

**VERSATILE FUNCTIONAL NUCLEIC
ACIDS AND THEIR APPLICATIONS IN
BIOSENSING**

**VERSATILE FUNCTIONAL NUCLEIC ACIDS AND
THEIR APPLICATIONS IN BIOSENSING**

By

Wenqing Zhang

B. HSc. McMaster University, Hamilton, Ontario, 2011

A Thesis

Submitted to the School of Graduate Studies

In partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

DOCTOR IN PHILOSOPHY (2015) McMaster University
(Biochemistry & Biomedical Sciences) Hamilton, Ontario

TITLE: Versatile Functional Nucleic Acids and Their Applications in
Biosensing

AUTHOR: Wenqing Zhang, B. HSc. (McMaster University)

SUPERVISOR: Dr. Yingfu Li

NUMBER OF PAGES: XI, 106

ABSTRACT

It is now widely known that some nucleic acid molecules, either DNA or RNA, are capable of forming intricate three-dimensional structures and carrying out functions of molecular recognition and catalysis. Most of known functional nucleic acids are isolated from DNA or RNA pools with random sequences using the technique of in vitro selection. With intensive research for the past three decades, a variety of functional nucleic acids have been discovered and examined for potential applications. The general objective of this thesis is to expand the repertoire of functional nucleic acids via new in vitro selection experiments and pursue their biosensing applications. I started by asking the question of whether it is possible to develop a new kind of functional nucleic acids: chimeric RNA/DNA substrates that have high activity for ribonuclease H2 from the important bacterial pathogen *Clostridium difficile* but much reduced activity towards the same enzymes from other bacterial species. The key rationale behind pursuing these special functional nucleic acids is my hypothesis that these molecules can eventually be developed into useful biosensors for diagnosing *Clostridium difficile* infection. For this reason, in my first project, I applied the in vitro selection technique to a random-sequence DNA pool, obtained several highly selective chimeric RNA/DNA substrates, and carried out in-depth analysis of their reactivities and their structural properties. During this study, I accidentally discovered a family of highly guanine-rich DNA molecules that are able to form an unusual guanine-quadruplex structure in 7 molar urea, a strong denaturing condition for nucleic acid structures. This discovery constitutes a novel observation and therefore, in my second project, I fully characterized the sequence and structural properties of these special DNA molecules and established

the conditions that allow these molecules to create stable structures in 7 molar urea. I then got interested in devising a unique application to take advantage of the urea-resistant property exhibited by these molecules. Towards this end, in my third project, I used one such DNA molecule to set up a DNA detection method capable of detecting single nucleotide polymorphism in very long DNA sequences, a desired application that has never been demonstrated before. The findings made in these projects contribute to the ever-growing appreciation of functional capability and practical utility of nucleic acids.

ACKNOWLEDGEMENT

My time spent in the Li Lab during the past few years has been incredible. I would like to express my sincere appreciation to my supervisor, Dr. Yingfu Li, for the opportunities he provided me to learn, grow, and use my creativity in the development of research projects. I have gratefully enjoyed the freedom offered by Dr. Li. I was given the opportunities to realize many of my scientific ideas. On the other hand, he was always there for me when I struggled with my weaknesses and failures. Besides scientific knowledge, I have gained significant improvement in many transferable skills, which will be more than valuable in my future endeavors.

I would also like to express my sincerely appreciation to my committee members, Dr. J. P. Xu, Dr. Christine Lee, and Dr. Murray Junop for their support, constructive criticisms, and suggestions.

I have had the privilege to befriend my fellow researchers in the Li Lab, who are brilliant, hard-working, passionate and kind. Our discussions on a wide range of topics were often insightful, inspirational, and kept all of us open minded.

Last but not least, I would like to express my gratitude to my family. I would like to thank my husband, Xiuhua Yan, for his years of support and understanding. I hope that my son, Alexander Yan, will someday read this thesis and find it inspiring. My father was the earliest audience of my “ideas” since I was very young. He has always encouraged me to think independently and critically. My mom has always motivated me with her own success and demonstrated the importance of continuous personal growth.

