

DIRECTED EVOLUTION OF A NOVEL HETEROCHIRAL RIBONUCLEASE  
RIBOZYME  
&  
KINETIC CHARACTERIZATION OF HETEROCHIRAL DNA STRAND  
DISPLACEMENT REACTIONS

A Dissertation

by

NANDINI KUNDU

Submitted to the Graduate and Professional School of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jonathan Szczepanski
Committee Members,	Tadhg Begley
	Wenshe Liu
	Dorothy Shippen
Head of Department,	Simon North

December 2021

Major Subject: Chemistry

Copyright 2021 Nandini Kundu

## ABSTRACT

Nature is inherently homochiral. L-DNA and L-RNA, the enantiomeric forms of native D-DNA and D-RNA respectively, do not occur naturally and are virtually bioorthogonal. Compared to common chemical modifications, L-DNA/RNA exhibit superior qualities such as increased biostability due to exceptional nuclease resistance, low immunogenicity, and minimal off-target effects. When interacting with an achiral physical or chemical environment, they behave similar to their native D-counterparts, facilitating easy optimization of designs based on L-DNA/RNA. However, while these “mirror-image” nucleic acids hybridize to each other, they are incapable of forming contiguous WC base-pairs with complementary native nucleic acids, a caveat that until recently precluded their use in applications at the interface of native biology. Redirecting the focus toward non-canonical heterochiral nucleic acid interactions based on non-WC hydrogen bonding, Van der Waals and hydrophobic interactions etc., overcomes this obstacle. This work focuses on exploring and expanding these novel recognition modalities with the aim of interfacing L-DNA/RNA with native nucleic acids using both structure and sequence-based approaches.

Complex structural interactions in RNA govern almost all aspects of gene expression. Targeting structured RNA is thus integral to our fundamental understanding of RNA biology as well as in the development of therapeutics. Several heterochiral L-DNA/RNA aptamers have been evolved to bind their respective RNA targets with high specificity and

affinity, in a structure-specific manner. In contrast, evolution of L-ribozymes is more challenging. RNA ligation and polymerization is the only reported example of heterochiral catalysis to date. In the first part of my thesis, I will discuss the *in vitro* evolution of a heterochiral ribonuclease ribozyme that interacts with a representative structured RNA target to mediate phosphodiester bond scission resulting in cleavage of the target. This opens a novel route to chemically target a specific RNA within its structural context while eliminating WC-based off-target hybridization. Furthermore, this approach is also promising for future therapeutic applications.

The second part of this thesis focuses on sequence-specific interfacing strategies. Our research group recently described a novel technique called “heterochiral” DNA strand displacement reactions that utilize an achiral peptide nucleic acid (PNA) mediator to exchange sequence information between D-DNA and L-DNA, thus providing a route to exploit the advantageous properties of L-DNA in dynamic DNA nanotechnology applications. In this work I present extensive kinetic characterization of these novel reactions by systematically varying key design parameters in order to establish a set of design principles that will facilitate the rational design of such devices for biomedical applications in the future. Additionally, investigation of the biophysical mechanism of these reactions reveals a novel stereochemical control over reaction kinetics, that adds to the versatility of future designs.

## DEDICATION

*This work is dedicated to my family.*

To my parents Dr. Rita Kundu and Dr. Anadi Kumar Kundu who have always had unwavering faith in my abilities and made sure I had access to every resource and opportunity. Their strength and resilience are my source of inspiration. I hope that has been reflected in everything I strive to do and the person I aspire to be.

To my beloved uncle Mr. Dhruva Jyoti Sen and grandmother Mrs. Pramila Sen. Their love and warmth keep me grounded and closer to home.

And finally, to the memory of my late uncle Dr. Shyamal Kumar Datta who was one of my biggest cheerleaders. He barely missed this milestone, and I know he would have been very proud.

## ACKNOWLEDGEMENTS

I would like to thank my Committee Chair Dr. Jonathan Szczepanski for his guidance and support, and my Committee Members Dr. Begley, Dr. Liu and Dr. Shippen for their valuable feedback on my research endeavors, especially during the candidacy examinations.

When I joined the PhD program at TAMU Chemistry, I mostly knew that I wanted to work for Jon, but that apart, I was entirely clueless. Jon took a chance on me and personally invested a lot of time in training me. Over the years, he has entrusted me with ambitious projects and lab responsibilities. He has also shown great faith in my scientific capabilities and research progress, even during the times when I wasn't too sure myself. He has provided every possible resource to me and the group in general, and always encouraged participation in extra-curricular activities. My time in his lab has definitely made me a more confident and self-assured individual. I am especially grateful to Jon for his understanding and flexibility this last year, and continued support toward my scientific endeavors.

I owe a great deal to my colleagues in lab, especially to some members who have now left the group but contributed immensely to my PhD experience. I would like to specially mention Brian Young for his mentorship, especially during our collaborative project, and for being so generous with his time over the years. He was always eager to discuss science which was very helpful on difficult days. He lent a helpful hand when an instrument broke, or I hit a major roadblock in my experiments. I am grateful for his

friendship and support in lab. Joanna San Pedro helped me master many of the techniques in lab, she is the reason I overcame my fear of casting giant gels before too long! I will remember the days she extended her lab hours to help me out, and has always been a good, empathetic friend. I would also like to thank Adam Kabza; his energy and motivation challenged me to strive and work harder during the very hectic first year of graduate school. I have been moved by his genuine friendship and care many times. Charles Deckard has been a role model in work ethic and discipline, and despite our research areas being different, we had many invigorating science discussions. His friendly presence in lab was a great comfort during the pandemic. Special thanks to Deb Ranjan Banerjee and Sougata Dey for their brotherly presence and affection. I have only recently realized what a privilege it was to be able to talk in my native tongue with them on a daily basis. I am also grateful to the current group members. They are all highly talented and motivated individuals, and I greatly appreciate all their help.

I would also like to recognize the fantastic work our graduate office does to support grad students. They make our life easy in so many ways. Sandy (Sandra Horton) has especially helped me out on countless occasions. Thanks to the healthcare professionals at TAMU as well for their dedicated care during the pandemic.

It certainly takes a village to produce a PhD. Some behind the scenes players deserve special mention. I have found an excellent community in College Station, and time spent with them will form some of the most cherished memories of grad school. Special thanks to Isita Jhulki, Anuvab Das and Akshi Singla who have been always there for me through good times and bad, and I couldn't have survived grad school without

them. Isita Di always had a hot cup of tea and comfort food to replenish my weary soul after a long day, and I have found a big sister in her. Anuvab, who has been a faithful friend since our undergraduate days, I thank for the example he set by being an excellent scientist and an equally excellent human being. Akshi has been another friend who I could always rely on, and I am truly grateful for the wonderful memories we have made together, and our shared perspective on many things and values in life. My childhood friends Tanuj Kar, Anwasha Bhattacharya and Samadrita Sengupta deserve special mention. Distance has never mattered in our friendships, and they make life better by just existing. I can't thank them enough for keeping me grounded through my graduate school years.

Words cannot capture the gratitude I have for my parents and my family. My parents are the reason I am here today. Their trajectories motivated me to pursue this path since I was a child and I look up to them for everything. They have set a high standard for me by their work ethic, integrity and ambitions. They have sacrificed many a pleasure to make sure I never wanted for anything and that I had every opportunity that they couldn't have. My family is my foundation, and their love and support has been instrumental in my growth as a scientist and an individual. Finally, this list remains incomplete if I don't mention my good boy Baltu who has been the perfect fur-baby. He has made my life so much better, and I truly appreciate his hugs and company during my writing sessions.

## CONTRIBUTORS AND FUNDING SOURCES

### **Contributors**

This work was supervised by a thesis (or) dissertation committee consisting of Professor Jonathan Szczepanski (Committee Chair), Professor Tadhg Begley and Professor Wenshe Liu of the Department of Chemistry, and Professor Shippen of the Department of Biochemistry & Biophysics.

Hypothesis and experiments of Chapter 3 were designed in collaboration with Dr. Brian Young. The analyses depicted in Chapter 3 were conducted in part by Dr. Brian Young of the Department of Chemistry (currently Department of Radiology, Stanford University) and were published in 2021.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

### **Funding Sources**

This work was supported by Cancer Prevention and Research Institute of Texas under Grant Number RR150038, National Institute of Biomedical Imaging and Bioengineering under Grant Number R21EB027855, National Institute of General Medical Sciences under Grant Number R35GM124974 and the National Science Foundation under Grant Number 2003534.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the respective funding agencies.



## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
DEDICATION .....	iv
ACKNOWLEDGEMENTS .....	v
CONTRIBUTORS AND FUNDING SOURCES.....	viii
TABLE OF CONTENTS .....	ix
LIST OF FIGURES.....	xii
LIST OF TABLES .....	xviii
1. INTRODUCTION.....	19
1.1. Prelude: The Basics of Nucleic Acids Biochemistry .....	19
1.2. The Structure-Function Relationship in RNA.....	22
1.3. Ribozymes - Nature's earliest catalysts .....	28
1.3.1. Origin of Life: The RNA World Theory .....	29
1.3.2. RNA cleaving DNazymes and Ribozymes .....	29
1.4. RNA Therapies for RNA Targets.....	32
1.4.1. RNA as a Target .....	32
1.4.2. Strategies for targeting RNA .....	33
1.4.3. Advantages and Challenges associated with RNA based therapeutics .....	37
1.5. Common Chemical Modifications in Nucleic Acids Based Strategies .....	40
1.5.1. Phosphate Backbone Modifications (Phosphorothioate) .....	40
1.5.2. Ribose Modifications .....	41
1.5.3. Alternative Backbone .....	42
1.5.4. Ribose Analogues.....	43
1.6. L-DNA as a Stereochemical Modification: Properties and Applications .....	43
1.7. Interfacing Native RNA with L-DNA/RNA .....	47
1.7.1. Structure-based Binding Interactions between Native D-RNA and L- Aptamers .....	47
1.7.2. Heterochiral Interactions in Catalysis of Chemical Reactions .....	48
1.7.3. Sequence-specific Interfacing by Heterochiral DNA Strand Displacement Reactions .....	49
1.8. DNA Strand Displacement Reactions .....	53

1.8.1. Basics and Mechanism .....	53
1.8.2. Applications in Dynamic DNA Nanodevices for Biological Applications ....	55
1.9. Goals of Research.....	55
1.9.1. Directed Evolution of a Novel Heterochiral Ribonuclease Ribozyme.....	55
1.9.2. Kinetic Characterization of Heterochiral Strand Displacement Reactions ....	57
1.10. References .....	59
2. DIRECTED EVOLUTION OF A NOVEL HETEROCHIRAL RIBONUCLEASE RIBOZYME AGAINST A STRUCTURED RNA TARGET .....	71
2.1. Introduction .....	71
2.2. Results and Discussion.....	75
2.2.1. Selection of <i>cis</i> -cleaving Heterochiral Ribonuclease Ribozymes .....	75
2.2.2. Selection for <i>trans</i> -cleaving Ribonuclease Ribozymes .....	89
2.3. Summary and Outlook .....	103
2.4. Materials and Methods .....	106
2.4.1. Materials .....	106
2.4.2. Analytical Services.....	107
2.4.3. Oligonucleotide Synthesis.....	107
2.4.4. Library Preparation.....	108
2.4.5. In vitro Selection .....	109
2.4.6. Cloning Sequences: .....	112
2.4.7. Preparation of Individual Clones/ Mutations/ Truncations for Characterization.....	112
2.4.8. Assaying of <i>cis</i> cleavage activity .....	113
2.4.9. Kinetic Analysis of <i>cis</i> cleavage.....	113
2.4.10. Characterization of Cleavage Sites and Cleavage Products .....	114
2.4.11. Trans Assays of R9t-24 and related truncations and mutations .....	114
2.4.12. Kinetic analysis of trans assays .....	115
2.5. References .....	116
3. KINETICS OF HETEROCHIRAL STRAND DISPLACEMENT FROM PNA-DNA HETERODUPLEXES .....	120
3.1. Introduction .....	120
3.2. Results and Discussion.....	125
3.2.1. Toehold length and stereochemistry modulate the kinetics of strand displacement from a PNA–DNA heteroduplex.....	125
3.2.2. Strand displacement from PNA–DNA heteroduplexes is sensitive to mismatches .....	129
3.2.3. Toehold stability is dependent on the reaction configuration .....	134
3.2.4. Strand displacement from PNA–DNA heteroduplexes is compatible with toehold exchange .....	138

3.2.5. RNA inputs accelerate the rate of strand displacement from PNA–DNA heteroduplexes .....	142
3.3. Summary and Outlook .....	146
3.4. Materials & Methods.....	149
3.4.1. Materials .....	149
3.4.2. Oligonucleotide synthesis and purification .....	150
3.4.3. Preparation and characterization of duplex reaction components .....	151
3.4.4. Monitoring strand displacement reactions by fluorimetry .....	152
3.4.5. Rate constant fitting procedure.....	153
3.4.6. Melting temperature analysis .....	155
3.5. References .....	157
4. CONCLUSIONS .....	164
4.1. The Bigger Picture .....	164
4.2. In vitro evolution of heterochiral ribonuclease ribozymes.....	166
4.2.1. Selection Methodology.....	166
4.2.2. Model System for Characterizing complex Heterochiral Interactions and possible applications.....	167
4.2.3. Unique dependence on 5'-triphosphate .....	169
4.2.4. Future Therapeutic Potential .....	170
4.3. Kinetic Characterization of Heterochiral DNA Strand Displacement Reactions	171
4.4. References .....	173
APPENDIX A .....	175
APPENDIX B.....	187

## LIST OF FIGURES

	Page
Figure 1-1 Central Dogma of Molecular Biology. As depicted by Francis Crick, this is the central dogma representation as was known in 1958. Solid arrows depict the known transfer of genetic information, dashed arrows represent the plausible transfer of information. Adapted by permission from Ref (1). Copyright © 1970, Nature Publishing Group.....	19
Figure 1-2 DNA Structure. Top left shows the double helical structure of DNA, with Watson-Crick base-pairing holding the two complementary strands in an anti-parallel direction. To the right is the molecular structure of a representative dinucleotide showing the individual components and the hydrogen bonding forming the WC base pairs. The chemistry of a representative nucleotide is elaborated in the bottom half. Adapted from <i>Openstax</i> . .....	20
Figure 1-3 Secondary Structures of RNA. a) Hierarchical organization in RNA starting from the primary nucleotide sequence directing formation of simple secondary structure motifs, which interact to form higher order structures. Quaternary assemblies represent the highest possible complexity. Left structure shows DNA chromatin structure wrapped around histone proteins. Source: Wikipedia Commons. b) RNA secondary structures color coded and depicted in the context of a representative RNA molecule. Adapted from Ref (12). Copyright © 2019, Springer Nature Limited.....	24
Figure 1-4 Structure-Function Relationship in RNA. Examples of RNA structure mediated biological functions some of which are discussed in the text of this section. Figure adapted from Ref (12). Copyright © 2019, Springer Nature Limited. ....	26
Figure 2-1 Design of Selection Target. a) Secondary structure representation of L-TARS showing a 9 base-pair long discontinuous helix with a 6-nucleotide unpaired loop. The 3'-end is functionalized with biotin for immobilization of the substrate on streptavidin coated beads and the 5'-end is equipped with a linker that helps in monitoring the desired cleavage reaction. b) Chemical structure of the 5'-linker.....	76
Figure 2-2 A detailed schematic of “on the bead” selection strategy. Step 1 involves extension of the random D-RNA library using a D-DNA primer to generate an RNA-DNA hybrid. There is a 4-nucleotide overhang on the 3'-terminus of the RNA strand in this duplex that enables ligation of the hybrid duplex to the L-TAR RNA substrate in Step 2. In the next Step 3, this whole ligated cis-assembly	

is immobilized on streptavidin coated beads using the biotin handle and the blocking DNA strand is removed using a denaturing. Step 4 is the actual selection step where the single-stranded assembly is incubated for the requisite amount of time in the 1X selection buffer activated with 20 mM MgCl<sub>2</sub>. Any RNA molecules falling off the beads are isolated and converted to cDNA in Step 5 followed by PCR amplification and transcription in Step 6 to generate the enriched library for the next iterative round of selection.....78

Figure 2-3 Denaturing PAGE analysis of cleavage products from Round 6: Lanes 1,4 and 5 are controls for the full-length *cis*-assembly, irradiated *cis*-assembly and the enriched RNA library alone, respectively. Lane 2 represents the cleavage products isolated from Round 6. The boxed area shows potential cleavage products migrating between the full-length assembly and the RNA library alone. Arrow marks indicate the two distinct bands emerging from potentially two separate cleavage sites on the substrate. Lane 3 is part of the sample run on Lane 2 but irradiated to remove the photocleavable linker. Faster migration in Lane 3 compared to Lane 2 confirms that cleavage occurs on the substrate as desired..82

Figure 2-4 Secondary structure representation of *cis*-cleaving ribozyme R8c-3: Short terminal stem-loop structures and a long discontinuous but stable stem affords the best *cis*-cleavage activity among all other ribozymes isolated.....83

Figure 2-5 Characterization of *cis*-cleaving ribonuclease ribozymes. a) Sequence information of the four unique active sequences after 8 rounds of iterative selection. b) a time-course of *cis*-cleavage activity shows the three clones R8c-1,3 and 6 having similar  $k_{obs}$  values but different equilibrium yields. c) denaturing PAGE analysis of cleavage products of the individual ribozyme sequences show diversity in the cleavage sites, all of them on the 6-nucleotide loop as indicated by the ladder generated by alkaline digestion of the substrate used for the assay.....84

Figure 2-6 Truncation studies on R8c-3. a) Secondary structural representation of R8c-3 after removal of 10 nucleotides from the 5'-terminus shows two long stem elements S1 and S2. Removal of the boxed part in dark grey only generates S1-T1, removal of boxed parts in both dark and light grey generates S1-T2. Similarly, removal of only boxed part in dark blue generates S2-T1, and removal of both dark and light blue colored boxes generates S2-T2. S2-T2 is only active when combined with S1-T2 but not with S1-T1. b) the minimal catalytic motif identified based on R8c-3.....88

Figure 2-7 Design of Reselection Target. a) Secondary structure representation of L-TAR<sub>L</sub> compared to L-TAR<sub>S</sub> showing an additional 5 base-pairs stabilizing the stem.

The 5'-end of L-TAR<sub>L</sub> is functionalized with biotin for immobilization of the substrate on streptavidin coated beads and the 3'-end is equipped with a linker that helps in monitoring the desired cleavage reaction. b) Chemical structure of the 3'-linker.....91

Figure 2-8 Active Ribozyme Sequences from Reselection Round 9. Comparison of sequence information from unique active ribozymes isolated from reselection to wild-type R8c-6.....93

Figure 2-9 Secondary structure representation of R9t-24. This *cis* and *trans*-active ribozyme has several stable stem and hairpin structures. The residues in red show the location of mutations compared to the WT R8c-6 and the table lists the residues they replace.....94

Figure 2-10 Characterization of R9t24. a) *cis* kinetics of R9t24 at 20mM and 50 mM Mg<sup>2+</sup> shows similar first order rate constant at lower Mg<sup>2+</sup> concentrations which doubles at 50 mM Mg<sup>2+</sup> b) comparison of equilibrium yields for the *cis* reactions of the three main families of ribozymes and two other unique sequences selected after 9 rounds of reselection at 20mM and 50 mM Mg<sup>2+</sup> shows much enhanced activities compared to the WT-clones. c) a time course of *trans* kinetics of R9t-24 at 500 nM of substrate, 5μM of the ribozyme and 50 mM Mg<sup>2+</sup> at room temperature, spanning a day (main graph) and a week (left top inset) indicates very weak *trans* activities. The kinetics is fitted to a pseudo-first order rate equation. Catalyzed reaction is ~20 times faster than the non-specific hydrolysis. d) comparison of *trans* cleavage yields keeping the substrate constant at 500 nM and varying only the concentration of R9t-24 over 2 orders of magnitude show saturation around 5μM.....97

Figure 2-11 Cleavage sites of R9t-24 *in cis* and *in trans*. Alkaline digestion ladder and RNase A and T1 digestions of the corresponding D-TAR<sub>L</sub> substrate shows that the cleavage site remains unchanged from that of WT R8c-6 between G16 and G17 (numbering is consistent with L-TAR<sub>L</sub> for ease of reference).....99

Figure 2-12 Dependence of *trans*-activity on Mg<sup>2+</sup> concentration and pH. a) Activity increases with increasing concentrations of Mg<sup>2+</sup> until saturation at 50 mM, after which there is a sharp decline most probably due to non-specific degradation. Right inset shows activity is highly specific for only Mg<sup>2+</sup> b) activity remains mostly constant from physiological pH to alkaline pH of 9.0 after which non-specific laddering takes over.....100

Figure 3-1 Toehold-mediated strand displacement reaction. DNA is depicted as lines with the half arrow indicating the 3' end throughout the text. A substrate strand consisting of a single-stranded toehold domain ( $t^*$ ) and a branch migration domain ( $a^*$ ) is initially hybridized to an incumbent strand (OUT) to form duplex A. The input (or invader) strand IN is complementary to both the toehold ( $t^*$ ) and branch migration domains ( $a^*$ ) of the substrate strand. Displacement is initiated by binding of IN to the toehold (via  $t/t^*$ ) followed by a three-way branch migration process in which base pairs between the incumbent and substrate strands dissociate and are replaced by base pairs with invader IN. The reaction is complete once the incumbent strand (OUT) is fully displaced from duplex A. .... 121

Figure 3-2 Strand displacement from a PNA–DNA heteroduplex is dependent on toehold length and stereochemistry. (A) Schematic illustration of the strand displacement reaction system depicting both the homo- and heterochiral pathways. D-DNA is shown in black; L-DNA is shown in blue, and PNA is shown in green throughout the text. The half-arrow denotes the C-terminus of the PNA strand. The sequences of all strands are depicted in Appendix Figure B1 and Appendix Table B1. (B, C) Fluorescence monitoring (Cy3) of the homochiral (B) and heterochiral (C) reaction pathways initiated with inputs (D-IN) having toehold domains varying in length from 0–10 nucleotides (nt). The length of the toehold is indicated on the right y-axis. The reactions depicted contained 150 nM D-IN, 100 nM D/L-Ai, 300 nM D/L-R, 300 mM NaCl, 1 mM EDTA, and 10 mM Tris (pH 7.6) and were carried out at 37°C. Fluorescence in all figures is reported in units such that 1.0 is the fluorescence of the maximally activated reporter control and 0.0 is the background of the quenched reporter complex (D/L-R). (D) Semilogarithmic plot showing the exponential dependence of calculated rate constants on toehold length. All calculated rate constants are listed in Appendix Table B2. .... 127

Figure 3-3 The position of a mismatch affects the rate of strand displacement from PNA–DNA heteroduplexes. (A) Schematic of the mismatched inputs used in this study. The identity of the mismatched nucleotide relative to the fully matched input ( $D-IN_{TH8}$ ) is shown above the strand. (B, C) Fluorescence monitoring (Cy3) of the homochiral (B) and heterochiral (C) reaction pathways initiated with different mismatched inputs in (A). The identity of the mismatched input ( $IN_{MMA-F}$ ) is indicated on the right y-axis. The reaction initiated with the fully matched input ( $D-IN_{TH8}$ ) is shown as a black dotted line. Reactions depicted here were carried out as described in Figure 3- 2. (D) Kinetic discrimination factors ( $DF = k_{match}/k_{mismatch}$ ) of hetero- and homo- chiral reactions towards inputs having different mismatches, where  $k_{match}$  and  $k_{mismatch}$  are the calculated rate constants for a

fully-matched and mismatched input, respectively. Error bars represent standard deviation from three independent experiments. .... 133

Figure 3-4 Characterization of homo and heterochiral toehold interactions. (A) The model system used to monitor toehold association based on fluorescence quenching. (B) Fluorescence melting curves for the homo and heterochiral toehold duplexes ( $C_i = 2 \mu\text{M}$ ). Fluorescence values were corrected ( $F_{\text{corr}}$ ) for background fluorescence and temperature-dependent effects as defined in eq 4 (see Materials and Methods). The  $T_m$  for each toehold configuration was averaged over three melting experiments. (C) Thermodynamic parameters for homo and heterochiral toehold association. .... 136

Figure 3-5 The length and nucleotide composition of the incumbent toehold affects toehold exchange on PNA–DNA heteroduplexes. (A) Schematic illustration of the toehold exchange mechanism for both the homo- and heterochiral reaction pathways. (B) Sequences of the incumbent toehold domains within D/L-Ai and its truncated version D/L-Ais. Individual incumbent toeholds are boxed based on their length and red letters emphasize G/C base pairs. Incumbent toeholds are produced by truncating the input strand by the corresponding length. (C) Calculated rate constant as a function of incumbent toehold length ( $k$ ) relative to the full-length input ( $k_0$ ) having no incumbent toehold. Error bars represent standard deviation from three independent experiments. (D, E) Fluorescence monitoring (Cy3) of toehold exchange for the homochiral (D) and heterochiral (E) reaction pathways. The length of the incumbent toehold is indicated on the right y-axis. Dotted lines indicate reactions carried out with the truncated PNA–DNA heteroduplex (Ais) for the indicated incumbent toehold lengths (black = 0-nt; red = 8-nt). Reactions depicted here were carried out as described in Figure 3-2. .... 140

Figure 3-6 Strand displacement from PNA–DNA heteroduplexes is faster for RNA inputs than DNA inputs. (A, B) Fluorescence monitoring (Cy3) of the homochiral (A) and heterochiral (B) reaction pathways initiated with inputs RNA inputs (D-IN<sub>RNA</sub>) having toehold domains varying in length from 6–10 nucleotides. The length of the toehold is indicated on the right y-axis. For reference, the reaction initiated with a DNA input having a 10-nucleotide toehold is shown as a black dotted line. The reactions depicted contained 30 nM D-IN, 20 nM D/L-Ai, 60 nM D/L-R, 300 mM NaCl, 1 mM EDTA and 10 mM Tris (pH 7.6) and were carried out at 37°C. (C) Calculated rate constant for RNA inputs as a function of toehold length ( $k_{\text{RNA}}$ ) relative to the DNA input ( $k_{\text{DNA}}$ ) having the same length toehold. Error bars represent standard deviation from three independent experiments. .... 144



Figure 3-7 Discrimination factors ( $DF = k_{\text{match}}/k_{\text{mismatch}}$ ) of PNA–DNA heteroduplexes towards RNA inputs having different mismatches, where  $k_{\text{match}}$  and  $k_{\text{mismatch}}$  are the calculated rate constants for a fully matched and mismatched RNA input, respectively. Error bars represent standard deviation from three independent experiments. Hetero and Homo refer to the heterochiral and homochiral reactions pathways, respectively. See Figure 3-3A for mismatch identity and position.....145

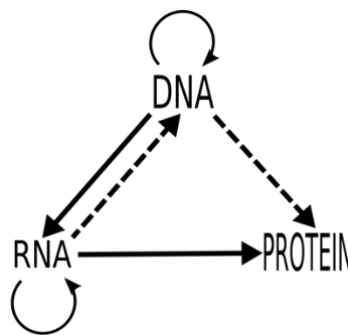
## LIST OF TABLES

	Page
Table 1 Comparison between Common Chemical Modifications to D-DNA. Advantages and disadvantages of their use in DNA strand displacement reactions. Reprinted from Ref (131) with permission. © 2021 Wiley Periodicals LLC. ....	50

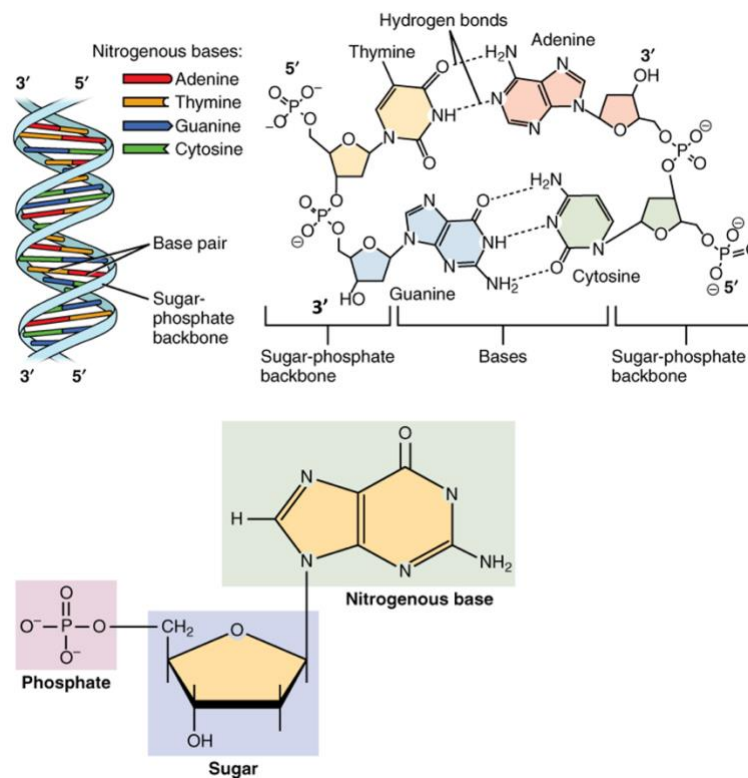
## 1. INTRODUCTION

### 1.1. Prelude: The Basics of Nucleic Acids Biochemistry

Every cellular function is ultimately the result of very specific sets of instructions pre-programmed into our genes. Genes are long segments of DNA whose unique nucleotide sequence directs the synthesis of a functional biological product, most commonly proteins. The central dogma of molecular biology underlines three main processes that facilitate the flow of this information inside the living cell (1). First, DNA undergoes “replication” to form an identical daughter DNA. Replication ensures preservation of genetic information during cell division as well as reproduction. Next, the sequence information of DNA is copied onto the corresponding messenger RNA (mRNA) through “transcription”. Finally, in a process called “translation”, specific amino acids corresponding to “codons”, which are groups of three nucleotides on the mRNA, are recruited and assembled by the ribosome to form the final protein product.



**Figure 1-1 Central Dogma of Molecular Biology.** As depicted by Francis Crick, this is the central dogma representation as was known in 1958. Solid arrows depict the known transfer of genetic information, dashed arrows represent the plausible transfer of information. Adapted by permission from Ref (1). Copyright © 1970, Nature Publishing Group.



**Figure 1-2 DNA Structure.** Top left shows the double helical structure of DNA, with Watson-Crick base-pairing holding the two complementary strands in an anti-parallel direction. To the right is the molecular structure of a representative dinucleotide showing the individual components and the hydrogen bonding forming the WC base pairs. The chemistry of a representative nucleotide is elaborated in the bottom half. Adapted from *Openstax*.

To understand how this intermolecular transfer of information takes place, it is necessary to look at the chemistry of the biopolymers deoxyribonucleic acid or DNA, and ribonucleic acid or RNA. DNA consists of monomeric units called nucleotides that are composed of a phosphorylated deoxyribose sugar covalently attached to either one of the four nitrogenous bases: adenine (A), thymine (T), guanine (G) or cytosine (C). The RNA nucleotide is very similar except that it has a ribose sugar and the base uracil (U) instead of thymine (T). The genetic alphabet is thus fundamentally simple but at the same time

capable of storing vast amounts of information in a highly diverse sequence space generated by combination of these four bases.

DNA was physically isolated for the first time in 1869, but it wasn't until the 1950s that its function as genetic material was established (2,3). The double-helical structure of DNA as we know today, was proposed in 1953 (4). This model is principally credited to James Watson and Francis Crick who later shared the Nobel Prize for this discovery with biophysicist Maurice Wilkins; Rosalind Franklin, who was studying DNA structure using X-Ray crystallography at the same time and whose "Photo 51" provided the final missing clue to the complementary base-pairing in DNA, had arrived at many of the same conclusions herself, but her contributions are lesser known and credited (5,6). In this double-helical structure, each single-stranded DNA was aligned around a central axis in a way that the deoxyribose-phosphate backbone faced the solvent and the hydrophobic nucleobases faced inwards. These nucleobases are engaged in two major interactions. Since they are mostly planar, the nitrogenous bases of A, T, G and C are "stacked", with the rings of any two adjacent nucleotides in the DNA strand parallel to each other. Mediated by Van der Waals and dipole-dipole interactions, these stacking interactions are responsible for the stabilization of the helical structure. The distance along the vertical axis between two adjacent bases is 3.4 Å and each helical turn comprises approximately 10.5 such bases in solution and thus spans 34 Å. The double-helical DNA is a stable, rigid structure but there can be variations in the periodicity, handedness of the helix as well as in the depth of the major and minor grooves that are formed by the helical turns. Depending on these parameters, there can be two right-handed DNA structures: the A-

form and the predominant B-form. The left-handed form of DNA or the Z-form is not naturally found.

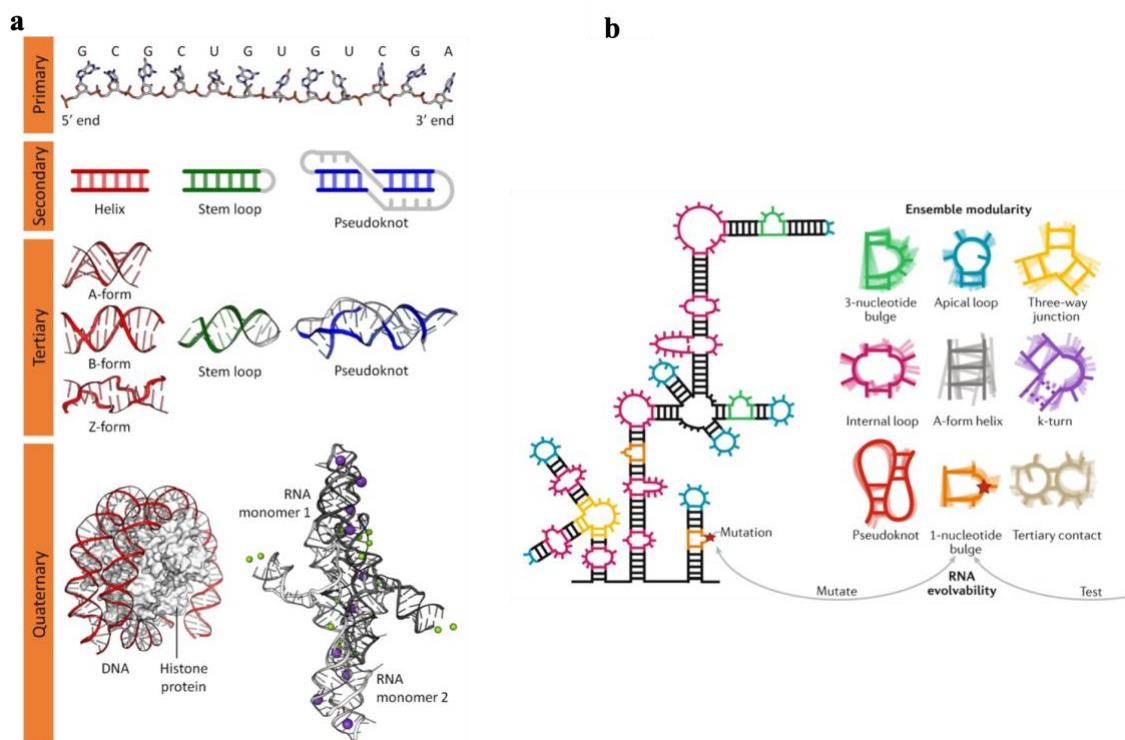
The second kind of interaction arises from the hydrogen-bonding interactions mediated by the ring nitrogen and the carbonyl and exocyclic amine functionalities of the nucleobases. The strongest and most stable pairing is found between the complementary bases of A and T (2 hydrogen bonds) and between G and C (three hydrogen bonds). This highly specific non-covalent interaction, known as the canonical Watson-Crick base-pairing, is the mechanism by which a unique DNA strand can form a fully base-paired double stranded structure with a perfectly matching complementary sequence of DNA or RNA. In this manner, any DNA-templated biological process such as replication and transcription can transmit the encoded message from the source DNA. Other interactions are possible, such as the non-canonical G-U wobble pair commonly found in RNA, and Hoogsteen base-pairing that can lead to the formation of triplexes (7-9).

## **1.2. The Structure-Function Relationship in RNA**

In contrast to DNA, RNA serves as much more than an inert vessel for genetic information and shows high structural as well as functional diversity. Chemically, RNA differs from DNA in that 1) it contains a 2'-OH instead of a 2'-H, and 2) it has the nucleobase uracil instead of thymidine. RNA is transcribed as a single-stranded polymer from the corresponding DNA template but can fold into complex shapes and structures. While RNA adopts a right-handed A-form helical conformation, it doesn't have a stable secondary structure, unlike the DNA double helix. Instead, RNA samples and adopts

multiple flexible conformations, facilitated by intermolecular and intramolecular WC base-pairing interactions. The 2'-hydroxyl group affords additional hydrogen bonding contacts and metal coordination sites. Additionally, due to nucleophilic attack of the 2'-hydroxyl group on the adjacent phosphodiester bond, RNA is susceptible to spontaneous hydrolytic cleavage. RNA is thus much less stable than DNA, probably one of the primary reasons why DNA emerged as the primary genetic material for higher organisms.

Some of the most ubiquitous secondary structural motifs found in RNA are hairpins, loops, internal bulges and pseudoknots. Hairpins are formed by self-complementary stretches of nucleotides forming a stable stem, capped by a short loop. Loops are flexible, consisting of a short stretch of unpaired nucleotides. Certain sequence motifs such as UUCG have been observed to form tight and stable loops in hairpins and are believed to facilitate RNA folding (10). GNRA tetraloops are also widely known to occur in RNA (11). Bulges are formed by unpaired nucleotides in an otherwise fully base-paired stem. The unique primary sequence of RNA thus directs specific intramolecular interactions to form different secondary structural motifs that can further engage in long range interactions to form a higher order RNA tertiary structure (12). Multiple possible conformations of a single RNA molecule can co-exist, and RNA function depends largely on stabilizing the active conformation in the equilibrium ensemble (12). Complex architecture of functional RNA often requires scaffolding on other nucleic acids and proteins; for example, ribosome or Ribonuclease P which are complex riboprotein complexes (13-16). These complexes, also referred to as quaternary structures, represent the highest level of structural organization in RNA.



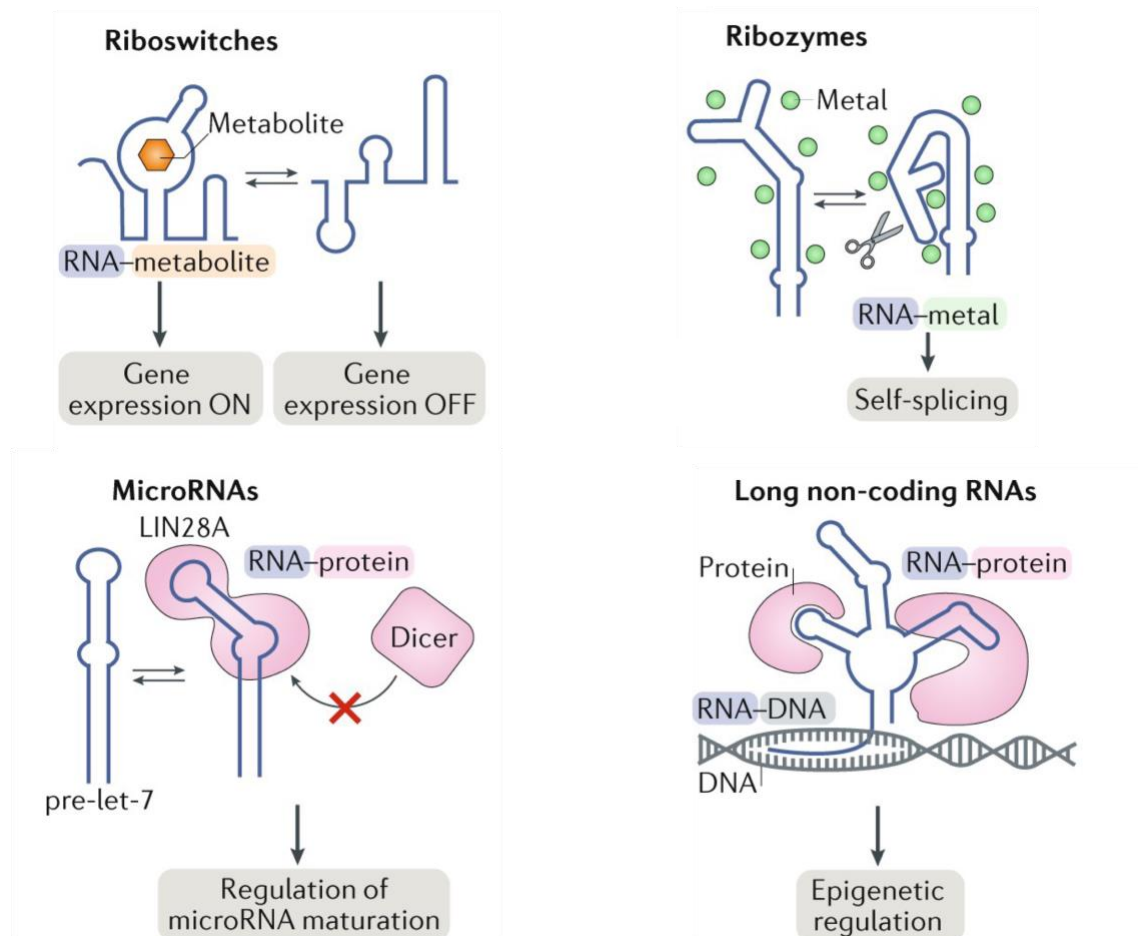
**Figure 1-3 Secondary Structures of RNA.** a) Hierarchical organization in RNA starting from the primary nucleotide sequence directing formation of simple secondary structure motifs, which interact to form higher order structures. Quaternary assemblies represent the highest possible complexity. Left structure shows DNA chromatin structure wrapped around histone proteins. Source: Wikipedia Commons. b) RNA secondary structures color coded and depicted in the context of a representative RNA molecule. Adapted from Ref (12). Copyright © 2019, Springer Nature Limited.

Modern structural biology has different tools at its disposal for the accurate elucidation of RNA structure and function. Traditional chemoenzymatic methods such as hydrolysis assays, ribonuclease mediated degradation assays or DMS profiling generally target unpaired residues and utilize the known specificities of these reagents to identify the nucleotide, thus generating a reliable structural map of RNA (17-19). Newer methods such as Selective 2' Hydroxyl Acylation analyzed by Primer Extension (SHAPE) targets



every 2'-OH position of any unpaired residue and is capable of generating highly accurate structures for long, complex RNA both *in vitro* as well as *in vivo* (20-21). Biophysical techniques such as X-Ray crystallography and mass-spectroscopic analysis can be used for certain targets, and Cryo-EM is slowly emerging as a powerful technique as well (22). In parallel, RNA sequencing technologies and bioinformatics have made great advances allowing high throughput analysis of multiple transcripts. Immunoprecipitation and cross-linking methods such as CLIP or capture-probe based methods such as CHIRP are being increasingly used in conjunction with high throughput sequencing to study the transcriptome (23). Decoding RNA function relevant to actual physiological conditions is thus becoming more accessible. Computational approaches are also exploiting this massive amount of data to refine their predictive algorithms (24). Machine learning is another powerful tool to be applied to problems related to RNA structure-function in recent times (25).

The diversity of RNA function parallels the complexity of RNA structure. There are two main classes of RNA: coding and non-coding RNAs. Coding RNAs are transcribed from gene segments in DNA as messenger RNAs (mRNAs) and encode the primary sequence of a protein that is synthesized during translation. Though relatively unstructured, the untranslated regions of mRNA called the 5'- and 3'-UTRs have highly structured motifs that play significant regulatory roles in gene expression, mRNA trafficking, localization, and stability (26). While the eukaryotic gene expression platform utilizes mainly protein factors for ribosome binding, the 5'-UTR of bacteria and both 5'- and 3'-UTRs of eukaryotes can also exert an allosteric control over gene expression in a



**Figure 1-4 Structure-Function Relationship in RNA.** Examples of RNA structure mediated biological functions some of which are discussed in the text of this section. Figure adapted from Ref (12). Copyright © 2019, Springer Nature Limited.

ligand gated manner via riboswitches that consist of an aptamer domain specific towards certain metabolites such as SAM or TPP. Ligand binding induces a conformational transition to control gene expression by either enhancing/ inhibiting ribosomal access and binding or by cleavage of the transcript (27,28).

