b4]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.746

0	20 nM [b5]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.682
0	20 nM [b6]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.179
0	20 nM [s1]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	1.032
0	20 nM [s2]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.329
0	20 nM [s3]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.338
0	20 nM [s4]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.139
0	20 nM [sb1]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.251
0	20 nM [sb2]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.065
0	20 nM [sb3]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.093
0	20 nM [sb4]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.054
0	20 nM [sb5]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.038
0	20 nM [sb6]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.042
0	20 nM [sb7]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.045
0	20 nM [sb8]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1]	40 nM [dq1]	0.029
0	20 nM [sb0]	2 $\mu$ M [i2]	40 nM [dq1]	16.498
0	20 nM [sb0]	2 $\mu$ M [i3]	40 nM [dq1]	3.998
0	20 nM [s4]	2 $\mu$ M [i2]	40 nM [dq1]	0.954
0	20 nM [s4]	2 $\mu$ M [i3]	40 nM [dq1]	0.437
0	20 nM [s2]	2 $\mu$ M [i2]	40 nM [dq1]	1.123
0	20 nM [sb0]	2 $\mu$ M [i3]	40 nM [dq1]	1.866
0	20 nM [sb4]	2 $\mu$ M [i2]	40 nM [dq1]	0.198
0	20 nM [sb4]	2 $\mu$ M [i3]	40 nM [dq1]	0.200
0	20 nM [sb2]	2 $\mu$ M [i2]	40 nM [dq1]	0.236
0	20 nM [sb2]	2 $\mu$ M [i3]	40 nM [dq1]	0.533
0	20 nM [b6]	2 $\mu$ M [i2]	40 nM [dq1]	6.893
0	20 nM [b6]	2 $\mu$ M [i3]	40 nM [dq1]	0.446
0	20 nM [sb0]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1d5]	40 nM [dq1]	1.312
0	20 nM [sb0]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1d3]	40 nM [dq1]	0.395
0	20 nM [sb0]	2 $\mu$ M, 4 $\mu$ M, and 6 $\mu$ M [i1db]	40 nM [dq1]	0.043
1	20 nM [sb0]	2 $\mu$ M [i-1]	40 nM [dq1]	6.798
1	20 nM [b6]	2 $\mu$ M [i-1]	40 nM [dq1]	1.980
2	20 nM [sb0]	200 nM [i-2]	40 nM [dq1]	1.860E+02
2	20 nM [b6]	200 nM [i-2]	40 nM [dq1]	9.712E+01
3	20 nM [sb0]	40 nM [i-3]	40 nM [dq1]	1.425E+03
3	20 nM [b6]	40 nM [i-3]	40 nM [dq1]	8.183E+02
4	20 nM [sb0]	40 nM [i-4]	40 nM [dq1]	1.488E+04
4	20 nM [b6]	40 nM [i-4]	40 nM [dq1]	9.307E+03
5	10 nM [sb0]	40 nM [i-5]	20 nM [dq2]	2.844E+05
5	10 nM [b6]	40 nM [i-5]	20 nM [dq2]	1.610E+05
6	10 nM [sb0]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	7.389E+05
6	10 nM [b1]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	8.678E+05
6	10 nM [b2]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	9.884E+05
6	10 nM [b3]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	9.061E+05
6	10 nM [b4]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	8.859E+05
6	10 nM [b5]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	8.936E+05
6	10 nM [b6]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	8.084E+05
6	10 nM [s1]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	9.502E+05
6	10 nM [s2]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	3.018E+05
6	10 nM [s3]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	3.100E+05
6	10 nM [s4]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	3.305E+05
6	10 nM [sb1]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	7.888E+05
6	10 nM [sb2]	20 nM, 40 nM, and 60 nM [i4]	20 nM [dq2]	1.298E+04
6	10 nM [sb3]	20 nM, 40 nM, and 60 nM [i4]	20 nM [dq2]	2.867E+04
6	10 nM [sb4]	40 nM, 60 nM, and 80 nM [i4]	20 nM [dq2]	1.160E+04
6	10 nM [sb5]	2 nM, 4 nM, and 6 nM [i4]	20 nM [dq2]	1.652E+05
6	10 nM [sb6]	40 nM, 60 nM, and 80 nM [i4]	20 nM [dq2]	2.701E+04
6	10 nM [sb7]	40 nM, 60 nM, and 80 nM [i4]	20 nM [dq2]	8.681E+03
6	10 nM [sb8]	120 nM, 200 nM, and 300 nM [i4]	20 nM [dq2]	4.310E+02
6	10 nM [sb0]	2 nM, 4 nM, and 6 nM [i6]	20 nM [dq1]	6.471E+05