“The only true wisdom is in knowing you know nothing.”
-Socrates

Table of Contents

Chapter 1	1
A General Introduction on Discovery of Functional Nucleic Acids and Their Applications in Biosensing	1
1.1 Author's Preface	1
1.2 History of Functional Nucleic Acids	2
1.3 Advantages of SELEX Derived Aptamers and Nucleic Enzymes in Applications ...	5
1.4 Applications of Nucleic Acids as Biosensor Components	8
1.4.1 DNA Used as the Recognition Element in Biosensors	8
1.4.2 DNA Used as the Signal Transducer in Biosensors	11
1.4.3 Combining Nucleic Acid-Based Recognition Elements and Signal Transducers.....	13
1.5 Considerations for In Vitro Selection Process	18
1.5.1 Library Design.....	19
1.5.2 Population Diversity.....	20
1.5.3 Selective-Pressure Manipulation.....	21
1.6 Thesis Objective and Outline	22
1.6.1 Development of A Homologue-Specific Substrate for Bacterial RNase H2	22
1.6.2 The Accidental Discovery of a Highly Stable G-Quadruplex with Resistance to 7M-Urea Gel Electrophoresis	23
1.6.3 Utilizing highly stable UD G-quadruplexes for DNA sensing and highly specific detection of single nucleotide polymorphism	24
1.7 References	25
Chapter 2	34
In vitro Selection of a Fluorogenic DNA/RNA molecule as the Substrate for Ribonuclease H2 from Clostridium difficile	34
2.1 Author's Preface	34
2.1 Introduction	35
2.2 Results and Discussion	39
2.2.1 In vitro Selection	39
2.2.2 Activity Assessment of Enriched Sequences	43
2.2.3 Secondary Structure Prediction.....	45
2.2.4 Secondary Structure Characterization of FDRC1	48
2.2.5 Requirement of F and Q.....	51
2.2.6 Monitoring cleavage reaction by real-time fluorescence	52
2.3 Concluding Remarks and Future Directions	53
2.4 Materials and Methods	56
2.4.1 Enzymes, chemicals, and other materials.....	56
2.4.2 Synthesis and purification of oligonucleotides.	56
2.4.3 Molecular Cloning and Expression of RNase H2.	56
2.4.4 Protein Purification.....	58
2.4.5 In vitro selection.....	58
2.4.6 Assessment of the activity of top 10 classes.	60
2.4.7 Assessment of Activity of FDRC1 and its variants.....	61
2.4.8. Activity Assessment of DRC1-3B.	61
2.5 Reference	62
Chapter 3	67

Serendipitous Discovery of a Guanine-rich DNA Molecule with a Highly Stable Structure in Urea	67
3.1 Author's Preface.....	67
3.2 Abstract.....	68
3.3 Introduction	68
3.4 Results and Discussion	69
3.5 Conclusion.....	78
3.6. Methods	79
3.6.1 Enzymes, chemicals, and other materials.....	79
3.6.2 Synthesis and purification of oligonucleotides	80
3.6.3 ³² P-labelling of oligonucleotides	80
3.6.4 In vitro selection procedures	81
3.6.5 General procedure for dPAGE-based folding analysis of urea-resistant DNA molecules....	82
3.7. References	84
3.8 Supplementary Tables and Figures	87
Chapter 4.....	92
A DNA Switch for Detecting Single Nucleotide Polymorphism within a Long DNA Sequence Under Denaturing Conditions	92
4.1 Author's Preface.....	92
4.2 Manuscript: A DNA Switch for Detecting Single Nucleotide Polymorphism within a Long DNA Sequence Under Denaturing Conditions	93
4.3 Supplementary Information: A DNA Switch for Detecting Single Nucleotide Polymorphism within a Long DNA Sequence Under Denaturing Conditions	98
Chapter 5.....	103
Conclusion and Outlooks	103

List of Figures

Figure 1-1. The General Schematics of in vitro Selection	3
Figure 1-2. Gold Nanoparticle Labeled DNA Probes for ssDNA or RNA Detection...	9
Figure 1-3. Chemical Structure of G Quartet	11
Figure 1-4. Various Placement of Fluorophore and Quencher Pairs on RNA-cleaving DNAzymes and Their Substrates	12
Figure 1-5. Molecular Beacon and Aptamer Beacon	14
Figure 1-6. Catalytic Beacon Constructed using PW17 G-quadruplex.....	15
Figure 1-7. Substrate of RNA-cleaving DNAzyme Incorporated into a Molecular Beacon	16
Figure 1-8. A Complex Biosensor Constructed with Nucleic Acids and Protein Enzyme	17
Figure 2-1. RNA cleavage by a target-induced RNA-cleaving DNAzyme	36
Figure 2-2. Fluorogenic DNA/RNA (FDR) Substrate for <i>C. difficile</i> RNase H2 (CDH2)	38
Figure 2-3. Sequence Alignment of RNase H2 of <i>E. coli</i> (EC), <i>S. typhimurium</i> (ST), and <i>C. difficile</i> (CD), and conserved binding sites and catalytic sites	39
Figure 2-4. In vitro Selection.....	41
Table 2-1. In vitro selection conditions and outcomes	41
Figure 2-5. Monitoring selection progress	43
Table 2-2. Top 10 sequences in G10 pool	44
Figure 2-6. (A) Relative activity and (B) selectivity of top 10 sequences	44
Figure 2-7. Structural models predicted with m-fold for (A) FDRC1, and (B) FDRC6	46
Figure 2-8. Comparing the similarity of the bottom half of the secondary structures predicted for FDRC1 and FDRC6	47
Figure 2-9. Structural models predicted with Mfold for the remaining 8 top 10 sequences.....	48
Figure 2-10. Assessing the significance of pairing regions (P6 and P7), loop (L1), and bulges (B5 and B6).....	49
Figure 2-11. FDRC1-3 Constructs with Nucleotides Truncated from the 3'-End of the Sequence.....	50
Figure 2-12. Assessment of the Significance of Bulges within FDRC1-3A	51