RNAs that do not code for any protein form the bulk percentage of the transcriptome and belong to the class of non-coding RNAs. Non-coding RNAs are highly

heterogeneous in biogenesis, structure, and functionality. The most familiar and well-studied non-coding RNAs are the highly structured transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), primarily responsible for protein synthesis. During translation, a tRNA carrying a specific amino acid, binds to the mRNA using the complementary “anticodon” sequence. The corresponding amino acid then gets recruited to the growing polypeptide chain on the ribosome.

Although the ribosome is a highly structured riboprotein complex, ribosomal RNA is the primary catalytic component involved in peptide bond formation (13-15). Catalytic RNAs are extant in both lower and higher organisms but are more limited in scope and efficiency when compared to protein catalysis.

Post-transcriptional regulation of gene expression is a common function of small non-coding RNAs such as miRNAs, shRNAs and piRNAs. For example, miRNAs can have multiple target mRNAs and often operate as part of a cluster (29,30). Since miRNA can be overexpressed in cancers, their detection and profiling are becoming increasingly crucial to cancer diagnosis (31-33).

Any non-coding RNAs of length more than 200 nucleotides are termed as long non-coding RNAs or lncRNAs and constitute a highly diverse group of RNA. These long RNA molecules are highly structured and engage in diverse functions such as chromatin remodeling and transcriptional control of gene expression (34). Research in this area lies at the frontiers of RNA biology, and new lncRNAs with novel functions are being discovered and annotated every day. Overall, the multifaceted nature of RNA structure and function has enriched our understanding of gene expression.

### **1.3. Ribozymes - Nature's earliest catalysts**

Ribozymes or RNA molecules capable of catalyzing chemical reactions, were first discovered in 1981 by Thomas Cech and Sidney Altman, an accomplishment that earned them the Nobel Prize (16,35). Unlike protein enzymes, the primary catalysts of modern biology, ribozymes have a limited source of chemical functionality in its four nucleobases. It is therefore unsurprising that ribozyme catalysis is restricted to mainly phosphotransferase reactions in the cleavage and ligation of RNA substrates (16,35-38). The peptide bond formation during protein synthesis is also catalyzed by the RNA component of the ribosome (13-15). The complex architecture of ribozymes and the intramolecular interactions between the secondary structural motifs are responsible for aligning the reaction site of the substrate RNA within the catalytic fold of the ribozyme. Often, these ribozymes are part of a more complex riboprotein complex that play a role in stabilizing the global architecture and facilitating interactions. RNase P involved in tRNA processing and the ribosome are two such examples. Although nature utilizes ribozymes for only a limited type of reactions, artificial selection has identified a broad range of catalytic properties in ribozymes, such as alkylation, acylation, Diels-Alder reaction, RNA polymerization etc (39-41). This underlines that the lack of functional diversity of ribozymes in extant biology is shaped more by evolutionary needs than inherent chemical limitations.

### **1.3.1. Origin of Life: The RNA World Theory**

Life as we know it, would have originated as a self-sustaining machinery capable of preserving and transmitting heritable genetic information while catalyzing biochemical processes. The origin of life and questions involving the prebiotic world have always captured the imagination of scientists. The RNA world theory posits that RNA predated DNA or proteins as the ancestral biomolecule (42). Discovery of ribozymes lent more credence to this theory. It is possible that in the prebiotic era, under the prevalent environmental conditions, RNA monomers came to exist and were capable of non-enzymatic polymerization to form short RNA oligomers which in turn ligated with each other to generate higher complexity and diversity (43). Eventually, weak catalytic properties emerged, with ribozymes finally acquiring self-replication, which is key to a sustainable life-system (44,45). In addition, these primitive catalytic molecules, though low in efficiency, were probably promiscuous in their choice of substrates and catalyzed more than one reaction (46). For example, some naturally occurring ribozymes such as the hairpin ribozyme catalyze phosphodiester bond scission, but at lower temperatures, the reverse reaction of ligation predominates (38). This generality would allow the same ribozyme to catalyze multiple essential reactions of life. Eventually, some evolutionary pressure would have led to the emergence of protein enzymes that outcompeted ribozymes in the long run with their greater chemical diversity and catalytic efficiency.

### **1.3.2. RNA cleaving DNazymes and Ribozymes**

One of the most well-studied and characterized classes of ribozyme catalyzed reactions are RNA strand scission reactions (47-49). The phosphodiester bond cleavage

usually occurs at a specific site, resulting from nucleophilic attack at the phosphate center that forms a 2'-3'- cyclic phosphate and a 5'-hydroxyl terminus (48-49). The cyclic phosphate can further hydrolyze in certain situations. The nucleophile is usually the 2'-hydroxyl adjacent to the scissile phosphate in self-cleaving ribozymes, but in more complex systems such as the RNase P, engages in an intermolecular attack at the cleavage site on the substrate. Found often in lower organisms and viral pathogens, the self-cleaving ribozymes such as hammerhead, hairpin, twister, pistol and hatchet etc. employ either general acid-base catalysis to deprotonate the nucleophilic hydroxyl to activate it or use metal cations for the same purpose (50-51). The structural and mechanistic diversity in these self-cleaving ribozymes together with their steady evolutionary maintenance in both prokaryotic and eukaryotic systems indicate that they can adapt to different physiological conditions. High resolution crystal structures have made it possible to identify the structural requirements of many of these ribozymes (52-53). This not only advances our understanding of ribozyme catalysis but also helps in the development of biosensors and other biotechnology tools based on these ribozymes. For example, most of these ribozymes have highly complex architectures *in vivo*, but can be truncated to minimal catalytic motifs without compromising the activity (51). This is ideal for repurposing ribozymes as regulators of gene expression. Although primarily *cis*-cleaving in nature, many of these ribozymes can be reprogrammed for *trans*-cleavage by separating the scissile bond containing well-defined secondary structure from the rest of the ribozyme (54). This usually results in a generalizable platform compatible with any target RNA. Often, metabolite or small molecule specific aptamer domains are connected by means of

a communication module, to the active ribozyme module, and precise conformational changes induced by ligand binding exerts an allosteric control of RNA cleavage (55,56). Hammerhead is a very well-known motif among self-cleaving RNA and has been used in therapeutic designs for *in vitro* applications involving cleavage of physiologically relevant RNA (54). However, *in vitro* selection could evolve completely new catalytic domains in trans-cleaving ribozymes that could serve the same purpose (57).

Although native DNA is primarily double stranded thus lacking folding and catalytic ability, artificial DNAzymes have been evolved by directed evolution. The most well-known class of DNAzymes are perhaps the  $Mg^{2+}$  dependent 10-23 DNAzymes. These selections are usually designed to have a single unpaired ribonucleotide in its RNA substrate while the rest is hybridized to the substrate binding arms of the DNAzyme (58-60). The easy programmability of adapting DNAzyme designs to RNA substrates by changing the substrate binding arms accordingly, have been used in numerous *in vitro* applications for inactivating pathogenic RNA (58). However, their practical utility is limited by biostability issues. In recent times, significant progress has been made in this area by use of XNA modifications with promising *in vivo* applications, and we can expect a revival of interest in RNA cleaving ribozymes and DNAzymes as potent biotechnology tools in the coming days (61,62). In fact, one DNAzyme against an oncogenic target is currently in clinical trials in the US (63,64).

## **1.4. RNA Therapies for RNA Targets**

### **1.4.1. RNA as a Target**

Proteins are primary targets for most of the drugs currently available. However, only a fraction of the human genome encodes for proteins of which approximately only 2000-3000 are considered to be druggable (65). The bulk of the RNA transcriptome is non-coding and plays key regulatory roles in cellular processes. During its lifetime, RNA undergoes multiple biochemical transformations through synthesis, processing, and post-transcriptional modifications, all of which rely heavily on both sequence and structural motifs. Hence, any aberrations in RNA sequence and structure manifest in significantly altered RNA biosynthetic and metabolic pathways and by extension RNA function, which ultimately lead to diseased states.

Sequence mutations in pre-mRNA transcripts can significantly alter the secondary structure and affect co-transcriptional processing and splicing. Aberrant splicing has been implicated in many liver diseases including cancer, as well as neurological disorders such as dementia and Parkinson's Disease (66,67). Improper ADAR-mediated RNA editing arising from sequence mutations strongly correlates with several breast, lung, liver and esophageal cancers (66). Synonymous mutations in the coding region of the mRNA preserves the nature of the encoded amino acid; however, an unintended replacement can still lead to protein misfolding and aggregation (68,69). Non-synonymous mutations have more severe consequences, often leading to production of toxic proteins, or deletion of key motifs as well as truncated proteins (69). Misfolded or destabilized structural motifs



in non-coding RNAs can disrupt interactions with effector molecules, which in turn lead to dysregulated gene expression. Similar correlations can be drawn in lower organisms. Pathogenic bacteria and viruses abundantly use relatively simple secondary structural motifs such as the TAR hairpin in HIV-1 and the IRES domains in HCV to regulate gene expression (70-72).

Associations between diseased states and functional RNA counterpart(s) are becoming more accurate and accessible in this modern era of RNA biology, aided by revolutionary advances in high throughput sequencing as well as experimental and computational methodologies. It is therefore no surprise that the scientific community is increasingly turning to the RNA transcriptome as a potent alternative source of druggable therapeutic targets.

#### **1.4.2. Strategies for targeting RNA**

RNA is very commonly targeted with small molecule drugs that bind secondary structures to inhibit function. While small molecule drugs are highly effective and have their own advantages in terms of bioavailability, ease of delivery and reliable technique transfer from protein targeted drug design, non-specific binding and related toxicity often offsets the benefits (65,73). Small molecule drug discovery relies on a combination of biochemical screening, *in silico* methods of rational design and lastly, experimental validation of binding and RNA function inhibition/ manipulation. Almost inevitably, each new RNA target requires iterative processes of optimization of drug design by trial and error to suit the particular situation.

In contrast, the accurate, predictable, and programmable nature of WC base-pairing mediated hybridization makes it conceptually straightforward to design and tailor a DNA/RNA therapeutic molecule against the desired target. Although secondary structures and complex RNA architectures might make it difficult to access the target area by simple hybridization-based probes/ therapeutics, structure-based interactions can be exploited to explore diverse and novel recognition modalities and mechanisms of action. Additionally, the ease of automated solid-phase synthesis as well as enzymatic routes to access nucleic acids carrying both natural and unnatural modifications, have enabled routine use of nucleic acids-based strategies and made the validation and optimization stages more cost and time-effective.

#### **1.4.2.1. Oligonucleotide Therapies for targeting RNA**

As of 2021, there are at least 13 FDA approved oligonucleotide-based drugs, the latest being a splice-modulating ASO drug for Duchenne muscular dystrophy, called casimersen marketed by Sarepta Therapeutics and hundreds of other drugs in clinical trials and the development pipeline (64,65,73). In the following few paragraphs, I will be outlining the most well-established oligonucleotide-based strategies for targeting RNA.

*Antisense Strategy:* Antisense oligonucleotides (ASOs) are short (usually not longer than ~30 nucleotides), synthetic, single-stranded DNA oligonucleotides that are designed to have complementary sequence to the target region of a particular mRNA. Upon hybridization, ASOs form a DNA-RNA hybrid that recruits RNase H, followed by degradation of the target transcript (73). Since this mechanism is prevalent in both the

nucleus as well as the cytoplasm, nuclear pre-mRNA transcripts can also be targeted. The first FDA approved oligonucleotide drug (1998) fomivirsen is an ASO of this type, designed for the treatment of cytomegalovirus retinitis (73). Steric block ASOs do not form RNase H substrates with target transcripts but either inhibit gene expression by steric inhibition of translation initiation sites or ribosomal binding sites on the target mRNA or blocking of polyadenylation sites to enhance mRNA decay. AntagomiRs or miRNA decoys operate in a similar way, using sequences complementary to the short mature miRNA sequences that can sequester the said miRNA and thus inhibit their function, for example miravirsen, that is complementary to mature miR122 whose function is integral to HCV infection.

Steric block ASOs are also capable of blocking target introns and exons to modulate splice junctions and sites. Alternatively, some ASOs even enhance gene expression by binding upstream of ORFs and destabilizing structures that negatively regulate gene expression. Since antisense ONs are recognized as foreign DNA, they are highly susceptible to degradation and are often heavily modified to have a phosphorothioate backbone or designed to have short DNA stretches or “gaps” flanked by nuclease resistant LNA modifications (for example in GapmRs), to increase stability as well as binding affinity.

*RNA Interference (siRNAs and shRNAs):* This strategy uses short RNA duplexes, consisting of a passenger strand or the sense strand and the guide strand or the strand complementary to the target. The siRNA duplex directs the Argonaute complex (Ago2) to

the target transcript, where the passenger strand is released, and the guide strand-target mRNA duplex is processed to result in target cleavage and degradation. This silencing pathway is the same as followed in miRNA mediated mRNA processing. Alternatively, shRNAs that mimic endogenous miRNAs can be expressed *in situ*, to ultimately be processed into an siRNA duplex. Most used modification in this case is the 2'-OMe on the ribose to increase stability and alleviate some of the immunogenic responses. Two FDA approved drugs based on this design are patisiran and givosiran targeting amyloidosis and hepatic porphyria respectively.

**DNAzymes and Ribozymes:** ASOs can be repurposed to incorporate a catalytic domain that can be adopted either from naturally occurring ribonuclease ribozymes such as the hammerhead ribozyme, or artificially selected DNAzymes such as the 10-23 and 8-17 (58,59). In the late 1990s and early 2000s, there was encouraging use of the hammerhead and the hairpin ribozymes in regulation of gene expression in certain pre-clinical diseased models with high efficiency; however, this strategy hasn't been as successful as other oligonucleotide-based therapies due to poor delivery, high dosage requirements, need for higher frequency of administration and toxicity.

The 10-23 DNAzyme that is highly generalizable and has been adapted to a variety of physiologically relevant RNA targets such as the HIV *gag*, *pol* mRNAs in *in vitro* studies, is a good therapeutic candidate, but suffers the obvious practical issues of product inhibition and nuclease degradation (63,64,74). Of late, there has been renewed interest in research looking to increase the efficiency of such DNAzymes in intracellular

environments in gene silencing using XNA modifications to overcome many of the existing challenges (61,62). Moreover, one DNAzyme-based drug against an oncological target is also in clinical trials. Hopefully, with newer developments in drug delivery and stability, this field will soon be revived.

**Other RNA Therapies:** For the sake of brevity, I have limited the discussion to the most prevalent and well-known RNA based therapies that work on RNA targets specifically. However, I would like to close by mentioning some alternative therapies that either have non-nucleic-acid targets, or work in combination with protein components.

One such example would be Pegaptinib, an RNA aptamer that binds a growth factor responsible for macular degeneration with high specificity and affinity (65,75). It's the only FDA approved RNA aptamer-based drug. In light of the ongoing COVID-19 pandemic, mRNA-based vaccines have not only been developed with unprecedented speed but have already saved millions of lives. Other mRNA-based approaches involve production of functional proteins for restoration of normal physiological levels. CRISPR-Cas systems have been another revolutionary progress in biomedicine in the past few years that now affords precise and accurate control over gene editing (65).

#### **1.4.3. Advantages and Challenges associated with RNA based therapeutics**

Compared to alternative strategies for targeting RNA such as small molecules and biologics, RNA based therapies afford many advantages. The precision of WC base-pairing offers a straightforward design principle as opposed to rigorous trial and optimization protocols required for other strategies. Additionally, a knowledge of the

specific cause of pathology (a mutation, an aberrant splice site) can be addressed with more precision and drugs can be tailored accordingly, which can even be extended to individual patients for personalized medicine. Although delivery of RNA based materials is tricky, some strategies such as RNAi are catalytic and show high efficiency, especially since it exploits the endogenous RISC silencing machinery, thus balancing out poor uptake. Commercially available phosphoramidites, especially specialty modifications are becoming more user-friendly and cost effective, which is making it possible to broaden the scope of such drugs. This also precludes the potential variability of biologics.

For all the benefits of RNA based therapies, there are some key challenges as well. Exogenous DNA or RNA are readily attacked by endogenous nucleases and cleared by renal filtration, thus limiting bioavailability. As a reference, short single-stranded components have a half-life on the order of minutes (76). Some strategies adopted to enhance stability include incorporating chemical modifications, increasing the complexity of the system such as tethering to nanostructures that are more stable towards nucleases, or by protection of vulnerable free ends by circularization, or end-capping, or installing hairpins (77-79). Of these strategies, chemical modifications will be discussed in detail in the next section.

As with other therapeutics, proper delivery of the drugs to the target location is a significant challenge. The most popular and commonly used method of cellular delivery is perhaps liposomal delivery by endocytosis (80,81). The main challenge in this case is often insufficient endosomal escape, which must be counterbalanced by delivery of high

amounts of the drug that could lead to toxicity. However, more refined technologies are being applied these days, such as conjugation with lipids, receptor targeting aptamers, cell-penetrating peptides etc. that can lead to more targeted delivery, or encapsulation in nanocarriers such as DNA nanostructures, lipid-based nanoparticles or exosomes (82-85).

Non-specific hybridization related toxicity is another major concern. Interaction with other cellular components such as protein effectors could similarly prove counterproductive. Exogenous and relatively short DNA as a “foreign material” stimulates the innate and adaptive immune system (86,87). Short, single-stranded DNA have been found to interact with multiple receptors including Toll-like receptors that can lead to inflammatory responses of varying severity (88). This is a major issue for the *in vivo* application of nucleic acids-based strategies. One of the simplest examples is perhaps the immune response to unmethylated CpG that originates in bacteria and is thus considered to be a threat (88,89). Additionally, many nucleic acid modifications that are essential to stability and target affinity, elicit strong immune responses which undermine their pharmacological benefits (87,90-91).

Finally, the effectiveness of the nucleic acids-based approach depends on the accessibility of the target RNA and highly structured regions are not very good candidates for hybridization-based probes. Additionally, RNA structural motifs and specific conformations serve as functional footprints, especially in non-coding RNAs that are constantly emerging as crucial players in most biochemical processes. Targeting specific structural features without relying too heavily on sequence would potentially mitigate

some of the off-target hybridization mediated toxicity and expand the “druggable” transcriptome.

### **1.5. Common Chemical Modifications in Nucleic Acids Based Strategies**

Incorporation of chemical modifications can improve biostability of DNA/RNA based therapies. These modifications can target different functional groups of the nucleotides such as phospho-sugar backbone or the 2'-hydroxyl group or the nucleobase itself. The most common modifications that are extensively used for intracellular and *in vivo* applications are discussed below.

#### **1.5.1. Phosphate Backbone Modifications (Phosphorothioate)**

This modification is achieved by replacing one of the non-bridging oxygens of the phosphodiester bond by sulfur. The advantage of phosphorothioate (PS) is twofold. First, the stability of the modified oligonucleotide is much enhanced, thus increasing bioavailability (92). Second, phosphorothioate is known to interact with many cellular receptor and plasma proteins thereby promoting uptake and circulation (93). This modification is compatible with RNase H degradation pathway and has been used generously in ASOs including gapmers. The major disadvantage is the reduced binding affinity of PS as well as slower hybridization kinetics. This can however be compensated by introduction of other modifications or by optimizing the number and positions of the modification. An oligonucleotide heavily modified with PS can also elicit a strong immune response which would require some changes in design to offset the negative effects. Another interesting feature is that PS can generate two stereoisomers  $R_p$  or  $S_p$



depending on which oxygen is replaced. This has implications in the melting temperature as well as nuclease resistance. Although not routine, it is possible to obtain PS modified oligonucleotides with predetermined stereochemistry. It is however not a guarantee of enhanced efficacy. One such stereopure drug was recently discontinued after no significant activity in a clinical trial. The stereoisomers thus exert an overall effect by maintaining a balance between stability, binding affinity and toxicity.

### **1.5.2. Ribose Modifications**

Locked nucleic acid is a bicyclic ribose analog consisting of a methylene bridge connecting the 2'-oxygen and 4'-oxygens. This modification renders high stability in terms of nuclease resistance and increases the melting temperature of duplexes formed with RNA, by locking the ribose in a C3'-endo pucker conformation (94). The melting temperature can be enhanced by as much as 10 °C for every LNA added (95). They are most abundantly used as end or capping modifications in ASO designs such as gapmers where as few as 2 or 3-nucleotides of LNAs in the termini can protect the internal DNA, usually PS modified, from degradation. Tiny LNAs or very short LNA ONs are also used to mask miRNA binding sites and are capable of simultaneously targeting multiple miRNAs with functional redundancies.

Another common site for modification is the 2'-OH on the ribose. 2'-*OMe*, 2'-*O*-methoxyethyl and 2'-*F* are extensively used to enhance nuclease resistance (73). Incompatible with RNase H mechanism, these modifications are more commonly used, sometimes in combination, in siRNA designs. For example, the siRNA-based drugs

patisiran and givosiran both have 2'-*OMe* modifications, and the RNA aptamer-based drug has a combination of 2'-*OMe* and 2'-*F*. Similar to LNA, 2'-*OMe* affords high thermostability especially in RNA duplexes, by adopting a C3'-endo pucker (96). The specificity of binding and hybridization kinetics are also highly favorable. Additionally, the low cost of synthesis and compatibility with enzymatic methods of incorporation make this modification a popular choice.

### 1.5.3. Alternative Backbone

Contrasting to PS, LNA and 2'-*OMe*, PNA is a modification that was designed *in silico* as a nucleic acid analogue preserving the structural parameters of DNA as accurately as possible (97). Originally intended to have anti-gene applications by forming a stable triplex with a regular DNA helix, PNA has since proved useful in an array of versatile applications such as antisense reagents and in gene editing. Instead of the phosphor-sugar backbone, PNA has a pseudo peptide backbone composed of *N*-(2'-aminoethyl)-glycine monomeric units, with the canonical nucleobases connected to the glycine nitrogen by a methylene carbonyl linkage (98). Apart from having an uncharged backbone that affords high thermodynamic stability of duplexes formed with DNA, RNA or other PNA molecules under different ionic strengths, PNA is also achiral and is virtually invisible to nucleases, proteases and peptidases (99-101). However, some of the issues are poor solubility, high chances of aggregation and intramolecular structures and poor uptake. However, hydrophobic and hydrophilic modifications at  $\beta$  and  $\gamma$  positions can help alleviate a lot of these problems.

#### **1.5.4. Ribose Analogues**

“Xeno” nucleic acids (XNA) are five or six membered congeners of the (deoxy) ribose moiety equipped with the canonical nucleobases (102). They are bioorthogonal, highly resistant to nucleases and often afford high stability of binding. Owing to the advances made in the development of polymerases that can accept different XNA monomers (103), in addition to post-synthetic modifications many functional molecules such as aptamers and enzymes having XNA backbones can be directly evolved (104,105). This is an emerging area of research with exciting implications for antisense and siRNA therapeutic applications.

Overall, use of these modifications can improve the pharmacological properties of therapeutic ONs when optimized and characterized thoroughly. Some of the other aspects pertaining more to the hybridization thermodynamics and kinetics will be addressed later in the introduction in the context of strand displacement.

#### **1.6. L-DNA as a Stereochemical Modification: Properties and Applications**

An ideal modification would be extremely resistant to endogenous nucleases, elicit minimal immune response and avoid crosstalk with endogenous biomolecules. Ease of delivery and enhanced binding affinity (for WC hybridization-based designs or binding based inhibition or activation strategies) are desirable properties as well. However, the major issue that compromises intracellular and *in vivo* efficiency of nucleic acids-based sensors, therapeutics and other devices is still poor biostability.

L-DNA or L-RNA, the enantiomeric form of native D-DNA or D-RNA, does not occur naturally. Their inverted chiral centers compared to native nucleic acids make these mirror-image DNA and RNA virtually invisible to endogenous nucleases (106-108). Unless a chiral environment is provided, such as circularly polarized light or a stereospecific protein enzyme, enantiomeric nucleic acids behave identically (109). L-DNA or L-RNA have the same physical properties as their native counterparts such as solubility, spectroscopic properties, hybridization kinetics and thermal stability of duplexes (106,107,109). This is especially useful since a homochiral L-DNA duplex will have the same thermodynamic and kinetic parameters of formation as its D-DNA counterpart despite forming an unnatural left-handed helix. This means that the same rules of WC base-pairing apply to the L-DNA duplex, and design principles can be first tested with the more easily accessible D-DNA duplex for certain applications. For example, DNA nanotechnology applications involving DNA origami structures can be constructed using either D-DNA or L-DNA based on the same optimized design. Due to the “chiral clash” between the opposite helicities however, L-DNA cannot form contiguous base-pairs with D-DNA (109-111). While this precludes the use of L-DNA in applications requiring hybridization to endogenous nucleic acids, this also helps avoid off-target hybridization related toxicity and immunogenic responses. From a synthetic perspective, standard phosphoramidite chemistry developed for regular D-DNA or D-RNA can be easily applied to the synthesis of L-DNA and L-RNA, and the cost of commercially sourcing L-nucleotides or L-oligonucleotides have decreased dramatically in the last few years.

For all these superior qualities biotechnological applications of enantiomeric nucleic acids were restricted to antiviral agents against diseases such as HBV, HCV and HDV for a long time. This is because the primary motivation in developing nucleic acids-based tools back then was for antisense applications and although early modeling and experimental studies suggested the formation of heterochiral complexes between homopurine or homopyrimidine L-DNA or RNA and their complementary natural counterparts, these complexes were consistently less stable and less predictable than native DNA or RNA duplexes. Further work with L-DNA and RNA oligonucleotides comprising all four bases in different permutations established clearly that enantiomeric nucleic acids are incompatible with canonical WC base-pairing interactions with native nucleic acids, thus precluding their use as antisense reagents (110). Initially this discovery curbed interest in this area; however, scientists quickly found that this very obstacle can be an asset in biological applications using interfacing strategies beyond traditional hybridization approaches.

L-DNA has found use in several exciting technologies with applications in research and medicine. Identical interactions of enantiomers with achiral entities make it possible for metal responsive DNAzymes to be synthesized with L-DNA backbone as a highly stable biosensor. Since L-DNA duplexes exhibit all the properties of D-DNA duplexes, a molecular beacon has been repurposed to a molecular thermometer. Additionally, L-DNA have also found uses as bio-orthogonal tags on native DNA for genotyping purposes. However, the most successful application of L-DNA and L-RNA to date, has been in the development of “mirror-image” aptamers also called “spiegelmers”. In general, aptamers

are single-stranded oligonucleotides that by virtue of their ability to fold into complex shapes, can be evolved through an iterative enrichment-amplification process called SELEX (systematic evolution of ligands by exponential enrichment) to bind a target (a small molecule, metal ion, nucleotide, amino acids, proteins enzymes, whole cells etc.) with high specificity and affinity (112-114). Since enantiomeric oligonucleotides cannot be amplified or manipulated by bioenzymatic methods usually employed in *in vitro* selections, a “mirror-image selection” procedure was developed in the early 1990s whereby an aptamer comprised fully of a L-DNA or RNA backbone can be selected against technically any target whose enantiomer can be synthesized. First, a D-aptamer is selected against the enantiomer of the intended target following which the exact sequence of the D-aptamer is synthesized using L-nucleotides for use against the natural target. According to the rules of chiral symmetry originally proposed by Pasteur and Fischer, a spiegelmer must bind to the original target with identical affinity with which the D-aptamer binds to the enantiomeric target. The first native targets that were used for mirror image selection of L-RNA aptamers were adenosine and arginine (115,116). Since then, the repertoire of Spiegelmers have grown, and several therapeutically relevant proteins have been accessed by this technology (117-119). In addition to high specificity, Spiegelmers show excellent pharmacological properties. An L-RNA aptamer selected against barnase, an RNase enzyme was shown to inhibit degradation of native D-RNA by binding to barnase, while being inert to the enzyme itself (120). This only goes to show the incredible stability of Spiegelmers and their potential as therapeutic reagents. Indeed, several Spiegelmer therapeutics are currently in phase I or phase II of clinical trials in the US.

## **1.7. Interfacing Native RNA with L-DNA/RNA**

### **1.7.1. Structure-based Binding Interactions between Native D-RNA and L-Aptamers**

Spiegelmer targets have been traditionally small molecules, peptides or proteins. RNA targets have been less explored in this sense despite their value as therapeutic targets. Previous work by Toulme and Famulok showed that in vitro selection of D-RNA aptamers against D-RNA targets such as the TAR hairpin yields aptamer-target binding based on “kissing-loop” interactions (121,122). Canonical WC base-pairing thus still dominates substrate recognition, eliminating the possibility of targeting the specific RNA sequence in its structural context. As I have elaborated earlier, structured RNA serves as valuable players in regulation of gene expression and cellular functions, and the intricate structure-function relationship makes it worthwhile to develop affinity reagents targeting functional structures as opposed to sequence alone.

Cross-chiral interactions between D-DNA/RNA and L-DNA/RNA involve non-canonical hydrogen bonded interactions as well as the general Van der Waals, electrostatic and hydrophobic interactions. In the absence of WC interactions, these cross-chiral non-covalent interactions can drive the evolution of highly efficient and specific aptamers against structured RNA. This was demonstrated for the first time in 2013 by Szczepanski and Joyce. An L-RNA aptamer selected against the native HIV-1 D-TAR RNA element showed binding capabilities potent enough to disrupt interactions of TAR with Tat protein under simulated physiological conditions. The repertoire has grown over the years with several L-RNA aptamers and more recently an L-DNA aptamer against different D-

miRNA targets. In all these cases, there are no appreciable sequence complementarity between the substrate-aptamer sequences and although interactions are mediated through mostly unpaired nucleotides, recognition is based on the structural signatures of the target rather than sequence.

### **1.7.2. Heterochiral Interactions in Catalysis of Chemical Reactions**

Ribozymes or catalytic RNA utilizes the chemical functionalities of its ribose hydroxyls or the exocyclic amines and carbonyls of the nucleobases to catalyze mainly phosphotransferase reactions and peptide bond formation in nature. Since enantiomeric nucleic acids retain these chemical functionalities, it's not unexpected that heterochiral interactions could result in reaction catalysis. At the same time, it is highly challenging since heterochiral ribozymes cannot rely on contiguous WC interaction for substrate recognition that is key to functioning of many natural ribozymes such as the HHR. Rather, the heterochiral interactions need to be strong enough for adequate substrate binding that could result in a successful reaction.

The first heterochiral ribozyme was reported in 2014 by Sczepanski and Joyce. This L-ribozyme can carry out templated ligation of two D-RNA substrates or polymerization of D-nucleotides on a D-RNA-primer complex (123). This discovery is a crucial step forward in L-nucleic acid research because 1) it demonstrates that not only can enantiomeric RNA interact to form complexes, but these interactions can also lead to chemical transformations mediated by an RNA molecule on its enantiomeric form. 2) It strengthens the argument for a prebiotic RNA world and presents a scenario for the evolution of homochirality in nature. It is possible that enantiomeric nucleic acids



coexisted at some point of time, and ribozymes could replicate their own sequence in the opposite handedness, but over time, evolution of protein enzymes led to a simplified homochiral biology. 3) this ribozyme provides a practical tool to assemble long L-RNA molecules by using the D-form of the ribozyme to catalyze ligations between multiple L-RNA fragments. This is especially important since evolution of protein enzyme catalyzed amplification and transcription of L-DNA to form L-RNA is still at a nascent stage (124,125). A more evolved version of these 83 nucleotides long cross-chiral RNA polymerase has been recently reported (126). This L-ribozyme shows enhanced polymerization activity with trinucleotide building blocks and more generalizability compared to the original ribozyme. Further evolution of this class of ribozymes could result in a true RNA based cross-chiral replicating system.

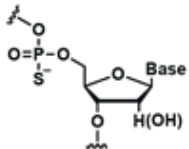
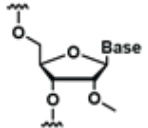
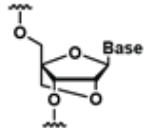
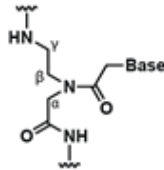
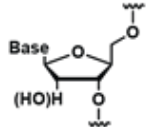
### **1.7.3. Sequence-specific Interfacing by Heterochiral DNA Strand Displacement**

#### **Reactions**

Detection and manipulation of RNA sequence is of great practical utility. For example, mature miRNAs are short RNA sequences that are implicated in a variety of cancers as well as other diseases. Abnormalities in mRNA sequences are also detrimental to protein synthesis, folding and aggregation (69). Sequence information is therefore critical to our understanding of diseased states as well as diagnostic and therapeutic development. In recent years, dynamic DNA nanodevices have seen a flurry of innovation and versatile applications in biomedical research. By made of primarily DNA/RNA, and

often modified by nucleic acid analogues, these devices are biocompatible in the sense that they can interact with the biological environment and report on its molecular state.

**Table 1 Comparison between Common Chemical Modifications to D-DNA.** Advantages and disadvantages of their use in DNA strand displacement reactions. Reprinted from Ref (131) with permission. © 2021 Wiley Periodicals LLC.

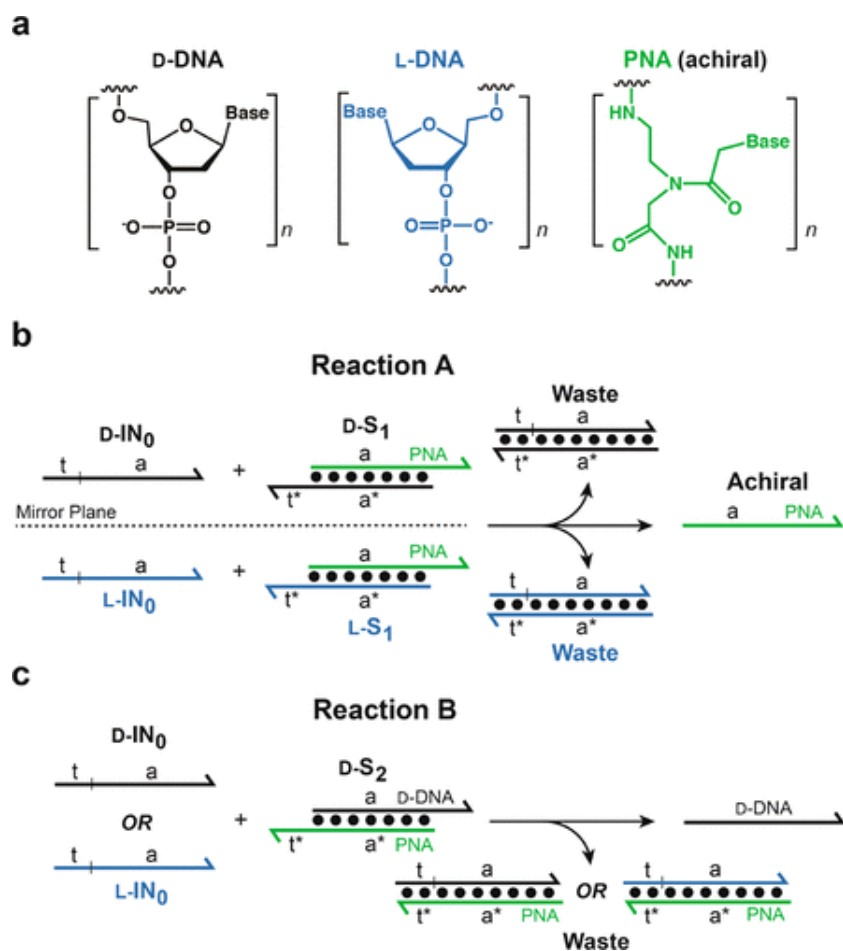
	 Phosphorothioate (PS)	 2'-O-methyl (2'-OMe)	 Locked Nucleic Acid (LNA)	 Peptide Nucleic Acid (PNA)	 L-(Deoxy)ribose (L-DNA/RNA)
Nuclease resistance	Partial, isomer dependent	Partial	Complete	Complete	Complete
Duplex Thermostability <sup>a</sup> ( $\Delta T_m$ /bp)	Decreased ( $\sim 0.5$ °C)	Increased ( $\sim 1$ °C)	Increased (2–8 °C)	Increased ( $\sim 1$ °C)	Identical
Toxicity	High	Moderate	Moderate	Low	Low
Commercial Cost <sup>b</sup> (\$USD/bp)	3.5 <sup>d</sup>	17 <sup>d</sup>	20 <sup>d</sup>	58 <sup>e</sup>	185 <sup>f</sup>
In-House Cost <sup>b,c</sup> (\$USD/bp)	0.5 <sup>g</sup>	1 <sup>g</sup>	1.5 <sup>g</sup>	2.5 <sup>h</sup>	2.7 <sup>g</sup>

<sup>a</sup>Relative to unmodified DNA  
<sup>b</sup>Commercial cost of native DNA (\$USD/bp): 0.7; in-house cost of native DNA (\$USD/bp): 0.25  
<sup>c</sup>Prepared using an Expedite 8909 DNA/RNA Synthesizer using manufacturer recommended protocols  
<sup>d</sup>Estimated cost if purchased from Integrated DNA Technologies, USA  
<sup>e</sup>Estimated cost if purchased from Panagene, South Korea  
<sup>f</sup>Estimated cost if purchased from Bio-Synthesis, USA  
<sup>g</sup>Estimated cost using monomers and reagents purchased from Glenn Research, USA  
<sup>h</sup>Estimated cost using monomers and reagents purchased from PolyOrg, USA

The well-known thermodynamics and kinetics of DNA hybridization, and the reliable programmability of WC base-pairing are instrumental to the enormous potential of such devices. Nevertheless, the intrinsic poor biostability of exogenously introduced DNA is the main deterrent in the successful applications of such devices *in vivo*. Use of chemical modifications is a useful strategy to significantly enhance biostability as well as improve thermostability of formed duplexes. However, thermodynamic and kinetic parameters can be significantly altered by these modifications which can convolute the desired interactions between the different modules of these devices inside cells (127,128). L-DNA/RNA would be the ideal choice for such applications because of its incredible

stability as well as unaltered thermodynamic and kinetics of hybridization. Its incapability of WC-base pairing with native D-DNA/RNA has however precluded its use in such DNA nanodevices.

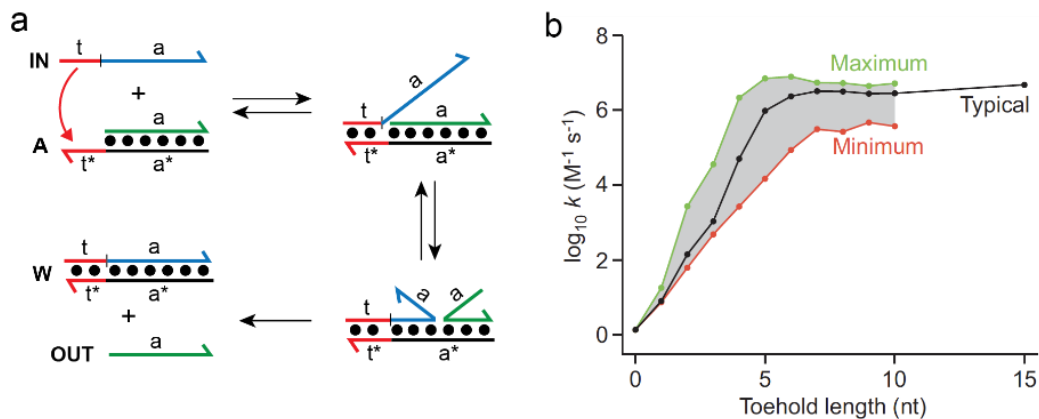
Our research group recently devised a heterochiral strand displacement strategy to alter the homochiral paradigm in dynamic DNA nanotechnology and lay the foundation for a second generation of exceptionally biostable devices with L-DNA/RNA backbone (129). Toehold mediated strand displacement reactions underline a vast majority of dynamic DNA nanodevices. They utilize single-stranded complementary domains on the invading strand and the substrate duplex to initiate strand displacement, eventually sequestering the invading strand in the reconfigured duplex and displacing the incumbent strand (130). Rationally designed DNA strand-displacement systems can thus be triggered by molecular triggers, namely nucleic acid sequences, and generate a specific molecular signal in the process. Instead of traditional DNA duplexes, our method uses a DNA-PNA hybrid as the substrate duplex with the toehold on the PNA. Since PNA is an achiral nucleic acid analogue, it hybridizes to both D-DNA and L-DNA. In the homochiral reaction pathway, where the invading strand and the incumbent strand are of the same chirality, the sequence information remains coupled with the stereochemical information. This information can be decoupled via the heterochiral pathway, where the invading strand and the incumbent strand are of opposite chirality. This particular pathway becomes incredibly useful in rationally engineering an exceptionally biostable dynamic DNA



**Figure 1-5 Schematics of Heterochiral Strand Displacement Reactions.** a) Chemical Structures of D-DNA (Black), L-DNA (Blue) and PNA (Green). b) Reaction A is a strand displacement reaction from a DNA-PNA heteroduplex, where the toehold is on the DNA strand. c) Reaction B is a strand displacement reaction occurring from a PNA-DNA heteroduplex, but from the achiral PNA toehold. Reprinted with permission from Ref (129). Copyright © 2017, American Chemical Society.

device responsive to a native D-DNA or RNA input that serves as the invading strand. The substrate duplex or the “inversion gate” as we will call it henceforth, is a L-DNA(RNA)-PNA hybrid that subsequently releases a L-DNA or RNA strand triggering further displacement cascades. The sequence information from the native input is thus

successfully retained in space and time, for sensing, amplification and transduction purposes.



**Figure 1-6 Toehold mediated DNA Strand Displacement Reaction.** a) An invader hybridizes to the toehold domain  $t^*$  by its complementary domain  $t$  (red), and proceeds to branch migrate, forming a three-stranded intermediate. Successful displacement of the incumbent strand and formation of a waste duplex occurs. b) Second-order rate constant of toehold mediated strand displacement can vary over a range of 6 orders of magnitude. Reprinted with permission from Ref (130). Copyright © 2011, Nature Publishing Group.

## 1.8. DNA Strand Displacement Reactions

### 1.8.1. Basics and Mechanism

The reaction is initiated through binding of the toehold-complementary domain on the invader to the single-stranded domain also referred to as the toehold, on the substrate duplex. This interaction initiates a three-way branch migration process that ultimately results in displacement of the incumbent strand. Three-way or four-way branch migrations are the most used strand displacement designs (81). Toeholds accelerate the rate of strand displacement by increasing the probability that the incumbent strand is successfully replaced by the invader once bound (132-134). Second-order rate of these reactions

depend on a number of parameters such as toehold length and sequence composition, ambient temperature and salt concentration. Toeholds in particular are a key design parameter, and in all-DNA strand displacements can be modulated for the rate to vary over 6 orders of magnitude (132-134).

Phenomenological modeling and biophysical studies have helped gain more insight into the individual steps involved in a successful strand displacement reaction: toehold nucleation and dissociation by fraying, initiation of branch migration and successful completion of branch migration process by random walk (132). Strand-exchange reactions i.e. in a 0-nucleotide toehold scenario are inherently very slow and occurs mainly due to the fraying and breathing at the duplex ends. The same mechanisms contribute to “leak” reactions in toehold-mediated strand displacements, and this is important in deciding the signal to noise ratio of such systems (135).

Toehold nucleation is essentially reversible, and once the toehold is bound, the stability of the three-stranded intermediate decides the fate of the reaction. From here it can either proceed to branch migration or the toehold can dissociate. It seems that the kinetic barrier is extremely high for the first step of branch migration that generates a single-stranded overhang. The sterics of this overhang generates a high entropic barrier. Once past this point, the activation energies associated with each step is consistently same. Three stranded intermediates with longer lifetimes are thus more prone to dissociate and lead to failed displacements. Some examples are systems with mismatches at the toehold junction.

## **1.8.2. Applications in Dynamic DNA Nanodevices for Biological Applications**

Strand displacements are ubiquitous in nature and are most common in the context of genetic recombination. It was described for the first time in 2000 in a design of DNA-fuelled nanomachine (136). Since then, dynamic DNA nanodevices based on this principle have acquired more complexity and versatility in applications. Molecular motors (137-139), switchable nanostructures (140,141), Boolean logic computation (142,143), and molecular sensors and amplifiers (144) have all been realized using dynamic DNA interactions. However, the true potential of these devices lies in biomedical applications. Detection, imaging, and analysis of physiologically relevant biomarkers in fixed cells or tissues, surface of cell membranes or even inside mammalian cells represent significant progress in this direction (145-147). Other innovations include *in vitro* signal amplification strategies for detection of RNA (144,148), conditional cargo delivery or gene silencing (82). There are numerous examples of DNA circuits and other nanodevices functioning as molecular sensors and/or effectors of endogenous nucleic acids and even proteins (147,149), with the ultimate aim to scale-up the complexity of such systems to mimic the intricate biological interactions.