6	10 nM [sb0]	2 nM, 4 nM, and 6 nM [i5]	20 nM [dq1]	6.029E+05
6	10 nM [sb1]	4 nM [i6]	20 nM [dq1]	7.364E+05
6	10 nM [sb1]	4 nM [i5]	20 nM [dq1]	8.319E+05
6	10 nM [sb2]	2 nM, 4 nM, and 6 nM [i6]	20 nM [dq1]	7.452E+05
6	10 nM [sb2]	2 nM, 4 nM, and 6 nM [i5]	20 nM [dq1]	9.532E+05
6	10 nM [sb3]	4 nM [i6]	20 nM [dq1]	3.117E+04
6	10 nM [sb3]	4 nM [i5]	20 nM [dq1]	4.212E+05
6	10 nM [sb4]	40 nM [i6]	20 nM [dq1]	5.901E+04
6	10 nM [sb4]	40 nM [i5]	20 nM [dq1]	4.230E+05
6	10 nM [sb5]	4 nM [i6]	20 nM [dq1]	7.220E+05
6	10 nM [sb5]	4 nM [i5]	20 nM [dq1]	7.252E+05
6	10 nM [sb5]	4 nM [i5]	20 nM [dq1]	7.252E+05
6	10 nM [sb6]	2 nM, 4 nM, and 6 nM [i6]	20 nM [dq1]	1.120E+06
6	10 nM [sb6]	2 nM, 4 nM, and 6 nM [i5]	20 nM [dq1]	1.230E+06
6	10 nM [sb7]	4 nM [i6]	20 nM [dq1]	8.101E+03
6	10 nM [sb7]	4 nM [i5]	20 nM [dq1]	3.902E+05
6	10 nM [sb8]	40 nM, 60 nM, and 80 nM [i6]	20 nM [dq1]	8.409E+03
6	10 nM [sb8]	2 nM, 4 nM, and 6 nM [i5]	20 nM [dq1]	5.325E+05
7	10 nM [sb0]	4 nM [i-7]	20 nM [dq2]	1.113E+06
7	10 nM [b6]	4 nM [i-7]	20 nM [dq2]	1.091E+06
7	10 nM [sb2]	4 nM [i-7]	20 nM [dq2]	2.771E+05
8	10 nM [sb0]	4 nM [i-8]	20 nM [dq2]	1.507E+06
8	10 nM [b6]	4 nM [i-8]	20 nM [dq2]	1.285E+06
8	10 nM [sb2]	4 nM [i-8]	20 nM [dq2]	4.338E+05
9	10 nM [sb0]	4 nM [i-9]	20 nM [dq2]	1.495E+06
9	10 nM [b6]	4 nM [i-9]	20 nM [dq2]	1.506E+06
9	10 nM [sb2]	2 nM, 4 nM, and 6nM [i-9]	20 nM [dq2]	9.939E+05
10	10 nM [sb0]	4 nM [i-10]	20 nM [dq2]	1.536E+06
10	10 nM [b6]	4 nM [i-10]	20 nM [dq2]	1.530E+06
10	10 nM [sb2]	4 nM [i-10]	20 nM [dq2]	7.796E+05
11	10 nM [sb0]	4 nM [i-11]	20 nM [dq2]	1.657E+06
11	10 nM [b6]	4 nM [i-11]	20 nM [dq2]	1.503E+06
14	10 nM [sb0]	4 nM [i-14]	20 nM [dq2]	1.730E+06
14	10 nM [b6]	4 nM [i-14]	20 nM [dq2]	1.621E+06