Figure 2-13. Removing Fluorescein and Dabcyl from FDR1	52
Figure 2-14. Assessing the Sensitivity of FDRC1-3B Using Fluorescence Spectrometer.....	53
Figure 3-1. Discovery of DNA sequences with urea-resistant structures.....	71
Figure 3-2. Metal ion requirements for the folding of U2T	73
Figure 3-3. Time for U2T folding	74
Figure 3-4. Testing variant sequences of U2R	75
Figure 3-5. A structure model for U2R1	77
Figure S3-1. In vitro selection	88
Figure S3-2. Effect of varying concentrations of K ⁺ , Li ⁺ and NH ₄ ⁺ on the folding of U2T.....	89
Figure S3-3. Examination of sequence truncation on the structure folding of UD2 ...	89
Figure S3-4. Methylation interference of guanine nucleotides in U2T.....	90
Figure S3-5	91
Figure 4-1. Reporting systems constructed with ULX.....	93
Figure 4-2. G4 structure formation by ULX.....	94
Figure 4-3. Design of a harpin structure with a G-blocker element	94
Figure 4-4. SNP detection using a hairpin/G4 structure switch	95
Figure 4-5. Four additional sensors with lengthened probe sequences	96
Figure 4-S1. Discrimination of R50A and R50C using UL20GBP50, under 0MU or 7MU.....	101
Figure 4-S2. Testing the importance of the G-blocker	101
Figure 4-S3. Testing the UL20GB probe using other DNA targets	102
Figure 4-S4. Testing the specificity of UL20GBP90 in the presence of excess mismatched targets	102

Chapter 1

A General Introduction on Discovery of Functional Nucleic Acids and Their Applications in Biosensing

1.1 Author's Preface

The following introductory sections are intended to provide an overall description of the discovery, development and applications of functional nucleic acids. Specifically, section 1.5 is a modified version of a section of a published paper, which I primarily wrote under the guidance from Dr. Yingfu Li (citation listed below).

Zhang, W., Feng, Q., Chang, D., Tram, K. & Li, Y. (2016). In vitro selection of RNA-cleaving DNazymes for bacterial detection. *Methods*. 106, 66-75.

1.2 History of Functional Nucleic Acids

Nucleic acids have been commonly understood as the storage material for genetic information; however, their functions beyond that have also been increasingly recognized. Single-stranded RNA and DNA molecules can adopt intricate structures, allowing them to perform complex functions such as target recognition and enzymatic catalysis. Single-stranded nucleic acids that can bind to a specific target are called aptamers, and those with catalytic properties are called ribozymes (RNA) and DNAzymes (DNA). They are all defined as functional nucleic acids (FNAs). While some FNAs are found in nature, such as riboswitches and the self-splicing group I and Group II intron ribozymes (1,2), many artificial FNAs have been developed through a laboratory technique named systematic evolution of ligands by exponential enrichment (SELEX), also known as *in vitro* selection (3–5). Similar to Darwinian evolution's principle of "selection for the fittest", *in vitro* selection adopts the principle that the molecules with the specifically required function(s) are the most likely to "evolve" from a randomized pool. More specifically, a set of selection criteria, such as binding to a target, or catalyzing a reaction, are applied to a large pool of nucleic acid molecules with randomized sequences. Species in the pool that can perform the function of interest will be separated from the non-functional ones, and subsequently amplified. This amplified population then becomes the new generation of pooled sequences and will be subjected to the next round of selection. This process is often repeated for many rounds to allow the "fittest" species of interest to dominate the pool (Figure 1-1).