## **1.9. Goals of Research**

### **1.9.1. Directed Evolution of a Novel Heterochiral Ribonuclease Ribozyme**

RNA function is tightly regulated by its structure. A key player in all major biochemical pathways, RNA employs diverse mechanisms to regulate gene expression. Approaches focused on manipulating RNA structural motifs is thus an attractive strategy

to disrupt its interactions with effector molecules and by extension, its normal function. Additionally, structure-based approaches have the potential of opening up novel and diverse routes to access exceptionally complex RNA with extensive secondary structures that can obstruct effective complementary base-pairing employed by hybridization-based effectors. While hybridization-based approaches for targeting RNA such as antisense strategy can be rationally designed following the rules of WC base pairing, structural interactions are difficult to predict and design due to the sheer diversity of non-canonical interactions that can arise in that specific structural context. Harnessing the power of *in vitro* selection, random RNA libraries can be evolved to acquire highly specific interactions with the target of interest due to its natural inclination to fold and form structured surfaces with well-defined topology. Ensuring structure-based recognition is not always possible for D-DNA/RNA based affinity reagents against D-RNA targets because of the natural bias toward WC base-pairing. Using a bio-orthogonal backbone such as L-DNA/RNA eliminates this outcome and affords structure dependent modalities. In Chapter 2, I describe the design of a selection experiment based on the mirror-image selection methodology for the directed evolution of a heterochiral ribozyme that uses non-canonical interactions to recognize and cleave a structured TAR RNA target of the opposite chirality. Ribozymes combine binding with chemical modification of the target that can irreversibly manipulate the RNA of interest, and thus can be more potent than affinity reagents. Following the isolation of active ribozymes that operate in the *cis* format, I carried out characterization of the efficiency of cleavage, the first-order reaction kinetics, nature and site of the cleavage products among the different clones. The



probable secondary structural features that contribute to the catalytic function was also identified by truncation studies. Based on these results, we redesigned the libraries based on mutagenized versions of the *cis*-active clones and carried out more iterations of the selection to obtain libraries highly enriched in new *cis*-ribozymes with enhanced efficiency but only one ribozyme sequence with *trans* activity. Detailed characterization of the *trans* activity under different concentration regimes, Mg concentrations etc. reveal very weak *trans*-activity with slow kinetics. Absence of true catalytic power (i.e inability to operate in a multiple turnover context) is indicative of weak target binding that can be improved by further evolution. Overall, this work adds to the complexity of RNA structural interactions and provides a model for future heterochiral selection experiments.

### **1.9.2. Kinetic Characterization of Heterochiral Strand Displacement Reactions**

The heterochiral DNA strand-displacement system is significantly different from the standard reaction in two ways 1) the strand displacement takes place from a DNA-PNA hybrid system with a toehold on the achiral PNA and 2) the stereochemistry of the invading strand is inverted with respect to the incumbent strand. This is a potential contributor to the overall slower kinetics of the heterochiral reaction pathway. Given the unique design of our strand displacement strategy, we realized the importance of extensive characterization for the purpose of laying down reliable design principles for devices based on this strategy. Chapter 3 of this thesis involves studying the effects of varying different parameters on the kinetics under simulated physiological conditions, given our goal of ultimately being able to translate these designs to fully functional dynamic devices in living systems. By basing our sequences on previous work (132), we have analyzed the

kinetics of both the homochiral and heterochiral reactions, thus enabling a direct comparison with all DNA systems as well as highlighting the differences arising from the stereochemical difference. First, we wanted to see how the kinetics vary and saturate depending on the toehold length. Secondly, since detection of mismatches is of significant interest in biomedicine, we looked at how well rationally positioned mismatches were tolerated by these systems. We also tested a subset of RNA inputs as RNA is the most likely single stranded nucleic acid target for biosensor development. Truncated inputs which lack the full complementary sequence to the branch migration domain enables us to test the possibility of using our system in modular toehold-exchange reactions and catalytic designs. Finally, we attempt at understanding the source of this intriguing kinetic penalty for heterochiral reactions over homochiral reactions. With such a thorough investigation of different parameters, this study will provide a blueprint to inform future designs of robust heterochiral systems with promising applications in research and medicine.

## 1.10. References

1. Crick, F. (1970) Central Dogma of Molecular Biology. *Nature*, 227, 561-563.
2. Maderspacher, F. (2004) Rags before the riches: Friedrich Miescher and the discovery of DNA. *Current Biology*, 14, R608.
3. Portin, P. (2014) The birth and development of the DNA theory of inheritance: sixty years since the discovery of the structure of DNA. *Journal of genetics*, 93, 293-302.
4. Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, 737-738.
5. Klug, A. (1968) Rosalind Franklin and the discovery of the structure of DNA. *Nature*, 219, 808-810.
6. Franklin, R.E. and Gosling, R.G. (1953) Molecular Configuration in Sodium Thymonucleate. *Nature*, 171, 740-741.
7. Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *Rna*, 7, 499-512.
8. Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q. and Fox, G.E. (2002) NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Res*, 30, 395-397.
9. Gautheret, D., Konings, D. and Gutell, R.R. (1995) GU base pairing motifs in ribosomal RNA. *Rna*, 1, 807-814.
10. Hannoush, R.N. and Damha, M.J. (2001) Remarkable Stability of Hairpins Containing 2',5'-Linked RNA Loops. *Journal of the American Chemical Society*, 123, 12368-12374.
11. Moore, P.B. (1999) Structural motifs in RNA. *Annual review of biochemistry*, 68, 287-300.
12. Ganser, L.R., Kelly, M.L., Herschlag, D. and Al-Hashimi, H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology*, 20, 474-489.

13. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, 289, 905-920.
14. Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vornheim, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, 407, 327-339.
15. Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D. and Noller, H.F. (2001) Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science*, 292, 883-896.
16. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35, 849-857.
17. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S. and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nature methods*, 14, 75-82.
18. Regulski, E.E. and Breaker, R.R. (2008) In Wilusz, J. (ed.), *Post-Transcriptional Gene Regulation*. Humana Press, Totowa, NJ, pp. 53-67.
19. Wan, Y., Qu, K., Ouyang, Z. and Chang, H.Y. (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nature protocols*, 8, 849-869.
20. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society*, 127, 4223-4231.
21. Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T. and Chang, H.Y. (2013) RNA SHAPE analysis in living cells. *Nat Chem Biol*, 9, 18-20.
22. Kappel, K., Zhang, K., Su, Z., Watkins, A.M., Kladwang, W., Li, S., Pintilie, G., Topkar, V.V., Rangan, R., Zheludev, I.N. et al. (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nature Methods*, 17, 699-707.
23. Chu, C., Spitale, R.C. and Chang, H.Y. (2015) Technologies to probe functions and mechanisms of long noncoding RNAs. *Nature Structural & Molecular Biology*, 22, 29-35.
24. 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.

25. Sato, K., Akiyama, M. and Sakakibara, Y. (2021) RNA secondary structure prediction using deep learning with thermodynamic integration. *Nature Communications*, 12, 941.
26. Gandhi, M., Caudron-Herger, M. and Diederichs, S. (2018) RNA motifs and combinatorial prediction of interactions, stability, and localization of noncoding RNAs. *Nature Structural & Molecular Biology*, 25, 1070-1076.
27. Scull, C.E., Dandpat, S.S., Romero, R.A. and Walter, N.G. (2021) Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control. *Frontiers in Molecular Biosciences*, 7, 607158.
28. Rinaldi, A.J., Lund, P.E., Blanco, M.R. and Walter, N.G. (2016) The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nature Communications*, 7, 8976.
29. Di Leva, G. and Croce, C.M. (2013) miRNA profiling of cancer. *Current opinion in genetics & development*, 23, 3-11.
30. Lewis, B.P., Shih, I.h., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of Mammalian MicroRNA Targets. *Cell*, 115, 787-798.
31. Hu, X., Schwarz, J.K., Lewis, J.S., Huettner, P.C., Rader, J.S., Deasy, J.O., Grigsby, P.W. and Wang, X. (2010) A microRNA expression signature for cervical cancer prognosis. *Cancer research*, 70, 1441-1448.
32. Visone, R. and Croce, C.M. (2009) MiRNAs and cancer. *The American journal of pathology*, 174, 1131-1138.
33. Iorio, M.V., Visone, R., Di Leva, G., Donati, V., Petrocca, F., Casalini, P., Taccioli, C., Volinia, S., Liu, C.-G. and Alder, H. (2007) MicroRNA signatures in human ovarian cancer. *Cancer research*, 67, 8699-8707.
34. Yao, R.-W., Wang, Y. and Chen, L.-L. (2019) Cellular functions of long noncoding RNAs. *Nature Cell Biology*, 21, 542-551.
35. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31, 147-157.
36. Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R. and Bruening, G. (1986) Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231, 1577-1580.
37. Saville, B.J. and Collins, R.A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria. *Cell*, 61, 685-696.

38. Kazakov, S.A., Balatskaya, S.V. and Johnston, B.H. (2006) Ligation of the hairpin ribozyme in cis induced by freezing and dehydration. *RNA*, 12, 446-456.
39. Ishida, S., Terasaka, N., Katoh, T. and Suga, H. (2020) An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nature Chemical Biology*, 16, 702-709.
40. Scheitl, C.P.M., Ghaem Maghami, M., Lenz, A.-K. and Höbartner, C. (2020) Site-specific RNA methylation by a methyltransferase ribozyme. *Nature*, 587, 663-667.
41. Tjhung, K.F., Shokhirev, M.N., Horning, D.P. and Joyce, G.F. (2020) An RNA polymerase ribozyme that synthesizes its own ancestor. *Proceedings of the National Academy of Sciences*, 117, 2906.
42. Joyce, G.F. (2009) Evolution in an RNA world. *Cold Spring Harbor Symposia on Quantitative Biology*, 74, 17-23.
43. Zhou, L., Kim, S.C., Ho, K.H., O'Flaherty, D.K., Giurgiu, C., Wright, T.H. and Szostak, J.W. (2019) Non-enzymatic primer extension with strand displacement. *Elife*, 8, e51888.
44. von Kiedrowski, G. (1986) A self-replicating hexadeoxynucleotide. *Angewandte Chemie International Edition in English*, 25, 932-935.
45. James, K.D. and Ellington, A.D. (1999) The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Origins of Life and Evolution of the Biosphere*, 29, 375-390.
46. Janzen, E., Blanco, C., Peng, H., Kenchel, J. and Chen, I.A. (2020) Promiscuous Ribozymes and Their Proposed Role in Prebiotic Evolution. *Chemical Reviews*, 120, 4879-4897.
47. Peng, H., Latifi, B., Müller, S., Lupták, A. and Chen, I.A. (2021) Self-cleaving ribozymes: substrate specificity and synthetic biology applications. *RSC Chemical Biology*, 2, 1370-1383.
48. Lilley, D.M.J. (2011) Mechanisms of RNA catalysis. *Philosophical Transactions of the Royal Society B Biological Sciences*, 366, 2910-2917.
49. Doherty, E.A. and Doudna, J.A. (2001) Ribozyme structures and mechanisms. *Annual review of biophysics and biomolecular structure*, 30, 457-475.
50. Nakano, S.-i., Chadalavada, D.M. and Bevilacqua, P.C. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, 287, 1493-1497.

51. Hammann, C., Luptak, A., Perreault, J. and De La Peña, M. (2012) The ubiquitous hammerhead ribozyme. *RNA*, 18, 871-885.
52. Nguyen, L.A., Wang, J. and Steitz, T.A. (2017) Crystal structure of Pistol, a class of self-cleaving ribozyme. *Proceedings of the National Academy of Sciences*, 114, 1021-1026.
53. Liu, Y., Wilson, T.J., McPhee, S.A. and Lilley, D.M. (2014) Crystal structure and mechanistic investigation of the twister ribozyme. *Nature chemical biology*, 10, 739-744.
54. Huang, X., Zhao, Y., Pu, Q., Liu, G., Peng, Y., Wang, F., Chen, G., Sun, M., Du, F. and Dong, J. (2019) Intracellular selection of trans-cleaving hammerhead ribozymes. *Nucleic acids research*, 47, 2514-2522.
55. Felletti, M., Stifel, J., Wurmthaler, L.A., Geiger, S. and Hartig, J.S. (2016) Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nature communications*, 7, 1-8.
56. Ausländer, S., Ketzer, P. and Hartig, J.S. (2010) A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Molecular BioSystems*, 6, 807-814.
57. Beaudry, A., DeFoe, J., Zinnen, S., Burgin, A. and Beigelman, L. (2000) In vitro selection of a novel nuclease-resistant RNA phosphodiesterase. *Chemistry & Biology*, 7, 323-334.
58. Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences*, 94, 4262-4266.
59. Breaker, R.R. and Joyce, G.F. (1994) A DNA enzyme that cleaves RNA. *Chemistry & Biology*, 1, 223-229.
60. Liu, H., Yu, X., Chen, Y., Zhang, J., Wu, B., Zheng, L., Haruehanroengra, P., Wang, R., Li, S., Lin, J. et al. (2017) Crystal structure of an RNA-cleaving DNAzyme. *Nature Communications*, 8, 2006.
61. Nguyen, K., Wang, Y., England, W.E., Chaput, J.C. and Spitale, R.C. (2021) Allele-Specific RNA Knockdown with a Biologically Stable and Catalytically Efficient XNAzyme. *Journal of the American Chemical Society*, 143, 4519-4523.
62. Wang, Y., Nguyen, K., Spitale, R.C. and Chaput, J.C. (2021) A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nature Chemistry*, 13, 319-326.

63. Fokina, A.A., Stetsenko, D.A. and François, J.C. (2015) DNA enzymes as potential therapeutics: towards clinical application of 10-23 DNazymes. *Expert opinion on biological therapy*, 15, 689-711.
64. Xiong, H., Veedu, R.N. and Diermeier, S.D. (2021) Recent Advances in Oligonucleotide Therapeutics in Oncology. *International journal of molecular sciences*, 22, 3295.
65. Lieberman, J. (2018) Tapping the RNA world for therapeutics. *Nature structural & molecular biology*, 25, 357-364.
66. Yang, H.D. and Nam, S.W. (2020) Pathogenic diversity of RNA variants and RNA variation-associated factors in cancer development. *Experimental & molecular medicine*, 52, 582-593.
67. Pickering, B.M. and Willis, A.E. (2005), *Seminars in cell & developmental biology*. Elsevier, Vol. 16, pp. 39-47.
68. Shao, Y. and Zhang, Q.C. (2020) Targeting RNA structures in diseases with small molecules. *Essays in Biochemistry*, 64, 955-966.
69. Robert, F. and Pelletier, J. (2018) Exploring the impact of single-nucleotide polymorphisms on translation. *Frontiers in genetics*, 9, 507.
70. Davidson, A., Leeper, T.C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J.A. and Varani, G. (2009) Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proceedings of the National Academy of Sciences*, 106, 11931.
71. Garbesi, A., Hamy, F., Maffini, M., Albrecht, G. and Klimkait, T. (1998) TAR-RNA binding by HIV-1 Tat protein is selectively inhibited by its L -enantiomer. *Nucleic Acids Research*, 26, 2886-2890.
72. Otto, G.A. and Puglisi, J.D. (2004) The pathway of HCV IRES-mediated translation initiation. *Cell*, 119, 369-380.
73. Roberts, T.C., Langer, R. and Wood, M.J. (2020) Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery*, 19, 673-694.
74. Shippy, R., Lockner, R., Farnsworth, M. and Hampel, A. (1999) The hairpin ribozyme. *Molecular Biotechnology*, 12, 117-129.
75. Gragoudas, E.S., Adamis, A.P., Cunningham Jr, E.T., Feinsod, M. and Guyer, D.R. (2004) Pegaptanib for neovascular age-related macular degeneration. *New England Journal of Medicine*, 351, 2805-2816.



76. Fisher, T.L., Terhorst, T., Cao, X. and Wagner, R.W. (1993) Intracellular disposition and metabolism of fluorescently-labeled unmodified and modified oligonucleotides microinjected into mammalian cells. *Nucleic Acids Research*, 21, 3857-3865.
77. Chandrasekaran, A.R. (2021) Nuclease resistance of DNA nanostructures. *Nature Reviews Chemistry*, 5, 225-239.
78. Dar, S.A., Thakur, A., Qureshi, A. and Kumar, M. (2016) siRNAmoD: a database of experimentally validated chemically modified siRNAs. *Scientific Reports*, 6, 1-8.
79. Fern, J. and Schulman, R. (2017) Design and Characterization of DNA Strand-Displacement Circuits in Serum-Supplemented Cell Medium. *ACS Synthetic Biology*, 6, 1774-1783.
80. Roberts, T.C., Langer, R. and Wood, M.J.A. (2020) Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery*, 19, 673-694.
81. Chen, Y.-J., Groves, B., Muscat, R.A. and Seelig, G. (2015) DNA nanotechnology from the test tube to the cell. *Nature Nanotechnology*, 10, 748-760.
82. Bujold, K.E., Hsu, J.C.C. and Sleiman, H.F. (2016) Optimized DNA “Nanosuitcases” for Encapsulation and Conditional Release of siRNA. *Journal of the American Chemical Society*, 138, 14030-14038.
83. Fu, J., Yang, Y.R., Johnson-Buck, A., Liu, M., Liu, Y., Walter, N.G., Woodbury, N.W. and Yan, H. (2014) Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nature Nanotechnology*, 9, 531-536.
84. Lee, H., Lytton-Jean, A.K.R., Chen, Y., Love, K.T., Park, A.I., Karagiannis, E.D., Sehgal, A., Querbes, W., Zurenko, C.S., Jayaraman, M. et al. (2012) Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery. *Nature Nanotechnology*, 7, 389-393.
85. Walsh, A.S., Yin, H., Erben, C.M., Wood, M.J. and Turberfield, A.J. (2011) DNA cage delivery to mammalian cells. *ACS Nano*, 5, 5427-5432.
86. Surana, S., Shenoy, A.R. and Krishnan, Y. (2015) Designing DNA nanodevices for compatibility with the immune system of higher organisms. *Nature Nanotechnology*, 10, 741-747.
87. Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K. and MacLachlan, I. (2005) Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nature Biotechnology*, 23, 457-462.

88. Vollmer, J., Weeratna, R.D., Jurk, M., Samulowitz, U., McCluskie, M.J., Payette, P., Davis, H.L., Schetter, C. and Krieg, A.M. (2004) Oligodeoxynucleotides lacking CpG dinucleotides mediate Toll-like receptor 9 dependent T helper type 2 biased immune stimulation. *Immunology*, 113, 212-223.
89. Qiuyan, Z., Dong, Y. and Agrawal, S. (1999) Site of chemical modifications in CpG containing phosphorothiate oligodeoxynucleotide modulates its immunostimulatory activity. *Bioorganic & Medicinal Chemistry Letters*, 9, 3453-3458.
90. Quemener, A.M., Bachelot, L., Forestier, A., Donnou-Fournet, E., Gilot, D. and Galibert, M.-D. (2020) The powerful world of antisense oligonucleotides: From bench to bedside. *WIREs RNA*, 11, e1594.
91. Frazier, K.S. (2015) Antisense oligonucleotide therapies: the promise and the challenges from a toxicologic pathologist's perspective. *Toxicologic pathology*, 43, 78-89.
92. Crooke, S.T., Vickers, T.A. and Liang, X.-h. (2020) Phosphorothioate modified oligonucleotide–protein interactions. *Nucleic Acids Research*, 48, 5235-5253.
93. Crooke, S.T., Wang, S., Vickers, T.A., Shen, W. and Liang, X.-h. (2017) Cellular uptake and trafficking of antisense oligonucleotides. *Nature Biotechnology*, 35, 230-237.
94. 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 puckering. *Tetrahedron Letters*, 38, 8735-8738.
95. 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.
96. Majlessi, M., Nelson, N.C. and Becker, M.M. (1998) Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Research*, 26, 2224-2229.
97. Nielsen, P.E., Egholm, M., Berg, R.H. and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, 254, 1497.
98. Nielsen, P.E. and Haaima, G. (1997) Peptide nucleic acid (PNA). A DNA mimic with a pseudopeptide backbone. *Chemical Society Reviews*, 26, 73-78.
99. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B. and Nielsen, P.E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature*, 365, 566-568.

100. Egholm, M., Buchardt, O., Nielsen, P.E. and Berg, R.H. (1992) Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone. *Journal of the American Chemical Society*, 114, 1895-1897.
101. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P.E., Nordén, B. and Gräslund, A. (1996) Ionic Effects on the Stability and Conformation of Peptide Nucleic Acid Complexes. *Journal of the American Chemical Society*, 118, 5544-5552.
102. Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S.H. et al. (2012) Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science*, 336, 341.
103. Mei, H., Wang, Y., Yik, E.J. and Chaput, J.C. (2021) Synthesis and polymerase recognition of a pyrrolocytidine TNA triphosphate. *Biopolymers*, 112, e23388.
104. Taylor, A.I., Pinheiro, V.B., Smola, M.J., Morgunov, A.S., Peak-Chew, S., Cozens, C., Weeks, K.M., Herdewijn, P. and Holliger, P. (2015) Catalysts from synthetic genetic polymers. *Nature*, 518, 427-430.
105. Dunn, M.R., McCloskey, C.M., Buckley, P., Rhea, K. and Chaput, J.C. (2020) Generating Biologically Stable TNA Aptamers that Function with High Affinity and Thermal Stability. *Journal of the American Chemical Society*, 142, 7721-7724.
106. Ashley, G.W. (1992) Modeling, synthesis, and hybridization properties of (L)-ribonucleic acid. *Journal of the American Chemical Society*, 114, 9731-9736.
107. Urata, H., Ogura, E., Shinohara, K., Ueda, Y. and Akagi, M. (1992) Synthesis and properties of mirror-image DNA. *Nucleic Acids Research*, 20, 3325-3332.
108. Hauser, N.C., Martinez, R., Jacob, A., Rupp, S., Hoheisel, J.D. and Matysiak, S. (2006) Utilising the left-helical conformation of L-DNA for analysing different marker types on a single universal microarray platform. *Nucleic Acids Research*, 34, 5101-5111.
109. Szabat, M., Gudanis, D., Kotkowiak, W., Gdaniec, Z., Kierzek, R. and Pasternak, A. (2016) Thermodynamic features of structural motifs formed by  $\beta$ -L-RNA. *PloS one*, 11, e0149478.
110. Garbesi, A., Capobianco, M., Colonna, F., Tondelli, L., Arcamone, F., Manzini, G., Hilbers, C., Aelen, J. and Blommers, M. (1993) L-DNAs as potential antimessenger oligonucleotides: A reassessment. *Nucleic Acids Research*, 21, 4159-4165.
111. Hoehlig, K., Bethge, L. and Klussmann, S. (2015) Stereospecificity of oligonucleotide interactions revisited: no evidence for heterochiral hybridization and ribozyme/DNAzyme activity. *PLoS One*, 10, e0115328.

112. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249, 505-510.
113. Ellington, A.D. and Szostak, J.W. (1992) Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*, 355, 850-852.
114. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346, 818-822.
115. Nolte, A., Klußmann, S., Bald, R., Erdmann, V.A. and Fürste, J.P. (1996) Mirror-design of L-oligonucleotide ligands binding to L-arginine. *Nature biotechnology*, 14, 1116-1119.
116. Klussmann, S., Nolte, A., Bald, R., Erdmann, V.A. and Fürste, J.P. (1996) Mirror-image RNA that binds D-adenosine. *Nature biotechnology*, 14, 1112-1115.
117. Eulberg, D. and Klussmann, S. (2003) Spiegelmers: Biostable Aptamers. *ChemBioChem*, 4, 979-983.
118. Vater, A. and Klussmann, S. (2003) Toward third-generation aptamers: Spiegelmers and their therapeutic prospects. *Current Opinion in Drug Discovery and Development*, 6, 253-261.
119. Vater, A. and Klussmann, S. (2015) Turning mirror-image oligonucleotides into drugs: the evolution of Spiegelmer® therapeutics. *Drug Discovery Today*, 20, 147-155.
120. Olea Jr, C., Weidmann, J., Dawson, P.E. and Joyce, G.F. (2015) An L-RNA aptamer that binds and inhibits RNase. *Chemistry & Biology*, 22, 1437-1441.
121. Boiziau, C., Dausse, E., Yurchenko, L. and Toulmé, J.-J. (1999) DNA Aptamers selected against the HIV-1 Trans-activation-responsive RNA element form RNA-DNA kissing complexes. *Journal of Biological Chemistry*, 274, 12730-12737.
122. Lünse, C.E., Michlewski, G., Hopp, C.S., Rentmeister, A., Cáceres, J.F., Famulok, M. and Mayer, G. (2010) An Aptamer Targeting the Apical-Loop Domain Modulates pri-miRNA Processing. *Angewandte Chemie International Edition*, 49, 4674-4677.
123. Sczepanski, J.T. and Joyce, G.F. (2014) A cross-chiral RNA polymerase ribozyme. *Nature*, 515, 440-442.
124. Jiang, W., Zhang, B., Fan, C., Wang, M., Wang, J., Deng, Q., Liu, X., Chen, J., Zheng, J., Liu, L. et al. (2017) Mirror-image polymerase chain reaction. *Cell Discovery*, 3, 17037.

125. Wang, M., Jiang, W., Liu, X., Wang, J., Zhang, B., Fan, C., Liu, L., Pena-Alcantara, G., Ling, J.-J., Chen, J. et al. (2019) Mirror-Image Gene Transcription and Reverse Transcription. *Chem*, 5, 848-857.
126. Tjhung, K.F., Sczepanski, J.T., Murtfeldt, E.R. and Joyce, G.F. (2020) RNA-catalyzed cross-chiral polymerization of RNA. *Journal of the American Chemical Society*, 142, 15331-15339.
127. Chen, Y.-I., Chang, Y.-J., Nguyen, T.D., Liu, C., Phillion, S., Kuo, Y.-A., Vu, H.T., Liu, A., Liu, Y.-L., Hong, S. et al. (2019) Measuring DNA Hybridization Kinetics in Live Cells Using a Time-Resolved 3D Single-Molecule Tracking Method. *Journal of the American Chemical Society*, 141, 15747-15750.
128. Datta, B. and Armitage, B.A. (2001) Hybridization of PNA to Structured DNA Targets: Quadruplex Invasion and the Overhang Effect. *Journal of the American Chemical Society*, 123, 9612-9619.
129. Kabza, A.M., Young, B.E. and Sczepanski, J.T. (2017) Heterochiral DNA Strand-Displacement Circuits. *Journal of the American Chemical Society*, 139, 17715-17718.
130. Zhang, D.Y. and Seelig, G. (2011) Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry*, 3, 103-113.
131. Kabza, A.M., Kundu, N., Zhong, W. and Sczepanski, J.T. Integration of chemically modified nucleotides with DNA strand displacement reactions for applications in living systems. *WIREs Nanomedicine and Nanobiotechnology*, n/a, e1743.
132. 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.
133. Srinivas, N., Ouldridge, T.E., Sulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res*, 41, 10641-10658.
134. Yurke, B. and Mills, A.P. (2003) Using DNA to Power Nanostructures. *Genetic Programming and Evolvable Machines*, 4, 111-122.
135. Wang, B., Thachuk, C., Ellington, A.D., Winfree, E. and Soloveichik, D. (2018) Effective design principles for leakless strand displacement systems. *Proceedings of the National Academy of Sciences*, 115, E12182.
136. 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.

137. Bath, J. and Turberfield, A.J. (2007) DNA nanomachines. *Nature Nanotechnology*, 2, 275-284.
138. Shin, J.-S. and Pierce, N.A. (2004) A Synthetic DNA Walker for Molecular Transport. *Journal of the American Chemical Society*, 126, 10834-10835.
139. Sherman, W.B. and Seeman, N.C. (2004) A Precisely Controlled DNA Biped Walking Device. *Nano Letters*, 4, 1203-1207.
140. Rothemund, P.W.K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, 440, 297-302.
141. Yin, P., Choi, H.M.T., Calvert, C.R. and Pierce, N.A. (2008) Programming biomolecular self-assembly pathways. *Nature*, 451, 318-322.
142. Qian, L. and Winfree, E. (2011) Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science*, 332, 1196.
143. Seelig, G., Soloveichik, D., Zhang, D.Y. and Winfree, E. (2006) Enzyme-Free Nucleic Acid Logic Circuits. *Science*, 314, 1585.
144. 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.
145. Hemphill, J. and Deiters, A. (2013) DNA Computation in Mammalian Cells: MicroRNA Logic Operations. *Journal of the American Chemical Society*, 135, 10512-10518.
146. Groves, B., Chen, Y.-J., Zurla, C., Pochekailov, S., Kirschman, J.L., Santangelo, P.J. and Seelig, G. (2016) Computing in mammalian cells with nucleic acid strand exchange. *Nature Nanotechnology*, 11, 287-294.
147. Chatterjee, G., Chen, Y.-J. and Seelig, G. (2018) Nucleic Acid Strand Displacement with Synthetic mRNA Inputs in Living Mammalian Cells. *ACS Synthetic Biology*, 7, 2737-2741.
148. Kabza, A.M. and Sczepanski, J.T. (2020) l-DNA-Based Catalytic Hairpin Assembly Circuit. *Molecules (Basel, Switzerland)*, 25.
149. Gavins, G.C., Gröger, K., Bartoschek, M.D., Wolf, P., Beck-Sickinger, A.G., Bultmann, S. and Seitz, O. (2021) Live cell PNA labelling enables erasable fluorescence imaging of membrane proteins. *Nature Chemistry*, 13, 15-23.

## 2. DIRECTED EVOLUTION OF A NOVEL HETEROCHIRAL RIBONUCLEASE RIBOZYME AGAINST A STRUCTURED RNA TARGET

### 2.1. Introduction

Ribonucleic acid (RNA) regulates almost all aspects of gene expression. Hierarchical organization in RNA generates highly complex structures which accurately determine its function by facilitating key intramolecular and intermolecular interactions. Canonical Watson-Crick (WC) base-pairing direct the formation of well-defined secondary structures; however, non-canonical base-pairs such as the G-U wobble-pair and structures such as G quadruplexes, stacking and hydrophobic interactions as well as post-transcriptional modifications are major contributors to the structural diversity as well as diverse modalities of interaction between functional RNA (2-4). Exploration of RNA-RNA interactions beyond canonical WC-mediated hybridization is thus fundamental to a more comprehensive understanding of RNA biology.

Ribozymes are representative examples of one of the earliest, complex RNA interactions evolved by Nature. They are a fascinating class of naturally occurring RNA, capable of folding into complex shapes to form binding pockets for their respective substrates and active conformations for reaction catalysis. The scope of catalysis in naturally occurring ribozymes is however limited to RNA cleavage and ligation, and protein synthesis (5-12). According to the RNA World theory for the origin of life, ribozymes are ancestral RNA that not only carried genetic information but were also capable of self-replication and other crucial life sustaining processes by virtue of their

catalytic functionality (13-14). Eventually, ribozymes were outcompeted by protein enzymes that evolved as the more efficient catalysts owing to greater chemical diversity. The chemical functionality of ribozymes can, however, be expanded by means of *in vitro* evolution. Today, there are diverse ribozymes that have been evolved with equally diverse mechanisms and an expanded substrate-reaction scope (5,7). Self-alkylation, triphosphorylation, aminoacylation, kinase, RNA polymerase and Diels-Alder catalyzing ribozymes are just a few examples (15-19).

The most well-characterized and well-studied system remains the RNA cleaving ribozymes. They exist in nature either as the highly complex *trans*-acting ribozymes such as the RNase P that are involved in the maturation of tRNAs, or the self-cleaving ribozymes such as the hammerhead (6,20). The latter class of ribozymes are commonly found in viral pathogens and regulate gene expression by cleaving off transcripts from rolling circle amplification (5). Of this, the hammerhead is a model system, with its mechanism and substrate interaction well characterized. These kinds of ribozymes have attracted significant interest in fundamental RNA biology as well as biomedical research. The hammerhead for example, can be reprogrammed to cleave any RNA target *in trans* by separating the catalytic domain and the substrate binding elements from the substrate itself. This is further exemplified by the large body of DNAzymes that are artificially selected, single-stranded DNA enzymes. The most well-known DNAzyme is perhaps the Mg-dependent 10-23 class of DNAzymes that were originally selected from a random library having a central “catalytic” core, flanked by substrate binding arms (21). The cleavage site is specified by the WC-mediated DNAzyme-substrate hybridization that



leave a single unpaired ribonucleotide to begin with. This design is highly valuable for biomedical applications. Post-selection, it can be easily redesigned to operate on an intermolecular format and generalized to any target of interest. Indeed, there are several instances of *in vitro* applications of DNAzymes for gene silencing, some of the latest being XNA modified DNAzymes against valuable therapeutic targets (22-23).

However, this approach has some inherent drawbacks. While the ease of WC programmability is advantageous for repurposing the same catalytic platform for different targets in a sequence specific manner, the rich structural information afforded by structured RNA targets remain mostly unexplored and unexploited. From a therapeutic standpoint as well, targeting structured RNA based on structure-specific interactions is valuable for disease diagnosis as well as interventions, providing new routes to access previously inaccessible or “undruggable” targets.

RNA aptamers that have been evolved to bind specific targets with high affinity and specificity, is an attractive structure-based approach. When it comes to native RNA targets however, RNA aptamer selections succumb to the “tyranny of WC base-pairing”. This has been demonstrated by previous selections of an RNA aptamer developed against native TAR hairpin, where the primary binding modality was based on WC base-pairing mediated “kissing-loop” interactions (24).

In this project, I have attempted to develop a heterochiral ribonuclease ribozyme that recognizes a structured RNA target based on non-canonical heterochiral nucleic acid interactions (i.e., non-WC based interactions between enantiomeric DNA/ RNA) and inactivates it by cleavage. The HIV-1 TAR hairpin regulates transcriptional elongation of

the viral genome by mediating crucial interactions with the Tat protein and other protein effectors and is thus a therapeutically relevant molecule (25-26). This element has served as the target for many small molecule-based approaches. It has also proved amenable to heterochiral interactions; an L-RNA aptamer has been shown to not only bind the TAR hairpin but also manipulate its biological interactions with the Tat protein (27).

L-DNA/ RNA, the enantiomeric form of native DNA and RNA does not occur naturally and is an ideal biorthogonal biomaterial for developing biotechnological tools. It circumvents the major issues that plague nucleic acids-based technologies. With its exceptional biostability stemming from nuclease resistance, low immunogenicity and toxicity due to minimal crosstalk with unintended targets, L-DNA/RNA is superior to many commonly used modifications (27,28). Its inability to form contiguous WC hybridization mediated base-pairs with complementary native nucleic acids precludes its use in antisense applications and other hybridization-based approaches (29). This very property however can be exploited to explore novel recognition modalities based on non-canonical hydrogen bonding mediated interactions between these mirror-image enantiomeric nucleic acids and native D-nucleic acids. Indeed, directed evolution has resulted in the selection of several L-RNA and L-DNA aptamers against structured RNA such as TAR RNA and several miRNAs (30-32). However, cross-chiral catalytic activity is more challenging to evolve and remains restricted to just one class of reactions: RNA ligation and polymerization. The first heterochiral ribozyme was reported in 2014 (33). This 83-nucleotide long L-ribozyme carries out templated ligation of RNA oligomers of opposite chirality and even possesses appreciable polymerization prowess. A more

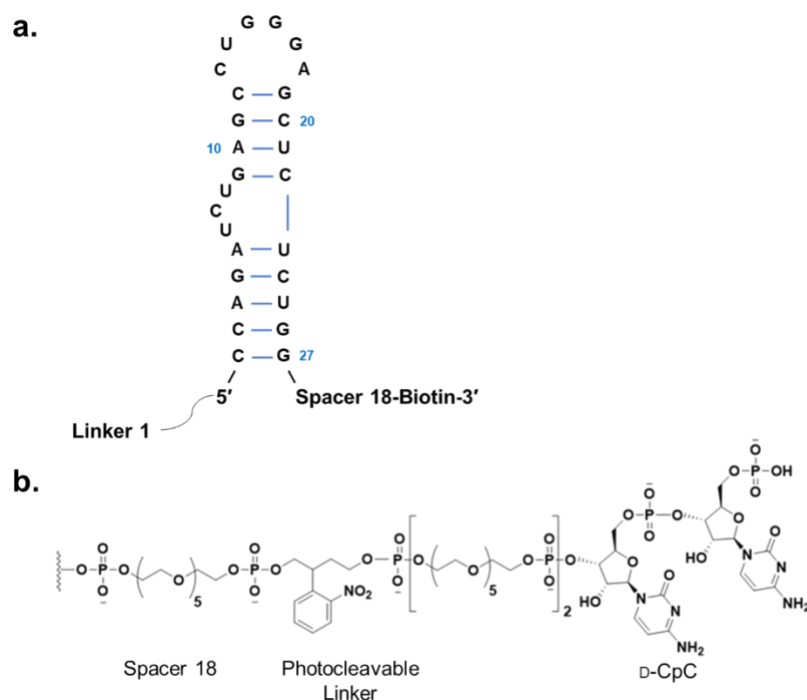
evolved version of this ribozyme can accept trinucleotide substrates in more efficient polymerization reactions (34). A heterochiral ribonuclease ribozyme evolved by *in vitro* evolution would thus expand the repertoire of heterochiral nucleic acids technologies and help explore one of the most common, naturally occurring RNA catalyzed reactions in the context of novel heterochiral, structure-dependent interactions, while holding potential for future therapeutic uses.

## **2.2. Results and Discussion**

### **2.2.1. Selection of *cis*-cleaving Heterochiral Ribonuclease Ribozymes**

#### **2.2.1.1. Target and Selection Design**

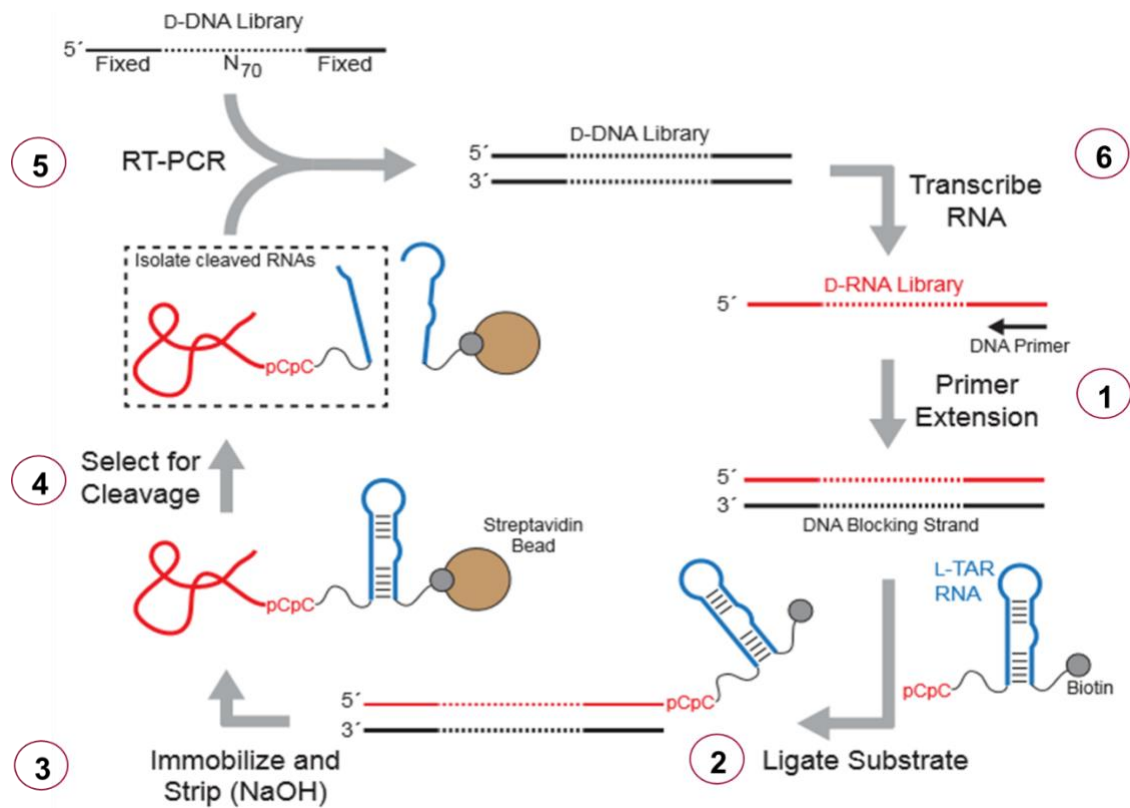
We first attempted to isolate a D-ribozyme capable of cleaving the HIV-1 L-TAR RNA element *in cis*, as a proof-of-concept that cross-chiral interactions between enantiomeric, structured RNA can result in phosphodiester bond scission. The “mirror-image selection” methodology was adopted similar to previous studies exploring cross-chiral interactions (30-34). For this purpose, a 27 nucleotides long truncated version of the physiological HIV-1 TAR hairpin was synthesized in the L-chirality with commercially available L-nucleoside phosphoramidites. This allows us to use a native D-RNA random library and use standard biochemical enzymatic processes such as reverse transcription, PCR amplification, transcription etc. for library manipulation. L-TARs (see Fig. 2-1 for detailed construct design) was biotinylated at its 3'--end to facilitate immobilization on



**Figure 2-1 Design of Selection Target.** a) Secondary structure representation of L-TAR<sub>5</sub> showing a 9 base-pair long discontinuous helix with a 6-nucleotide unpaired loop. The 3'-end is functionalized with biotin for immobilization of the substrate on streptavidin coated beads and the 5'-end is equipped with a linker that helps in monitoring the desired cleavage reaction. b) Chemical structure of the 5'-linker.

streptavidin-coated beads during selection. On the 5'-end, the L-TARs nucleotide sequence was further extended to incorporate a photocleavable linker and a monophosphorylated D-CpC RNA dinucleotide, interspaced with flexible polyethylene glycol (PEG) spacers. While the biotin enables an “on the bead” selection strategy, the incorporation of the photocleavable linker aids in an additional gel selection step in the later rounds of selection (discussed in more detail below) for further vetting of sequences with cleaving properties. A modified version of the L-TARs substrate equipped with an internal 6-Fluorescein (6-FAM) was also prepared for assaying cleavage activities of the evolving ribozyme libraries and individual clones, using fluorescence.

A ssDNA library R161 (refer to sequence Table A-1 in *Appendix A*) containing a 70 nucleotide long random region flanked with 19 nucleotides long primers on either side, was used as the template for primer extension, followed by *in vitro* transcription to generate the random RNA library ( $\sim 10^{14}$  unique sequences) for the first round of selection. Traditionally, *in vitro* selections of ribozymes employ longer random regions compared to aptamer or DNAzyme selections. Empirically, longer sequences have a higher probability of generating a larger and more diverse array of secondary structures. The D-RNA library was first extended with a DNA primer to generate an RNA-DNA hybrid. This suppresses self-hydrolysis of the RNA library and prevents the loss of substrate cleaving sequences before the actual selection step has been initiated. The primer is designed in such a way that the RNA-DNA hybrid has a 4-nucleotide single-stranded overhang on the 3'-end of the RNA, which was essential to T4 RNA Ligase mediated ligation in the final step of library preparation. The 3'-hydroxylated end of the RNA library is finally ligated with the L-TARs construct *via* the 5'-monophosphorylated D-CpC RNA dinucleotide. A detailed scheme outlining this process and the selection strategy is presented in Fig. 2-2.



**Figure 2-2** A detailed schematic of “on the bead” selection strategy. Step 1 involves extension of the random D-RNA library using a D-DNA primer to generate an RNA-DNA hybrid. There is a 4-nucleotide overhang on the 3'-terminus of the RNA strand in this duplex that enables ligation of the hybrid duplex to the L-TAR RNA substrate in Step 2. In the next Step 3, this whole ligated cis-assembly is immobilized on streptavidin coated beads using the biotin handle and the blocking DNA strand is removed using a denaturing. Step 4 is the actual selection step where the single-stranded assembly is incubated for the requisite amount of time in the 1X selection buffer activated with 20 mM MgCl<sub>2</sub>. Any RNA molecules falling off the beads are isolated and converted to cDNA in Step 5 followed by PCR amplification and transcription in Step 6 to generate the enriched library for the next iterative round of selection.