### 3.7.3 Kinetics of Reporters

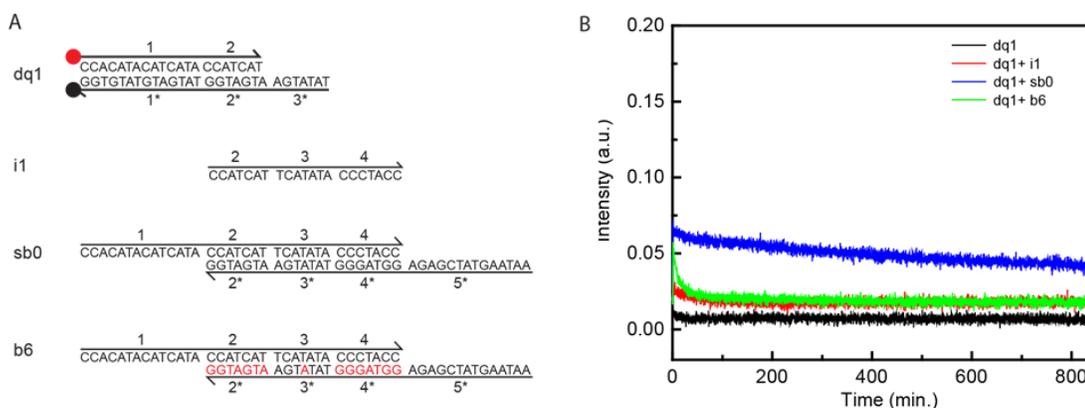
The reaction kinetics between Reporters dq1 and dq2 and the signal strands are shown in Figure 3.8. The reaction was fit to a second order reaction in equation (34). The rate constants for Reporters dq1 and dq2, which reacted with the signal strand, were much faster than the rate constants of the leakage and invading reactions (Table 3.4). Reporter dq1 was used to monitor the leakage and invading reactions when the invader toehold length was less than 5 nt. In comparison, Reporter dq2 was used to monitor the invading reactions when the invader toehold length was  $\geq 5$  nt.



**Figure 3.8.** The reaction kinetics between Reporters dq1 and dq2 and the signal strand. (A) Reporters and signal strand in sequence level. (B) Reporter dq1 kinetics. The final fluorescence intensity was normalized to one arbitrary units (a.u). The red dotted line is the fit. (C) Reporter dq2 kinetics. The final fluorescence intensity was normalized to 1 arbitrary units (a.u). The red dotted line is the fit.

### 3.7.4 Control experiments

The stability of the Reporter in the experimental buffer and the background reactions were monitored and are shown in Figure 3.9. The Reporter was stable in the reaction buffer, as shown by the black line. There was no detectable leakage between the Reporter and Substrates for the experimental conditions. The intensity of the Reporter reacting with the Substrate sb0 (blue line) went down during the reaction while the intensity of the Reporter reacting with the Substrate b6 (green line) only slightly decreased at the very beginning of the reaction and then stabilized (Figure 3.9B). This effect is not significant for the Substrate b6 since b6 with 15-LNA at the bottom strand is more stable than sb0.



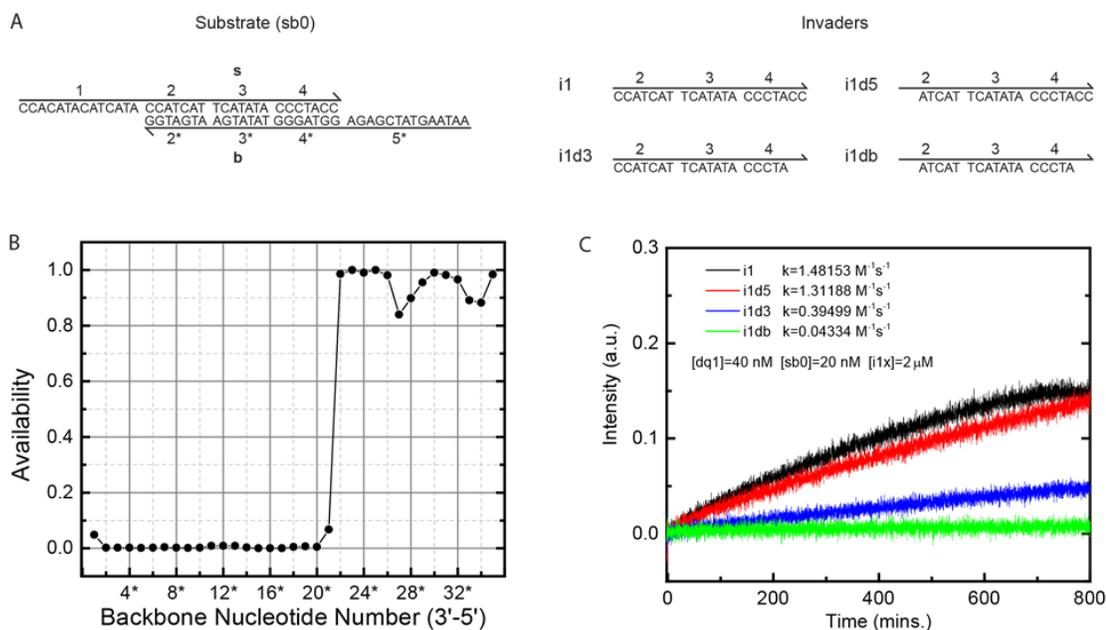
**Figure 3.9.** (A) Reporter (dq1), invader (i1), Substrates (sb0 and b6) are depicted at the sequence level. (B) Leakage controls. The black line denotes 40 nM Reporter dq1. The red line denotes the reaction of 40 nM Reporter dq1 and 6  $\mu$ M zero toehold invader i1. The blue line denotes the reaction of 40 nM Reporter dq1 and 20 nM original Substrate (sb0). The green line denotes the reaction of 40 nM Reporter dq1 and 20 nM LNA substituted Substrate b6.