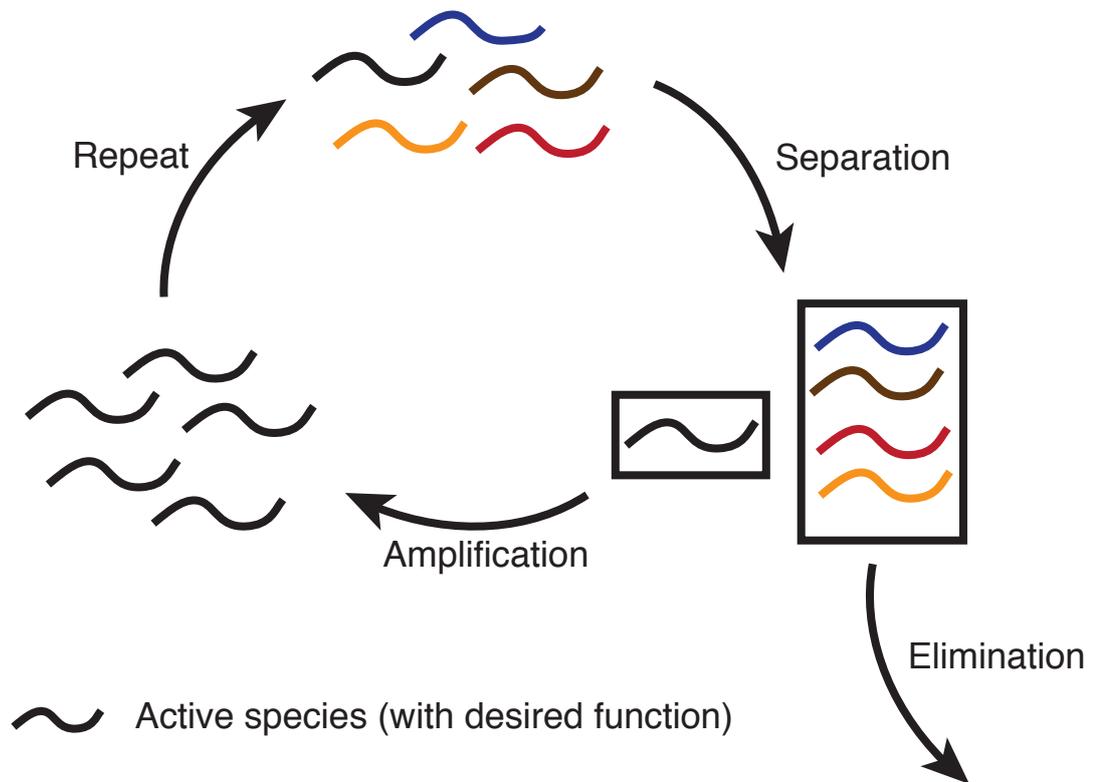


Figure 1-1. The General Schematics of in vitro Selection. They typically initiate with a pool of nucleic acid sequences with random sequences. By exerting a selective pressure on the population, and with an appropriate method of separation, species that do not meet the requirement of selective pressure will be eliminated, while the “fitted” or active sequences will be separated for amplification and the subsequent round of selection.

Test tube evolution was first attempted as early as the 1960s. By liberating a library of RNA and RNA-dependent RNA polymerase from cells into a test tube, Spiegelman and co-workers imposed a single selective pressure, namely, time of reaction, on the population. This led to the observation that only short RNA molecules were duplicated when time was limited (6). Later, the discovery of natural ribozymes inspired tremendous interests in developing FNAs using in vitro selection experiments (3–5). To date, in vitro selection-generated ribozymes demonstrate a wider range of functions in comparison to those of naturally occurring ribozymes (7). Since in vitro selection can start with a pool containing as many as 10^{16} of random sequences, it boasts an immense diversity of structures that could encode a large variation of functions; it is

assumed that, in this large pool, at least one variant can perform a desired function.

Early research interests for FNAs largely focused on RNA, but not DNA. One reason is that the 2'-OH group on RNA was shown to be important for the naturally occurring ribozymes such as the RNase P (8), and the absence of it on deoxyribose was thought to render DNA incapable of forming flexible structures (9). Another reason is the fact that no DNAzymes have yet been discovered in nature. Following the development of the SELEX technique in 1990 for isolating RNA aptamers, however, similar techniques were developed in 1992 that enabled the isolation of DNA aptamers from random-sequence DNA pools (10,11). Two years later, the Joyce group reported the first-ever DNAzyme. With an initial randomized DNA pool, they generated a Pb^{2+} dependent RNA-cleaving DNAzyme with only five rounds of selection (12,13). Shortly after, they derived a highly proficient RNA-cleaving DNAzyme, named 10-23, with a catalytic efficiency (k_{cat}/K_M) not inferior to their RNA or protein counterparts (14). These discoveries led to the boom of research interests in developing DNA aptamers and DNAzymes, not only because DNA is much more stable than RNA, but also because the in vitro selection process with DNA is much less laborious.

To date, thousands of aptamers have been developed for a wide range of targets, including metal ions, small molecules, peptides, proteins, viruses, bacteria, and even whole cells (15,16), along with the development and optimization of various strategies of in vitro selection (17,18). Several databases have been established to keep track of the development of aptamers and their characteristics, as well as experimental conditions (19–21).