The success of the selection experiment is incumbent upon evolution of a) cross-chiral interactions mediated formation of D-ribozyme-L-TAR complex that has a dissociation rate slower than the rate of a phosphodiester cleavage reaction, and b) an active catalytic motif capable of inducing cleavage at specific sites of the target at a faster rate compared to non-specific hydrolysis. It has been previously shown that the TAR hairpin is amenable to cross-chiral interactions, which was a significant encouragement for pursuing this target (31). However, combining two functionalities of binding and cleavage in a cross-chiral ribozyme may not be as straightforward. Therefore, the physical tether between the RNA library and the substrate, while flexible enough to allow the RNA library to sample different conformations, was designed to be short in order to promote physical proximity.

#### **2.2.1.2. On the bead Selection Strategy**

A fairly large amount of RNA library (24 nmols) was ligated with L-TARs (6 nmols) for the first round of selection. After immobilization on streptavidin coated beads (agarose beads in the first round and then magnetic beads in successive rounds), the RNA-DNA hybrid was denatured by alternate cold washes with buffers containing 25 mM NaOH and 8M Urea as denaturants, respectively. While the blocking DNA strand was removed on the beads in the first round, the ligated library was purified by denaturing PAGE for the successive rounds. Stringent washes were done between steps to remove all unbound material, sometimes in the presence of 1 mg/ml tRNA to prevent non-specific binding of RNA fragments to the beads. All the steps were carried out under the buffer conditions of 100 mM NaCl, 25 mM TRIS-HCl (pH 7.6), 1mM EDTA and 0.05% Tween 20. The final positive selection step was initiated by the addition of 20 mM Mg<sup>2+</sup> as a

divalent metal cofactor. It is to be noted that we use a higher concentration of  $Mg^{2+}$  than the intracellular levels of  $\sim 1$  mM. This is not unreasonable since naturally occurring ribozymes have been shown to use amino acid coordinated  $Mg^{2+}$  for their functions in addition to free  $Mg^{2+}$ , thus utilizing higher concentrations of  $Mg^{2+}$  to enhance folding and formation of complex motifs (35). Another deviation from the physiological conditions is the use of room temperature, since there is a higher chance of metal-mediated degradation of RNA at higher temperatures of  $37$  °C.

Following incubation in the positive selection buffer containing  $Mg^{2+}$  for the respective allotted time in each round (for detailed conditions, refer Table A-2 in the *Appendix A*), the supernatant containing any RNA sequences that cleave the substrate and thus fall off the bead, was collected, reverse-transcribed to their respective cDNAs and PCR amplified to generate sufficient template for transcription. The extension-ligation protocol was followed to generate the *input* library for the next round.

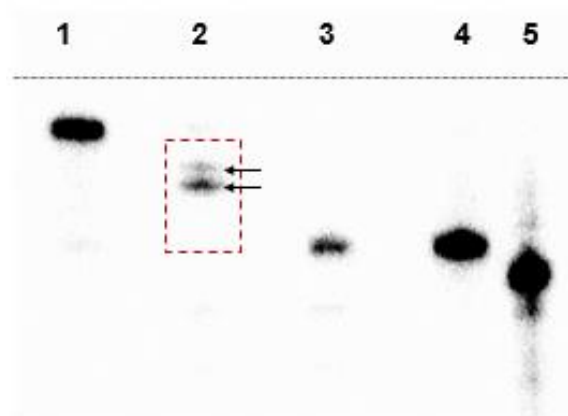
### **2.2.1.3. Length based Selection Strategy by PAGE purification**

In the initial rounds, when the library is not yet expected to have been enriched in ribonuclease ribozymes, there is a high chance that self-cleaving sequences and fragments resulting from spontaneous hydrolysis will dominate the pool of isolated sequences. Therefore, an additional PAGE based selection step was introduced from the second round onwards (See Figure A-1 in *Appendix*). For this purpose, we radiolabeled the RNA library generated from the first round by the incorporation of  $\alpha$ - $^{32}P$ -ATP during *in vitro* transcription. The high sensitivity of radiolabeling made it possible to visualize the



cleavage products on denaturing PAGE by autoradiography. Any bands migrating faster than the full-length single-stranded ligated *cis*-assembly but slower than the RNA library alone, were isolated and eluted. To further ensure that the cleavage was occurring from the substrate itself, half of the products were irradiated at 300-360 nm wavelength UV light for ~10-15 mins to remove the photocleavable group. Any radiolabeled RNA that had cleaved the substrate would still have part of the L-TARs sequence attached which gets released when the internal photocleaving group is removed upon irradiation. The irradiated fraction of active ribonuclease ribozymes thus runs faster than the non-irradiated fraction on denaturing PAGE. Alternatively, if the bands were resulting from self-cleaving RNA sequences that had acquired extra nucleotides during the extension step in between the rounds, (as was observed in a previous selection, data not included here), the migration of the cleavage products would not change upon irradiation.

We believe that this strategy was instrumental in the success of the selection, and clear enrichment in substrate cleaving ribozymes was observed in Round 6 (Fig.2-3 shows the enrichment of cleaving ribozymes). From Round 6 onwards, only the irradiated fraction was further processed for the next rounds. The main selective pressure throughout was the time allowed for the cleavage reaction to take place. Starting at 6 hours for the first two rounds, the positive selection step was reduced to 3 hours for the next few rounds, before allowing the reaction to last for just 2 minutes and 30 seconds respectively, in the last two rounds. A cleavage assay was carried out on the libraries isolated after Rounds 6 and 8 ligated with the FAM labeled L-TARs. Two distinct bands in the cleavage products



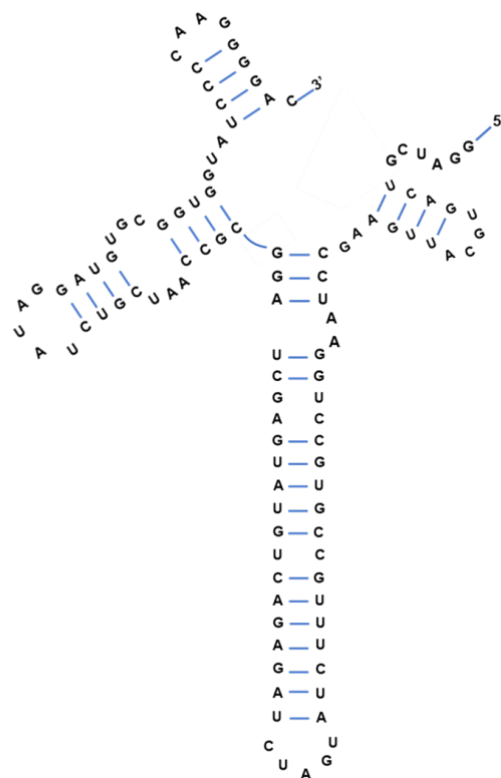
**Figure 2-3 Denaturing PAGE analysis of cleavage products from Round 6:** Lanes 1,4 and 5 are controls for the full-length *cis*-assembly, irradiated *cis*-assembly and the enriched RNA library alone, respectively. Lane 2 represents the cleavage products isolated from Round 6. The boxed area shows potential cleavage products migrating between the full-length assembly and the RNA library alone. Arrow marks indicate the two distinct bands emerging from potentially two separate cleavage sites on the substrate. Lane 3 is part of the sample run on Lane 2 but irradiated to remove the photocleavable linker. Faster migration in Lane 3 compared to Lane 2 confirms that cleavage occurs on the substrate as desired.

confirmed ribonuclease activity and indicated potential diversity in the cleavage sites. The enriched library from Round 8 was cloned and sequenced.

#### **2.2.1.4. Characterization of active clones from Round 8**

Of the 15 clones that were sequenced, one clone that we will refer to as R8c-3, dominated the pool, being 38% abundant (Fig. 2-4 and Fig.2-5a). All other clones were unique sequences, with a few of them being almost identical except for single mutations or deletions. One such example is clone R8c-1. Each of the sequences were then tested for

the *in cis* cleavage activity under the same conditions of buffer strength, salt concentrations and temperature as used in the selection experiment. Although the active



**Figure 2-4. Secondary structure representation of *cis*-cleaving ribozyme R8c-3:** Short terminal stem-loop structures and a long discontinuous but stable stem affords the best *cis*-cleavage activity among all other ribozymes isolated.

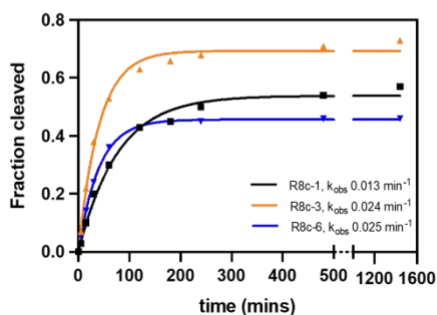
clones were also tested with lower  $Mg^{2+}$  concentrations of 10 mM, the activity was found to be much reduced.

Clones R8c-1, R8c-3, R8c-6 and R8c-13 were the only four sequences that had appreciable cleavage activity. These clones have no significant sequence homology, but according to structure prediction softwares Nupack and RNAfold (XX), have at least one long stem-like motif and several short hairpin-like structures (Fig.A-2 in Appendix).

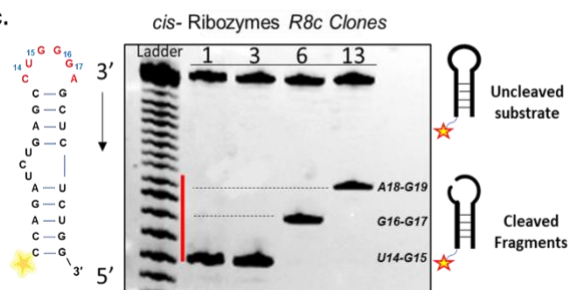
a.

**R8c-1:** GGATCGTCAGTGCATTGAGATCATGGCTGTGCTAAGGTCCGGAATATTTCTGAGCTAGCATATCAGTTTCGCATAAATTGGACTGGTTGAGGGTGGTATCCCCAAGGGGAC  
**R8c-3:** GGATCGTCAGTGCATTGAGACCTAAGTCCGTGCCGTTTCTATGATCTAGAGACTGTATGAGCTAGGCGCCAATCGTCTATAGGATGTCCGGTGGTATCCCCAAGGGGAC  
**R8c-6:** GGATCGTCAGTGCATTGAGATGGGGTGGAAATCCCGAGTTGGTCTCCTGTTTACCGTGCTCGGCTCGGGTTCGGTTCGGAATCTAGACGGAGGGTGGTATCCCCAAGGGGAC  
**R8c-13:** GGATCGTCAGTGCATTGAGAGACAAGGGTCTCGATGCTTTTGCACGGGTTGATCTGGAATGATGGAAGCCTGAAGGGGAAGCATGGACAGGTGGTATCCCCAAGGGGAC

b.



c.



**Figure 2-5 Characterization of *cis*-cleaving ribonuclease ribozymes.** a) Sequence information of the four unique active sequences after 8 rounds of iterative selection. b) a time-course of *cis*-cleavage activity shows the three clones R8c-1,3 and 6 having similar  $k_{obs}$  values but different equilibrium yields. c) denaturing PAGE analysis of cleavage products of the individual ribozyme sequences show diversity in the cleavage sites, all of them on the 6-nucleotide loop as indicated by the ladder generated by alkaline digestion of the substrate used for the assay.

A time course of *cis* cleavage activity (Fig. 2-5b) under selection conditions show that R8c-3 is the most efficient with a  $k_{obs}$  of  $0.024 \text{ min}^{-1}$ , which is not surprising, given its dominance in the enriched library. Although R8-6c also exhibits a similar initial activity profile, the activity plateaus at a little higher than 40% while R8c-3 has an equilibrium

yield of 70% substrate cleaved. R8c-1 is the slowest of the three characterized clones, with a  $k_{\text{obs}}$  of  $0.013 \text{ min}^{-1}$  and a maximum activity of ~50%.

Compared to the saturation kinetics observed in the hammerhead ribozyme (HHR), one of the most well-characterized naturally occurring self-cleaving ribozymes, where the  $k_{\text{obs}}$  of some of the fastest variants can be as high as  $1 \text{ min}^{-1}$  under single turnover conditions, these cross-chiral *cis*-ribonuclease ribozymes operate at kinetics slower by at least two orders of magnitude. However, the equilibrium yield of percent substrate cleaved is still impressively high. The slower kinetics is not surprising since while the HHR has the advantage of WC base-pairing for binding to the substrate, cross-chiral interactions for substrate recognition could be expected to have a high kinetic barrier.

The multiple helices of the active sequences probably have significant roles to play in the substrate binding as well as catalysis of phosphodiester cleavage. R8c-1 and R8c-3, according to Nupack, are 10kcal/ mol less stable than R8c-6. Even though we cannot draw any definitive conclusions about the mechanism in absence of knowledge regarding tertiary structural interactions, it is possible that these ribozymes utilize their stable stems to bind the substrate through non-canonical hydrogen bonded mediated interactions, following which there is a conformational change to form the active catalytic motif. R8c-3 is predicted to have short stem-loop structures on both termini, which are both crucial to the ribozyme activity according to the truncation studies that I am going to discuss next (Fig.A-2 in *Appendix*). These short hairpins could be playing a role in conformational transitions, pushing the ensemble equilibrium to the correctly folded catalytic form, thus

resulting in 70% of substrate cleavage. On the other hand, R8-6 that has a lower free energy, is more stable and would not be as easily amenable to conformational changes which could probably explain the lower yield of ~45%. However, these are just speculations, given that the individual ribozymes could be utilizing significantly different mechanisms/ transition states to achieve cleavage.

This is further exemplified by the different cleavage sites of the four active ribozymes, although all of them lie on the 6 nucleotide long unpaired loop region (Fig 2-5c). Whereas R8c-1 and R8c-3 both cleave between U14 and G15, R8c-6 cleaves between G16 and G17 and R8c-13, with the weakest activity of all, cleaves between A18 and G19 (Fig 2-5c). Hence, in the same enriched library, different sequences with similar kinetic profiles but different cleavage sites were able to evolve simultaneously. This is another advantage of sequence independent binding modality. Although less efficient than canonical WC binding, this strategy doesn't have a predetermined site of reaction, as is most often the case with in vitro selections of DNAzymes or ribozymes where the unpaired nucleotide (s) are designed into the selection by adjacent substrate-binding arms. Therefore, there is greater opportunity for more diverse site selection by the catalytic molecules. Another interesting observation is that when the substrate was redesigned to have a longer stem by 5 base-pairs, in addition to losing much of their activity, all three ribozymes now cleaved between G16 and G17 (See Fig.A-3 in *Appendix*). A longer stem probably has more steric requirements, and in order to maintain optimal substrate binding, the active catalytic motif can access only one phosphodiester bond, that between G16 and G17. Not only that, the equilibrium reaction yields for all three ribozymes are now the

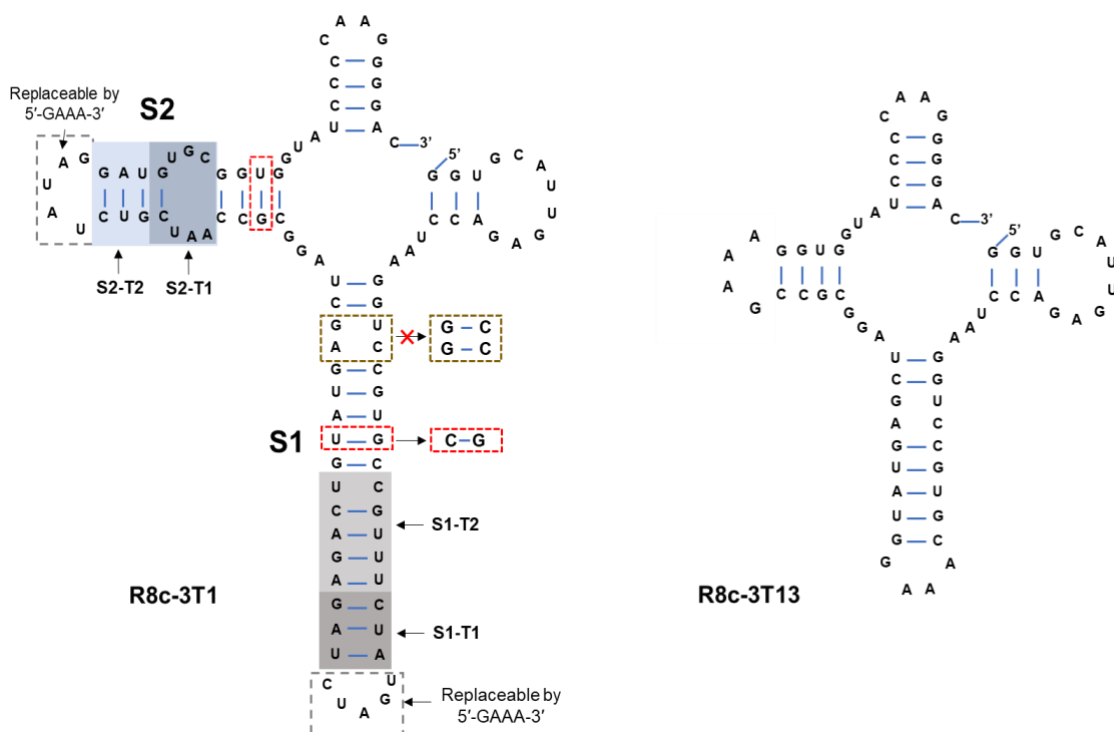
same (~25% in 3 hours) with the faster clones R8c-1 and R8c-3 being more severely inhibited. It is likely then that the choice of cleavage site was not arbitrary; rather there is a correlation between ribozyme conformational flexibility, accessibility of the cleavage sites and the intrinsic efficiency of cleavage at U14-G15 compared to G16-G17.

#### **2.2.1.5. Truncation Studies of R8c-3 and R8c-6**

Next, we focused on designing different truncated variants of the wild-type ribozymes R8c-3 and R8c-6 respectively, with the goal of removing extraneous elements and identifying an optimal minimal ribozyme sequence. The short hairpin in the predicted secondary structures of the 3'-termini of both Rc8-3 and Rc8-6 proved to be critical for its activity; removal of this motif renders both ribozymes entirely inactive, it probably engages in tertiary interactions with other motifs in the ribozyme sequences or facilitate those interactions by conformational rearrangement. The 5'-terminus of Rc8-6 tolerates no truncations either, indicating that the sequence-structure-function of this ribozyme is very tightly regulated. The more efficient Rc8-3 ribozyme can however be truncated by a maximum of 10 nucleotides from its 5'-terminus without significantly compromising activity.

The secondary structure predictions indicate the presence of two distinct and stable stems in Rc8-3, annotated S1 and S2 respectively (Fig.2-4). S1 is the longer and the more stable stem and can be progressively shortened by removal of 3 base-pairs and then 4 more base-pairs in combination with a C-U mismatch, without significantly impacting

equilibrium cleavage yields (Fig.2-6a). The resulting short stem S1-T2 has a C-A mismatch that when corrected to canonical C-G inhibits the cleavage activity completely. Therefore, this element favors a discontinuous stem over a highly stable and rigid stem, most probably to afford higher flexibility and stem-loop transitions. The shorter stem S2 has an internal bulge that can be removed to connect the two discontinuous parts of the



**Figure 2-6 Truncation studies on R8c-3.** a) Secondary structural representation of R8c-3 after removal of 10 nucleotides from the 5'-terminus shows two long stem elements S1 and S2. Removal of the boxed part in dark grey only generates S1-T1, removal of boxed parts in both dark and light grey generates S1-T2. Similarly, removal of only boxed part in dark blue generates S2-T1, and removal of both dark and light blue colored boxes generates S2-T2. S2-T2 is only active when combined with S1-T2 but not with S1-T1. b) the minimal catalytic motif identified based on R8c-3.



stem with some reduction in activity; moreover, S2 can be severely truncated to just a 4 base-pair stem provided S1 is simultaneously shortened to S1-T2. Shortening of S2 is incompatible when the longer stem of S1-T1 is retained, indicating complex synergistic interactions between S1 and S2, rather than a mere structural role of a stem. Both loops capping S1 and S2 and their shortened versions can be replaced by a general GAAA tetraloop. Figure 2-6b shows the minimal active motif for R8c-3 identified by mutation studies.

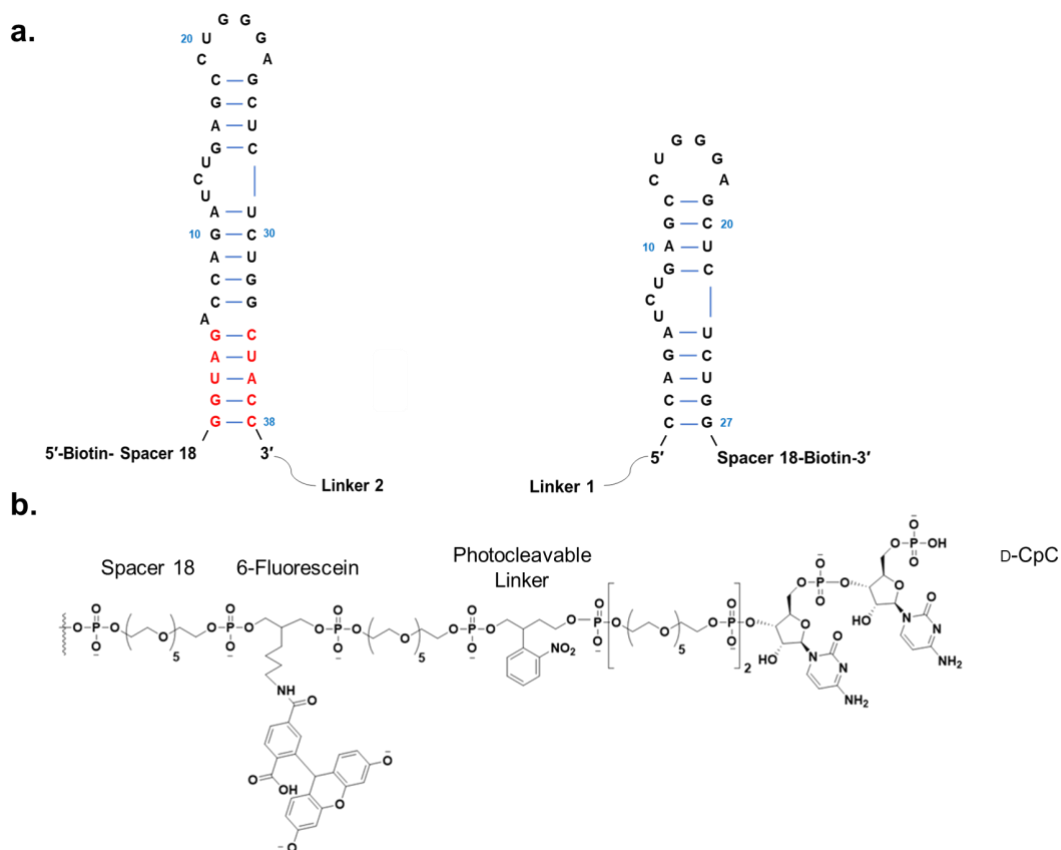
In case of R8c-6, some internal truncations in the long stem by the removal of 6 or 10 base-pairs resulted in the decrease of *cis* activity by 50% and 75% respectively (data not shown here). A detailed account of all the truncations tested on R8c-3 and their respective activities can be found in Table A-3 in the *Appendix*.

## **2.2.2. Selection for *trans*-cleaving Ribonuclease Ribozymes**

### **2.2.2.1. Reselection Substrate and Library Design based on truncated R8c-3**

Naturally occurring, self-cleaving ribozymes such as the hammerhead (HHR) have been repurposed to cleave RNA substrates *in trans*, usually by separating the covalent linkage between the substrate and the substrate-binding arm. DNAzymes evolved *via* directed evolution, although covalently linked to their substrate during the selection experiment, operate *in trans*, often with high turnover. This is possible because canonical WC base-pairing is still the premise of substrate recognition. In our case however, obtaining *trans* activity from a ribozyme originally selected for *cis* cleavage is less

plausible. Indeed, the *cis*-active ribozymes R8c-1, R8c-3 and R8c-6 did not show any *trans* cleavage even under saturating conditions of the ribozymes and high  $Mg^{2+}$  concentrations (data not shown). It is evident that in the absence of sequence guided substrate-binding through canonical WC base-pairing, removal of the physical tether compromises the co-localization of the substrate-ribozyme pair. A reselection was thus necessary to evolve for enhanced cross-chiral recognition between the D-ribozyme and the L-TAR hairpin followed by phosphodiester bond scission ability that the *cis* ribozymes have already acquired.



**Figure 2-7 Design of Reselection Target.** a) Secondary structure representation of L-TAR<sub>L</sub> compared to L-TAR<sub>S</sub> showing an additional 5 base-pairs stabilizing the stem. The 5'-end of L-TAR<sub>L</sub> is functionalized with biotin for immobilization of the substrate on streptavidin coated beads and the 3'-end is equipped with a linker that helps in monitoring the desired cleavage reaction. b) Chemical structure of the 3'-linker.

Retaining the same selection strategy as before, we redesigned the substrate to maximize the chances of isolating functional ribozymes with better on rates and (or) slower off rates. A longer version of the TAR hairpin was synthesized in the L-chirality, consisting of 38 nucleotides. This version that we will refer to as L-TAR<sub>L</sub> had a stem that was longer by 5 base-pairs and thus more stable compared to L-TAR<sub>S</sub> (See Fig.2-7). Since cross-chiral interactions rely heavily on structure, it is common that binding is facilitated by flexible, unpaired regions in the substrate that can form different structural motifs. On the other hand, rigid structures such as long and stable stems would be more challenging for binding. Indeed, when this longer substrate was tested with the *cis* active ribozymes, activities decreased dramatically (*Appendix Fig.A-3*). In addition, we reversed the polarity of attachment of the physical linker, conjugating the ribozyme library *via* the 3'-terminus of the substrate instead of the 5'-terminus as in the *cis* target. We reasoned that these two changes in substrate design would facilitate the enrichment of sequences that show enhanced binding to the substrate without significant dependence on the linker.

Two new random libraries were generated by replacing either the 5'-GAAA-3' tetraloop capping stem S1 or S2 respectively of the minimally truncated ribozyme R8c-3T13 by a 25 nucleotide long random region (See *Appendix Fig. A-4*). These two libraries were further mutagenized to generate more diversity. In addition, the primer regions were

also redesigned. These four libraries showed very low (<1%) cleavage activity *in cis* with the newly synthesized substrate. This was expected since R8c-3T13 itself is only marginally active with the new longer substrate. However, we anticipated recovery of activity following the evolution of these new libraries. Five rounds of selection were carried out, under the same experimental conditions of buffer composition, Mg<sup>2+</sup> concentration and temperature. The incubation times of each iteration of the selection experiment were decreased following the same progression as before. However, we saw no sufficient enrichment of the random library by Round 5. Since the activity was almost negligible, it was evident that the addition of the random domains was not beneficial to the binding and(or) folding of the active form.

#### **2.2.2.2. Revised Reselection Strategy based on Mutagenized R8c1, R8c-3 and R8c-6 Libraries**

Following the unsuccessful first attempt at reselection, we focused our attention on the wild type *cis*-ribozymes R8c-1, R8c-3 and R8c-6. They were validated against the L-TAR<sub>L</sub> substrate with R8c-3 having the most reduced activity compared to that *in cis* (only 16 % compared to ~70% in 3 hours) (*Appendix Fig.A-5*). Three new libraries were generated by mutagenizing the core sequence of R8c-1, R8c-3 and R8c-6 with 15% degeneracy while retaining the primer sequences of the *cis* selection library. In order to avoid enrichment of the wild-type ribozymes, the selective pressure was increased quite early in the selection. We started with a generous 3 hours of incubation time in Round 1 but rapidly decreased that to 30 mins in Round 3 and eventually on the order of a few seconds in Round 6 (See *Appendix Table A-2*). No appreciable activity was observed in

the R8c-1 library from Round 3 onwards and the rest of the selection was performed with the R8c-3 and R8c-6 libraries. A minimal *trans* activity was observed in the R8c-6 library, but not in R8c-3 at the end of Round 6; both libraries were subjected to a further three rounds of selection, with incubation times as short as manually possible. Each enriched library isolated from Round 6 onwards was assayed for *trans* activity. PCR amplified products from both the enriched libraries isolated from Round 6 and Round 9 were cloned into TOPO vector and about 20 unique colonies were sequenced (See *Appendix* Figure A-6 for detailed sequence information).

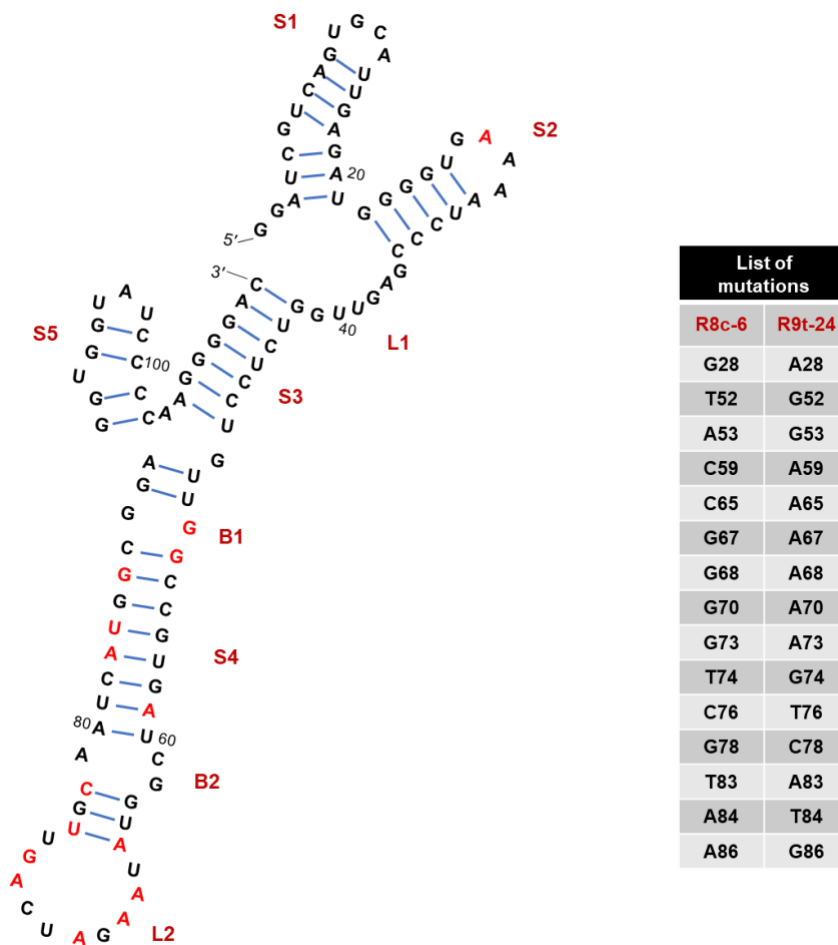
### 2.2.2.3. Active Ribozymes Isolated from Reselection: Sequence and Secondary Structure Considerations

Sequencing data from R8c-3 library revealed that there was no enrichment of any active sequences either in Round 6 or Round 9, and this was experimentally verified by assaying the *cis*-cleaving activity of this library, which was negligible compared to the wild-type ribozyme (data not shown). The library based on the R8c-6 library on the other hand, was enriched in three very closely related families of sequences (Fig 2-8).

	1	10	20	30	40	50	60	70	80	90	100	110	% Abundance
R8c-6 (WT)	GGATCGTCAGTGCATTGAGATGGGGTGGAAATCCCGAGTTGGTCTCCTGTTTACCGTCTCGGTCGTTGGGGTTCGGGAATCTAGACGGAGTGGTATCCCCAAGGGGAC												0
R9t-24	GGATCGTCAGTGCATTGAGA-----A-----GG-----A-----A-AA-A--AG-T-C---AT-G---GGTGGTATCCCCAAGGGGAC												28
R9t-23	GGATCGTCAGTGCATTGAGA-----G-----A--A-AA-A--A--A---A-AC-G---GGTGGTATCCCCAAGGGGAC												33
R9t-25	GGATCGTCAGTGCATTGAGA-----GG-----A--A-AA-A---T---G-AG-G---GGTGGTATCCCCAAGGGGAC												19
R9t-22	GGATCGTCAGTGCATTGAGA-----C-----A--A-G--C-A-G---T---TA--A-G-AGT---GGTGGTATCCCCAAGGGGAC												9

**Figure 2-8 Active Ribozyme Sequences from Reselection Round 9.** Comparison of sequence information from unique active ribozymes isolated from reselection to wild-type R8c-6.

To evaluate the *cis* and *trans* cleavage activity of the unique sequence R9t-21 and the other representative sequences, R9t-23, R9t-24 and R9t-25, we resynthesized the L-TAR<sub>L</sub> substrate to have the linker in the same orientation as in L-TAR<sub>S</sub> i.e. attached to the 5'-terminus of the substrate. This is to ensure a fair comparison to the



**Figure 2-9 Secondary structure representation of R9t-24.** This *cis* and *trans*-active ribozyme has several stable stem and hairpin structures. The residues in red show the location of mutations compared to the WT R8c-6 and the table lists the residues they replace

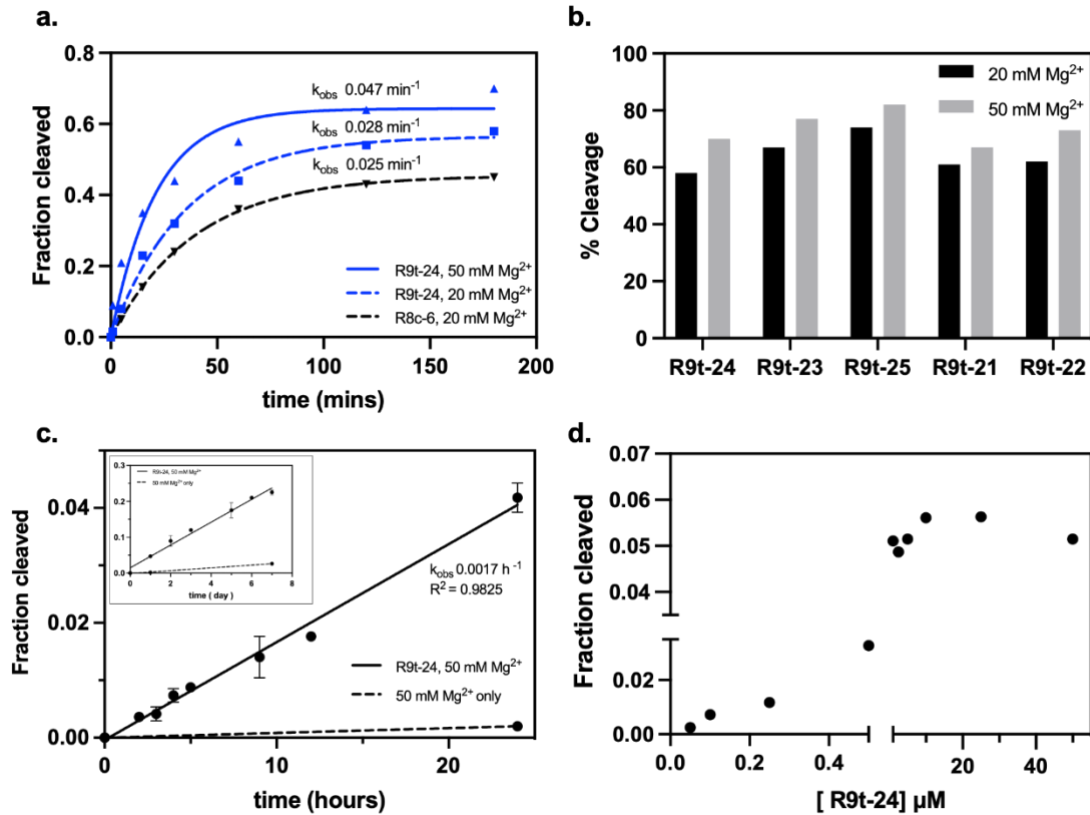
Wild type *cis*-ribozyme. According to the predicted secondary structure, a 5'-AUAAGA-3' motif that is predicted to be a loop (annotated L2 in Fig. 2-9) replaces the wild-type 5'-CUGGGG-3' and is highly conserved. Present in all but one sequence evolved in Round 9, this motif could have a role in enhancing the efficiency of the phosphodiester cleavage reaction, since these three families have very high *cis*-activities with equilibrium yields between 60-70 % and 67-82 % at 3 hours, at 20 mM and 50 mM Mg<sup>2+</sup> respectively (Fig 2-10b). The unique sequence R9t-22 lacking this motif, however, still has comparably high *cis*-activity. It is also possible that this motif enhances binding to the substrate. Indeed, preliminary in-line probing experiments indicate that this motif is increasingly protected from spontaneous hydrolysis under saturating conditions of the substrate (*Appendix* Fig.A-7). A few other mutations seem to be highly favored. One such example is two adjacent mutations at T52A53 to either G52G53 or G52 followed by deletion. Another is replacement of A84 to G84 in all the sequences. However, it is to be noted that most of the mutations accumulate after the first 50 nucleotides from the 5'-terminus, and even though there are subtle differences between the three main families in the nature of the mutations, the locations of the mutations are very similar.

#### **2.2.2.4. Characterization of *cis* and *trans*-cleaving activity of R9t-24**

All the sequences obtained from Round 9 of the reselection were *cis*-active but only R9t-24 had a weak *trans*-activity. Under the optimal conditions of 500 nM of substrate, 5 μM R9t-24 and 50 mM Mg<sup>2+</sup> at room temperature, 5% cleavage of either FAM labeled or radiolabeled L-TAR<sub>L</sub> was observed after 24 hours, with a pseudo-first order rate constant  $k_{obs}$  of 0.0017 h<sup>-1</sup> (See Fig. 2-10c). At prolonged reaction times spanning a week,

~20% of cleavage was observed. Although the *trans*-ribozyme is very slow at this point, it is encouraging that Mg<sup>2+</sup> mediated non-specific hydrolysis at the same site is 20 times slower. The activity is only slightly increased with the shorter substrate L-TARs and the



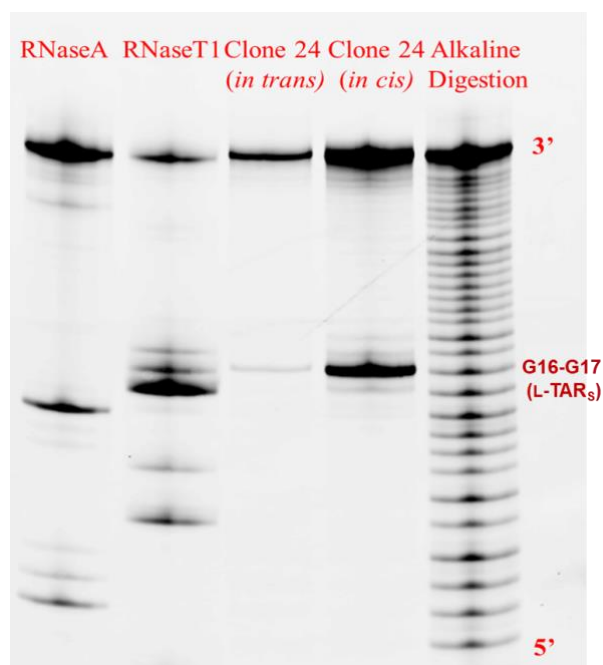


**Figure 2-10 Characterization of R9t24.** a) cis kinetics of R9t24 at 20mM and 50 mM Mg<sup>2+</sup> shows similar first order rate constant at lower Mg<sup>2+</sup> concentrations which doubles at 50 mM Mg<sup>2+</sup> b) comparison of equilibrium yields for the cis reactions of the three main families of ribozymes and two other unique sequences selected after 9 rounds of reselection at 20mM and 50 mM Mg<sup>2+</sup> shows much enhanced activities compared to the WT-clones. c) a time course of trans kinetics of R9t-24 at 500 nM of substrate, 5 μM of the ribozyme and 50 mM Mg<sup>2+</sup> at room temperature, spanning a day (main graph) and a week (left top inset) indicates very weak trans activities. The kinetics is fitted to a pseudo-first order rate equation. Catalyzed reaction is ~20 times faster than the non-specific hydrolysis. d) comparison of trans cleavage yields keeping the substrate constant at 500 nM and varying only the concentration of R9t-24 over 2 orders of magnitude show saturation around 5 μM.

presence or absence of the linker doesn't seem to make much of a difference either (data not shown here). This sequence has 15 mutated residues compared to the wild-type R8c-6 and shares many conserved point mutations and mutated sequence motifs with the other *cis*-active ribozymes. It does have a G to A transition mutation early in the 28th position

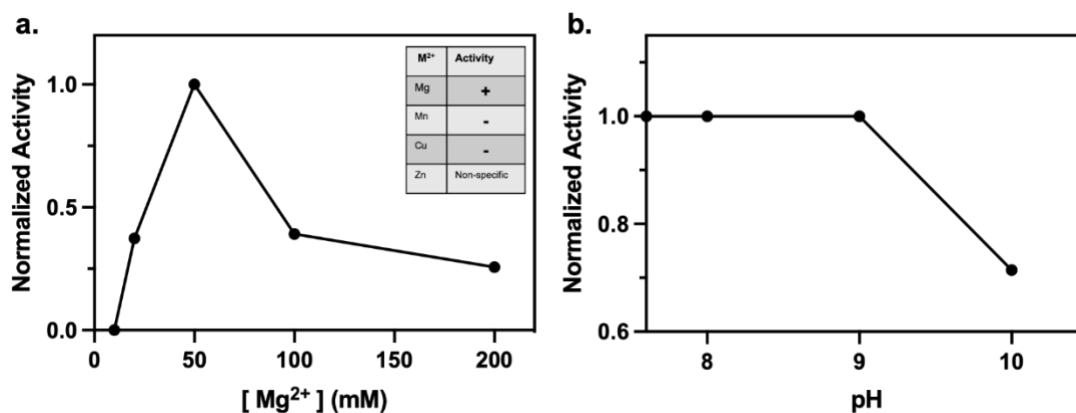
and has similarities as well as subtle differences in the pattern of mutations between nucleotides 70-86 with the *cis*-cleaving ribozymes. Therefore, it is difficult to specify which mutations were crucial for evolving *trans* activity. The *cis* cleaving activity of R9t-24 as well as the other representative sequences was much enhanced compared to the wild-type R8c-6 (Fig2-10a). Compared to a  $k_{\text{obs}}$  of  $0.025 \text{ min}^{-1}$  for the wild-type R8c-6 at 20 mM  $\text{Mg}^{2+}$  with the shorter substrate, R9t-24 had only slightly higher  $k_{\text{obs}}$  of  $0.028 \text{ min}^{-1}$  under the same conditions, albeit with the longer and more stable substrate. The enhanced efficiency is also reflected by the higher fraction of substrate cleaved. The  $k_{\text{obs}}$  almost doubles at 50 mM  $\text{Mg}^{2+}$  to  $0.047 \text{ min}^{-1}$  with a much higher equilibrium yield of 70 %.

The cleavage site for both the *cis* and *trans* activity of R9t-24 is same as the wild-type *cis* cleaving ribozyme R8c-6 i.e between G16 and G17. As mentioned before, this is also the same site of cleavage for the other *cis* ribozymes R8c-1 and R8c-3 against L-TAR<sub>L</sub> as well as all other *cis* active ribozymes isolated from the Round 9 of reselection, all of which are mutated forms of R8c-6 (Fig 2-11). The cleavage products obtained from the *cis* reaction between R9t-24 and L-TAR<sub>L</sub>, were characterized by mass-spectroscopy and the respective masses of each fragment were consistent with the formation of a terminal 2'-3'-cyclic monophosphate (See Fig. A-8 in *Appendix*). However, it is not possible to say from only mass data whether this is finally hydrolyzed into a 3'-monophosphate and a 2'-hydroxyl.



**Figure 2-11 Cleavage sites of R9t-24 *in cis* and *in trans*.** Alkaline digestion ladder and RNase A and T1 digestions of the corresponding D-TAR<sub>L</sub> substrate shows that the cleavage site remains unchanged from that of WT R8c-6 between G16 and G17 (numbering is consistent with L-TAR<sub>L</sub> for ease of reference).