### 3.7.5 Validating the source of leakage

To quantify the effects of secondary structure on the leakage rates of Substrates s1 and s3, the probability that a base was unpaired at equilibrium was calculated for the original Substrate (sb0) using NUPACK – which is a thermodynamic design and analysis tool for nucleic acid systems that was created at Caltech.<sup>1,2</sup> Presented here, our NUPACK software settings closely mirrored the experimental conditions of the study. For example, the operational temperature was set to 25 °C, the Substrate concentration was set to 10 nM, and the  $Mg^{2+}$  concentration was set to 11.5 nM. While  $Na^+$  was not experimentally added to the buffer, the  $Na^+$  concentration was set to NUPACK’s minimum required value of 0.05 nM. Finally, the maximum allowed components in the Substrate was set to a 2-component complex and the dangle option was set to “some” because the Substrate lacked a nick site but had extended tails.

As shown in (C) of Figure 3.10, the availabilities of bases 1\* and 21\* are 0.0488 and 0.0675, respectively. For reference, the susceptibility of the backbone at base 21\*

corresponds to the 3' end of the invader. The larger availability indicates that the right versus the left side of Substrate *sb0* is more susceptible to leakage. The near-zero availabilities, found within the interior of the duplex, also indicate that leakage is more probable at the terminal ends of the Substrate.



**Figure 3.10. Leakage kinetics.** 1 a.u. is corresponding to 20 nM signal strand. (A) The original sequence-level Substrate (*sb0*) and invaders. (B) Availability of Substrate (*sb0*) backbone. (C) Leakage kinetics with zero toehold invaders with deletions. Invader *i1d5* denotes zero toehold invader with 2 nt deletion at 5' end. Invader *i1d3* denotes zero toehold invader with 2 nt deletion at 3' end. Invader *i1db* denotes zero toehold invader with 2 nt deletion at both 5' and 3' ends.

Experimental results in (C) of Figure 3.10 confirm that leakage is faster on the right versus the left side of the Substrate (*sb0*). More specifically, the original invader (*i1*) was modified by deleting two nucleotides at its: (1) 5' end to create *i1d5*, (2) 3' end to create *i1d3*, and (3) both its 5' and 3' ends to create *i1db*. Compared to the original invader, *i1d3* and *i1d5* caused a 4-fold and marginal leakage reduction, respectively – confirming that the leakage pathway is dominated by the 3' end of the invader nucleating

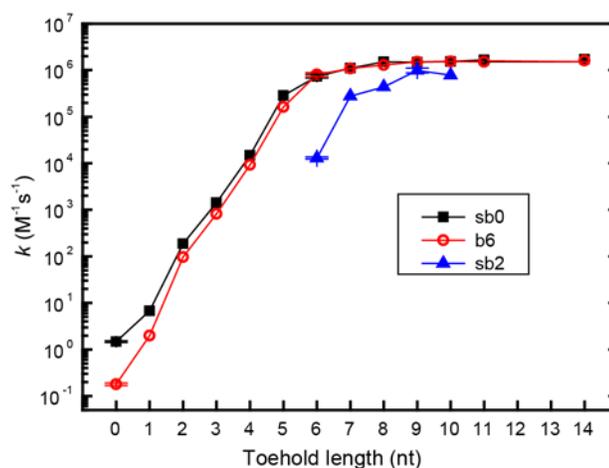
on the right side of Substrate sb0. As expected, deletions at both ends of the invader (**i1db**) showed the strongest leakage suppression.