Thus far, most DNAzymes developed through in vitro selection catalyze reactions that involve nucleic acids, such as cleavage and ligation of DNA or RNA

(Gysbers et al, 2015). DNAzymes that can catalyze RNA cleavage have been the most popular, and a large number of RNA-cleaving DNAzymes have been isolated (12,14,30–37,22–29). The best-known of them are 10-23, MgZ, and also 8-17. Other reactions for which a DNAzyme has been isolated include DNA hydrolysis (38–41), RNA ligation (42–44), RNA branching (45,46), RNA lariat formation (47), RNA phosphorylation (48), DNA adenylation (49), DNA ligation (50,51), oxidative DNA cleavage (13,52), DNA glycosylation (53), thymine dimer repair (54), phosphoramidate cleavage (55), porphyrin metallation (56), aldol reaction (57), depurination (58), ester and amide hydrolysis (59), and self-phosphorylation (60).

1.3 Advantages of SELEX Derived Aptamers and Nucleic Enzymes in Applications

Functional nucleic acids are highly suitable for both biotechnological and clinical applications, attributed to their chemical, functional, and structural properties (61,62). Many publications have discussed the application of aptamers in clinical diagnosis and therapy (63–69), bioassays (70–72), and food safety and environmental monitoring (73,74). Applications of DNAzymes have been recognized in the area of biosensors, site-specific RNA labeling (75,76), logic gates, computing circuits and switches (77–79), and biofuel cell catalysis (80). In addition, ongoing efforts are being put into *in vivo* mRNA degradation (14,26,28,81–87), to explore the potential of DNAzymes in clinical therapies. Due to functional similarities, aptamers are often compared to antibodies, while ribozymes and DNAzymes are compared to protein enzymes.

First of all, nucleic acids offer high stability, short production time, minimum batch variation, and cost-effectiveness. FNAs, especially DNA aptamers and DNAzymes, are highly stable, thus offering long shelf-life. They can also be denatured

by heat and renatured without losing their function, hence they can be reused if necessary. On the other hand, protein molecules require careful handling, proper storage, and most of them cannot be renatured after denaturation. FNAs can be chemically synthesized within a day at a low cost. Because they have defined sequences, there is very limited batch-to-batch variations. In comparison, antibodies are generated by introducing a target into an animal. The production is time-consuming, expensive, and one specific batch of antibodies cannot be highly accurately reproduced. In addition, nucleic acids are also compatible with many chemical modifications, which can be incorporated to enhance the performance of FNAs, such as enhancing their stability in biological systems, or to provide ways of signal transduction for bioanalytical applications. (88). All the above chemical properties, and the non-immunogenic property of FNAs, have promoted the research of their applications as diagnostic or therapeutic agents (63–69).

Secondly, the method of *in vitro* selection provides FNAs with unlimited possibilities of functions. From the perspective of molecular recognition, aptamers and nucleic acid enzymes can bind to or respond to a wide range of ligands, from ions to complex biological matrices. This is superior compare to antibodies, because immune systems cannot generate antibodies to some analytes, such as toxic ions and toxins. Aptamers can be evolved to have superb binding affinity and specificity. For example, the DNA aptamer against the B-chain of the platelet-derived growth factor (PDGF) shows ~370-fold higher binding affinity for the PDGF B chain homodimer than the PDGF A Chain homodimer (7). The aptamer for theophylline binds theophylline >10,000 times better than caffeine, which only differs by one methyl group (89). Alternatively, our group has demonstrated an example of the evolution of specific pH-dependent DNazymes through the manipulation of solution acidity (36). Many

aptamers have dissociation constants (K_d) that range from pico- to nanomolar. Numerous aptamer-based sensors have been developed, with many of them showing excellent detection sensitivity (7,72,88,90–92).

Similarly, catalytic nucleic acids are believed to offer certain advantage over their protein counterpart, as pointed out by Dr. Scott Silverman recently (62). The methodology of *in vitro* selection allows the selection of a particular catalytic DNA or RNA from a complete randomized pool, which originally has no observed catalytic activity. This provides much more room for acquiring activity, compared to protein activity which can only be optimized through directed evolution, the screening of variants generated from partial mutations (93,94). Although a protein enzyme can be computationally designed, it has to be embedded into a known protein structure and optimized through direct evolution (95). This has the potential to alter the reaction mechanism of the protein enzyme (96). Comparably, nucleic acid enzymes, for any specific requirement, can be straightforwardly generated through *in vitro* selection, and the design of nucleic acid constructs are easier and the change in catalytic activity is less likely.