Next, we focused on extensive characterization of R9t-24 and studied the effect of systematically varying Mg<sup>2+</sup> concentrations, temperature, pH, saturating concentrations of either the ribozyme or the substrate on the *trans* activity. Although the reselection used 20 mM Mg<sup>2+</sup> throughout, the ribozyme was found to have a higher requirement of 50 mM Mg<sup>2+</sup> for an optimal *trans* activity, most probably to enhance the formation of the properly folded active ribozyme (Fig 2-12a). At higher concentrations of 100 mM Mg<sup>2+</sup>, metal mediated degradation sets in and leads to lower yield of the specific cleavage product. It is also to be noted that Mg<sup>2+</sup> when added during the annealing of the substrate with the



**Figure 2-12 Dependence of *trans*-activity on Mg<sup>2+</sup> concentration and pH.** a) Activity increases with increasing concentrations of Mg<sup>2+</sup> until saturation at 50 mM, after which there is a sharp decline most probably due to non-specific degradation. Right inset shows activity is highly specific for only Mg<sup>2+</sup> b) activity remains mostly constant from physiological pH to alkaline pH of 9.0 after which non-specific laddering takes over.

ribozyme also leads to lower yield, which could be due to once again, spontaneous hydrolysis at higher temperatures used for annealing. R89t thus has quite a narrow window of Mg<sup>2+</sup> for maintaining an optimal activity. As far as the metal cofactor goes, the ribozyme is active with only Mg<sup>2+</sup>. Other divalent cations that are commonly used by cleaving ribozymes/ DNAzymes such as Mn<sup>2+</sup>, Cu<sup>2+</sup> failed to elicit any activity whereas Zn<sup>2+</sup> led to non-specific laddering. The addition of crowding agents such as PEG 8K which are known to stabilize complex RNA architecture, did not have any effect on the activity either. Room temperature, which was maintained throughout the selection was appropriate for *trans* cleavage; lowering the temperatures to 16 °C did not affect the equilibrium yield

much, but higher temperatures of 37°C led to much lower activity. This was not entirely unexpected since metal mediated spontaneous hydrolysis will be faster at higher temperatures. Next, we varied the pH of the buffer system, keeping the buffer strength the same, from physiological 7.6 to 10. Interestingly, the fraction cleaved remained the same throughout until it started decreasing after 9 due to background hydrolysis (Fig 2-12b). Finally, we wanted to see how the activity varies under single turnover as well as multiple turnover conditions. As we can see in the left half of the x axis in Fig. 2-10d, when the substrate is fixed at 500 nM, keeping other conditions constant, there is very little cleavage under single turnover conditions of ribozyme concentrations between 0-500 nM. When the substrate is saturated with the enzyme, activity starts increasing until it saturates and plateaus off at 10 times excess of the ribozyme at 5µM. The  $K_m$  or the binding affinity of the ribozyme for the substrate is thus very low at this point.

This is still an encouraging result, since an indirect strategy designed to obtain *cis* cleavage is still able to evolve cross-chiral interactions mediated *trans* recognition and cleavage activity. Preliminary structural studies using in line probing techniques reveal certain key residues involved in interaction with the substrate.

#### **2.2.2.5. Effect of Triphosphorylation on activity**

A final interesting and serendipitous discovery was made during attempts to assemble the 110 nucleotide long L-R9t-24 (not discussed here). Although the mass-spectrometric data showed the formation of the desired ribozyme in L-chirality, there was no discernible cleavage either *in cis* or *in trans*. Since, the only difference between solid-

phase synthesized (and the individual fragments of the L-ribozyme ligated) L-ribozyme and the *in vitro* transcribed D-ribozyme would be the presence of the 5'-triphosphate in the D-ribozyme, we decided to see if that was essential to the ribozyme activity. Much to our surprise, the absence of the 5'-triphosphate eliminates all *trans* activity and severely inhibits *cis* activity of R9t-24 (data in Figure A-9 in *Appendix*). The presence of a 5'-monophosphate rescues the *cis* activity to a certain extent but fails to restore the *trans* activity. In contrast, the wild-type *cis*-ribozymes have no significant dependence on the triphosphate group. The 5'-triphosphate could coordinate  $Mg^{2+}$  ions and thus contribute to the proper folding of the ribozyme, which is probably more crucial to the weak *trans* activity. However, it seems to have a requirement of the triphosphate for only the *cis*-ribozymes evolved during the reselection, which is surprising given the availability of the 5'-triphosphate and the use of exact same selection strategy in both cases. The longer substrate in the reselection could have been the selective pressure responsible for this certain evolutionary feature.

#### **2.2.2.6. Truncation Studies of R9t-24**

As with the wild type R8c-6 *cis*-ribozyme, the more evolved R9t-24 was resistant to terminal truncations. This is perhaps not unexpected, given that the predicted secondary structure remains mostly unchanged. Internal truncations aimed at systematically removing or shortening predicted secondary structure motifs such as the short hairpins S1, S2 or S5 or the stem regions of S3 and S4 rendered the ribozyme inactive. However, several A-U base-pairs in these stems when replaced by G-C base-pairs can stabilize the stem further, especially when introduced in the middle of the stem and enhance the *trans*

activity to 10% compared to the original 5% activity. Some of these mutations can be combined for the same yield. However, as is the case with complex ribozyme structures, there is a delicate thermodynamic balance between the different structural elements and their disruption inhibits function. For example, the bulge B3 must require some flexibility, evident by the inactive mutant generated by replacing U60-A80 by a stronger G-C base-pair. Similarly, some of these mutations forming more stable stems when designed into the ribozyme in certain combinations, completely inhibits activity such as two A-U to 2 G-C changes in stems S3 and S4 respectively. Some of these truncations are depicted in Fig A-9 in the *Appendix*. Therefore, R9t-24 is a complex ribozyme employing multiple structural interactions to carry out catalysis of phosphodiester bond cleavage.

### **2.3. Summary and Outlook**

In conclusion, heterochiral ribonuclease ribozymes were evolved to cleave a structured RNA target, the HIV-1 TAR RNA element of the opposite chirality. Using the “mirror-image” selection methodology, a random D-RNA library consisting of approximately  $10^{14}$  unique sequences were subjected to a total of 8 rounds of *in vitro* selection in the first phase to yield 4 unique *cis*-active ribozymes. Biotin-streptavidin interactions were exploited to physically sequester the inactive RNA pool on streptavidin coated magnetic beads whereas the active ribozyme sequences fell off the beads after cleavage and were collected from the supernatant. This straightforward selection strategy, coupled with an additional denaturing PAGE analysis of the cleaved fragments allowed targeted amplification of site-specific target cleaving ribozymes. The selection used 20

mM Mg<sup>2+</sup> as a divalent metal cofactor, and the main selective pressure throughout was decreasing times of incubation with Mg containing buffer.

It took only 6 rounds in the first *cis* selection for sufficient enrichment of the RNA library, and by round 8 we had increased the stringency to the maximum, by decreasing the incubation time to a matter of seconds. There were 3 unique sequences evolved at this stage, having high equilibrium yields and first-order rate kinetics in the range of 0.013-0.025 min<sup>-1</sup>. A very interesting observation was the diversity in site selection. The unpaired loop is obviously the most accessible site on the TAR hairpin and have been previously shown to be amenable to heterochiral interactions with the corresponding cross-chiral RNA aptamer. The three most efficient *cis*-ribonucleases in our case were also found to cleave this 6-nt distal loop, but very interestingly, at two distinct sites. One of them is a pyrimidine-purine junction (U-G) and the other is a purine-purine (G-G) junction. Previously selected DNAzymes often have biases towards the cleavage junction, in part because the cleavage junction is specified as part of the selection design. The advantage of our approach is that we introduced no such bias and allowed the RNA library to sample multiple conformations with the flexibility to have the entire unpaired loop available. Thus, evolution of two distinct cleavage sites poses the question of whether this was a chance outcome, or a natural consequence of specific structural interactions in the immediate environment of the cleavage sites, modulated by the electrostatic and steric effects associated with the specific dinucleotide junction under consideration. Moreover, truncations of even a few nucleotides from each end, and some internal truncations



eliminate activity completely. Overall, this does reveal complex structural interactions at play that have been so far unexplored.

The same *cis*-selection strategy on the other hand yields highly efficient *cis*-ribonucleases from mutagenized libraries based on the wild-type clones. What is interesting is that the three main sequences emerging from this reselection share some well-conserved motifs. There is however, only one *trans*-cleaving ribozyme and it is not clear what led to the evolution of such unique behavior. At this stage, it is difficult to pinpoint exactly which motifs are responsible for the *trans* activity. Although this activity is very weak at this point, with the equilibrium yield reaching a mere 10% in 24 hours under the optimized conditions with the best variant of the *trans*-ribozyme, the ribozyme is not a true catalyst at this point being unable to operate under multiple turnover conditions. This is most probably due to very weak affinity towards the substrate which can be improved by further evolution through mutagenesis or incorporating the cross-chiral TAR aptamer in the library design.

We conclude therefore, that heterochiral interactions when applied to the selection of a heterochiral ribozyme with common functionalities such as RNA cleavage, can reveal unique modes of recognition that can be exploited to address fundamental questions of RNA-RNA interactions. Moreover, “flipping the mirror” to synthesize the L-ribonuclease ribozyme would be a valuable biomedical tool against the native HIV-1 TAR RNA. Overall, as the second known example of a true heterochiral ribozyme, this work represents a significant and exciting addition to the repertoire of heterochiral nucleic acids-based technology with promising future applications in biomedicine and nanotechnology.

## 2.4. Materials and Methods

### 2.4.1. Materials

Oligonucleotides were either purchased from Integrated DNA Technologies (IDT, Coralville, IA) or purchased in-house, using standard solid-phase oligonucleotide chemistry on an Expedite 8909 DNA/RNAsynthesizer (Biolytic). Synthesizer reagents as well as D-nucleosides and specialty phosphoramidites such as 6-FAM, PC Linker, Spacer 18, 3'--Biotin, 3'--Phosphate resin, 5'--Biotin phosphoramidite and chemical phosphorylation reagent were purchased from Glen Research (Sterling, Va). L-nucleoside phosphoramidites were purchased from ChemGenes Corp. (Wilmington, MA). Histidine-tagged T7 RNA polymerase was purified from *E. coli* strain BL21 containing plasmid pBH161 (generously provided by Dr. Gerald Joyce, The Salk Institute of Biological Studies). *Thermus aquaticus* (Taq) DNA polymerase was prepared as described previously. Superscript II reverse transcriptase (nuclease deficient strain) and streptavidin coated High-Capacity Agarose Beads and magnetic beads (MyOne Streptavidin C1 Dynabeads X) were purchased from Life Technologies (Carlsbad, CA). T4 RNA Ligase 1 was purchased from NEB. [ $\gamma$ -<sup>32</sup>-P] ATP was purchased from Perkin Elmer (Waltham, MA). dNTPs and NTPs were purchased from Sigma Aldrich (St. Louis, MO). Deprotection reagents i.e aqueous Ammonium hydroxide, methylamine, inorganic Pyrophosphatase (IPP), glycogen and tRNA were also purchased from Sigma Aldrich. Electrophoresis reagents 40% 19:1 Acrylamide/ Bisacrylamide, APS and TEMED were purchased from Bio-Rad Laboratories (Hercules, CA).

### **2.4.2. Analytical Services**

Mass spectrometric analysis of oligonucleotides was performed by Novatia LLC (Newton, PA). Enriched RNA libraries were first cloned using TOPO TA Vector Kit purchased from Invitrogen (Waltham, CA) and sequenced by Eton Biosciences, Inc. (San Diego, CA).

### **2.4.3. Oligonucleotide Synthesis**

Oligonucleotides purchased from the IDT (except random DNA libraries) were purified by 20% or 10% 19:1 denaturing polyacrylamide gel electrophoresis (PAGE), eluted overnight from the gel in 1X Elution Buffer containing 100 mM NaCl, 10 mM TRIS (pH 7.5) and 10 mM EDTA. Reaction mixtures from transcription and ligation were also similarly purified using denaturing PAGE to get rid of excess NTPs and aborted transcripts. Following elution, the solution was filtered through sterile 0.2  $\mu\text{m}$  syringe filters and the eluate was either directly precipitated or after Amicon filtration, using a final 0.3M sodium acetate buffer (pH 5.3) and 3 volumes of 100% molecular biology grade ethanol. The pellet obtained after spinning down the precipitated solution was reconstituted to the desired working concentration using Milli-Q water. Quantification was done by measuring absorbance at 260 nm using the NanoDrop Spectrophotometer (Thermo Scientific).

Any oligonucleotides containing specialty amidites were synthesized on the Expedite 8909 and deprotected following the manufacturers' recommended coupling times and deprotection conditions. A standard 2.5- 3mins of coupling time was used for

D-DNA and L-DNA phosphoramidite coupling and an extended 15 min coupling time was used for D-RNA, L-RNA or any other specialty phosphoramidites used. Light-sensitive phosphoramidites such as the FAM and the PC Linker were protected from light as and when required. Mutagenesis libraries for reselection were prepared in house following literature protocols.

#### **2.4.4. Library Preparation**

##### **2.4.4.1. Transcription of Random D-DNA Library to D-RNA Library**

A ds-DNA library was first generated by a templated extension of 1.5 nmols of the reverse primer Rev161 on 1nmol of the template R161. The extension reaction was performed in a 1 ml volume containing 10U/ul Superscript II reverse transcriptase, 3mM MgCl<sub>2</sub>, 75 mM KCl, 10 mM DTT, 50 mM Tris (pH 8.3) and 0.5 mM each of the four dNTPs for 1 hour at 42 °C. The reaction mixture was ethanol precipitated and the pellet was reconstituted in double-distilled water to be directly used in a 2 ml transcription reaction containing 10 U/ul T7 RNA polymerase, 0.001 U/ul inorganic pyrophosphatase (IPP), 25 mM MgCl<sub>2</sub>, 2 mM spermidine, 10 mM DTT, 40 mM Tris (pH 7.9) and 5 mM each of the four NTPs. The reaction mixture was incubated for 2.5 hours at 37 °C, pelleted after ethanol precipitation and purified by denaturing PAGE.

##### **2.4.4.2. Extension of D-RNA library and Ligation to L-RNA substrate**

24 nmols of the *in vitro* transcribed RNA from the previous step was used as a template for extension by 48 nmols of Rev161\_Ext, in a 4.5 ml reaction using the same

conditions as described in the previous paragraph. The extended RNA library is then concentrated and desalted by Amicon filtration to eliminate excess dNTPs and the Rev161\_Ext primer. Next, the D-DNA-RNA hybrid was ligated with 6 nmols of the L-TAR RNA substrate functionalized with a 5'--phosphate in a 2ml reaction mixture containing 10 U/ul T4 RNA Ligase 1, 50 $\mu$ M ATP, 10mM MgCl<sub>2</sub>, 10% DMSO, 1mM DTT and 50 mM TRIS-HCl (pH 7.6) for 16 hours at room temperature. At the end of 16 hours of reaction, the Mg was quenched by addition of 40  $\mu$ l of 0.5M EDTA, concentrated and desalted to a final volume of 250 $\mu$ l by Amicon filtration.

#### **2.4.5. In vitro Selection**

In the first round of selection the D-RNA-L-RNA ligated assembly was added to a mixture of 300  $\mu$ l of High-Capacity Streptavidin Agarose Beads and 1.2 ml of 1X Selection Buffer (1X SB) containing 1mM EDTA, 100mM NaCl, 0.05 % Tween 20 and 25 mM TRIS-HCl (pH 7.6), and incubated at room temperature on a tilter for ~2 hours. The immobilization of the biotin functionalized assembly on the streptavidin coated beads was checked by running a small aliquot of the settled solution (without beads) before and after the incubation. After 2 hrs, there was a negligible amount of unbound substrate-library. The mixture was transferred to a poly prep column (Bio-Rad) and the solution allowed to settle and the flow-through was collected. The beads were washed successively with 2 ml of 1X SB buffer three times, 1 ml of double-distilled water two times and again with 1 ml of 1X SB buffer. This ensures complete removal of unbound material. The beads were next washed with alternate washes of 1 ml of freshly prepared, ice-cold NaOH buffer

(25 mM NaOH and 1mM EDTA) and 1 ml of freshly prepared, ice-cold Urea Wash Buffer (8M Urea, 1mM EDTA, 0.05% Tween 20 and 25 mM TRIS-HCl pH 7.6), up to a total of 4 washes. This leads to the stripping away of the DNA strand of the RNA-DNA hybrid, while the D-RNA library ligated to the L-TAR substrate remains attached. The beads are further washed with 1 ml of 1X SB containing tRNA at a final concentration of 1mg/ml. This step is to prevent the stripped ssDNA to non-specifically bind to the agarose beads. The final washes are with 1 ml of 1X SB thrice and 1 ml of double-distilled water twice. It is to be noted that all washes are done by gentle pipetting of the solution a few times in order to disturb the beads from the settled solution homogeneously. Thus prepared, the beads are finally incubated with 1 ml of 1X SB adjusted to a final concentration of 20 mM  $Mg^{2+}$  by the addition of the appropriate amount of 1M  $MgCl_2$  for 3 hours, taking care to protect the material from light. This first eluate, containing any cleaved RNA from the selection step is removed after 3 hours, and the beads are incubated with a fresh 1 ml of 1X SB adjusted to a final concentration of 20 mM  $Mg^{2+}$  with 1M  $MgCl_2$  for 3 more hours. The beads are then washed with 2 consecutive washes of 500  $\mu$ l of the ice-cold Urea buffer to ensure denaturation and complete removal of cleaved RNA still held together by WC base-pairing. The eluates are combined, quenched with equivalent amounts of EDTA and ethanol-precipitated in the presence of 5  $\mu$ l glycogen (1mg/ ml, Roche Diagnostics, Indianapolis, IN).

The pellet is resuspended in 30  $\mu$ l of Milli-Q water and reverse-transcribed with R161 primer in a 50  $\mu$ l reaction using the same conditions as described before. This time the RT enzyme was heat-inactivated by incubating the reaction mixture at 65 °C for 10

mins. The reaction mixture was then directly carried into a 1 ml PCR reaction. The PCR conditions were adjusted to Step 1: incubation at 95 °C for 3 minutes, Step 2: 95 °C for 30 seconds, Step 3: primer annealing at 50 °C for 30 secs, Step 5: extension by polymerase at 72 °C for 1 minute, Step 6: GOTO Step 2 for  $N-1$  times, where  $N$  cycles were used to reach the desired amplification. The reaction was precipitated and resuspended in 50  $\mu$ l of 10 mM TRIS-HCl pH 7.6 and 1 mM EDTA (TE Buffer). 25  $\mu$ l of the PCR products were transcribed in a 100  $\mu$ l transcription reaction using the same conditions as described before, with the addition of 4  $\mu$ l of  $\alpha$ -<sup>32</sup>P-ATP to uniformly radiolabel the transcribed RNA. The transcribed RNA was purified by 10% denaturing PAGE, eluted, precipitated and quantified. From Round 2 onwards, 4 nmols of the transcribed RNA from the previous round was extended with 8 nmols of Rev161\_Ext primer in a 500  $\mu$ l reaction, concentrated and desalted by Amicon filtration, and ligated to 1 nmol of the L-TAR<sub>s</sub>/ L-TAR<sub>L</sub> RNA substrate in a 200  $\mu$ l ligation reaction using identical conditions as described before. Unlike in the first round, all the ligated libraries were first purified by denaturing PAGE and the band containing the full-length ligated library was cut out and eluted, concentrated by Amicon filtration, buffer-exchanged by washing twice with 1X SB and finally resuspended in 1X SB.

From Round 2 onwards, the gel purified ligated RNA library (~200-300 pmols, depending on the ligation efficiency) was used for selection on 250  $\mu$ l of MyOne C1 Streptavidin Magnetic beads. The procedure of the selection experiment was similar to that described in the first round except that all the incubations and individual washes were in 500  $\mu$ l volume. The individual round conditions are detailed in the *Appendix*.

#### *Gel based Selection:*

The eluted RNA from the selection experiment was precipitated as described, redissolved in formamide buffer, and half was irradiated with a UV lamp at 360 nm for 15 mins. Both the irradiated and non-irradiated fragments were run on the same denaturing PAGE gel and their migrations compared to markers corresponding to the single-stranded full-length L-TAR/ D-RNA ligation product and the D-RNA library alone. The bands corresponding to both irradiated and non-irradiated fragments were cut out and eluted together until otherwise mentioned.

#### **2.4.6. Cloning Sequences:**

PCR products from enriched libraries were transformed into the DH5a strain of competent E. coli cells using the TOPO Vector Kit and grown on agar plates containing 50 µg/ml Kanamycin. Individual colonies were picked and sent for sequencing. The data was analyzed and aligned using the open access softwares Geneious and Mega7.

#### **2.4.7. Preparation of Individual Clones/ Mutations/ Truncations for Characterization**

Once the sequences of the individual clones were known, the respective ssDNA templates were ordered from the IDT and PCR amplified. For mutations and truncations, single-stranded cross-extension templates were designed and ordered from the IDT.



#### **2.4.8. Assaying of *cis* cleavage activity**

For evaluation of *cis* cleaving activities, the enriched D-RNA library or the individual D-ribozyme sequence is first ligated to the FAM labeled L-TARs as described before and resuspended in 1X SB Buffer after purification. Concentrations are maintained at approximately 500 nM to 1  $\mu$ M. The solution is annealed by heating at 72°C for 2 mins and then slowly cooled to room temperature. Addition of 1M MgCl<sub>2</sub> was done according to the desired final concentration of Mg<sup>2+</sup>. At predetermined time intervals, 2  $\mu$ l of the reaction mixture is aliquoted into 8  $\mu$ l of the formamide buffer and is run down denaturing PAGE. Alternatively, the same sample can be photocleaved and then run down PAGE. An aliquot of the reaction prior to Mg addition serves as the time 0 control. The percent or fraction of substrate cleaved can be determined by comparing the signal-intensities of the full-length ligation product (or L-TAR<sub>L</sub> from the irradiated fragment) and the cleaved fragment (or only the cleaved substrate after removal of the D-RNA library by photocleavage) by ImageQuant Software.

#### **2.4.9. Kinetic Analysis of *cis* cleavage**

The experiment was set up exactly as described above, and aliquots corresponding to different time points were collected. After quantification of the signal intensities from the corresponding bands on PAGE, the fraction cleaved or percent cleaved was calculated and plotted against time. The rate constant of the first-order rate equation was derived by fitting the data to the equation (1)

$$P_t = P_f (1 - e^{-k_{obs}t}) \quad (1)$$

Where  $P_t$  is the fraction cleaved at time  $t$ ,  $P_f$  is the plateau value of the fitted curve and  $k_{obs}$  is the first-order rate constant.

#### **2.4.10. Characterization of Cleavage Sites and Cleavage Products**

Cleavage products obtained from either *cis* or *trans* cleavage reaction were first photolyzed at 360 nm wavelength for 15 mins and run down 20 % denaturing PAGE. Alkaline digestion of the same substrate as that used in the assay was performed by heating the substrate in 50 mM sodium carbonate buffer (pH 9.0) at 95°C for 5 mins in the presence of excess tRNA, and then quenching by the addition of 1M TRIS-HCl (pH 7.6), 0.3 M NaOAc and 3 volume equivalents of 100 % EtOH. Similarly, RNase A and T1 ladders were generated by digestion of the identical D-TAR substrate using standard protocols. A large-scale *cis* reaction was carried out on 2 nmols of L-TAR<sub>L</sub> (modified linker) and the cleavage products resolved on PAGE and eluted. Both fragments were sent for mass spectrometric analysis.

#### **2.4.11. Trans Assays of R9t-24 and related truncations and mutations**

For kinetic analysis of R9t-24, approximately 1 μM of L-TAR<sub>L</sub> was dissolved in 5 μl of 1X SB buffer, and 10 μM of the *trans* ribozyme was dissolved in 5 μl and then the two solutions were separately annealed. Then both solutions were mixed and activated by the addition of 1M MgCl<sub>2</sub> to the desired final concentration. Aliquots were drawn at regular intervals, quenched with EDTA and formamide and resolved on denaturing PAGE.

Fraction cleaved was calculated as described before. For pH studies, the 1X SB buffer was simply adjusted to the different values by titrating TRIS-HCl with 1M NaOH to the desired value. To test the effect of different metals, the metal cofactor was replaced by the corresponding chloride salt of either  $Mn^{2+}$ ,  $Cu^{2+}$  and  $Zn^{2+}$ .

#### **2.4.12. Kinetic analysis of trans assays**

The pseudo first-order rate constant was calculated by fitting the time-course data of *trans* assays to the equation (2)

$$P_t = k'_{obs}t + c \quad (2)$$

where  $P_t$  is the fraction cleaved at time  $t$ ,  $k'_{obs}$  is the pseudo first-order rate constant and  $c$  is the y intercept determined by fitting the data to equation (2).

## 2.5. References

1. Ganser, L.R., Kelly, M.L., Herschlag, D. and Al-Hashimi, H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology*, 20, 474-489.
2. Gautheret, D., Konings, D. and Gutell, R.R. (1995) GU base pairing motifs in ribosomal RNA. *RNA*, 1, 807-814.
3. Moore, P.B. (1999) Structural motifs in RNA. *Annual Review of Biochemistry*, 68, 287-300.
4. Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q. and Fox, G.E. (2002) NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Research*, 30, 395-397.
5. Doherty, E.A. and Doudna, J.A. (2001) Ribozyme structures and mechanisms. *Annual Review of Biophysics and Biomolecular Structure*, 30, 457-475.
6. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31, 147-157.
7. Lilley, D.M.J. (2011) Mechanisms of RNA catalysis. *Philosophical Transactions of the Royal Society B Biological Sciences*, 366, 2910-2917.
8. Nakano, S.-i., Chadalavada, D.M. and Bevilacqua, P.C. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, 287, 1493-1497.
9. Peng, H., Latifi, B., Müller, S., Lupták, A. and Chen, I.A. (2021) Self-cleaving ribozymes: substrate specificity and synthetic biology applications. *RSC Chemical Biology*, 2, 1370-1383.
10. Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R. and Bruening, G. (1986) Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231, 1577-1580.
11. Shippy, R., Lockner, R., Farnsworth, M. and Hampel, A. (1999) The hairpin ribozyme. *Molecular Biotechnology*, 12, 117-129.
12. Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vornheim, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, 407, 327-339.

13. Joyce, G.F. (2009) Evolution in an RNA world. *Cold Spring Harbor Symposia Quantitative Biology*, 74, 17-23.
14. von Kiedrowski, G. (1986) A self-replicating hexadeoxynucleotide. *Angewandte Chemie International Edition in English*, 25, 932-935.
15. Akoopie, A. and Müller, U.F. (2016) Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *Physical Chemistry Chemical Physics*, 18, 20118-20125.
16. Ishida, S., Terasaka, N., Katoh, T. and Suga, H. (2020) An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nature Chemical Biology*, 16, 702-709.
17. Moretti, J. and Müller, U. (2014) A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. *Nucleic Acids Research*, 42, 4767-4778.
18. Scheitl, C.P.M., Ghaem Maghami, M., Lenz, A.-K. and Höbartner, C. (2020) Site-specific RNA methylation by a methyltransferase ribozyme. *Nature*, 587, 663-667.
19. Tjhung, K.F., Shokhirev, M.N., Horning, D.P. and Joyce, G.F. (2020) An RNA polymerase ribozyme that synthesizes its own ancestor. *Proceedings of the National Academy of Sciences*, 117, 2906-2913.
20. Hammann, C., Luptak, A., Perreault, J. and De La Peña, M. (2012) The ubiquitous hammerhead ribozyme. *RNA*, 18, 871-885.
21. Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences*, 94, 4262-4266.
22. Nguyen, K., Wang, Y., England, W.E., Chaput, J.C. and Spitale, R.C. (2021) Allele-Specific RNA Knockdown with a Biologically Stable and Catalytically Efficient XNAzyme. *Journal of the American Chemical Society*, 143, 4519-4523.
23. Wang, Y., Nguyen, K., Spitale, R.C. and Chaput, J.C. (2021) A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nature Chemistry*, 13, 319-326.
24. Boiziau, C., Dausse, E., Yurchenko, L. and Toulmé, J.-J. (1999) DNA Aptamers selected against the HIV-1trans-activation-responsive RNA element form RNA-DNA kissing complexes. *Journal of Biological Chemistry*, 274, 12730-12737.
25. Davidson, A., Leeper, T.C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J.A. and Varani, G. (2009) Simultaneous recognition of HIV-1 TAR RNA

bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proceedings of the National Academy of Sciences*, 106, 11931-11936.

26. Garbesi, A., Hamy, F., Maffini, M., Albrecht, G. and Klimkait, T. (1998) TAR-RNA binding by HIV-1 Tat protein is selectively inhibited by its L -enantiomer. *Nucleic Acids Research*, 26, 2886-2890.

27. Young, B.E., Kundu, N. and Sczepanski, J.T. (2019) Mirror-Image Oligonucleotides: History and Emerging Applications. *Chemistry*, 25, 7981-7990.

28. Kabza, A.M., Kundu, N., Zhong, W. and Sczepanski, J.T. (2021) Integration of chemically modified nucleotides with DNA strand displacement reactions for applications in living systems. *WIREs Nanomedicine and Nanobiotechnology*, n/a, e1743.

29. Garbesi, A., Capobianco, M., Colonna, F., Tondelli, L., Arcamone, F., Manzini, G., Hilbers, C., Aelen, J. and Blommers, M. (1993) L-DNAs as potential antimessenger oligonucleotides: A reassessment. *Nucleic Acids Research*, 21, 4159-4165.

30. Kabza, A.M. and Sczepanski, J.T. (2017) An l-RNA Aptamer with Expanded Chemical Functionality that Inhibits MicroRNA Biogenesis. *ChemBioChem*, 18, 1824-1827.

31. Sczepanski, J.T. and Joyce, G.F. (2013) Binding of a Structured d-RNA Molecule by an l-RNA Aptamer. *Journal of the American Chemical Society*, 135, 13290-13293.

32. Sczepanski, J.T. and Joyce, G.F. (2015) Specific Inhibition of MicroRNA Processing Using l-RNA Aptamers. *Journal of the American Chemical Society*, 137, 16032-16037.

33. Sczepanski, J.T. and Joyce, G.F. (2014) A cross-chiral RNA polymerase ribozyme. *Nature*, 515, 440-442.

34. Tjhung, K.F., Sczepanski, J.T., Murtfeldt, E.R. and Joyce, G.F. (2020) RNA-catalyzed cross-chiral polymerization of RNA. *Journal of the American Chemical Society*, 142, 15331-15339.

35. Yamagami, R., Bingaman, J.L., Frankel, E.A. and Bevilacqua, P.C. (2018) Cellular conditions of weakly chelated magnesium ions strongly promote RNA stability and catalysis. *Nature Communications*, 9, 2149.

36. 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.

37. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31, 3406-3415.
38. Ausländer, S., Ketzer, P. and Hartig, J.S. (2010) A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Molecular BioSystems*, 6, 807-814.
39. Huang, X., Zhao, Y., Pu, Q., Liu, G., Peng, Y., Wang, F., Chen, G., Sun, M., Du, F. and Dong, J. (2019) Intracellular selection of trans-cleaving hammerhead ribozymes. *Nucleic Acids Research*, 47, 2514-2522.
40. Wang, Y., Ngor, A.K., Nikoomanzar, A. and Chaput, J.C. (2018) Evolution of a General RNA-Cleaving FANA Enzyme. *Nature Communications*, 9, 5067.
41. Breaker, R.R. and Joyce, G.F. (1994) A DNA enzyme that cleaves RNA. *Chemistry & Biology*, 1, 223-229.

### 3. KINETICS OF HETEROCHIRAL STRAND DISPLACEMENT FROM PNA-DNA HETERODUPLEXES<sup>1</sup>

#### 3.1. Introduction

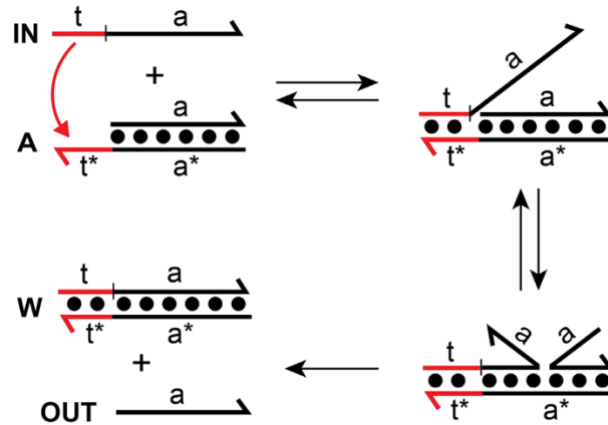
Rationally engineered, DNA-based molecular devices with reconfigurable parts constitute the core of dynamic DNA nanotechnology. By design, the different modules of these devices exist in a state of non-equilibrium. Upon perturbation by specific molecular signals, the modules interact with each other and their environment via programmed Watson-Crick (WC) base-pairing interactions. In particular, a molecular mechanism referred to as toehold-mediated strand displacement underlies the operation of most dynamic DNA-based devices reported to date (Figure 3-1) (1,2). Over two decades of research have established fundamental mechanisms and design principles for DNA strand displacement systems that has fueled the development of an impressive repertoire of molecular devices with complex functionalities, including motor activity (3–6), structural reconfiguration (7–10), Boolean logic computation (11,12), spatiotemporal signal resolution (13) and enzyme-free catalytic amplification (14,15).

Applications at the interface with biology represent the primary motivation behind the development of many dynamic DNA nanodevices. By being constructed of DNA, RNA or related analogues, these devices are inherently compatible with cellular nucleic

---

<sup>1</sup> Reprinted from “**Kinetics of heterochiral strand displacement from PNA–DNA heteroduplexes**” by Nandini Kundu<sup>†</sup>, Brian E. Young<sup>†</sup> and Jonathan T. Szepanski<sup>‡</sup>, **2021**. *Nucleic Acids Research*, Volume **49** (11), Pages 6114-6127. Open access article distributed under the terms of Creative Commons CC BY license. Copyright [2021] by Oxford University Press. (<sup>†</sup> Equal Contribution, <sup>‡</sup> Corresponding Author)





**Figure 3-1 Toehold-mediated strand displacement reaction.** DNA is depicted as lines with the half arrow indicating the 3' end throughout the text. A substrate strand consisting of a single-stranded toehold domain ( $t^*$ ) and a branch migration domain ( $a^*$ ) is initially hybridized to an incumbent strand (OUT) to form duplex A. The input (or invader) strand IN is complementary to both the toehold ( $t^*$ ) and branch migration domains ( $a^*$ ) of the substrate strand. Displacement is initiated by binding of IN to the toehold (via  $t/t^*$ ) followed by a three-way branch migration process in which base pairs between the incumbent and substrate strands dissociate and are replaced by base pairs with invader IN. The reaction is complete once the incumbent strand (OUT) is fully displaced from duplex A.

acids through WC base pairing interactions, thereby facilitating the interception and/or manipulation of molecular information in living systems. Use of DNA nanodevices for molecular sensing, imaging, and analysis of physiologically relevant biomarkers in fixed cells or tissues (16–18), surface of cell membranes (19,20), and even inside mammalian cells (21–23) has already been reported, representing significant progress in this direction. Yet, implementing DNA-based nanodevices within biological environments, and in particular, live cells, remains an ambitious undertaking. Exogenously delivered DNA has a cellular half-life on the order of minutes and is susceptible to unintended interactions with endogenous macromolecules (24), all of which adversely affect the performance of

the device. While use of chemical modifications, such as 2'-*O*-methyl ribonucleotides (25,26), locked nucleic acids (25) and phosphorothioate linkages (26) can confer nuclease stability, they also alter duplex thermostability (27) and hybridization kinetics (28,29) in an unpredictable manner. Indeed, compared to native DNA, strand displacement reactions involving chemically modified nucleic acids are poorly characterized, making the design of corresponding devices extremely challenging (22). Additionally, modified nucleotides can be toxic and tend to have adverse effects on cell viability (30,31). Due to these issues, modification-independent approaches for improving stability and retaining native hybridization parameters have also been explored, including the ligation of vulnerable free DNA ends and the use of more robust DNA architectures (32–34). However, these approaches have found only limited success. Importantly, none of the above approaches address potential off-target interactions of DNA-based devices with abundant cellular nucleic acids or other macromolecules, which further erode performance. Thus, there remains a need for new strategies aimed at improving the performance and reliability of DNA strand displacement systems within harsh biological environments.

L-DNA and L-RNA, the enantiomers (i.e., mirror images) of native D-nucleic acids, have recently emerged as promising alternatives to chemical modification for the development of biocompatible nucleic acid-based technologies (35). Due to the inverted stereochemistry of the (deoxy) ribose moiety, L-oligonucleotides are mostly orthogonal to the stereospecific environment of natural biology. Consequently, L-DNA and L-RNA are highly resistant to nuclease degradation and less susceptible to non-specific interactions with other proteins and cellular macromolecules (36–38). They also avoid off-target

hybridization with abundant cellular nucleic acids because oligonucleotides of opposite chirality (D versus L) are incapable of forming contiguous WC base pairs with each other (39–41). Importantly, as enantiomers, D and L-nucleic acids have the same physical properties, including solubility, hybridization kinetics, and duplex thermal stability, making them identical from a design perspective (36,37,42). Based on these characteristics, use of L-nucleic acids as alternative materials for constructing strand displacement systems circumvents many of the drawbacks associated with implementing this technology in biological matrices.

Inspired by this idea, we recently developed a toehold mediated strand displacement methodology for transferring sequence information between otherwise orthogonal oligonucleotide enantiomers (43). Our approach, termed ‘heterochiral’ strand displacement, relies on a heteroduplex between a *chiral* strand of L-DNA and an *achiral* strand of peptide nucleic acid (PNA), which hybridizes to DNA/RNA irrespective of chirality (Figure 3-2A). We refer to this complex (L-Ai) as the ‘inversion gate’. During the reaction, the D-nucleic acid input strand (D-IN) hybridizes to the inversion gate via the achiral toehold domain (1\*) on the PNA strand, leading to the displacement of the incumbent L-DNA strand (L-OUT) in the process. In this way, the sequence information within the D-input, and specifically domains 2 and 3, is ‘inverted’ into L-DNA. In principle, the inversion gate allows for any D-nucleic acid input signal to be sequence specifically interfaced with a robust nanodevice composed of bio-orthogonal L-DNA/RNA. For example, this approach has been used to interface microRNAs with L-

DNA-based logic circuits and catalytic amplifiers *in vitro* (43,44), and with an L-RNA - based fluorescent biosensor in live cells (45).

Although heterochiral strand displacement systems have shown remarkable promise, the key reaction strand displacement from a PNA–DNA heteroduplex remains poorly characterized, potentially limiting design capabilities. Indeed, a detailed understanding of all-DNA strand displacement kinetics and underlying biophysical mechanisms has greatly aided the rational design of dynamic DNA nanodevices having diverse behaviors. In the same way, establishment of a well-understood kinetic model of strand displacement from PNA–DNA heteroduplexes will be important for the rational design and optimization of dynamic L-DNA/RNA nanodevices that can be reliably interfaced with native biology. To this end, we determined the impact of several common design parameters, including toehold length and mismatches, on the kinetics of strand displacement from PNA–DNA heteroduplexes. To better understand how stereochemistry contributes to reaction kinetics, we directly compared homo and heterochiral reaction pathways (Figure 3-2A), wherein the input strand (D-IN) has the same or opposite stereochemistry as the PNA–DNA heteroduplex D-Ai and L-Ai, respectively. We show that the rate of strand displacement from PNA–DNA heteroduplexes can be tuned across several orders of magnitude based on the length of the PNA toehold, mismatch position, and stereochemical configuration of the reaction. Notably, heterochiral strand displacement reactions are slower than their homochiral equivalent, despite the overall change in free energy being identical. We experimentally investigate the source of this intriguing kinetic penalty and demonstrate how stereochemistry can be used to control

strand displacement kinetics. Despite having a slower rate, heterochiral strand displacement is highly sensitive to mismatches on the input strand, especially within the toehold domain, providing a potential advantage for engineering nucleic acid-based probes. Furthermore, heterochiral strand displacement rates are substantially enhanced when using RNA inputs, which represent the most common type of nucleic acid target for biosensing applications. Overall, this work establishes a basic set of design considerations to guide the future development of robust heterochiral strand displacement systems, thereby broadening the scope and applicability of L-DNA/RNA nanodevices for practical biomedical applications.

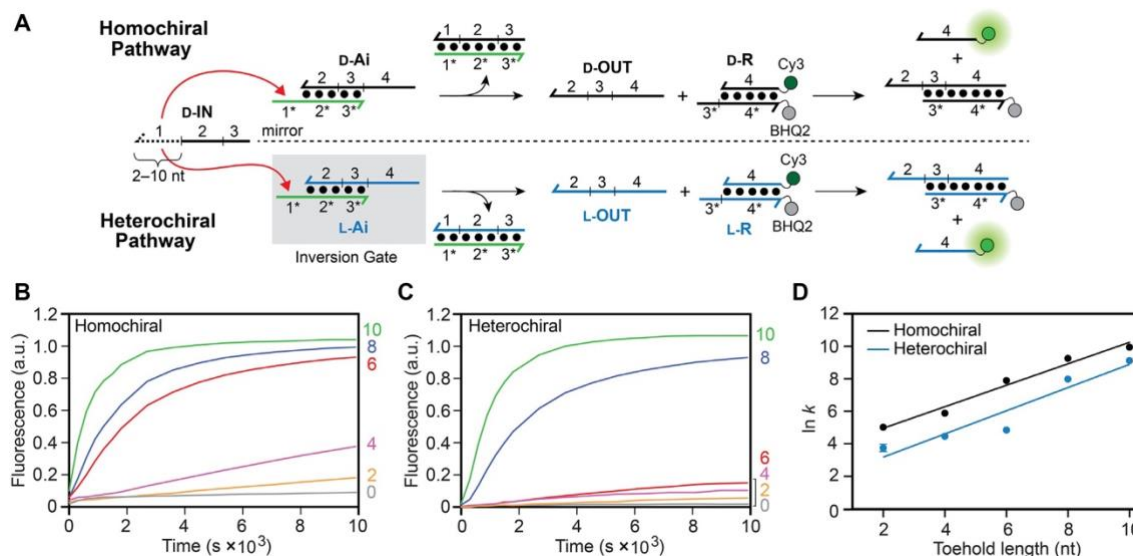
## **3.2. Results and Discussion**

### **3.2.1. Toehold length and stereochemistry modulate the kinetics of strand displacement from a PNA–DNA heteroduplex**

Toehold-mediated strand displacement reactions are initiated by binding of the input strand (also referred to as an ‘invader’) to the toehold domain. Toeholds increase the rate of strand displacement by increasing the probability that the incumbent strand is successfully replaced by the input strand once bound. As a result, the overall rate of the reaction, which can be approximated as being second order, is strongly dependent on toehold stability (46,52). For an all-DNA system, the rate of strand displacement can be adjusted over 6 orders of magnitude by simply changing the length and sequence of the toehold (46,53). Thus, the toeholds represent a key design parameter for kinetic control over engineering dynamic DNA devices.

To investigate the effects of toehold length on strand displacement from a PNA–DNA heteroduplex (D/L-Ai), we utilized a reaction system originally reported by Zhang and Winfree (Figure 3-2A) (46). The sequences were chosen (and confirmed via NUPACK) (54) to have no secondary structure within single-stranded regions to ensure decoupling of the second order process of strand displacement from the first order process of secondary structure unfolding. The reactions were monitored using an indirect reporter system in which the displaced incumbent strand (D/L-OUT) in turn displaces a fluorophore (Cy3)-labeled strand from the reporter duplex (D/L-R). This indirect strategy avoids the need to label primary reaction components, which could have unpredictable effects on their interactions (55,56). D-DNA inputs (D-IN) were used for all strand displacement reactions. Toehold length was varied by sequentially truncating the toehold domain (1) on the input strand from 10 to 0 nucleotides, whereas the length of the toehold domain (1\*) on the PNA strand of the heteroduplex (D/L-Ai) remained constant at 10 nucleotides. The chirality of the incumbent strand within the PNA–DNA heteroduplex was either D-DNA (D-OUT; for the homochiral reaction pathway) or L-DNA (L-OUT; for the heterochiral reaction pathway), and the two reaction pathways were monitored separately using either D-R or L-R reporter complex, respectively (Figure 3-2A). Because our goal is to provide design principles for heterochiral strand displacement devices that are compatible with living systems, and in particular human cells, all reactions were carried out under simulated physiological conditions (300 mM NaCl, pH 7.6 and 37°C).