In addition, when invader **i1d5** reacted with Substrate sb0 the leakage rate constant was nearly the same as when invader **i1** reacted with Substrate s1. This indicates that a 2 nt deletion at the 5' end of invader **i1d5** is equivalent to a 2 LNA substitution on domain 2 of signal **s1**. In contrast, when invader **i1d3** reacted with Substrate sb0 the leakage rate constant was almost the same as when invader **i1** reacted with Substrate s3. This indicates that the effect of a 2 nt deletion at the 3' end of invader **i1d3** is equivalent to a 2 LNA substitution on domain 4 of signal **s3**. In summary, nucleotide deletions at the ends of an invader likely reduce the probability that nucleation will occur between the invader and its Substrate. In addition, incorporating LNA substitutions on the Substrate can have a similar effect.

### 3.7.6 The effect of toehold length on reaction kinetics

Invasion rate constants are logarithmically plotted for three Substrates (sb0, b6, and sb2) versus different toehold length invaders (Figure 3.11). Substrates sb0 and b6 show similar kinetic performance even though they have 0 and 15 LNAs on their backbone strands, respectively. Both Substrates exhibit an initial slope region between 0-6 nt toehold invaders and a saturated plateau between 7-14 nt toehold invaders. The transition point indicates that a 6 nt toehold is sufficiently long to form a stable nuclei between an invader and both Substrates. If the toehold length is less than 6 nt, the invasion of Substrate b6 is slightly slower than the original Substrate (sb0). This performance change may be attributed to the additional energy that is required to break LNA-DNA base pairs during branch migration. In comparison, a longer toehold (i.e. 9 nt)

is required for Substrate sb2 to reach its saturation plateau and hence achieve equivalent performance as sb0 and b6. Unlike b6, the leakage reaction is thermodynamically less favorable because the LNA-LNA base pairs are replaced by DNA-LNA base pairs in the Waste complex. As a consequence, a larger toehold is likely required to overcome the enthalpic loss.

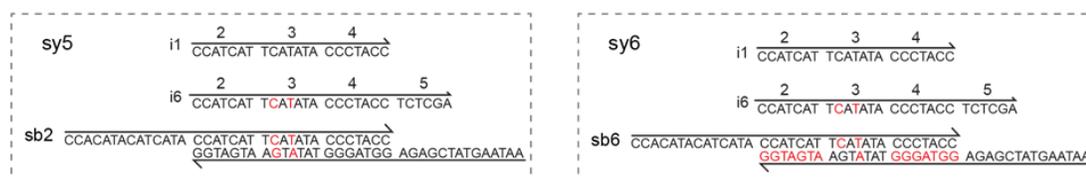


**Figure 3.11. Strand displacement rate constants and the toehold length of DNA invaders. sb0 is the original Substrate. b6 represents LNA substituted Substrate with 15-LNA on backbone strand. The rate constants are plotted against different toehold lengths of DNA invaders. Error bars are the standard deviation of three reactions with three different concentrations.**

### 3.7.7 Hybrid systems

To explore effective practices for incorporating LNA substitutions into DNA strand displacement systems, two additional systems were investigated in Figure 3.12 that were not included in Figure 3.6. For both systems, the leakage rate and invasion rate constants were measured and then compared to the DNA control (sy1) in Figure 3.6. For all systems in Figures 3.6 and 3.12, the performance enhancement ratio was calculated from the ratio of the invasion rate to the leakage rate constants. The performance enhancement ratio of sy1 was normalized to one unit and showed the largest leakage rate

(Figure 3.12). In comparison, systems sy5 and sy6 took full advantage of LNA substitutions during the leakage and invasion reactions. As shown in Figure 3.12, there are no LNA substitutions on the 0 nt toehold invader. More specifically, the leakage reactions are between DNA invader i1 and Substrates sb2 and sb6 with LNA substitutions on both systems. The dramatic leakage reduction likely comes from the thermodynamic penalty of the reaction and the kinetic barrier of branch migration. For example, the leakage reaction is thermodynamically unfavorable because the LNA-LNA base pairs in the Substrate are replaced by DNA-LNA base pairs in the Waste complex. In addition, kinetic hindrance may originate from: (1) a lower probability of hybridization between the zero toehold invader and the Substrates, and (2) the high energy barrier imposed by LNA-DNA and LNA-LNA base pairs in the Substrates during branch migration. In contrast, the invasion reactions are between hybrid DNA/LNA invaders and the Substrates. The invasion rate constants are the same as in the pure DNA control system (sy1). The elevated invasion rate is likely because of toehold base pairing gains and an unbiased random walk during branch migration. As a consequence, the performance enhancement of sy5 and sy6 was improved by a factor of 23 and 53, respectively. While not shown here, the design of hybrid DNA/LNA or pure LNA systems provide insights for future reaction networks made from nucleic acids.