Last but not least, DNA and RNA secondary structures are much more stable and predictable compare to protein structures (97). Common protein structures such as alpha-helices and beta-sheets are unstable without being in the context of a tertiary structure. Therefore, covalent or noncovalent linkages are often required to be incorporated in order to stabilize small synthetic peptides (98). Although FNAs fold into intricate structures and bind to their targets through many kinds of chemical forces including van der Waals forces, hydrogen bonding, electrostatic interactions, stacking of aromatic moieties, and shape complementarity (70,99), a significant degree of structure prediction and manipulation can be achieved based on the Watson-Crick

interactions between two complementary segments. This enables the rational design of many input-dependent biosensors and therapeutic agents (88).

1.4 Applications of Nucleic Acids as Biosensor Components

Extensive research effort has been put into building biosensors using FNAs. A biosensor can be defined as an analytical device that combines biological recognition with a signal transducer. Common examples of recognition elements include enzymes, antibodies, DNA, cells, and biomimetics (i.e. aptamers). Association with targets by the recognition element can be transduced through electrochemical, piezoelectric, calorimetric, or optical signals (100,101). Nucleic acids are highly versatile when used as biosensors, especially DNAs, for its better stability and cost-efficiency compared to RNAs. The following section will focus on DNA as building blocks for biosensors. As the recognition component, they can recognize analytes through Watson-Crick hybridization, aptamer-target complex formation, or the dependency on certain metal ions or solution conditions for DNAzymes. They can be used as signal transducers in many flexible and creative ways, owing to the catalytic ability of some DNAzymes, their compatibility with chemical modifications, and possibility of being rationally designed to generate conformational change with the addition of the analyte. Moreover, some aptazymes are capable of both recognition and catalysis. The following sections will provide examples of all above listed nucleic acid sensors according to the roles they play in biosensors.

1.4.1 DNA Used as the Recognition Element in Biosensors

DNA can be used to detect other nucleic acids directly; aptamers can recognize a wide range of non-nucleic acid analytes; and DNAzymes can be responsive towards specific metal ions or other changes in the environment.

Watson-Crick interactions refer to the hydrogen bond between two single-stranded nucleic acids that are complementary to each other, and this working principle promoted ssDNA to be a recognition tool for the highly specific detection of human and pathogen genes. The event of hybridization to a target can be measured electrochemically if the sensing strand can be immobilized to a surface (88,102,103), or detected optically through fluorescence, surface plasmon resonance (SPR), chemiluminescence, colorimetry, interferometry, or surface-enhanced Raman scattering (SERS) spectroscopy (104). One simple way of visualizing DNA hybridization is through the use of two gold nanoparticles-labeled strands, which can be joined by the target strand. Gold nanoparticles are pulled closer together resulting in a color change (105) (Figure1-2).

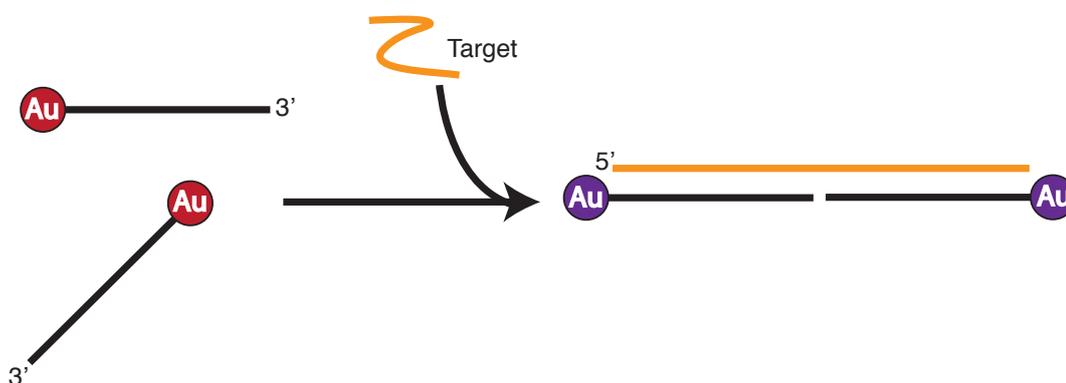


Figure 1-2. Gold Nanoparticle Labeled DNA Probes for ssDNA or RNA Detection. This detection systems uses two ssDNA probes, each with a gold nanoparticle attached to the 5' end. The two probes are complimentary to the 5' portion and the 3' portion of the target, respectively. Upon addition of the target, two probes both anneal to the target, thus creating a close proximity of the two gold nanoparticles. This leads to the color change from red to purple.