Figure 3-2 shows how the rate of both homo and heterochiral strand displacement from a PNA–DNA heteroduplex depends on toehold length. For both reaction configurations, the rate dependency on toehold length is roughly exponential (Figure 3-2D), varying by up to two orders of magnitude for toeholds between 2 and 10 nucleotides long. Consistent with our previous observations (43), the homochiral reaction pathway is



**Figure 3-2 Strand displacement from a PNA–DNA heteroduplex is dependent on toehold length and stereochemistry.** (A) Schematic illustration of the strand displacement reaction system depicting both the homo- and heterochiral pathways. D-DNA is shown in black; L-DNA is shown in blue, and PNA is shown in green throughout the text. The half-arrow denotes the C-terminus of the PNA strand. The sequences of all strands are depicted in Appendix Figure B1 and Appendix Table B1. (B, C) Fluorescence monitoring (Cy3) of the homochiral (B) and heterochiral (C) reaction pathways initiated with inputs (D-IN) having toehold domains varying in length from 0–10 nucleotides (nt). The length of the toehold is indicated on the right y-axis. The reactions depicted contained 150 nM D-IN, 100 nM D/L-Ai, 300 nM D/L-R, 300 mM NaCl, 1 mM EDTA, and 10 mM Tris (pH 7.6) and were carried out at 37°C. Fluorescence in all figures is reported in units such that 1.0 is the fluorescence of the maximally activated reporter control and 0.0 is the background of the quenched reporter complex (D/L-R). (D) Semilogarithmic plot showing the exponential dependence of calculated rate constants on toehold length. All calculated rate constants are listed in Appendix Table B2.

faster than the heterochiral reaction pathway for any given toehold length. Taking the 6-nt toehold as an example, the homochiral strand displacement reaction is more than an order of magnitude faster than its heterochiral counterpart, despite their sequences and overall change in free energy being identical. This observation implies that helical inversion of the PNA–DNA heteroduplex during heterochiral strand displacement imposes an additional kinetic barrier relative to the homochiral reaction configuration, which may be due, in part, to weaker toehold interaction between the D-DNA input strand and L-Ai. We directly address the contribution of toehold binding on the observed reaction kinetics below. We note that no appreciable reaction occurred in the absence of an input (*Appendix Figure B3*).

Rate constants calculated for strand displacement from PNA–DNA heteroduplexes are considerably slower than for those previously reported for all DNA reactions for the same toehold length, regardless of reaction configuration (*Appendix Table B2*). Moreover, the rate fails to saturate for longer toeholds. For comparison, rate enhancement plateaus once the toehold becomes longer than ~6- nucleotides for traditional all DNA strand displacement reactions carried out at 23°C (46,52,53). We interpret these results as being partly due to the increased temperature at which our reactions were carried out (37°C) (57), and partly due to the greater enthalpic cost associated with disrupting PNA–DNA base pairs compared to DNA-DNA base pairs (58,59), which imposes a higher penalty for initiation and propagation of branch migration (52). Nevertheless, rate constants for longer PNA toeholds under these conditions still approach the lower bounds for what has been observed for all DNA strand-displacement reactions ( $\sim 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ), with potential for



further improvement. For example, the use of stronger toehold sequences (i.e., higher G/C content) and/or introduction of mismatches into the PNA–DNA heteroduplex (60) could be used to increase strand displacement kinetics without further increasing toehold length and will be the subject of future investigations.

Collectively, these data indicate that the rate constant for strand displacement from a PNA–DNA heteroduplex can be predictably tuned by adjusting toehold length and reaction configuration. Importantly, PNA toeholds  $\geq 8$  nucleotides long provide reaction rates that are sufficiently fast for most in vitro and intracellular applications of heterochiral strand displacement. Caution must be used when designing longer PNA toeholds to avoid undesirable secondary structures that impede binding to the input strand, as well as purine rich sequences that might promote aggregation (61).

### **3.2.2. Strand displacement from PNA–DNA heteroduplexes is sensitive to mismatches**

The rate of DNA strand displacement reactions can be modulated over several orders of magnitude by introducing one or more mismatches between the input strand and the target duplex (48,60). Rational positioning of mismatches provides a useful control mechanism for competitive reaction networks (60,62) and enables the design of strand displacement-based nucleic acid probes having a high degree of mismatch discrimination (63,64). A common practical application is the detection of single nucleotide polymorphisms (SNPs), which are of great diagnostic value (65–67). Given the design capabilities enabled by the incorporation of mismatches into all-DNA strand displacement

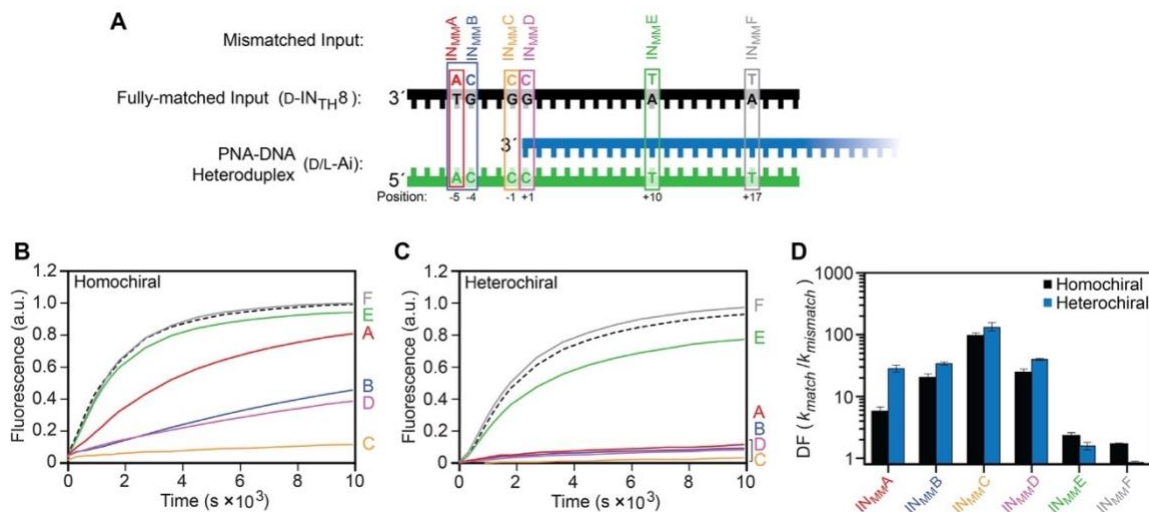
systems, we sought to characterize the influence of mismatches on the rate of homo and heterochiral strand displacement from a PNA–DNA heteroduplex. A series of inputs containing mismatches at a single position (and one input containing two mismatches within the toehold binding domain) were generated from input strand (D-IN<sub>TH8</sub>) having an 8-nucleotide toehold domain (Figure 3-3A). Placement of the mismatches were, in part, informed by previous literature on the effects of mismatches on rates of all DNA strand displacement reactions, as well as PNA–DNA hybridization (48,58,62,63).

We first considered inputs containing mismatches within the toehold domain. Compared to the homochiral reaction pathway, heterochiral strand displacement from a PNA–DNA heteroduplex is much more sensitive to a single mismatch positioned near the middle of the toehold domain (IN<sub>MM A</sub>), both in terms of rate and equilibrium yield (i.e., fraction D/L-OUT displaced) (Figure 3-3B, C and Appendix Table B2). This observation likely reflects an overall weaker toehold interaction between the input strand and the PNA–DNA heteroduplex of opposite chirality during the heterochiral reaction. The kinetic discrimination factors (DF) ( $DF = k_{\text{match}} / k_{\text{mismatch}}$ ) for IN<sub>MM A</sub> are 5.8 and 29.4 for the homo and heterochiral reactions, respectively (Figure 3-3D). For comparison, a DF of  $\sim 2$  has been reported for an all-DNA strand displacement system with a similarly positioned mismatch (63). This suggests that, regardless of the reaction configuration, strand displacement from PNA–DNA heteroduplexes is highly sensitive to single mismatches within the toehold domain. When two mismatches were present in the toehold (IN<sub>MM B</sub>), the DF for the homochiral reaction was greatly improved (DF = 20.8) but was still less than the heterochiral reaction (DF = 36.1) (Figure 3-3D). Interestingly, a mismatch

positioned immediately before the branch migration domain (IN<sub>MM</sub>C; position -1) almost completely inhibits strand displacement for both homo and heterochiral reaction pathways (DF > 100 for both reaction configurations). Computational studies suggest that coaxial stacking between the invading (input) and incumbent duplexes at the branch point (i.e. -1/+1) play an important role in the mechanism of branch migration initiation (52), as well as the kinetics of DNA hybridization (57). Therefore, a mismatch at position -1 (IN<sub>MM</sub>C) is expected to largely disrupt these interactions, greatly raising the activation barrier for initiating branch migration, and reducing the overall reaction rate.

We next examined inputs containing mismatches within the branch migration domain (Figure 3-3A). For both homo and heterochiral reaction pathways, introduction of a mismatch immediately adjacent to the toehold domain (+1; IN<sub>MM</sub>D) greatly impeded strand displacement (Figure 3-3B, C), which is consistent with all-DNA strand displacement reactions (48,56,62). While this mismatch does not compromise toehold stability, the system must enter an energetically less-favorable state to initiate branch migration because the input strand must immediately enclose a mismatch. This greatly reduces the probability that the input will successfully displace the incumbent strand prior to spontaneous detachment from the toehold, and thus, reduces the overall reaction rate. Again, the greater sensitivity of the heterochiral reaction configuration to this mismatch (IN<sub>MM</sub>D) is likely due to weaker toehold binding. Relative to IN<sub>MM</sub>D, mismatches within the middle or near the end of the branch migration domain (IN<sub>MM</sub> E and IN<sub>MM</sub> F) had only a modest effect on the rate of strand displacement for both reaction configurations (Figure 3-3D). Although inputs IN<sub>MM</sub> E and IN<sub>MM</sub> D must overcome similar energy barriers at the

site of the mismatch,  $IN_{MM E}$  has a more stable pre-mismatch state because it is able to form more base pairs with the PNA prior to encountering the mismatch. This increases the probability of successful displacement by  $IN_{MM E}$  relative to  $IN_{MM D}$ , leading to faster overall reaction kinetics and reduced mismatch discrimination. The inability of D/L-Ai to discriminate against  $IN_{MM F}$  can be explained by an alternative dissociation pathway for release of the incumbent strand as the branch point approaches the end of the branch migration domain, wherein spontaneous melting of the remaining base pairs provides a ‘shortcut’ for successful displacement (48,56). This process allows for displacement of the incumbent strand by  $IN_{MM F}$  before it encloses the mismatch site, and thus, its rate constant is expected to be similar to the perfectly matched input. We note that this trend for mismatch discrimination in the branch migration domain ( $DF = IN_{MM D} > IN_{MM E} > IN_{MM F}$ ) parallels what has been observed for all-DNA reactions systems (48,56,62).



**Figure 3-3 The position of a mismatch affects the rate of strand displacement from PNA–DNA heteroduplexes.** (A) Schematic of the mismatched inputs used in this study. The identity of the mismatched nucleotide relative to the fully matched input (D- $IN_{TH}8$ ) is shown above the strand. (B, C) Fluorescence monitoring (Cy3) of the homochiral (B) and heterochiral (C) reaction pathways initiated with different mismatched inputs in (A). The identity of the mismatched input ( $IN_{MM}A$ –F) is indicated on the right y-axis. The reaction initiated with the fully matched input (D- $IN_{TH}8$ ) is shown as a black dotted line. Reactions depicted here were carried out as described in Figure 3- 2. (D) Kinetic discrimination factors ( $DF = k_{match}/k_{mismatch}$ ) of hetero- and homo- chiral reactions towards inputs having different mismatches, where  $k_{match}$  and  $k_{mismatch}$  are the calculated rate constants for a fully-matched and mismatched input, respectively. Error bars represent standard deviation from three independent experiments.

Overall, these data demonstrate that the kinetics of strand displacement from PNA–DNA heteroduplexes can be controlled by mismatches within the input strand, and the relative reaction rate constants are dependent on the position of the mismatch and the stereochemical configuration of the reaction. Notably, for the sequence and toehold length examined herein, the heterochiral reaction configuration is intrinsically more sensitive to single mismatches within the toehold than its homochiral counterpart. Mismatches within

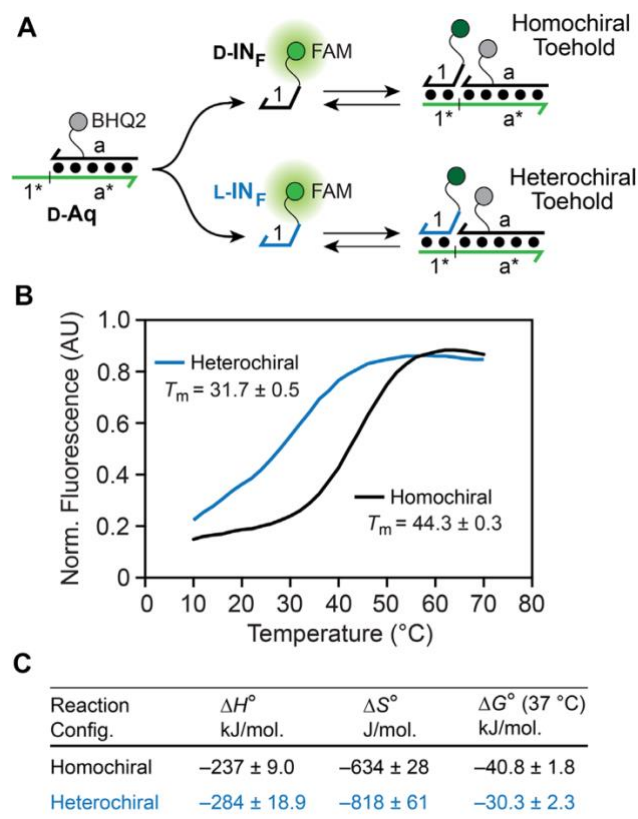
the toehold are rarely employed for kinetic control of strand displacement systems (68), including nucleic acid probes designed to detect SNPs. Instead, additional activation barriers are often introduced in the form of ‘remote toehold’ (68) or ‘toehold exchange’ strategies (69) to improve mismatch discrimination. Thus, our findings expand the design capabilities for kinetic control over strand displacement. Importantly, we expect that the improved mismatch discrimination observed for heterochiral strand displacement will confer high selectivity onto corresponding heterochiral DNA devices and probes—selectivity that can be further reinforced through the data provided herein.

### **3.2.3. Toehold stability is dependent on the reaction configuration**

In the current reaction system (Figure 3-2A), as well as those reported previously (43), we found that the rate of heterochiral strand displacement from PNA–DNA heteroduplexes is significantly slower than the corresponding homochiral reaction. Given that the overall reaction rate of toehold mediated strand displacement is strongly dependent on toehold stability, we hypothesized that differences in toehold binding energies between the homo and heterochiral reaction configurations are the major contributor to this rate disparity. To test this, we designed an experimental system that allowed us to monitor toehold binding in isolation from branch migration (Figure 3- 4A). The input strand (D/L- $IN_F$ ) lacks the branch migration domain found in D-IN (domains 2 and 3), instead containing five dT residues and a 5'-6-carboxyfluorescein (6-FAM) dye (*Appendix* Figure B1 and *Appendix* Table B1). Thus, binding of either D- $IN_F$  or L- $IN_F$  to the toehold of the PNA–DNA heteroduplex D-Aq results in formation of a three-strand complex that is

unable to proceed forward to displacement. Importantly, the D-DNA incumbent strand within heteroduplex D-Aq is labeled internally with a quencher (BHQ2) such that binding of either D- or L-IN<sub>F</sub> to the toehold domain (1\*) can be measured through fluorescence quenching.

Using this model system, we first determined the melting temperature ( $T_m$ ) of D-IN<sub>F</sub> and L-IN<sub>F</sub> with D-Aq, which correspond to the homo and heterochiral toehold configuration, respectively (Figure 3-4B). Remarkably, these data revealed that the  $T_m$  of the heterochiral toehold configuration (L-IN<sub>F</sub> + D-Aq) was  $\sim 12^\circ\text{C}$  lower than the corresponding homochiral toehold configuration (D-IN<sub>F</sub> + D-Aq) (Figure 3- 4B), despite their only difference being stereochemistry. In order to gain further insights, we determined the thermodynamics of toehold association for each configuration based on the concentration dependence of their melting profiles (50). The reciprocal  $T_m$  was plotted against the  $\ln C_t$  (total strand concentration) and fit to a linear relationship from which  $\Delta H^\circ$ ,  $\Delta S^\circ$  and  $\Delta G^\circ$  were derived according to established methods (*Appendix* Figure B4). As observed previously (70), formation of PNA–DNA toehold complex was accompanied by large enthalpy gains and entropy losses, in agreement with the formation of a more rigid duplex structure (Figure 3-4C). Notably, formation of the heterochiral toehold was associated with much greater entropy losses as compared to the homochiral toehold.



**Figure 3-4 Characterization of homo and heterochiral toehold interactions.** (A) The model system used to monitor toehold association based on fluorescence quenching. (B) Fluorescence melting curves for the homo and heterochiral toehold duplexes ( $C_t = 2 \mu\text{M}$ ). Fluorescence values were corrected ( $F_{\text{corr}}$ ) for background fluorescence and temperature-dependent effects as defined in eq 4 (see Materials and Methods). The  $T_m$  for each toehold configuration was averaged over three melting experiments. (C) Thermodynamic parameters for homo and heterochiral toehold association.

Together, these data clearly show that thermal stability of the toehold duplex (e.g., the duplex formed between domains 1 and 1\* on D/L-IN and D-Ai, respectively) is highly dependent on the toehold configuration, with the heterochiral toehold forming a far less stable complex with the input strand than in the homochiral toehold. Given the direct relationship between toehold stability and the rate of strand displacement (46), these



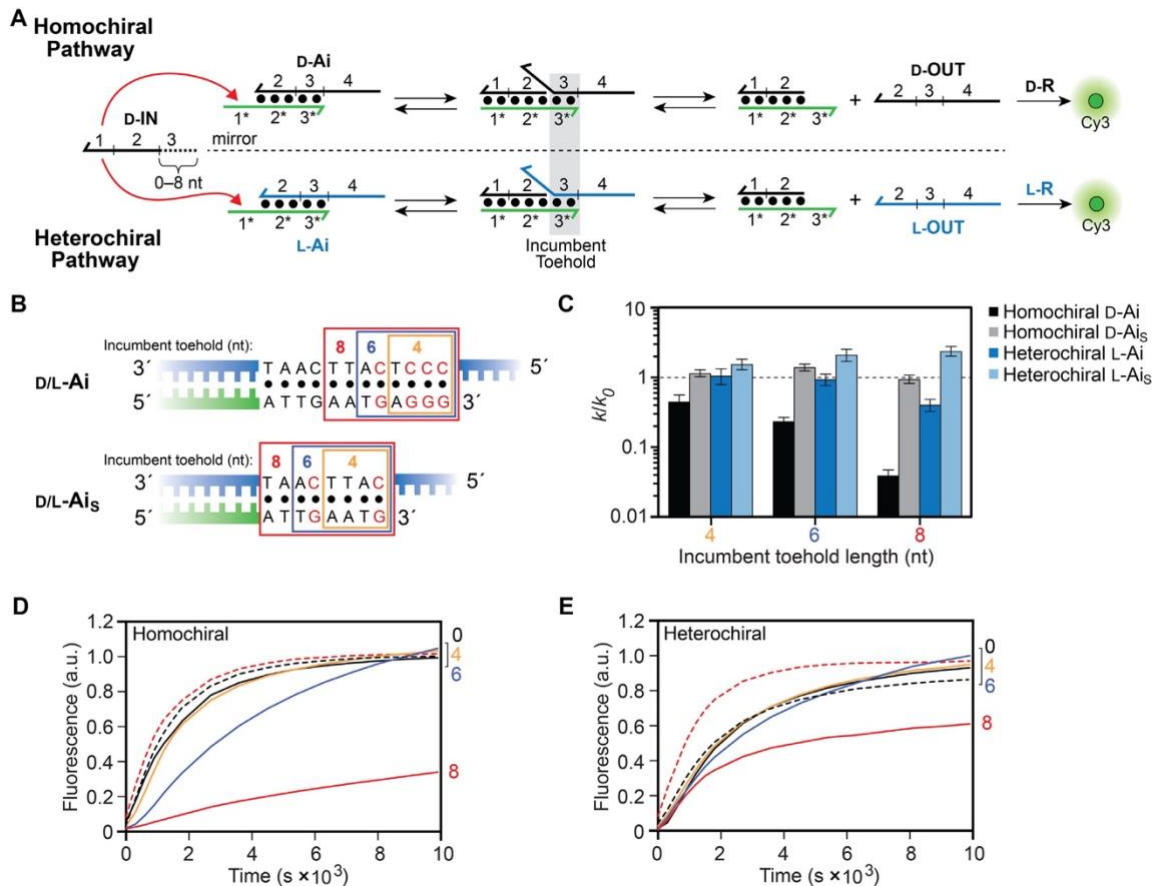
results explain, in part, why heterochiral strand displacement from PNA–DNA heteroduplexes is generally slower than the corresponding homochiral reaction and confirm toehold stability as a major contributor. This behavior may be rationalized according to the following considerations. Although PNA is achiral, upon hybridization to a chiral strand of D-DNA or L-DNA, the PNA will assume a right-handed or left-handed helical conformation, respectively (45,71). In the case of the left-handed PNA–DNA heteroduplex L-A<sub>i</sub>, the induced left-handedness in the PNA will propagate into the single stranded toehold domain through base stacking interactions (72,73). This gives rise to a ‘chiral conflict’ in which the right-handed input strand (D-IN) and left-handed PNA toehold domain in L-A<sub>i</sub> are unfavorably preorganized for binding. Consistently, our thermodynamic data indicates that the less-favorable energy for heterochiral toehold association stems predominantly from the entropic term, which is expected for the highly ordered transition state that would accompany helical inversion of the PNA strand. This is also in agreement with prior studies showing that PNAs with an induced left-handedness bind to native D-DNA more weakly than PNAs with induced right-handedness, which is attributed to the structural organization of the PNA (entropy effects) (70,73,74). In the context of the strand displacement reactions herein, these effects are manifested through slower reaction kinetics in the heterochiral configuration.

### **3.2.4. Strand displacement from PNA–DNA heteroduplexes is compatible with toehold exchange**

We sought to evaluate the compatibility of strand displacement from PNA–DNA heteroduplexes with the ‘toehold exchange’ mechanism (Figure 3- 5A) (15). In this type of reaction, a truncated input strand having an incomplete branch migration domain carries out only partial displacement of the incumbent strand. The remaining base pairs, referred to as the incumbent toehold (domains 3/3\* in Figure 3-5A), must then spontaneously dissociate for the reaction to complete. This mechanism has been evaluated extensively in the context of all-DNA strand displacement reactions (46) and provides improved control over strand displacement kinetics. The toehold exchange mechanism can also be exploited for catalysis (48,60). In presence of a fuel strand that can react with the incumbent toehold, the input strand can be regenerated over multiple turnovers. We envision those similar catalytic designs could be emulated using PNA–DNA heteroduplexes to develop heterochiral strand displacement devices and probes capable of signal amplification. Moreover, because the sequence of the incumbent toehold domain is independent of the input strand sequence, this approach may allow for construction of universal downstream reaction components (i.e., a modular design).

To demonstrate the potential for such designs using PNA–DNA heteroduplexes, we examined the rate of strand displacement initiated with versions of input D- IN<sub>TH</sub>8 (Figure 3-3A) that had been truncated by 4, 6 and 8 nucleotides from their 5' ends, resulting in incumbent toeholds (3\*) of corresponding length (Figure 3-5A, B). A four base pair incumbent toehold had little effect on the rate of either homo or heterochiral

reaction pathways relative to no incumbent toehold (i.e., a full-length input strand) (Figure 3-5C–E). This suggests that spontaneous detachment of the incumbent strand (D/L-OUT) likely occurs before the input makes significant contacts with the final four PNA–DNA base pairs of the heteroduplex (D/L-Ai). Increasing the length of the incumbent toehold further led to a decrease in strand displacement kinetics for both reaction configurations, especially for the 8 base pair incumbent toehold, which is the same length as the input toehold (1\*). This observation is consistent with model studies carried out on all DNA toehold exchange reactions: As the length of the incumbent toehold (3\*) approaches that of the input toehold (1\*), and the relative binding energies of both toeholds become similar, the probability of the input strand displacing the incumbent strand decreases (along with the rate) (46). Interestingly, the heterochiral reaction pathway was far less sensitive to the 8 bp incumbent toehold than its homochiral equivalent, with rate constants for strand displacement decreasing by 30-fold and 2.5-fold for the homo and heterochiral reactions, respectively, relative to no incumbent toehold (Figure 3-5C). We attribute this result to weakened toehold-binding interactions for the heterochiral reaction configuration, which also extends to the incumbent toehold. Destabilization of the incumbent toehold is expected to promote displacement of the incumbent strand (L-OUT) and impede its reassociation with the incumbent toehold following displacement of the incumbent (i.e., the reverse reaction), together leading to overall faster forward displacement kinetics.



**Figure 3-5 The length and nucleotide composition of the incumbent toehold affects toehold exchange on PNA–DNA heteroduplexes.** (A) Schematic illustration of the toehold exchange mechanism for both the homo- and heterochiral reaction pathways. (B) Sequences of the incumbent toehold domains within D/L-Ai and its truncated version D/L-Ai<sub>s</sub>. Individual incumbent toeholds are boxed based on their length and red letters emphasize G/C base pairs. Incumbent toeholds are produced by truncating the input strand by the corresponding length. (C) Calculated rate constant as a function of incumbent toehold length ( $k$ ) relative to the full-length input ( $k_0$ ) having no incumbent toehold. Error bars represent standard deviation from three independent experiments. (D, E) Fluorescence monitoring (Cy3) of toehold exchange for the homochiral (D) and heterochiral (E) reaction pathways. The length of the incumbent toehold is indicated on the right y-axis. Dotted lines indicate reactions carried out with the truncated PNA–DNA heteroduplex (Ais) for the indicated incumbent toehold lengths (black = 0-nt; red = 8-nt). Reactions depicted here were carried out as described in Figure 3-2.

Given the inverse relationship between strand displacement rate constants and the stability of the incumbent toehold (46), we hypothesized that the rate could be accelerated by reducing the G/C content within an incumbent toehold without shortening it. To test this, we truncated the branch migration domain of PNA–DNA heteroduplex D/L-Ai by four base pairs, resulting in a new, shorter PNA–DNA heteroduplex (D/L-Ai<sub>S</sub>) having two less G/C base pairs within each of the corresponding incumbent toehold domains (3/3\*) (Figure 3- 5B). Consistent with our hypothesis, the rate of strand displacement from D/L-Ai<sub>S</sub> for all incumbent toehold lengths (4, 6 and 8 nt) was at least as fast as the reaction in the absence of the incumbent toehold (Figure 3-5C). Notably, for the eight base pair incumbent toehold, the rate of the heterochiral reaction actually increased by ~2-fold as a result of the reduced G/C content. Because our approach for decreasing G/C content within the incumbent toehold also shortens the branch migration domain of the PNA–DNA heteroduplex, we sought to demonstrate that this truncation does not play a substantial role in the observed rates of toehold exchange. We calculated the ratio of rate constants for strand displacement from the short (D/L- Ai<sub>S</sub>) and long (D/L-Ai) heteroduplexes for each of the given inputs ( $k_{\text{short}}/k_{\text{long}}$ , *Appendix* Figure B5). For the full-length input strand (D-IN<sub>TH</sub> 8), the ratios of  $k_{\text{short}}/k_{\text{long}}$  are close to one for both reaction configurations, indicating that the rate of strand displacement is similar for both branch migration domain lengths. However, as the length of the incumbent toehold increases,  $k_{\text{short}}/k_{\text{long}}$  ratios become much larger than one (*Appendix* Figure B5). This strongly suggests that the sequence content of the incumbent toehold, not the length of the branch migration domain, is the primary contributor to the increased rate of toehold exchange observed for the shorter heteroduplex

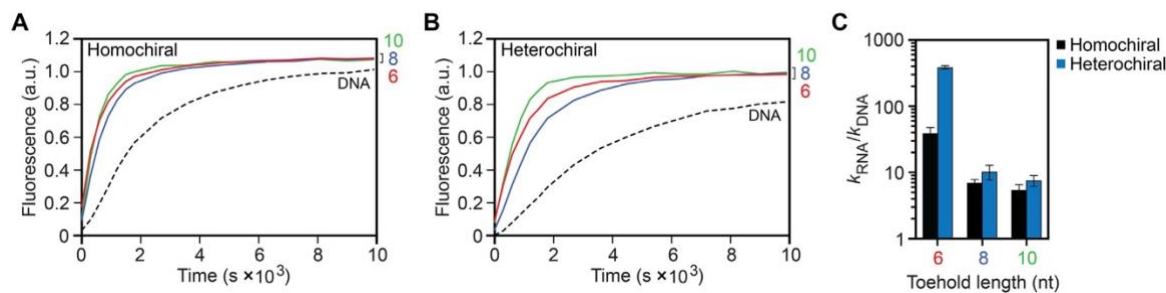
(D/L-Ai<sub>S</sub>). Thus, in addition to their relative lengths, the nucleotide content of the input and incumbent toeholds represents a key design parameter for kinetic control over toehold exchange on PNA–DNA heteroduplexes.

### **3.2.5. RNA inputs accelerate the rate of strand displacement from PNA–DNA heteroduplexes**

Detection of nucleic acid biomarkers, and in particular, RNA, has widespread applications in research and medicine (75,76). Dynamic nucleic acid devices based on DNA strand displacement have been previously repurposed for sensing RNA *in vitro* and in live cells, providing a foundation for future application in bio-imaging and disease diagnosis (22,23,77). However, the kinetics of strand displacement using RNA remains relatively unexplored compared to DNA (78). In the previous sections, we have enumerated on how strand displacement from PNA–DNA heteroduplexes can have markedly different kinetic properties from previously studied all DNA reaction systems (46,52). Therefore, it was imperative that we characterize strand displacement from PNA–DNA heteroduplexes using RNA inputs in order to establish design principles suitable for RNA detection and analysis under physiological conditions. The heterochiral reaction pathway is of particular interest in this regard because it provides the critical interface between endogenous RNA biomarkers (e.g., mRNA, microRNAs, viral RNAs, etc.) and molecular devices constructed from robust L-DNA.

We first examined the effects of toehold length on strand- displacement using RNA versions of D-IN (D-IN<sub>RNA</sub>) with toehold lengths varying from 6 to 10 nucleotides (Figure 3-6A, B). As before, all reactions were carried out under simulated physiological

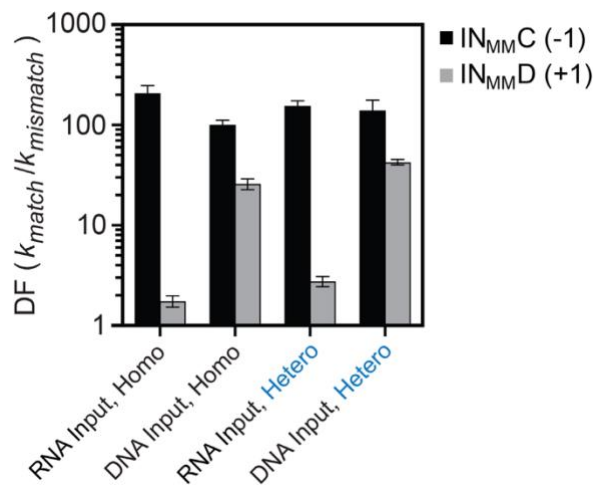
conditions. For both homo and heterochiral reaction configurations, strand displacement involving RNA inputs was drastically faster than their DNA counterparts (Figure 3-6C). Taking the 6-nucleotide toehold as an example, strand displacement rate constants for the RNA input were more than an order of magnitude faster than the DNA input for both reaction configurations. This was not completely unexpected given the increased stability and faster hybridization kinetics of PNA–RNA duplexes compared to PNA–DNA duplexes (58). Nevertheless, as previously observed for DNA inputs, the heterochiral reaction pathway using RNA inputs is slower than the homochiral reaction pathway for all toehold lengths tested (*Appendix* Table B2). Interestingly, for both reaction configurations, the rate constant was greater for the 6-nucleotide RNA toehold than for the 8-nucleotide RNA toehold. Further examination of the RNA input revealed that it is capable of folding into a hairpin structure (*Appendix* Figure B6), which is more stable for inputs with an 8- and 10-nucleotide toehold domain compared to the input with the 6-nucleotide toehold. Thus, the reduced rate of strand displacement for the RNA input with the 8-nucleotide toehold relative to the 6-nucleotide is likely due to the additional kinetic barrier of unfolding this RNA structure in some fraction of the input strand population. In the case of the 10-nucleotide toehold, the overall longer toehold domain appears to compensate for the presence of secondary structure, resulting in fast kinetics. In the context of RNA detection, this result not only demonstrates the ability of heterochiral strand displacement systems to overcome secondary structures within RNA targets, it also suggests that, through careful design, such systems may be capable of discriminating between different RNA structural conformations.



**Figure 3-6 Strand displacement from PNA–DNA heteroduplexes is faster for RNA inputs than DNA inputs.** (A, B) Fluorescence monitoring (Cy3) of the homochiral (A) and heterochiral (B) reaction pathways initiated with inputs RNA inputs ( $D\text{-IN}_{RNA}$ ) having toehold domains varying in length from 6–10 nucleotides. The length of the toehold is indicated on the right y-axis. For reference, the reaction initiated with a DNA input having a 10-nucleotide toehold is shown as a black dotted line. The reactions depicted contained 30 nM  $D\text{-IN}$ , 20 nM  $D/L\text{-Ai}$ , 60 nM  $D/L\text{-R}$ , 300 mM NaCl, 1 mM EDTA and 10 mM Tris (pH 7.6) and were carried out at 37°C. (C) Calculated rate constant for RNA inputs as a function of toehold length ( $k_{RNA}$ ) relative to the DNA input ( $k_{DNA}$ ) having the same length toehold. Error bars represent standard deviation from three independent experiments.

We also investigated the effect of mismatches between an RNA input strand and the PNA–DNA heteroduplex, focusing on mismatches positioned at the junction between the toehold and branch migration domains (i.e., positions –1 and +1). Mismatches at these two positions resulted in the greatest impact on strand displacement kinetics using DNA inputs, especially for the heterochiral reaction configuration. In contrast to the DNA input with a mismatch at position +1 ( $D\text{-IN}_{MMD}$ ), the RNA input with a mismatch at this position only modestly reduced the rate of strand displacement relative to the fully matched input for both reaction configurations (Figure 3-7).





**Figure 3-7 Discrimination factors (DF =  $k_{\text{match}}/k_{\text{mismatch}}$ ) of PNA–DNA heteroduplexes towards RNA inputs** having different mismatches, where  $k_{\text{match}}$  and  $k_{\text{mismatch}}$  are the calculated rate constants for a fully matched and mismatched RNA input, respectively. Error bars represent standard deviation from three independent experiments. Hetero and Homo refer to the heterochiral and homochiral reactions pathways, respectively. See Figure 3-3A for mismatch identity and position.

Thus, the stronger toehold binding interaction of the RNA input is potentially able to compensate for the increased activation energy associated with a mismatch positioned immediately adjacent to the toehold. However, the mismatch at position  $-1$  (D-IN<sub>MM</sub>C) within the toehold domain of the RNA input retains its strong inhibitory effect, decreasing the rate of strand displacement by at least two orders of magnitude relative to the fully-matched RNA input (Figure 3-7). Thus, a mismatch at this position ( $-1$ ) will be useful for designing kinetic probes capable of discriminating between RNAs based on SNPs.

Together, these studies demonstrate that strand displacement rates for PNA–DNA heteroduplexes are substantially enhanced, by up to two orders of magnitude, upon substitution of RNA inputs for DNA inputs. Furthermore, high sensitivity to single-

nucleotide mismatches, when appropriately positioned, can be maintained. Regarding heterochiral strand displacement, the increased reaction kinetics using RNA inputs will greatly benefit future applications aimed at interfacing endogenous RNAs with molecular devices and sensors constructed from biostable L-DNA. Moreover, the potential sensitivity of heterochiral strand displacement to RNA secondary structures is interesting and deserves to be explored further. We anticipate that discrimination between RNA structural conformations could be exploited in a broad range of applications beyond primary sequence detection.

### **3.3. Summary and Outlook**

Through detailed experimental analysis, we have demonstrated that several common design parameters for controlling DNA strand displacement kinetics, including toehold length and mismatches, can also be applied to strand displacement from PNA–DNA heteroduplexes. Although important differences exist, we found that the rate of PNA–DNA strand displacement reactions can be tuned in a manner that is mostly analogous to traditional all-DNA reactions. For example, reaction rates increase proportionally with toehold length and mismatches positioned proximal to the toehold domain strongly inhibit strand displacement. These similarities suggest that strand displacement systems based on PNA–DNA heteroduplexes can be predictably engineered to undergo similar dynamic behaviors as those constructed solely from DNA. Indeed, we showed that PNA–DNA heteroduplexes are compatible with toehold exchange. Importantly, our careful parameterization of the heterochiral strand displacement pathway, which serves as the key interface between L-DNA and endogenous D-nucleic acids,

provides an important contribution to the rational design and optimization of dynamic L-DNA based circuits and nanodevices capable of interfacing with biological systems.

In addition to their similarities, we find that strand displacement reactions from PNA–DNA heteroduplexes exhibit several unique, and potentially advantageous, characteristics relative to their all-DNA counterparts. In particular, we show that stereochemistry, a parameter unique to our system, offers an additional layer of kinetic control not possible using conventional all-DNA strand displacement. We anticipate that this capability will greatly expand the types of dynamic behaviors that can be programmed into nucleic acid-based devices. For example, one could imagine building a kinetic ‘thresholding’ gate based on the rate discrepancy between identical homochiral and heterochiral strand displacement reactions as a straightforward alternative to previous designs based on toehold length and/or composition (12). Toehold exchange reactions using PNA–DNA heteroduplexes also exhibit unique kinetic behaviors, especially for the heterochiral reaction configuration. We show that heterochiral toehold exchange is mostly insensitive to the length of the incumbent toehold (3\*), even as its length approaches that of the input toehold domain (1\*). This is in direct contrast to the expected trends for introducing incumbent toeholds into all-DNA systems (46) and could be exploited to develop heterochiral strand displacement devices and probes capable of rapid signal amplification.

During this work, we explored the underlying cause of the observed rate discrepancies between homo and heterochiral strand displacement reactions, focusing on the stability of the corresponding toehold domains. Melting temperature analysis revealed

that the toehold domain in the heterochiral reaction configuration forms a far less stable complex with the input strand than does the toehold domain in the homochiral reaction configuration. Thermodynamic data indicated that the less-favorable free energy for heterochiral toehold association stems predominantly from the entropic term, which we attribute to unfavorable preorganization between the induced left-handed PNA toehold and right-handed input strand. This observation not only provides important insights into the biophysical mechanisms of heterochiral strand displacement, but also into the broader role of molecular organization in controlling strand displacement kinetics, which could be exploited elsewhere as a tool for modulating reaction rates. For example, our results suggest that the rate of heterochiral strand displacement could be increased by enforcing a right-handed helical conformation on the toehold domain, possibly by using *chiral* PNA monomers such as those containing modifications at the gamma-position of the backbone (73). It is also worth noting that this study provides further experimental evidence supporting the extended propagation of induced helicity through single-stranded, achiral PNA.

Most importantly, heterochiral strand displacement reactions exhibited fast reaction kinetics with RNA inputs and maintained a high level of mismatch discrimination when appropriately positioned. This further demonstrates the potential applications of this technology for the detection and manipulation of biologically relevant RNA molecules. Indeed, we have previously shown heterochiral strand displacement circuits composed of L-DNA /RNA greatly outperform their all-DNA counterparts in living cells and can be directly interfaced with endogenous RNAs (45,79). We are continuing to pursue routes to

increase the performance of heterochiral strand displacement systems, both *in vitro* and *in vivo*. This work now provides a more solid foundation from which to base future designs. For example, we found that mismatch discrimination is dependent on whether the input strand was composed of DNA or RNA (Figure 3-7). This not only highlights the importance of studying both types of inputs, which is seldomly done, but also provides valuable information for engineering heterochiral strand displacement probes with increased selectivity for RNA based on SNPs. It should be mentioned that the rate of heterochiral strand displacement, especially for RNA inputs, may vary significantly depending on the sequence (and secondary structure), and it will be important to further characterize these effects in the future.

Overall, this work establishes a basic set of design considerations to guide future development of strand displacement systems based on the unique properties of PNA–DNA heteroduplexes. In particular, we expect that the detailed characterization of heterochiral strand displacement kinetics provided herein, along with the increasing availability of L-oligonucleotides, will broaden the scope and applicability of L-DNA /RNA-based circuits and other nanodevices for practical applications at the interface with biology.

### **3.4. Materials & Methods**

#### **3.4.1. Materials**

Oligonucleotide synthesis reagents, D-nucleoside phosphoramidites, 6-Fluorescein phosphoramidite (6-FAM) and the Cyanine 3 (Cy3) phosphoramidite were purchased from Glen Research (Sterling, VA). L-Nucleoside phosphoramidites were purchased from ChemGenes (Wilmington, MA). Black Hole Quencher 2 (BHQ2) CPG resins were

purchased from LGC Biosearch Technologies (Petaluma, CA). Peptide nucleic acids (PNAs) were purchased from Pan- gene (Daejeon, South Korea) at 99.9% purity and were not purified further. All other reagents were purchased from Sigma Aldrich (St. Louis, MO).

### **3.4.2. Oligonucleotide synthesis and purification**

Unmodified D-oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), and all L- oligonucleotides were synthesized in house using an Expedite 8909 DNA/RNA synthesizer. Terminal labeling of the 5' end with either Cy3 or 6-carboxyfluorescein (6-FAM) was carried out using the corresponding phosphoramidites, which were coupled based on the manufacturers recommended protocols. All 3'-BHQ2 modified oligonucleotides were obtained by conducting the synthesis on the corresponding BHQ2 CPG resin. Following synthesis and deprotection, single-stranded oligonucleotides were purified by 20% denaturing polyacrylamide gel electrophoresis (PAGE; 19:1 acrylamide: bisacrylamide). Purified oligonucleotides were excised from the gel and eluted overnight at room temperature in a buffer consisting of 200 mM NaCl, 10 mM EDTA, and 10 mM Tris (pH 7.6). The solution was filtered to remove gel fragments, then desalted by ethanol precipitation. The obtained pellet was resuspended in water and quantified by measuring the absorbance at 260 nm using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Commercial PNAs, obtained as lyophilized solids, were reconstituted at 100  $\mu$ M in water and used without further purification. Individual strands were quantified using the extinction coefficients provided by the manufacturer.

### 3.4.3. Preparation and characterization of duplex reaction components

In order to form complexes D/L-Ai, D/L-R and D-Aq the corresponding oligonucleotides (Supplementary Figure S1 and Supplementary Table S1) were annealed in a reaction mixture containing the appropriate amount of each strand (see below), 300 mM NaCl, 1 mM EDTA, 10 mM Tris pH 7.6) and were heated to 90°C for 3 min then cooled slowly to room temperature over 1 h. For D/L-Ai, 50  $\mu$ M of PNA<sub>L</sub> was annealed to 75  $\mu$ M D/L-OUT. For D/L-R, 100  $\mu$ M D/L-F was annealed to 150  $\mu$ M D/L-Q. For D-Aq, 10  $\mu$ M of PNA<sub>S</sub> was annealed to 10.5  $\mu$ M D-OUT<sub>Q</sub>. Whereas D-Aq was used directly, complexes D/L-Ai and D/L-R were further purified by 20% native PAGE (19:1 acrylamide:bisacrylamide) and single bands were carefully excised from the gel. The gel fragments were crushed and eluted overnight at room temperature in the same buffer that was used for annealing, and the suspension was filtered through a 0.2  $\mu$ m filter. The concentration was estimated from UV absorbance at 260 nm using the combined extinction coefficients of the individual strands comprising the duplex.