**Figure 3.12. Optimized hybrid DNA/LNA systems. Strands and sequences for sy5 and sy6 systems. Black and red letters respectively denote DNA and LNA.**

## REFERENCES

1. Dirks, R.M., Bois, J.S., Schaeffer, J.M., Winfree, E. and Pierce, N.A. (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49, 65-88.
2. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: Analysis and Design of Nucleic Acid Systems. *Journal of Computational Chemistry*, 32, 170-173.

## CHAPTER FOUR: FINAL CONCLUSIONS

Nucleic acids as information-dense programming materials are the fundamental building blocks for non-equilibrium dynamic DNA circuits. This dissertation discussed the kinetic control of the dynamic circuits based on toehold mediated strand displacement systems. The reaction kinetics can be tuned over three orders of magnitude by site-specifically incorporating defects into dynamic systems. Mismatched base pairs and chemically modified LNAs were strategically included into the circuits to decrease the unwanted background leakage and increase the desired reactions. In addition, availability as a design metric was used to quantify secondary structures of single strands and reveal vulnerable breathing positions of duplexes. Mutual availability was also used to facilitate the rational design of high performance nucleic acid circuits. The detailed discussion of availability and mutual availability and mismatch defects was given in Chapter 2 of this dissertation. Strategically incorporating LNAs into DNA strand displacement systems was discussed in Chapter 3.

In Chapter 2, kinetics barriers were introduced using mismatched base pairs between the fuel and backbone of the substrate. The leakage and catalyzed reaction rates for each fuel modification at the critical 5', 3', and nick sites were tested systematically. We found that mismatches were not the only factor that affects kinetics rates. Changes in the availability of each base on the fuel strand, induced by mismatch modifications, had a stronger effect on the kinetics of the leakage and catalytic reactions than mismatch identity itself. An exponential correlation between the mutual availability and the leakage

rate was observed independent of the mismatch identities. The leakage rate was reduced by a factor of 100 in a catalytic DNA strand displacement system by modifying the fuel strand at multiple sites to create mismatches with the backbone strand of the substrate. The best performance improvement of the catalytic system was a factor of 4 by using a single site fuel mismatch modification that was decoupled from the catalyzed rate.

Design recommendations for improving total performance can be summarized as following: (1) calculate the availability and mutual availability; (2) select substrate sequences that lower the availability at the breathing sites (i.e. 5', 3' and nick sites); (3) incorporate mismatch modifications into the fuel to increase the secondary structure ensemble and decrease its availability, especially at nucleation sites; (4) choose mismatch locations that do not occur at a fuel toehold or catalyst release site, and (5) choose mismatch locations that do not produce secondary structure interference at the fuel toehold.

Based on these recommendations, hybrid DNA/LNA systems were designed to optimize the performance of nucleic acid strand displacement circuits in Chapter 3. During the sequence design process, the base availability of the invader strand was maximized to minimize the unwanted secondary structure and ensure fast invasion kinetics. The base availabilities of the terminal ends of the substrate were minimized to mitigate fraying. In addition, LNA substitutions were strategically introduced to the substrate to further stabilize the duplex while maintaining the original sequence design.

For the leakage reaction, the probability of nucleation between the zero nt toehold invader and substrate were dramatically reduced due to the thermomechanical stability of LNA-DNA or LNA-LNA base pairs at the terminus of the substrate. LNAs in the central

region of the substrate rendered a biased random walk, which discouraged the invader from binding to the backbone. The leakage rate constant was reduced to  $0.03 \text{ M}^{-1}\text{s}^{-1}$  in a hybrid DNA/LNA system. For the invasion reaction, the availability of invader toehold was remained high to ensure fast toehold hybridizing with the backbone. High probability of strand displacement was sustained by extending the toehold length, substituting LNAs into the invader, or only incorporating LNAs into the backbone of the substrate. Overall, the performance enhancement ratio was improved over 50 folds by using the novel design metric of availability and site-specifically substituting LNAs into DNA strand displacement systems.

Future work can explore ways to combine mismatched base pairs and modified nucleic acids into dynamic nucleic acid systems to improve the performance of sensitivity, selectivity, scalability and stability. In addition, NUPACK as a design tool can be further developed to include LNA-DNA and LNA-LNA thermodynamics, interactions of mismatched base pairs and pseudo-knots. The study of kinetic control of strand displacement by defect engineering will make contributions to the rational design of dynamic DNA circuits and broaden the application of DNA nanotechnology into new fields.