Besides the well-known Watson-Crick duplex structure, ssDNA and ssRNA based aptamers can also fold into 3D structures and recognize specific ligands. Many RNA aptamers are found in nature and they act as intracellular sensors that respond to environmental changes (106,107), Starting around 2004, SELEX derived aptamers attracted lots of research interest as recognition elements in biosensors (72,90,91). Their

targets have ranged from inorganic ions, small molecules, biomacromolecules, to whole organisms. Aptamer-based biosensing strategies have also expanded from the use of a few model targets, such as ATP (108) and thrombin (109), to practically significant ones such as toxins, drugs, antibiotics, insecticides, tumor markers, biomarkers, cells, and pathogenic bacteria (88) (7,72,88,90–92).

If the target of interest is a specific ion, a DNAzyme might be a better candidate as the recognition element than an aptamer, because ions are difficult to be immobilized. DNAzymes often require divalent metal ions as their cofactors, and this allows them to act as sensors for such metal ions (110). Two classes of DNAzymes are most commonly considered as metal ion sensors, the RNA-cleaving DNAzymes and G-quadruplexes that demonstrate peroxidase activity (111–114). RNA-cleaving DNAzymes are often designed to cleave the phosphodiester bond of the (mostly single) RNA embedded within its substrate. However, their selectivity for metal ion cofactors vary. For instance, while the 8-17 RNA-cleaving DNAzyme can utilize a wide range of divalent metal ions such as Zn^{2+} , Mn^{2+} , Cd^{2+} , and Co^{2+} as its cofactor, another RNA-cleaving DNAzyme named GR5 shows a great selectivity towards lead ion over other ions, hence making it a great sensor for lead ion (115). A number of metal ion sensors have been made with RNA-cleaving DNAzymes, including those responding to Pb^{2+} (116), Mg^{2+} (117), Ca^{2+} (117), Zn^{2+} (26), UO_2^{2+} (118), Co^{2+} (119), and Cu^{2+} (120).

Another class of DNAzyme, G-quadruplexes, can also be used for ion sensing. This class of DNA molecules represent an unconventional structural arrangement of nucleic acids. They feature G-quartets, stacks of four guanines in a plane that are held together by 8 hydrogen bonds (Figure 1-3a). Many naturally-occurring and man-made G-rich DNA and RNA sequences have been shown to form G-quadruplexes (121–124). Their structures require monovalent or divalent ions to form, therefore can be used as

sensors for ions. Examples of divalent metal ions detected using G-quadruplexes include Ca^{2+} (125), Sr^{2+} (126,127), Hg^{2+} (128) and Pb^{2+} (129,130).

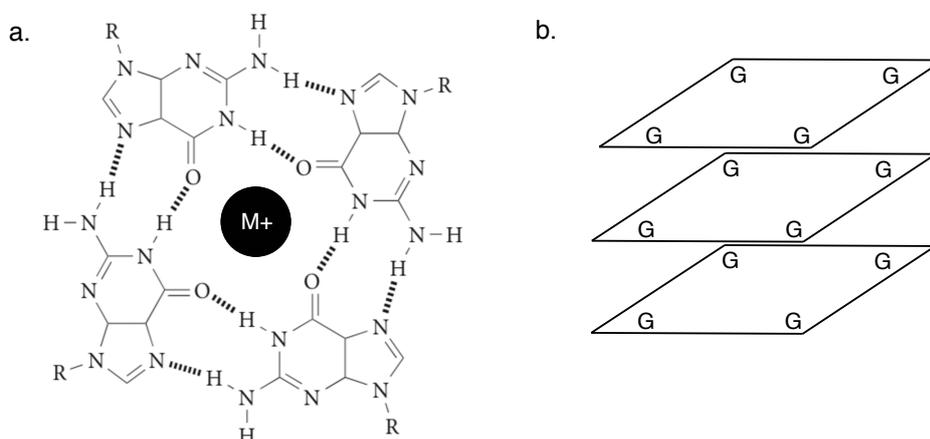


Figure 1-3. Chemical Structure of G Quartet. a) G-tetrad. Four guanine bases associate through Hoogsteen hydrogen bonding, which involves both the pyrimidine and the imidazole ring. A metal ion is often stabilized in the middle of the tetrad. b) Stacked G-tetrads

1.4.2 DNA Used as the Signal Transducer in Biosensors

One great advantage of using DNAzymes as metal ions sensors is that with their catalytic activity, a signal can be readily produced. RNA-cleaving DNAzymes and G-quadruplexes also serve as signal transducers in many FNA-based sensors. The cleavage of RNA provides an opportunity to design fluorescence-based biosensors. This can be done conveniently through the use of a pair of fluorophore and quencher to be placed in close proximity. In the uncleaved state, efficient fluorescence quenching is achieved. Upon RNA cleavage, the fluorophore is separated from the quencher, resulting in strong fluorescence enhancement (131,132). There are several ways to place a fluorophore relative to its matching quencher, as illustrated in Figure 1-4. Most of the reported RNA-cleaving DNAzymes adopt a three-way junction based secondary structure.