For a more accurate determination of the concentration of duplex components D/L-Ai and D/L-R, a calibration curve of Cy3 fluorescence of free strand D/L-F was generated over a range of concentrations from 100 nM to 300 nM, as measured at excitation/emission wavelength of 520 nm/580–640 nm (bandpass filter). These measurements were taken at 37°C in a buffer containing 300 mM NaCl, 1 mM EDTA, 10 mM Tris (pH 7.6). This calibration curve (i.e. the linear relationship between the concentration of strand D/L-F and Cy3 fluorescence) was then used to determine the concentration of components D/L-

Ai and D/L-R based on Cy3 fluorescence following a strand displacement reaction under the same conditions. For example, the concentration of D/L-R was determined by first reacting 100 nM of D/L-R (estimated) with a large excess of D/L-OUT in order to drive the strand displacement reaction to completion. The Cy3 signal generated from this reaction was then compared to the calibration curve to determine the amount of strand D/L-F present in D/L-R, and thus, its concentration. Prior to use, fresh dilutions of each complex were prepared in the presence of 10  $\mu$ M poly[T] carrier oligonucleotide to prevent loss of material from sticking to plastic surfaces of tubes and pipette tips (46).

#### **3.4.4. Monitoring strand displacement reactions by fluorimetry**

Strand displacement reactions were monitored using a Glomax Discover multi-well plate reader (Promega Corp.) using excitation/emission wavelengths 520 nm/580–640 nm (bandpass filter for Cy3). Reaction mixtures contained either 60 nM or 300 nM D/L-R, either 30 nM or 150 nM of the indicated input strand, 300 mM NaCl, 1 mM EDTA, and 10 mM Tris (pH 7.6) and were initiated by the addition of either 20 nM or 100 nM of D/L-Ai, respectively. The negative control contained no input. Each experiment was run side-by-side with a ‘pre-activated’ reaction mixture that contained 1  $\mu$ M of D/L-IN<sub>TH</sub>10 (Supplementary Table S1), either 20 nM or 100 nM D/L-Ai, either 60 nM or 300 nM D/L-R, 300 mM NaCl, 1 mM EDTA and 10 mM Tris (pH 7.6). Pre-activated reaction mixtures were incubated for 10 minutes prior to use in order to fully activate the reporter complex. Each reaction mixture was in a final volume of 30  $\mu$ L and carried out at 37°C unless otherwise specified.



All strand displacement reactions were normalized to the signal from the pre-activated reaction representing the maximum achievable fluorescence using Equation (1):

$$F_n = \frac{F - F_0}{F_c - F_0} \quad (1)$$

where  $F_n$  is the normalized fluorescence intensity,  $F$  is the measured fluorescence,  $F_0$  is the fluorescence of the quenched reporter complex and  $F_c$  is the fluorescence of the activated reporter complex at the time a measurement was taken. Because the PNA–DNA heteroduplex (D/L-Ai) is the limiting reagent in all cases, 1.0 normalized fluorescence units (FU) corresponds to either 20 or 100 nM of activated reporter D/L-R (i.e., free strand D/L-F).

### 3.4.5. Rate constant fitting procedure

All strand displacement reactions performed for this study are assumed to be second-order reactions with respect to the input strand and PNA–DNA heteroduplex Ai as described previously (43,46,47). As indicated above, two concentration regimes of strand displacement components were used to extract rate constants: (A) 150 nM D-IN, 100 nM D/L-Ai and 300 nM D/L-R and (B) 30 nM D-IN, 20 nM D/L-Ai and 60 nM D/L-R.

Very slow strand displacement reactions that are unlikely to go to completion, such as short heterochiral toeholds (2– 4 nt), and certain mismatches, were fit using Equation (2) (47):

$$\frac{([Ai]_0 - [OUT])[IN]_0}{([IN]_0 - [OUT])[Ai]_0} = ([Ai]_0 - [IN]_0)kt \quad (2)$$

where [OUT] is the concentration of displaced D/L-OUT at time  $t$ , and [IN]<sub>0</sub> and [Ai]<sub>0</sub> are the initial concentrations of the input strand and the PNA–DNA heteroduplex, respectively. The concentration of displaced incumbent strand [OUT] at any given time  $t$  was obtained by multiplying the respective normalized fluorescence intensity by [Ai]<sub>0</sub>. The left-hand side of the Equation (2) was plotted versus  $t$ , and the respective rate constant ( $k$ ) was extracted from the slope of the linear fit to the plot (Supplementary Figure S2; Representative fits in the SI). We note that minor deviations from second-order kinetics were observed for some very slow reactions, and in particular, those resulting from short toeholds in the heterochiral configuration (Supplementary Figure S2b). This is not unexpected (48). However, for the sake of a uniform analysis of reaction rates, we have reported all rate constants as second-order.

Fast reactions were fit using Equation (3), which is rearranged from Equation (2):

$$[\text{OUT}]_n = \frac{[\text{IN}]_0(1 - \exp^{[kt([\text{Ai}]_0 - [\text{IN}]_0)])}}{[\text{IN}]_0 - [\text{Ai}]_0 \exp^{[kt([\text{Ai}]_0 - [\text{IN}]_0)]}} \quad (3)$$

where [OUT]<sub>n</sub> is the normalized fluorescent intensity at time  $t$ , [IN]<sub>0</sub> and [Ai]<sub>0</sub> are the initial concentrations of the input strand and the PNA–DNA heteroduplex respectively. To extract rate constants ( $k$ ), Equation (3) was fit to all the data points.

All rate constants are reported as the mean value of at least three replicates, and the corresponding standard deviation has been used as a measure of the error. The data was plotted in GraphPad Prism and the extracted rate constants are listed in Supplementary Table S2 for all strand displacement reactions reported herein.

### 3.4.6. Melting temperature analysis

Melting experiments were performed on a Bio-Rad CFX96 Touch Real-Time PCR instrument and fluorescence was measured using excitation/emission wavelengths of 520 nm/580–640 nm (bandpass filter for Cy3). Reaction mixtures contained an equimolar ratio of either D- or L-IN<sub>F</sub> and D-Aq within a buffer comprising 300 mM NaCl, 1 mM EDTA, and 10 mM Tris (pH 7.6). D/L-IN<sub>F</sub> lacks the branch migration domains found in D-IN (domains 2 and 3), instead containing five dT residues and a 5'-6-FAM dye (Supplementary Figure S1 and Supplementary Table S1). The incumbent strand (D-OUT<sub>Q</sub>) within heteroduplex D-Aq was labeled internally with a quencher (BHQ2) such that binding of either D- or L-IN<sub>F</sub> to the toehold domain (1\*) resulted in fluorescence quenching. D/L-IN<sub>F</sub> and D-Aq were annealed at 6 concentrations ( $C_t$ ) of 0.5, 1, 2, 3, 4 and 6  $\mu$ M, where  $C_t$  represents the combined concentration of D/L-IN<sub>F</sub> and D-Aq. All reactions also contained 10  $\mu$ M poly[T] carrier oligonucleotide to prevent loss of material from sticking to plastic surfaces of tubes and pipette tips (46). Individual reaction mixtures containing the indicated concentrations of either D- or L-IN<sub>F</sub> and D-Aq were incubated at 10° C for 10 min then heated to 70° C in 2° C increments with an equilibration time of 5 min for each step. Each melting reaction was run side-by-side with a positive control containing only D-IN<sub>F</sub> or L-IN<sub>F</sub> under identical conditions. Fluorescence readings were taken at the end of each equilibration period, prior to the next 2° C temperature increase. The 70° C maximum temperature is ~20° C above the predicted melting temperature of the toehold duplex (1/1\*) (49).

To construct a temperature dependent melting profile at each concentration, the measured fluorescence signal was first corrected for (1) background and (2) temperature- dependent changes in fluorescence using Equation (4):

$$F_{\text{corr}} = \frac{F_{\text{melt}} - F_{\text{bkgnd}}}{F_{\text{pos}} - F_{\text{bkgnd}}} \quad (4)$$

where  $F_{\text{corr}}$  is the corrected fluorescence,  $F_{\text{melt}}$  is the measured fluorescence of the melting reaction (D- or L-IN<sub>F</sub> and D-Aq),  $F_{\text{pos}}$  is the measured fluorescence of the positive control (D- or L-IN<sub>F</sub> only) and  $F_{\text{bkgnd}}$  is the measured background fluorescence of an empty well.  $F_{\text{corr}}$  was plotted against temperature to derive the melting profile. The first and second derivatives of these curves were approximated in Microsoft Excel, and the value at which the second derivative intersected the x-axis was considered the melting temperature for the corresponding concentration ( $C_t$ ). The inverse of the melting temperature ( $1/T_m$ ) was plotted against  $\ln C_t$  (Annexure Figure B4), generating a line from which thermodynamic parameters  $\Delta H^\circ$  and  $\Delta S^\circ$  of toehold binding can be calculated according to Equation (5) (50,51):

$$\frac{1}{T_m} = \frac{(n-1)R}{\Delta H^\circ} \ln C_t + \frac{[\Delta S^\circ - (n-1)R \ln 2n]}{\Delta H^\circ} \quad (5)$$

where  $T_m$  is the melting temperature,  $C_t$  is the combined concentration of D/L-IN<sub>F</sub> and D-Aq,  $n$  is the molecularity of the binding reaction (assumed to be  $n = 2$  for D/L-IN<sub>F</sub> + D-Aq) and  $\Delta H^\circ$  and  $\Delta S^\circ$  are the enthalpy and entropy changes associated with the toehold binding event, respectively.

### 3.5. References

1. Zhang, D.Y. and Seelig, G. (2011) Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.*, 3, 103–113.
2. Simmel, F.C., Yurke, B. and Singh, H.R. (2019) Principles and applications of nucleic acid strand displacement reactions. *Chem. Rev.*, 119, 6326–6369.
3. Sherman, W.B. and Seeman, N.C. (2004) A precisely controlled DNA biped walking device. *Nano Lett.*, 4, 1203–1207.
4. Shin, J.-S. and Pierce, N.A. (2004) A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.*, 126, 10834–10835.
5. Bath, J. and Turberfield, A.J. (2007) DNA nanomachines. *Nat. Nanotechnol.*, 2, 275–284.
6. Kay, E. R., Leigh, D.A. and Zerbetto, F. (2007) Synthetic molecular motors and mechanical machines. *Angew. Chem. Int. Ed.*, 46, 72–191.
7. Rothemund, P.W.K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, 440, 297–302.
8. Goodman, R.P., Heilemann, M., Doose, S., Erben, C.M., Kapanidis, A.N. and Turberfield, A.J. (2008) Reconfigurable, braced, three-dimensional DNA nanostructures. *Nat. Nanotechnol.*, 3, 93–96.
9. Yin, P., Choi, H.M.T., Calvert, C.R. and Pierce, N.A. (2008) Programming biomolecular self-assembly pathways. *Nature*, 451, 318–322.
10. Sadowski, J.P., Calvert, C.R., Zhang, D.Y., Pierce, N.A. and Yin, P. (2014) Developmental self-assembly of a DNA tetrahedron. *ACS Nano*, 8, 3251–3259.
11. Seelig, G., Soloveichik, D., Zhang, D.Y. and Winfree, E. (2006) Enzyme-free nucleic acid logic circuits. *Science*, 314, 1585–1588.
12. Qian, L. and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332, 1196–1201.
13. Scheible, M.B., Pardatscher, G., Kuzyk, A. and Simmel, F.C. (2014) Single molecule characterization of DNA binding and strand displacement reactions on lithographic DNA origami microarrays. *Nano Lett.*, 14, 1627–1633.

14. Dirks,R.M. and Pierce,N.A. (2004) Triggered amplification by hybridization chain reaction. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 15275–15278.
15. 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.
16. Duose,D.Y., Schweller,R.M., Zimak,J., Rogers,A.R., Hittelman,W.N. and Diehl,M.R. (2011) Configuring robust DNA strand displacement reactions for in situ molecular analyses. *Nucleic Acids Res.*, 40, 3289–3298.
17. Levesque,M.J., Ginart,P., Wei,Y. and Raj,A. (2013) Visualizing SNVs to quantify allele-specific expression in single cells. *Nat. Methods*, 10, 865–867.
18. Choi,H.M.T., Beck,V.A. and Pierce,N.A. (2014) Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano*, 8, 4284–4294.
19. Douglas,S.M., Bachelet,I. and Church,G.M. (2012) A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335, 831–834.
20. Rudchenko,M., Taylor,S., Pallavi,P., Dechkovskaia,A., Khan,S., Butler,V.P. Jr, Rudchenko,S. and Stojanovic,M.N. (2013) Autonomous molecular cascades for evaluation of cell surfaces. *Nat. Nanotechnol.*, 8, 580–586.
21. Hemphill,J. and Deiters,A. (2013) DNA computation in mammalian cells: microRNA logic operations. *J. Am. Chem. Soc.*, 135, 10512–10518.
22. Groves,B., Chen,Y.-J., Zurla,C., Pochekaïlov,S., Kirschman,J.L., Santangelo,P.J. and Seelig,G. (2016) Computing in mammalian cells with nucleic acid strand exchange. *Nat. Nanotechnol.*, 11, 287–294.
23. Chatterjee,G., Chen,Y.-J. and Seelig,G. (2018) Nucleic acid strand displacement with synthetic mRNA inputs in living mammalian cells. *ACS Synth. Biol.*, 7, 2737–2741.
24. Fisher,T.L., Terhorst,T., Cao,X. and Wagner,R.W. (1993) Intracellular disposition and metabolism of fluorescently-labeled unmodified and modified oligonucleotides microinjected into mammalian cells. *Nucleic Acids Res.*, 21, 3857–3865.
25. Molenaar,C., Marras,S.A., Slats,J.C.M., Truffert,J.C., Lemaître,M., Raap,A.K., Dirks,R.W. and Tanke,H.J. (2001) Linear 2'-O-methyl RNA probes for the visualization of RNA in living cells. *Nucleic Acids Res.*, 29, e89.

26. Amarzguioui, M., Holen, T., Babaie, E. and Prydz, H. (2003) Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Res.*, 31, 589–595.
27. Freier, S.M. and Altmann, K.H. (1997) The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.*, 25, 4429–4443.
28. Schoen, I., Krammer, H. and Braun, D. (2009) Hybridization kinetics is different inside cells. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 21649–21654.
29. Chen, Y.-I., Chang, Y.-J., Nguyen, T.D., Liu, C., Phillion, S., Kuo, Y.-A., Vu, H.T., Liu, A., Liu, Y.-L., Hong, S. et al. (2019) Measuring DNA hybridization kinetics in live cells using a time-resolved 3D single-molecule tracking method. *J. Am. Chem. Soc.*, 141, 15747–15750.
30. Bramsen, J.B., Laursen, M.B., Nielsen, A.F., Hansen, T.B., Bus, C., Langkjær, N., Babu, B.R., Højland, T., Abramov, M., Van Aerschot, A. et al. (2009) A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res.*, 37, 2867–2881.
31. Watts, J.K., Deleavey, G.F. and Damha, M.J. (2008) Chemically modified siRNA: tools and applications. *Drug Discov. Today*, 13, 842–855.
32. Fern, J. and Schulman, R. (2017) Design and characterization of DNA strand-displacement circuits in serum-supplemented cell medium. *ACS Synth. Biol.*, 6, 1774–1783.
33. Chen, Y.-J., Groves, B., Muscat, R.A. and Seelig, G. (2015) DNA nanotechnology from the test tube to the cell. *Nat. Nanotechnol.*, 10, 748–760.
34. Conway, J.W., McLaughlin, C.K., Castor, K.J. and Sleiman, H. (2013) DNA nanostructure serum stability: greater than the sum of its parts. *Chem. Commun.*, 49, 1172–1174.
35. Young, B.E., Kundu, N. and Sczepanski, J.T. (2019) Mirror-Image oligonucleotides: history and emerging applications. *Chem. Eur. J.*, 25, 7981–7990.
36. Ashley, G.W. (1992) Modeling, synthesis, and hybridization properties of (L)-ribonucleic acid. *J. Am. Chem. Soc.*, 114, 9732–9736.
37. Urata, H., Ogura, E., Shinohara, K., Ueda, Y. and Akagi, M. (1992) Synthesis and properties of mirror-image DNA. *Nucleic Acids Res.*, 20, 3325–3332.

38. Hauser, N.C., Martinez, R., Jacob, A., Rupp, S., Hoheisel, J.D. and Matysiak, S. (2006) Utilising the left-helical conformation of L-DNA for analysing different marker types on a single universal microarray platform. *Nucleic Acids Res.*, 34, 5101–5111.
39. Hoehlig, K., Bethge, L. and Klussmann, S. (2015) Stereospecificity of oligonucleotide interactions revisited: No evidence for heterochiral hybridization and ribozyme/DNAzyme activity. *PLoS One*, 10, e0115328.
40. Szabat, M., Gudanis, D., Kotkowiak, W., Gdaniec, Z., Kierzek, R. and Pasternak, A. (2016) Thermodynamic features of structural motifs formed by  $\square$ -L-RNA. *PLoS One*, 11, e0149478.
41. Garbesi, A., Capobianco, M.L., Colonna, F.P., Tondelli, L., Arcamone, F., Manzini, G., Hilbers, C.W., Aelen, J.M.E. and Blommers, M.J.J. (1993) L-DNA s as potential antimessenger oligonucleotides: a reassessment. *Nucleic Acids Res.*, 21, 4159–4165.
42. Urata, H., Shinohara, K., Ogura, E., Ueda, Y. and Akagi, M. (1991) Mirror image DNA. *J. Am. Chem. Soc.*, 113, 8174–8175.
43. Kabza, A.M., Young, B.E. and Sczepanski, J.T. (2017) Heterochiral DNA strand-displacement circuits. *J. Am. Chem. Soc.*, 139, 17715–17718.
44. Kabza, A.M. and Sczepanski, J.T. (2020) L-DNA -based catalytic hairpin assembly circuit. *Molecules*, 25, 947.
45. Zhong, W. and Sczepanski, J.T. (2019) A mirror image fluorogenic aptamer sensor for live-cell imaging of microRNAs. *ACS Sens.*, 4, 566–570.
46. Zhang, D.Y. and Winfree, E. (2009) Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.*, 131, 17303–17314.
47. Olson, X., Kotani, S., Yurke, B., Graugnard, E. and Hughes, W.L. (2017) Kinetics of DNA strand displacement systems with locked nucleic acids. *J. Phys. Chem. B*, 121, 2594–2602.
48. Machinek, R.R.F., Ouldrige, T.E., Haley, N.E.C., Bath, J. and Turberfield, A.J. (2014) Programmable energy landscapes for kinetic control of DNA strand displacement. *Nat. Commun.*, 5, 5324.
49. Giesen, U., Kleider, W., Berding, C., Geiger, A., Orum, H. and Nielsen, P.E. (1998) A formula for thermal stability ( $T_m$ ) prediction of PNA/DNA duplexes. *Nucleic Acids Res.*, 26, 5004–5006.



50. Markey, L.A. and Breslauer, K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, 26, 1601–1620.
51. You, Y., Tataurov, A.V. and Owczarzy, R. (2011) Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers*, 95, 472–486.
52. Srinivas, N., Ouldrige, T.E., Sulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P. and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.*, 41, 10641–10658.
53. Yurke, B. and Mills, A.P. (2003) Using DNA to power nanostructures. *Genet. Program. Evolvable Mach.*, 4, 111–122.
54. 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. *J. Comput. Chem.*, 32, 170–173.
55. Moreira, B.G., You, Y., Behlke, M.A. and Owczarzy, R. (2005) Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. *Biochem. Biophys. Res. Commun.*, 327, 473–484.
56. Broadwater, D.W.B. Jr and Kim, H.D. (2016) The effect of basepair mismatch on DNA strand displacement. *Biophys. J.*, 110, 1476–1484.
57. Ouldrige, T.E., Sulc, P., Romano, F., Doye, J.P.K. and Louis, A.A. (2013) DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.*, 41, 8886–8895.
58. Jensen, K.K., Ørum, H., Nielsen, P.E. and Nordén, B. (1997) Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry*, 36, 5072–5077.
59. Ratilainen, T., Holmén, A., Tuite, E., Haaima, G., Christensen, L., Nielsen, P.E. and Nordén, B. (1998) Hybridization of peptide nucleic acid. *Biochemistry*, 37, 12331–12342.
60. Haley, N.E.C., Ouldrige, T.E., Mullor Ruiz, I., Geraldini, A., Louis, A.A., Bath, J. and Turberfield, A.J. (2020) Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement. *Nat. Commun.*, 11, 2562.
61. Braasch, D.A. and Corey, D.R. (2001) Synthesis, analysis, purification, and intracellular delivery of peptide nucleic acids. *Methods*, 23, 97–107.

62. Irmisch,P., Ouldrige,T.E. and Seidel,R. (2020) Modeling DNA-strand displacement reactions in the presence of base-pair mismatches. *J. Am. Chem. Soc.*, 142, 11451–11463.
63. Chen,S.X. and Seelig,G. (2016) An engineered kinetic amplification mechanism for single nucleotide variant discrimination by DNA hybridization probes. *J. Am. Chem. Soc.*, 138, 5076–5086.
64. Xiao,X., Wu,T., Xu,L., Chen,W. and Zhao,M. (2017) A branch-migration based fluorescent probe for straightforward, sensitive and specific discrimination of DNA mutations. *Nucleic Acids Res.*, 45, e90.
65. Hwang,M.T., Landon,P.B., Lee,J., Choi,D., Mo,A.H., Glinsky,G. and Lal,R. (2016) Highly specific SNP detection using 2D graphene electronics and DNA strand displacement. *Proc. Natl. Acad. Sci. U.S.A.*, 113, 7088–7093.
66. Khodakov, D.A., Khodakova, A.S., Linacre,A. and Ellis,A.V. (2013) Toehold-mediated nonenzymatic DNA strand displacement as a platform for DNA genotyping. *J. Am. Chem. Soc.*, 135, 5612–5619.
67. Hwang, M.T., Wang, Z., Ping, J., Ban, D.K., Shiah, Z.C., Antonschmidt, L., Lee, J., Liu, Y., Karkisaval, A.G., Johnson, A.T.C. et al. (2018) DNA nanotweezers and graphene transistor enable label-free genotyping. *Adv. Mater.*, 30, 1802440.
68. Li, C., Li, Y., Chen, Y., Lin, R., Li, T., Liu, F. and Li, N. (2016) Modulating the DNA strand-displacement kinetics with the one-sided remote toehold design for differentiation of single-base mismatched DNA. *RSC Adv.*, 6, 74913–74916.
69. Zhang, D.Y., Chen, S.X. and Yin, P. (2012) Optimizing the specificity of nucleic acid hybridization. *Nat. Chem.*, 4, 208–214.
70. Sforza, S., Haaima, G., Marchelli, R. and Nielsen, P.E. (1999) Chiral peptide nucleic acids (PNAs): Helix handedness and DNA recognition. *Euro. J. Org. Chem.*, 1999, 197–204.
71. Eriksson,M. and Nielsen,P.E. (1996) Solution structure of a peptide nucleic acid–DNA duplex. *Nat. Struct. Biol.*, 3, 410–413.
72. Datta,B. and Armitage,B.A. (2001) Hybridization of PNA to structured DNA targets: Quadruplex invasion and the overhang effect. *J. Am. Chem. Soc.*, 123, 9612–9619.
73. Dragulescu-Andrasi,A., Rapireddy,S., Frezza,B.M., Gayathri,C., Gil,R.R. and Ly,D.H. (2006) A simple  $\alpha$ -backbone modification preorganizes peptide nucleic acid into a helical structure. *J. Am. Chem. Soc.*, 128, 10258–10267.

74. Sforza,S., Tedeschi,T., Corradini,R. and Marchelli,R. (2007) Induction of helical handedness and DNA binding properties of peptide nucleic acids (PNAs) with two stereogenic centres. *Euro. J. Org. Chem.*, 2007, 5879–5885.
75. Chen,L., Heikkinen,L., Wang,C., Yang,Y., Sun,H. and Wong,G. (2019) Trends in the development of miRNA bioinformatics tools. *Brief. Bioinform.*, 20, 1836–1852.
76. Gebert,L.F.R. and MacRae,I.J. (2019) Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.*, 20, 21–37.
77. Li,M.-X., Xu,C.-H., Zhang,N., Qian,G.-S., Zhao,W., Xu,J.-J. and Chen,H.-Y. (2018) Exploration of the kinetics of toehold-mediated strand displacement via plasmon rulers. *ACS Nano*, 12, 3341–3350.
78. S̃ ulc,P., Ouldrige,T.E., Romano,F., Doye,J.P.K. and Louis,A.A. (2015) Modelling toehold-mediated RNA strand displacement. *Biophys. J.*, 108, 1238–1247.
79. Zhong,W. and Sczepanski,J.T. (2021) Direct comparison of D-DNA and L-DNA strand-displacement reactions in living mammalian cells. *ACS Synth. Biol.*, 10, 209–212.

## 4. CONCLUSIONS

### 4.1. The Bigger Picture

The unifying theme of the research presented in this thesis is the exploration of non-canonical interactions between native D-DNA/RNA, and its mirror-image counterparts L-DNA/RNA for the development of novel interfacing platforms. This in turn would allow us to exploit the many superior properties of L-DNA as a novel stereochemical modification such as nuclease resistance, minimal toxicity and immunogenicity, and identical physical properties as native nucleic acids, in biological applications directed towards interrogation and manipulation of endogenous nucleic acids. Beyond direct applications as biotechnological tools, L-DNA/RNA based strategies can also be valuable in exploring complex, dynamic interactions in RNA, thus addressing fundamental questions in RNA biology.

In Chapter 2, we take a structure-based approach and expand on the heterochiral structure-specific interactions between enantiomeric RNA to evolve catalytic, site-specific phosphodiester cleavage. The power of *in vitro* evolution, together with the heterochiral nature of the evolved recognition modalities resulted in the isolation of several unique ribonuclease ribozymes with many novel properties. Following the *in vitro* selection experiment, both the *cis*- and *trans*-active ribozymes were extensively characterized through analysis of their sequence information and predicted secondary structures by truncation-mutation studies and the activity profiles were obtained by systematically varying parameters such as  $Mg^{2+}$  concentration, pH, temperature, concentrations of the TAR substrate and the respective ribozyme etc. Additionally, the

kinetics of the cleavage reaction *in cis* as well as *in trans*, and the analysis of the site selection for cleavage and the nature of the cleavage product provide unique insight into the complex nature of the heterochiral interactions involved in this reaction. This establishes a model system for more detailed characterization of heterochiral interactions that can further expand the reaction and substrate scope in heterochiral catalysis. This is only the second known example of heterochiral ribozyme catalysis and represents a significant step forward in unlocking hitherto unexplored routes to access RNA structure.

In Chapter 3, I present detailed kinetic characterization of heterochiral DNA strand displacement reactions. This work clearly outlines a set of design principles based on parameters such as toehold length, mismatch positions, length of incumbent toehold and nature of the invading strand that will improve the rational design of nanodevices with an L-DNA/RNA backbone, operating via this reaction. Additionally, the opposing chiralities of the invading and the incumbent strand in heterochiral displacements enable a unique stereochemical control over the kinetics of these reactions that can add versatility and complexity to the design of such devices. This is valuable from a biomedical perspective, since it allows to harness the biostability and other advantageous properties of L-DNA as a biomaterial, while maximizing the kinetic efficiency and versatility in applications by modulation of different parameters.

From a broader perspective, novel interactions between biomolecules can shift the paradigm of how we perceive and investigate extant biology, which in turn, influence the development of technologies at the biological interface. There has always been sustained interest in the discovery and evolution of novel bioorthogonal as well as biocompatible materials and technologies. This is evident from the rigor of contemporary research in

expansion of the genetic code by designing artificial nucleobases or unnatural amino acids (1). We can expect that further research into the mirror-image world of L-nucleic acids will not only address fundamental questions at the interface of chemistry and biology, but also facilitate practical tools in biomedicine and nanobiotechnology.

## **4.2. In vitro evolution of heterochiral ribonuclease ribozymes**

### **4.2.1. Selection Methodology**

An “on the bead” selection strategy successfully yielded several unique ribonuclease ribozymes with *cis*-cleaving activities, and a single *trans*-cleaving ribozyme. Despite the inherent challenge of evolving catalysis based on heterochiral interactions it took only 8 and 9 rounds to evolve for *cis*-activity and *trans*-activity (reselecting from mutagenized *cis*-ribozymes) respectively. In particular, the gel-based selection strategy as an added stringency to the selection process was instrumental to the success of the selection experiment. However, the affinity of the single *trans*-ribozyme for the substrate is very low. This is evident from the slow kinetics of *trans*-cleavage with a  $k_{\text{obs}}$  of only  $0.0017 \text{ h}^{-1}$ . This can be improved by further evolution by 1) mutagenesis of the *trans*-ribozyme, 2) incorporating random domains in the new selection libraries, 3) incorporating the cross-chiral TAR aptamer sequence in the sequence library for enhancing binding (2) or 4) direct selection for *trans*-activity by methods that facilitate linking of phenotype to genotype such as oil-water emulsions etc. (3). Another interesting experiment would be targeted mutagenesis of the conserved motifs obtained after reselection and monitoring of the evolutionary progress. High throughput sequencing could be valuable for a better understanding of the fitness landscape in general.

#### 4.2.2. Model System for Characterizing complex Heterochiral Interactions and possible applications

The *cis*-ribonucleases from the first selection have similar predicted secondary structures with stable stems and several short stem-loop structures that could be amenable to conformational rearrangements. The three most efficient ribozymes have range of equilibrium yields ranging from ~40-70% in 3 hours at 20 mM Mg, with first-order rate constants in the range 0.013-0.025 min<sup>-1</sup>. Compared to the single-turnover kinetics of some of the fastest variants of the hammerhead ribozyme, these ribozymes are slower by two orders of magnitude. This is however only a crude measure of their efficiency since salt and buffer conditions also have a strong impact. Since, our focus was on obtaining *trans*-activity, detailed kinetic characterization of the *cis*-activity was not done. It is expected that increasing Mg<sup>2+</sup> concentrations or pH could enhance the kinetics. Comparison of the kinetics at varying pH values could also indicate the mechanism of cleavage (4-6). Given that the HHR and other naturally occurring ribozymes have evolved over millions of years and have conserved catalytic domains across different variants, it is not unexpected for them to have fast kinetics. Unique ribozymes evolved in selection experiments in the test tube on the other hand, have a low probability of being enriched, given the diversity of sample space, slight changes in selection conditions and day-to-day variations. Our results are thus promising in the sense that not only were we able to isolate several heterochiral ribonucleases, but they also had unique sequences with distinctly different cleavage sites. The cleavage sites though specific for each ribozyme, and on the unpaired loop region of the TAR hairpin, are different for the first generation of ribozymes selected; while R8c-1 and 3 cleave between a U and G with high yields, R8c-6 cleave

between a G and G with comparably lower yields. It is also particularly interesting that upon increasing the stability of the target stem, this diversity is lost. It will be interesting to further investigate the structural interactions that are making the G-G dinucleotide junction the only accessible site when the stem is extended, hindering access to the obviously more preferred U-G site that is associated with higher cleavage yields. It is known that the U and G are dynamic residues that engage in transient looping in and out motions which could explain higher cleavage yield (7). On the other hand, one of the residues of the G-G dinucleotide is engaged in a dynamic G-C WC base-pair which could explain the lower yield at this site (7). It is also known that the dynamic nature of the distal loop is influenced by its interactions with the long stem extending down from the UCU bulge (7) and this could explain why the G-G is the only accessible site for cleavage with the longer target. Thus, these ribonucleases have highly specific, structure-based interactions, and are sensitive to dynamic changes in the substrates. Novel tools based on these *cis*-ribozymes can thus be developed as structure sensors, and even for identification of TAR mutants that maintain the overall native structure that enables interactions with the Tat protein and other elements crucial for the HIV transcription elongation.

It is now known that there are many complex interactions involving both the UCU bulge and the 6-nt apical loop that mediate further binding to the Tat protein and other effectors and that targeting the bulge alone is not sufficient to disrupt the viral life cycle (8,9). Inhibitors for such native interactions have been a major focus of HIV drug discovery projects. Now, since L-TAR RNA has been shown to be a decoy for the native Tat protein despite the opposite stereochemistry (10), it is possible that interactions of the apical loop in L-chirality could also mimic the native interactions of the D-TAR substrate.



A screening assay for inhibitors can thus be devised with the *cis*-ribonucleases, where an effective interaction of the inhibitor with the loop will prevent the ribozyme mediated cleavage at that site.

The diversity in cleavage sites clearly shows the advantage of selection of catalytic behavior based on heterochiral structural interactions. In naturally occurring ribonuclease ribozymes, complex interactions between multiple secondary structures are responsible for aligning the substrate within the catalytic domain that influences site selection; in artificial selections, the site is predetermined by substrate binding arms. A cross-chiral ribozyme thus captures the complexity of naturally evolving systems more accurately by providing more conformational and structural freedom. This is also clear from the truncation studies on the *cis*-cleaving ribozymes where terminal truncations are not well-tolerated, and internal truncations aimed at shortening redundant stem regions always don't have the expected outcome. The different elements thus seem to have synergistic interactions that contribute to efficiency of cleavage, rate kinetics as well as site specificity. It is imperative therefore to carry out detailed structure probing of the ribozymes either in the presence or absence of the target by SHAPE profiling for example (11).

#### **4.2.3. Unique dependence on 5'-triphosphate**

The highly efficient ribozymes evolved from reselection were found to develop a dependence on the 5'-triphosphate that is always introduced during *in vitro* transcription. Removal of this group completely eliminates the weak *trans* activity, and severely inhibits the *cis* activity. The 5'-triphosphate is introduced during *in vitro* transcription and is

present in all ribozymes evolved through *in vitro* selection. The direct participation of this group in the catalytic reaction is not usually critical except in cases where phosphate transfer is involved, for example in capping reactions (12). In our case, the triphosphate could be involved in metal-ion co-ordination. This could be tested by substitution of phosphodiester with phosphorothioate that would increase the softness of the anion, and thus affect co-ordination with  $Mg^{2+}$ , potentially leading to reduced activity. Substitution by monophosphate instead of triphosphate was found to rescue some of the *cis* activity, making the role in coordination more plausible. The alternative role of phosphate exchange could be tested by radiolabeling the terminal triphosphate and examining the fate of the label following the cleavage reaction.

#### **4.2.4. Future Therapeutic Potential**

The TAR hairpin is a valuable therapeutic target; the apical loop along with the internal bulge are signature motifs required for interaction with other effectors to mediate transcription elongation of the HIV-1 genome. Targeted cleavage at the loop would thus irreversibly inactivate the TAR element and disrupt the viral replication process. This would require the *trans*-active ribozyme to be first synthesized in the L-chirality. For this purpose, I have employed different strategies of which the heterochiral RNA ligase mediated assembly of shorter fragments was most successful (13). However, the requirement of the 5'-triphosphate further complicated the accessibility of the 110 nucleotide long *trans*-ribozyme in sufficient quantities. Ongoing work is being focused on alternative routes to assemble both the *cis* and *trans*-ribozymes in the L-chirality.

### **4.3. Kinetic Characterization of Heterochiral DNA Strand Displacement Reactions**

The main goal of this part of my thesis was to investigate in detail the effect of systematic changes in key parameters that have been used routinely and extensively in literature for modulating strand displacement kinetics. We demonstrated that like all-DNA strand displacement reactions, the kinetics of strand displacements from a PNA-DNA heteroduplex depends strongly on a few factors. However, important differences exist between the two systems and also within the homochiral and heterochiral DNA strand displacement reactions from a PNA-DNA heteroduplex. We have shown that the second-order rate constant varies over two orders of magnitude when the toehold length is varied from 0- 10 nucleotides but unlike all-DNA strand displacement reactions, the rate doesn't saturate at even that high toehold length, which is evidently because of the higher PNA-DNA bond strength compared to DNA-DNA. This can be further explored in future designs to identify the toehold length at which saturation is achieved, as well as the effect of sequence composition on the kinetics.

We have shown that heterochiral strand displacement reactions in general are highly sensitive to mismatches within the PNA-DNA duplex especially at the junction positions. Such unique insights into position dependent kinetic modulation could be exploited in more complex reaction networks by incorporating mismatch correction or formation as additional parameters (14,15). The unique toehold exchange mechanism that favors the heterochiral reaction can enable modular reporter design and catalytic systems. Catalytic designs will be important in detection of endogenous D-DNA/ RNA with low abundance. In this context, the kinetics with RNA invaders show great enhancement compared to the DNA counterparts. Additionally, preliminary results show that RNA

structure in the toehold could have some subtle effects on kinetics that should be probed further by explicit design of structured toeholds. Finally, the unique stereochemical control of kinetics affords higher versatility in design of such systems. So far, we have focused on studying the decreased toehold stability in heterochiral displacements as a significant retardant of kinetics. The next obvious step would be to look at the branch migration step, separate from the toehold nucleation. Preliminary experiments looking at decreasing the “chiral conflict” in the toehold by systematic and incremental introduction of chiral PNA monomers are underway. Results suggest that indeed, this strategy enhances the rate kinetics appreciably, making the implementation of such nanodevices in biological applications more facile. However, this does compromise mismatch sensitivity which is expected to stem from the now reduced kinetic barrier. Overall, this work helps establish a set of design principles for dynamic DNA nanodevices made of L-DNA backbone that can be adapted to different applications at the interface of endogenous nucleic acids.

#### 4.4. References

1. Hoshika, S., Leal, N.A., Kim, M.-J., Kim, M.-S., Karalkar, N.B., Kim, H.-J., Bates, A.M., Watkins, N.E., SantaLucia, H.A. and Meyer, A.J. (2019) Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science*, **363**, 884-887.
2. Sczepanski, J.T. and Joyce, G.F. (2013) Binding of a structured D-RNA molecule by an L-RNA aptamer. *Journal of the American Chemical Society*, **135**, 13290-13293.
3. Lu, W.-C. and Ellington, A.D. (2013) In vitro selection of proteins via emulsion compartments. *Methods*, **60**, 75-80.
4. Doherty, E.A. and Doudna, J.A. (2001) Ribozyme structures and mechanisms. *Annual Review of Biophysics and Biomolecular Structure*, **30**, 457-475.
5. Lilley, D.M.J. (2011) Mechanisms of RNA catalysis. *Philosophical Transactions of the Royal Society B Biological Sciences*, **366**, 2910-2917.
6. Liu, Y., Wilson, T.J., McPhee, S.A. and Lilley, D.M. (2014) Crystal structure and mechanistic investigation of the twister ribozyme. *Nature Chemical Biology*, **10**, 739-744.
7. Dethoff, E.A., Hansen, A.L., Musselman, C., Watt, E.D., Andricioaei, I. and Al-Hashimi, H.M. (2008) Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophysical Journal*, **95**, 3906-3915.
8. Schulze-Gahmen, U. and Hurley, J.H. (2018) Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. *Proceedings of the National Academy of Sciences*, **115**, 12973-12978.
9. Davidson, A., Leeper, T.C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J.A. and Varani, G. (2009) Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proceedings of the National Academy of Sciences*, **106**, 11931-11936.
10. Garbesi, A., Hamy, F., Maffini, M., Albrecht, G. and Klimkait, T. (1998) TAR-RNA binding by HIV-1 Tat protein is selectively inhibited by its L-enantiomer. *Nucleic Acids Research*, **26**, 2886-2890.
11. Smola, M.J. and Weeks, K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nature Protocols*, **13**, 1181-1195.

12. Zaher, H.S., Watkins, R.A. and Unrau, P.J. (2006) Two independently selected capping ribozymes share similar substrate requirements. *RNA*, **12**, 1949-1958.
13. Szcepanski, J.T. and Joyce, G.F. (2014) A cross-chiral RNA polymerase ribozyme. *Nature*, **515**, 440-442.
14. Haley, N.E., Ouldrige, T.E., Ruiz, I.M., Geraldini, A., Louis, A.A., Bath, J. and Turberfield, A.J. (2020) Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement. *Nature Communications*, **11**, 1-11.
15. Irmisch, P., Ouldrige, T.E. and Seidel, R. (2020) Modeling DNA-strand displacement reactions in the presence of base-pair mismatches. *Journal of the American Chemical Society*, **142**, 11451-11463.

APPENDIX A

**Table A-1: Sequences of targets, libraries and *cis*-ribozymes**

<b>Sample Name</b>	<b>Sequence</b>
L-TAR <sub>S</sub>	<b>5'-P-rCrC-Sp18-Sp18-hv-Sp18-CCAGAUCUGAGCCUGGGAGCUCUCUGG-Sp18-Biotin-3'</b>
L-TAR <sub>S</sub> (FAM)	<b>5'-P-rCrC-Sp18-hv-Sp18-Fluorescein-Sp18-CCAGAUCUGAGCCUGGGAGCUCUCUGG-3'</b>
L-TAR <sub>L</sub>	<b>5'-Biotin-Sp18-GGU UAG ACC AGA UCU GAG CCU GGG AGC UCU CUG GCU AAC C-3'-Sp18-FAM-Sp18-PC linker-Sp18-Sp18-rC-rC-P-5'</b>
D-R161	<b>5'-TTCTAATACGACTCACTATAGGATCGTCAGTGCATTGAGA-N70-GGTGGTATCCCCAAGGGGAC-3'</b>
Rev 161	<b>5'-GTCCCCTTGGGGATAACCACC-3'</b>
Rev161_Ext	<b>5'-CCTTGGGGATAACCACC-3'</b>
Fwd161	<b>5'-TTCTAATACGACTCACTATAGGATCGTCAGTGCATTGAGA-3'</b>

Table A-1 Continued

Sample Name	Sequence
R8c-1 (also R8c-11 with a mutation and a deletion)	GGATCGTCAGTGCATTGAGATCATGGCTGTGCTAAGGTCCGGAATAT TTCTGAGCTAGCATATCAGTTCGCATAAATTGGACTGGTTGAGGGTG GTATCCCCAAGGGGAC
R8c-3 (also R8c-2,8,12 and 10 with a deletion)	GGATCGTCAGTGCATTGAGACCTAAGGTCCGTGCCGTTTCTATGATCT AGAGACTGTATGAGCTAGGCGCCAATCGTCTATAGGATGTGCGGTGG TATCCCCAAGGGGAC
R8c-6	GGATCGTCAGTGCATTGAGATGGGGTGGAATCCCGAGTTGGTCTCC TGTTTACCGTGCTCGGTCTGGGGTCGTTCCGGAATCTAGACGGAGGTG GTATCCCCAAGGGGAC
R8c-7	GGATCGTCAGTGCATTGAGACACGCGCTGTTGATCCCATCCATTGTGC CCTCATTGGGTGTAATGAGGTCCGGGAGAATACCAGCTGTGGGGTGG TATCCCCAAGGGGAC
R8c-9 (also R8c-14 with a mutation)	GGATCGTCAGTGCATTGAGAGCAACGTCTATGGTGCTTCTCTTCTCA CTGATCCGGTATAATTCCGAGGTCCGGGGTGCGCAACGGTTAGGTGG TATCCCCAAGGGGAC
R8c-13	GGATCGTCAGTGCATTGAGAGACAAGGGTCTCGATGCTTTTGCGACG GGTTGATCTGGAATGATGGAAAGCCTGAAGGGGAAGCATGGACAGG TGGTATCCCCAAGGGGAC
R8c-15	GGATCGTCAGTGCATTGAGAGTCGAAGGTCCGCGGCTAATACCGTGA GCTCGACACTAGTGAAAAGCGGCTTGCGGGGC AGGCTGTGATCGGTGGTATCCCCAAGGGGAC



**Table A-2. Conditions for a) *cis*-Selection and b) *trans*-Selection**

**a.**

Individual round conditions for *cis*-Selection<sup>a</sup>

Round #	Positive Selection Incubation Time
1-2	6 hours
3-6	3 hours
7	2 minutes
8	30 seconds

<sup>a</sup>Constant conditions for both selections:

100 mM NaCl, 25 mM TRIS-HCl ( pH 7.6), 1mM EDTA, 0.05% Tween 20 and 20 mM Mg<sup>2+</sup>, room temperature.

**b.**

Individual round conditions for *trans*-Selection<sup>a</sup>

Round #	Positive Selection Incubation Time
1-2	2 hours
3	30 minutes
4	5 minutes
5	1 minute
6	10 seconds
7-9	3 seconds (instantaneous)

**Table A-3:**

<b>X (R8c-3X)</b>	<b>Description</b>	<b>%Activity</b>
T1	-10 nts from 5'-terminal	50
T7	S1-T1	40
T9	S1-T2	30
T13	S1-T2 in combination with S2-T2	27
T16	S1-T1 in combination with S2-T2	16
T17	3T9, but the end loop of TATAG replaced by more generic GAAA	40
3T18	Essentially 3T16 but two G- U wobbles replaced by G-C as in some mutants of 3T13	18
3T19	3T16, with the extra stem of T7 (compared to T9) mismatches C-U and G-U corrected	0
3T21	Clone T17 (i.e Clone T9 with GAAA tetraloop capping the shorter stem) but with two extra base- pairs. No Clone T13 178	23

**Table A-4 Effect of 5'-triphosphate on *cis* and *trans*-activities of R8c-3T13 and a more efficient version of R9t-24 (Q). All conditions are same as the characterization assays described in the text.**

Clone Variant	20mM Mg 3h	50mM Mg 3h	20mM Mg ov.	50mM Mg ov.
D-Clone T13	22%	30%		
Dephosphorylated D-T13		20%		
Monophosphorylated D-T13	20%	22%	24%	30%
D-Q (in cis)	53%	60%	65%	73%
Dephosphorylated D-Q (in cis)	14%	10%	20%	30%
All forms of mutant D-CloneT13-d	40%	40%	40%	40%

**Fig A-1 GEL BASED STRATEGY.**

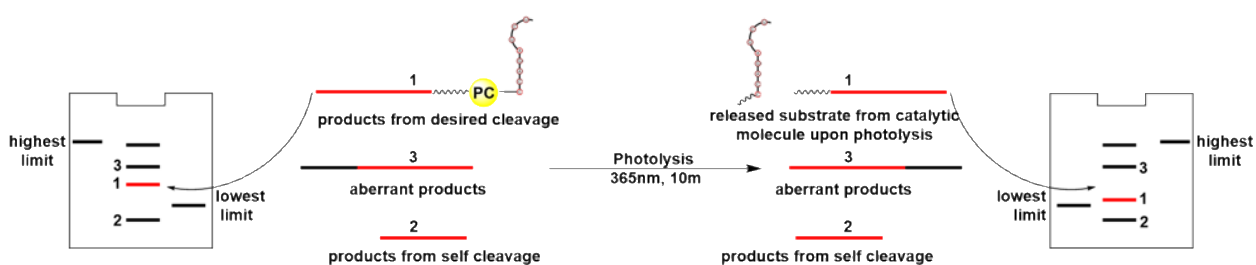
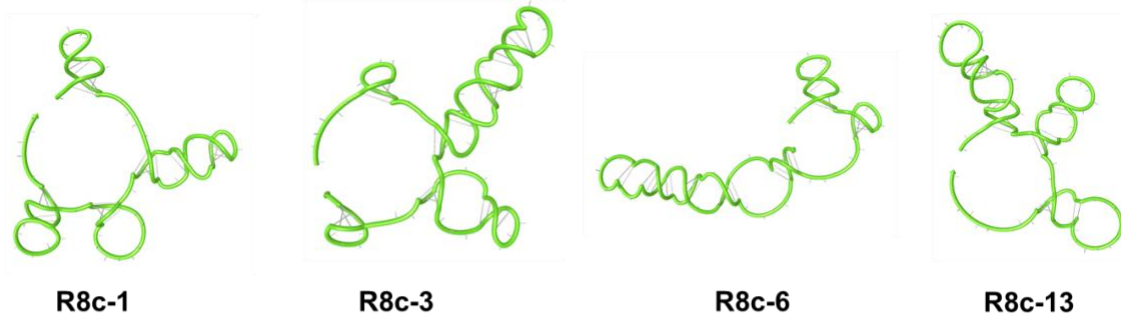


Fig A-2 Secondary Structures of *cis*-cleaving R8c-1, 3, 6 and 13.

a.



b.

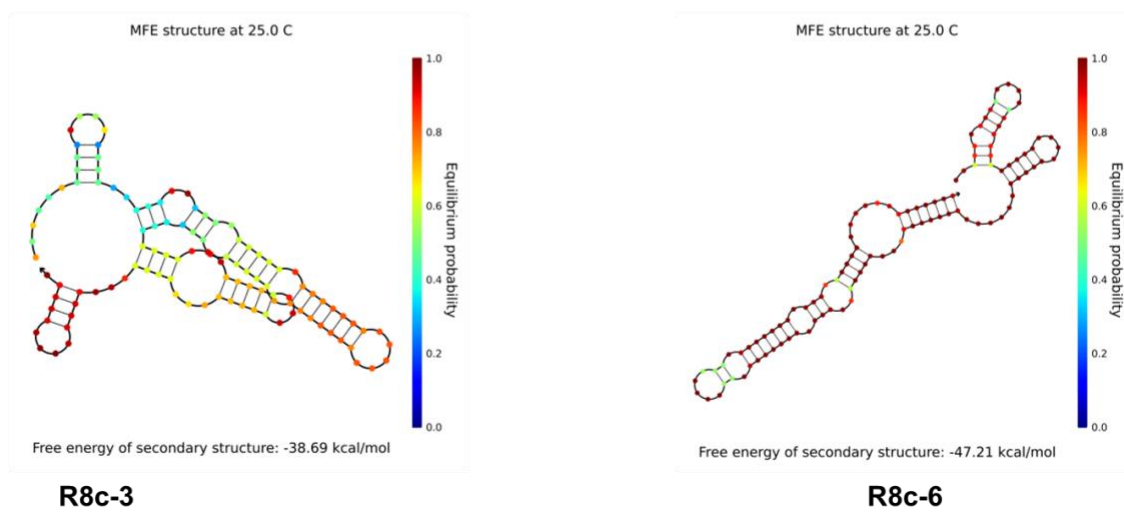


Fig. A-3 Cleavage sites are identical for R8c-1,3,6 and 13 and R9t-21-25 with L-TAR<sub>L</sub> at G16-G17 (w.r.t to L-TAR<sub>S</sub>)

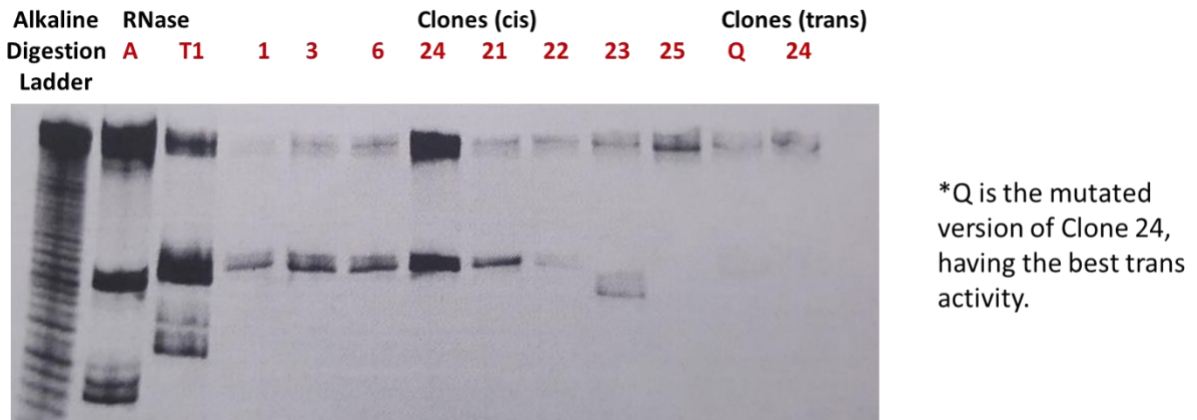


Fig. A-4 Libraries for the 1<sup>st</sup> Reselection Strategy based on R8c-3T13.

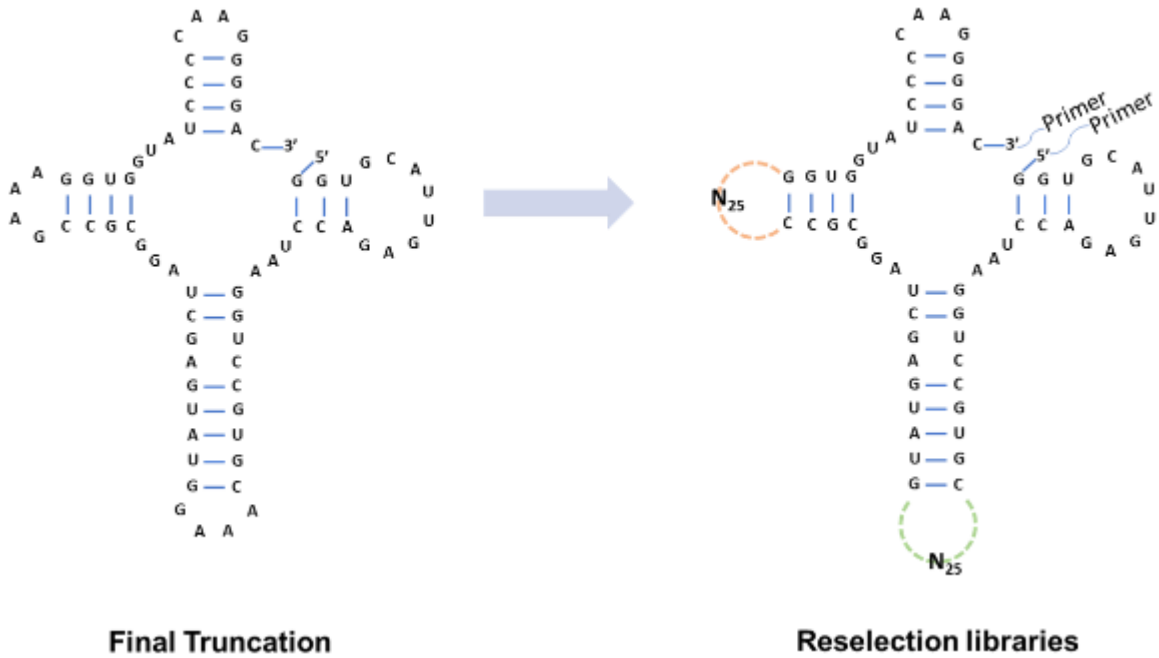


Fig. A-5 Cleavage activities of *cis*-ribonuclease ribozymes with L-TAR<sub>L</sub>.

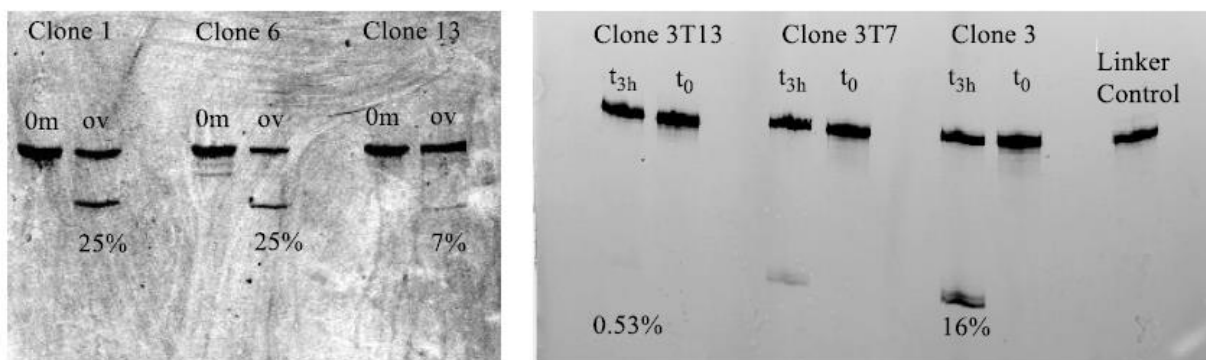


Fig. A-6. Sequencing data from Rounds 6 and 9 from the library based on R8c-6.

6 GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTTACCGTGCTCGGTCTGGGGTCGTTCCGGAATCTAGACGGAGGTGGTATCCCCAAGGGGAC WT

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGATTATCGTGCTAGCTTGGGGTCGT AATCTAGACGAAAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTTACCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTACCGTGCTCGGTCTGAGGTCGCTCGGAATGTGGACGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTCTCTACCTAGCTTGGTATAAGATCGTTTGAATCTAGGCAGGGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATAGGGGGAAAAATCCCGAGTTGGACTGGTATTACCGTGCTACGACTGCGGTCGTTCCGGAATCTAGACGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTCTACAGTGCTCTGTCAGGGGTCGTTAGGAATCTTGCAGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTCCCGTGGTAGGTATACGA TCATTAGGAAACACGGCGGAGGTGGTATCCCCGGGAGGGC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTGCCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGA GGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTTATCGTGTAGGTCTGGGGTCGTTCCGGAATCCAGACGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC

GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC

	1	10	20	30	40	50	60	70	80	90	100	110
6	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTTACCGTGCTCGGTCTGGGGTCGTTCCGGAATCTAGACGGAGGTGGTATCCCCAAGGGGAC											WT
23	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
24	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
25	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
26	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTAACAGGGCCAGGCTGGTGTGTTTAGAAACGAAGTGGAGGTGGTATCCCCAAGGGGAC											
2	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
4	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
5	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
6	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
8	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
9	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
10	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
11	CGATCTCAAAGGGTAAATTAACCTTAGCCGCCGATTTCCCGTGTCCCATGGGAATCGATAAATGATATATAAAGTGAATAGGATACCGATACGGCACAAGGAAAGGAC											
12	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
13	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
14	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
17	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
18	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTGGCCGTGATCGGTATAAGATCAGTTGCAATCATGGCGGAGGTGGTATCCCCAAGGGGAC											
19	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
20	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	G	CCGTGCTAGGTATAAGATCATTAGGAAACACGGCGGAGGTGGTATCCCCAAGGGGAC									
21	AGATTGGTCAAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGT	AACAGGTCCAGTATAAGATCGTTT	AGAAACGAGGCGGAGGTGGTATCCCCAAGGGGAC									
22	GGATCGTCAGTGCATTGAGATGGGGTGAAAAATCCCGAGTTGGTCTCCTGTTAACAGGGCCAGCTGGTGTGTTT	AGAAACGAAAGTGGAGGTGGTATCCCCAAGGGGAC										

Fig. A-7. In line probing analysis of R9t-24 under saturating conditions of Ligand (L-TAR<sub>L</sub>) Reduced intensities of hydrolysis products with increasing concentrations of the TAR substrate indicates residues at those points may be involved in substrate recognition. Secondary structural elements B1, B2 and L2 seem to be the motifs with the clearest pattern of increased protection from substrate interaction.

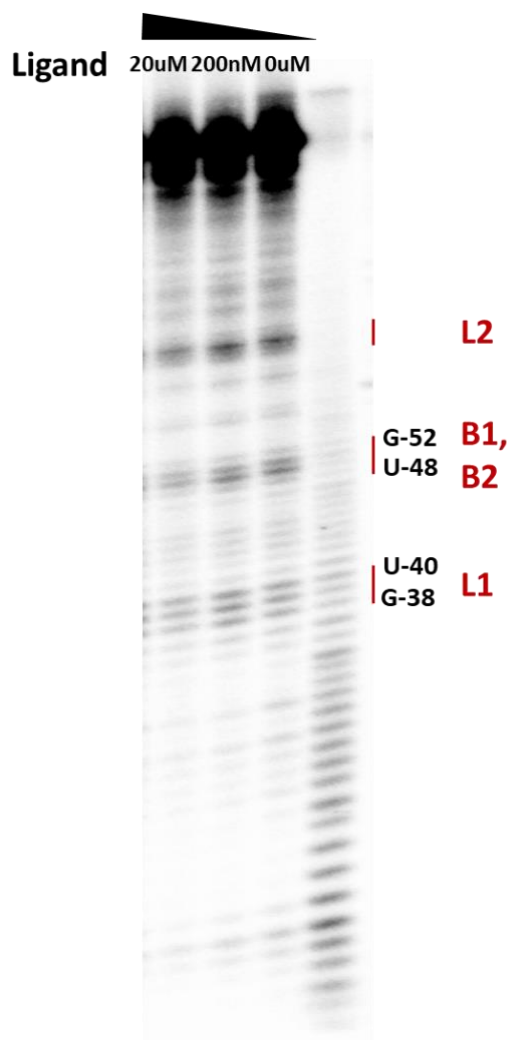
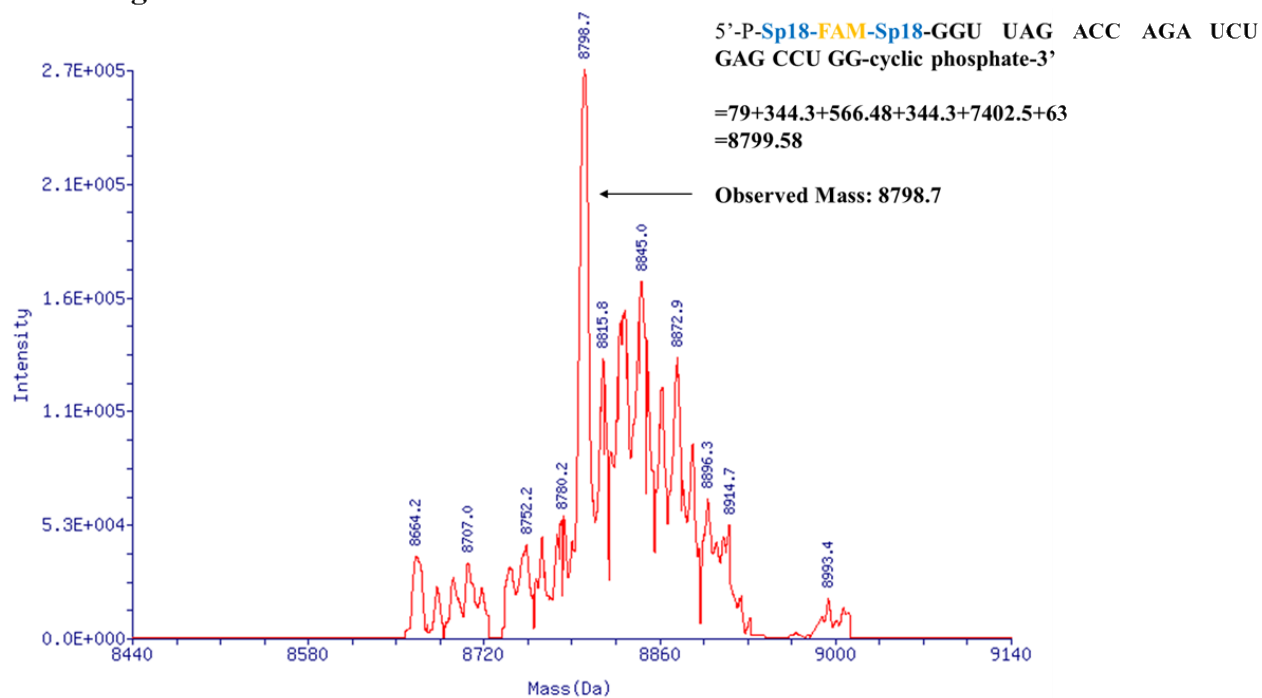




Fig. A-8 Mass Spectroscopic Characterization of Cleavage Products from the *cis*-reaction of R9t-24 with L-TAR<sub>L</sub>. (linker attachment in this version of L-TAR<sub>L</sub> same as in original L-TAR<sub>S</sub>)

**a. 5'-Fragment**



**b. 3'-Fragment**

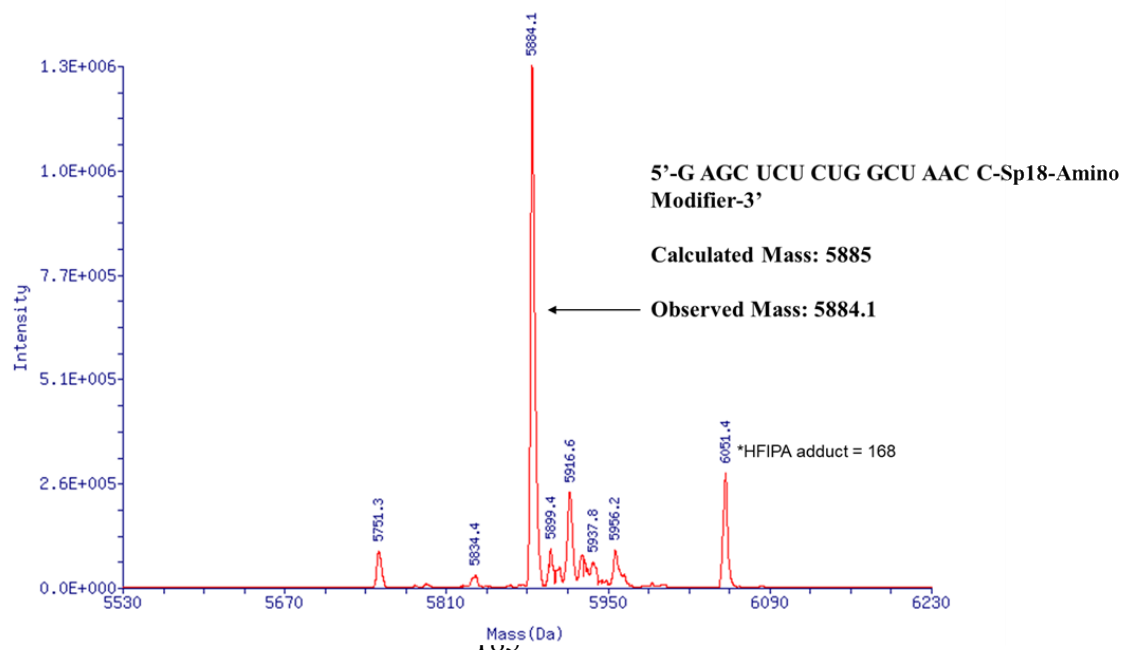
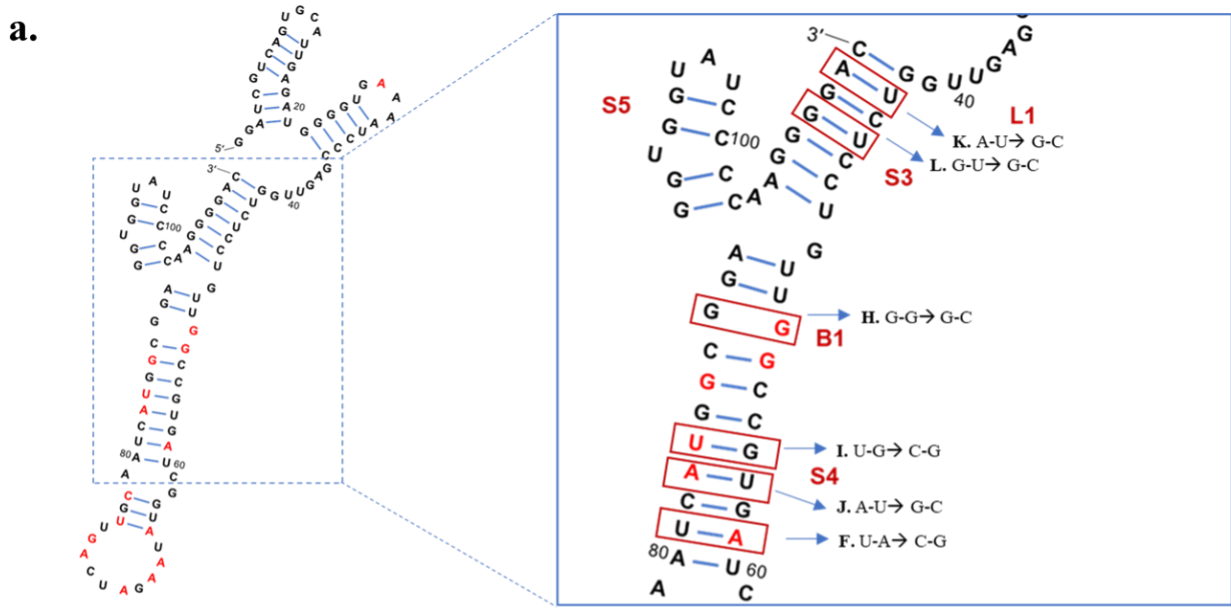
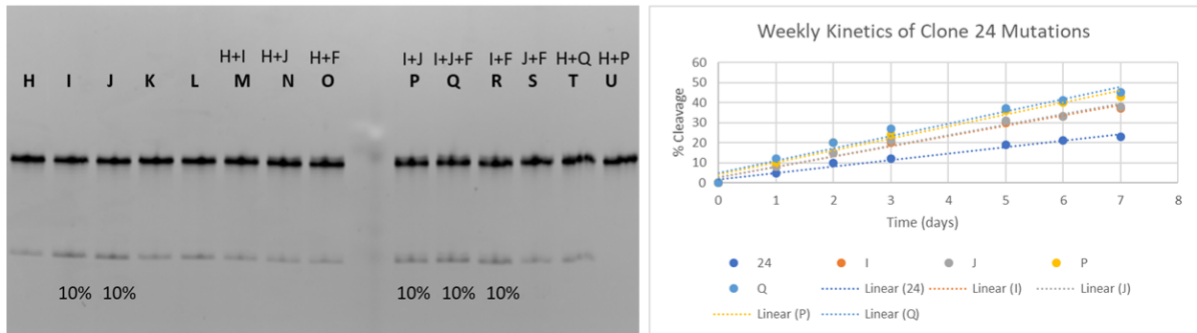


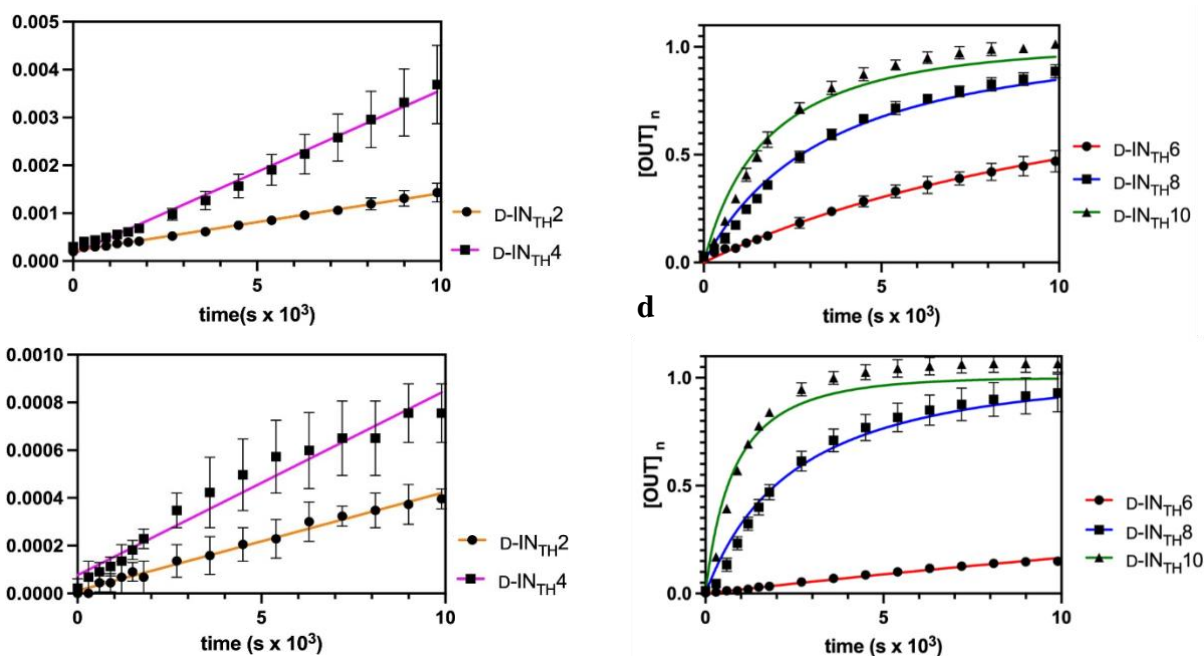
Fig. A-9 Mutation analysis of R9t-24 showing enhanced *trans*-activity by A-U to G-C replacement in certain stem elements.



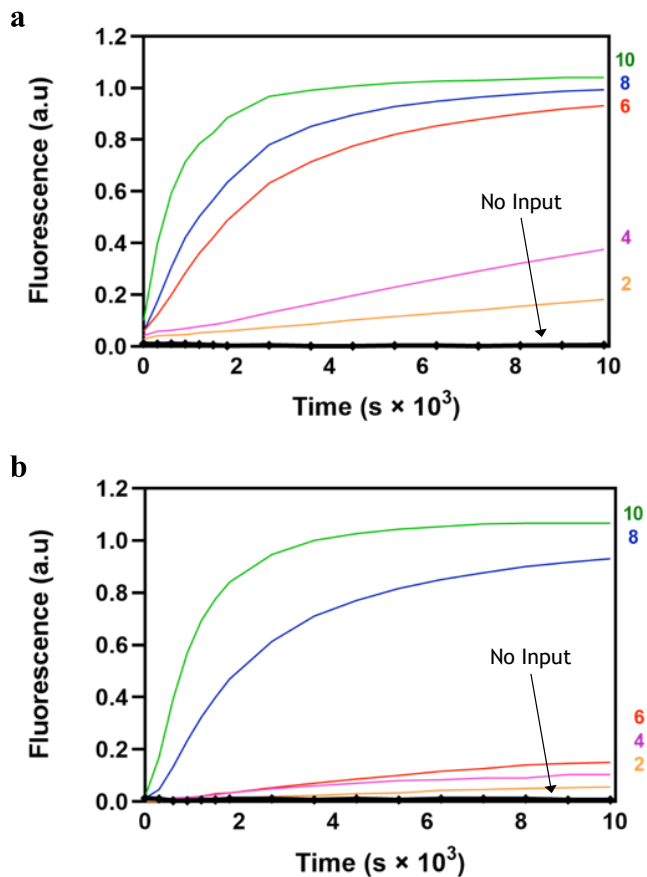
**b.**



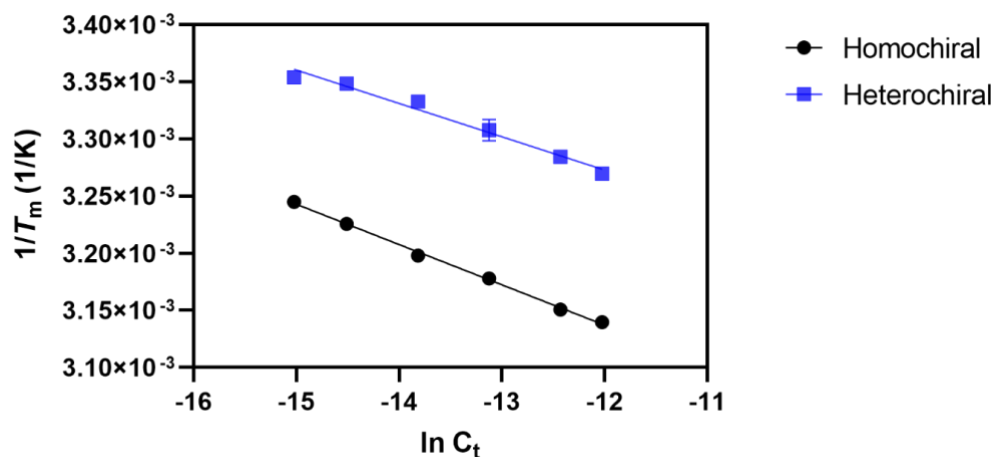




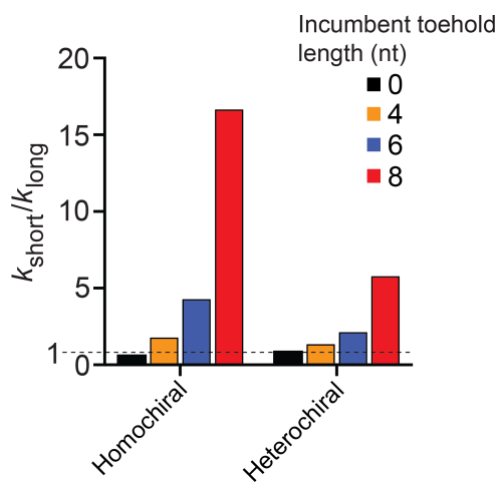
**Figure B-2.** Representative sets of data and fits used to derive second-order rate constants for homo- and heterochiral strand displacement reactions with different toehold lengths (See Materials and Methods for details). (a,b) Representative linear fits for slow homochiral (a) and heterochiral (b) strand displacement reactions with inputs containing 2 or 4 nucleotide toeholds (D-IN<sub>TH</sub>2 and D-IN<sub>TH</sub>4, respectively). The left-hand side of equation 2 (Materials and Methods) was plotted on the y-axis against time. Rate constants (Table S2) were extracted from the slope of the fit line. (c,d) Representative fits for fast homochiral (c) and heterochiral (d) strand displacement reactions with inputs containing 6, 8 and 10 nucleotide toeholds (D-IN<sub>TH</sub>6, D-IN<sub>TH</sub>8 and D-IN<sub>TH</sub>10, respectively). Rate constants (Table S2) were extracted by fitting equation 3 (Materials and Methods) to all the data points.



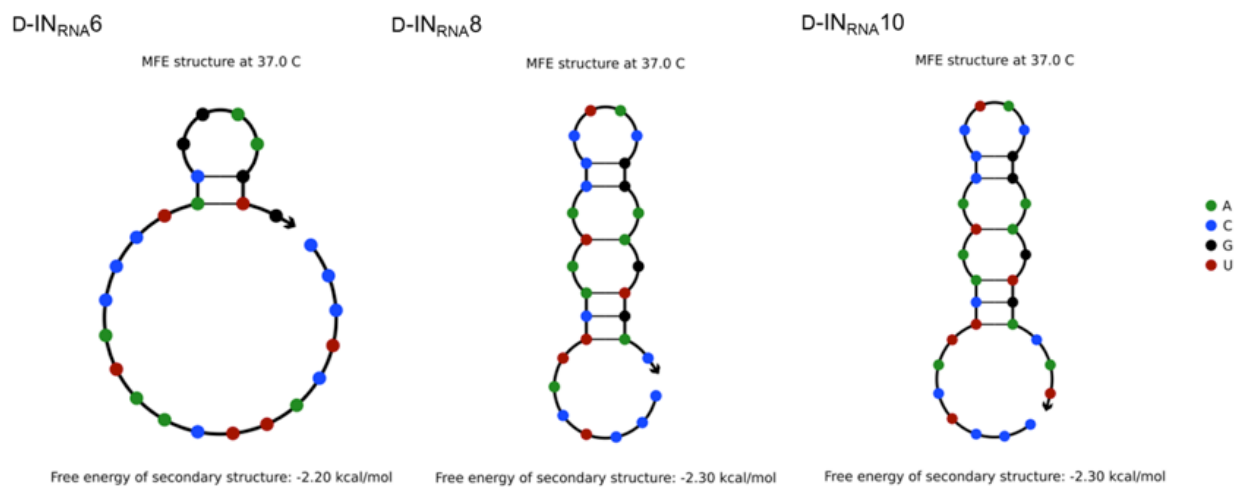
**Figure B-3.** Strand displacement from a PNA-DNA heteroduplex is strongly dependent on toehold length. Under the reaction conditions used in this work, essentially no fluorescence signal was observed in the absence of an input for either the homochiral (a) or heterochiral (b) reactions. The length of the toehold is indicated on the right y-axis. The depicted reactions contained either 0 nM or 150 nM input strand, 100 nM D/L-Ai, 300 nM D/L-R, 300 mM NaCl, 1 mM EDTA, and 10 mM Tris (pH 7.6) and were carried out at 37 °C.



**Figure B-4.** Thermodynamic parameters  $\Delta H^\circ$  and  $\Delta S^\circ$  were extracted from the linear best fit line of either the reciprocal melting temperature ( $1/T_m$ ) of the homochiral or heterochiral melting data against  $\ln C_t$ , where  $C_t$  is the total component concentration (strand D/L- $IN_F$  + D-Aq) used in each experiment.



**Figure B-5.** Ratio of rate constants for strand displacement from the short (D/L- $Ai_S$ ) and long (D/L- $Ai$ ) heteroduplexes for each of the given incumbent toehold lengths.



**Figure B-6.** RNA inputs with 8 and 10 nucleotide toeholds (IN<sub>RNA8</sub> and IN<sub>RNA10</sub>, respectively) are predicted to form more extensive secondary structure than the 6 toehold RNA input (IN<sub>RNA6</sub>). See Table S1 for sequences. RNA secondary structure was predicted using NUPACK.

**Table B-1.** Names, sequences, and chirality of strands used in this work. Dye and quencher modifications are indicated in bold, while mismatches within indicated strands are italicized. D-DNA (black), L-DNA (blue), PNA (green), and D-RNA (red) are indicated by color.

Sequence Name	Sequence Identity 5'→3'	Oligomer Stereochemistry
PNA <sub>L</sub>	<b>ATGTCACTTCCGTAGGGTATTGAATGAGGG</b>	Achiral
PNA <sub>S</sub>	<b>ATGTCACTTCCGTAGGGTATTGAATG</b>	Achiral
D-OUT	CCACATACATCATATTCCCTCATTCAATACCCTACG	D
D-F	<b>Cy3</b> -CACATCATATTCCCTCATTC	D
D-Q	AGGGTATTGAATGAGGGAATATGATGTG- <b>BHQ2</b>	D
D-IN <sub>TH10</sub>	CCCTCATTCAATACCCTACGGAAGTGACAT	D
D-IN <sub>TH8</sub>	CCCTCATTCAATACCCTACGGAAGTGAC	D
D-IN <sub>TH6</sub>	CCCTCATTCAATACCCTACGGAAGTG	D
D-IN <sub>TH4</sub>	CCCTCATTCAATACCCTACGGAAG	D
D-IN <sub>TH2</sub>	CCCTCATTCAATACCCTACGGA	D
D-IN <sub>MM A</sub>	CCCTCATTCAATACCCTACGGAAGAGAC	D
D-IN <sub>MM B</sub>	CCCTCATTCAATACCCTACGGAACAGAC	D
D-IN <sub>MM C</sub>	CCCTCATTCAATACCCTACGCAACTGAC	D
D-IN <sub>MM D</sub>	CCCTCATTCAATACCCTACCGAAGTGAC	D
D-IN <sub>MM E</sub>	CCCTCATTCAATACCCTACGGAAGTGAC	D
D-IN <sub>MM F</sub>	CCGTCATTCAATACCCTACGGAAGTGAC	D
D-IN <sub>INC8</sub>	CAATACCCTACGGAAGTGAC	D
D-IN <sub>INC6</sub>	TTCAATACCCTACGGAAGTGAC	D
D-IN <sub>INC4</sub>	CATTCAATACCCTACGGAAGTGAC	D
D-IN <sub>INC8(S)</sub>	ACCCTACGGAAGTGAC	D
D-IN <sub>INC6(S)</sub>	ATACCCTACGGAAGTGAC	D
D-IN <sub>INC4(S)</sub>	CAATACCCTACGGAAGTGAC	D
D-IN <sub>RNA10</sub>	<b>CCCUCAUUCAAUACCCUACGGAAGUGACAU</b>	D
D-IN <sub>RNA8</sub>	<b>CCCUCAUUCAAUACCCUACGGAAGUGAC</b>	D
D-IN <sub>RNA6</sub>	<b>CCCUCAUUCAAUACCCUACGGAAGUG</b>	D
D-IN <sub>RNA C</sub>	<b>CCCUCAUUCAAUACCCUACGCAACUGAC</b>	D
D-IN <sub>RNA D</sub>	<b>CCCUCAUUCAAUACCCUACCGAAGUGAC</b>	D
D-IN <sub>F</sub>	<b>FAM</b> -TTTTTGAAGTGACAT	D
D-OUT <sub>q</sub>	CCACATACATCATATTCCCTCATTCAATACCC[T( <b>BHQ2</b> )] ACG	D
L-OUT	<b>CCACATACATCATATTCCCTCATTCAATACCCTACG</b>	L
L-F	<b>Cy3</b> -CACATCATATTCCCTCATTC	L
L-Q	AGGGTATTGAATGAGGGAATATGATGTG- <b>BHQ2</b>	L
L-IN <sub>F</sub>	<b>FAM</b> -TTTTTGAAGTGACAT	L
L-IN <sub>TH10</sub>	<b>CCCTCATTCAATACCCTACGGAAGTGACAT</b>	L
L-IN <sub>RNA10</sub>	<b>CCCUCAUUCAAUACCCUACGGAAGUGACAU</b>	L



Input	Description	Heterochiral $k$ ( $M^{-1}s^{-1}$ )	Homochiral $k$ ( $M^{-1}s^{-1}$ )
D-IN <sub>TH</sub> 10	10 nt toehold	9194 ± 399.34 <sup>b</sup>	20953.33 ± 1597.55 <sup>b</sup>
D-IN <sub>TH</sub> 8	8 nt toehold	2971.33 ± 474.71 <sup>b</sup>	10546 ± 720.11 <sup>b</sup>
D-IN <sub>TH</sub> 6	6 nt toehold	127.30 ± 1.39 <sup>b</sup>	2704.67 ± 310.84 <sup>a</sup>
D-IN <sub>TH</sub> 4	4 nt toehold	88.97 ± 15.69 <sup>a</sup>	365.23 ± 57.99 <sup>a</sup>
D-IN <sub>TH</sub> 2	2 nt toehold	43.09 ± 7.67 <sup>a</sup>	152.70 ± 11.91 <sup>a</sup>
D-IN <sub>TH</sub> 0	0 nt toehold	7 ± 1.28 <sup>a</sup>	37.97 ± 1.80 <sup>a</sup>
D-IN <sub>MM</sub> A	Mismatched input A	100.99 ± 16.29 <sup>a</sup>	1810 ± 338.91 <sup>a</sup>
D-IN <sub>MM</sub> B	Mismatched input B	82.22 ± 6.29 <sup>a</sup>	505.93 ± 77.08 <sup>a</sup>
D-IN <sub>MM</sub> C	Mismatched input C	22.43 ± 5.41 <sup>a</sup>	105.37 ± 9.18 <sup>a</sup>
D-IN <sub>MM</sub> D	Mismatched input D	69.90 ± 3.56 <sup>a</sup>	411.13 ± 41.99 <sup>a</sup>
D-IN <sub>MM</sub> E	Mismatched input E	1837 ± 405.33 <sup>b</sup>	4348.33 ± 426.92 <sup>b</sup>
D-IN <sub>MM</sub> F	Mismatched input F	3507 ± 425.97 <sup>b</sup>	5866 ± 90.71 <sup>b</sup>
D-IN <sub>INC</sub> 8	8 nt incumbent toehold on D/L-Ai	1208.33 ± 43.44 <sup>b</sup>	417 ± 35.25 <sup>b</sup>
D-IN <sub>INC</sub> 6	6 nt incumbent toehold on D/L-Ai	2799.67 ± 74.82 <sup>b</sup>	2493 ± 231.67 <sup>b</sup>
D-IN <sub>INC</sub> 4	4 nt incumbent toehold on D/L-Ai	3155.67 ± 298.27 <sup>b</sup>	4748.67 ± 196.54 <sup>b</sup>
D-IN <sub>TH</sub> 8 (S)	8 nt toehold on D/L-Ai <sub>5</sub>	2966.33 ± 370.46 <sup>b</sup>	7805.33 ± 280.55 <sup>b</sup>
D-IN <sub>INC</sub> 8(S)	8 nt incumbent toehold on D/L-Ai <sub>5</sub>	7112.67 ± 221.61 <sup>b</sup>	6989 ± 530.47 <sup>b</sup>
D-IN <sub>INC</sub> 6(S)	6 nt incumbent toehold on D/L-Ai <sub>5</sub>	6282.33 ± 454.18 <sup>b</sup>	10967.67 ± 1271.73 <sup>b</sup>
D-IN <sub>INC</sub> 4(S)	4 nt incumbent toehold on D/L-Ai <sub>5</sub>	4618.33 ± 218.14 <sup>b</sup>	9008.67 ± 464.88 <sup>b</sup>
D-IN <sub>RNA</sub> 10	10 nt toehold with RNA input	69572.67 ± 9911.76 <sup>b</sup>	114444.33 ± 13865 <sup>b</sup>
D-IN <sub>RNA</sub> 8	8 nt toehold with RNA input	30520.67 ± 2762.89 <sup>b</sup>	73907.67 ± 3600.21 <sup>b</sup>
D-IN <sub>RNA</sub> 6	6 nt toehold with RNA input	49470.33 ± 2267.36 <sup>b</sup>	106228.3 ± 10418.23 <sup>b</sup>

D-IN <sub>RNA</sub> C	RNA input with mismatch C	196 ± 6.24 <sup>a</sup>	352.17 ± 47.00 <sup>a</sup>
D-IN <sub>RNA</sub> D	RNA input with mismatch D	11039.67 ± 267.09 <sup>b</sup>	42165.67 ± 3370.70 <sup>b</sup>

<sup>a</sup>Rate constant extracted using Equation 2 as described in Materials and Methods Section

<sup>b</sup>Rate constant extracted using Equation 3 as described in Materials and Methods Section

**Table B-2.** Calculated rate constants for homochiral and heterochiral reactions described in the text. Rates represent the average of 3 fittings and their standard deviation.