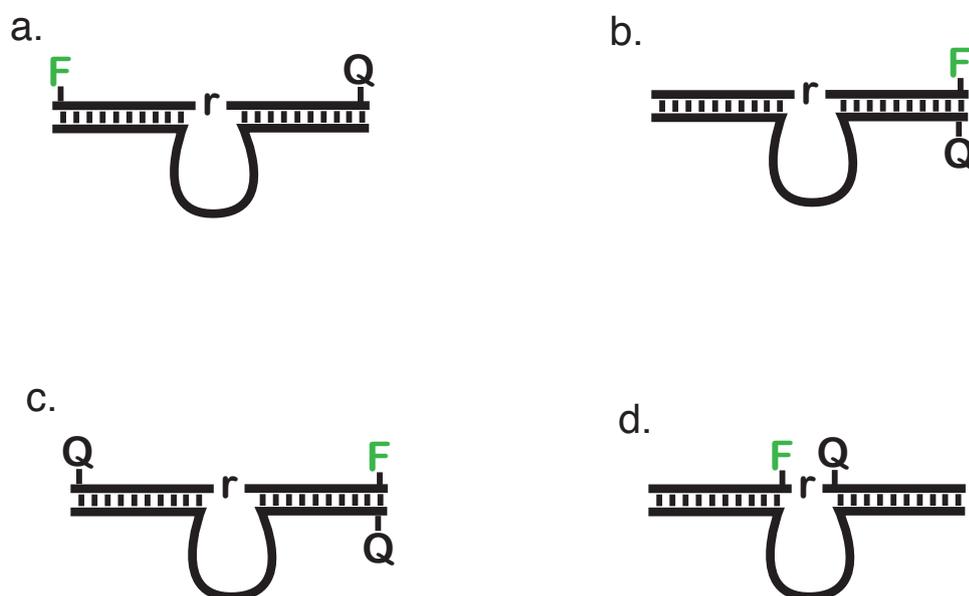


Figure 1-4. Various Placement of Fluorophore and Quencher Pairs on RNA-cleaving DNAzymes and Their Substrates.

a, b, and c describe the placement of modifications on the ends of nucleic acid sequences. In contrast, d describes the RNA cleavage site that is sandwiched by nucleotides modified with fluorophore and quencher, respectively.

G-quadruplexes possess peroxidase properties that can generate colorimetric signals. One interesting G-rich sequence was selected as a porphyrin binding DNA aptamer and was found to form a G-quadruplex structure (133). Interestingly, upon binding hemin (one type of porphyrin), this DNA molecule was shown to exhibit peroxidase-like activity (134), meaning it can catalyze the oxidation reaction of colorimetric dyes by H_2O_2 , such as 2,2'-azino-bis(3-ethylbenzthiazoline)-6-sulfonic acid (ABTS) and 3,3',5,5'-Tetramethylbenzidine (TMB). Alternatively, G-quadruplexes of different topologies may selectively bind different fluorescent dyes. For example, parallel G-quadruplexes can selectively bind protoporphyrin IX (PPIX) (135), and anti-parallel G-quadruplexes can be probed by crystal violet.

1.4.3 Combining Nucleic Acid-Based Recognition Elements and Signal Transducers

The combination of a nucleic acid recognition element and a nucleic acid signal transducer provides ease of synthesis and simplicity in preparation. Owing to the predictability and flexibility of its structure, a nucleic acid sensor can be designed so that the event of target recognition is coupled with a conformational change, which will enable downstream signalling directly, through DNAzyme catalysis, or via additional signaling cascades. Many probes are designed such that two strands, with partial or full complementarity, hybridized to each other, and the target is added to displace one or more pre-hybridized strands. It is required that the probe has a single-stranded domain, known as the toehold (136). The released strand can act as a secondary input for downstream signaling or a cascade of signal amplification (137). In this section, nucleic acid-based signalling in biosensors will be introduced according to the complexity of the design.

First, the event of recognition can be directly reported by chemical labelling. The aforementioned gold nanoparticle labeled ssDNA probes is one example that generates colorimetric signal directly. Alternatively, the invention of molecular beacon provides an immediate fluorescent signal upon recognition. As shown in Figure 1-5a, a molecular beacon forms a hairpin structure, with two ends labeled with a fluorophore and quencher, respectively. It is used to detect ssRNA or ssDNA. Once the target is bound, the hairpin opens. As a result, the fluorophore and quencher separates to generate a fluorescent signal (138). The probe region of a beacon can be replaced with an aptamer, and with careful tuning of the hairpin stability (often via changing the stem length), the system can be modified into an effective “aptamer-beacon” (Figure 1-5b).

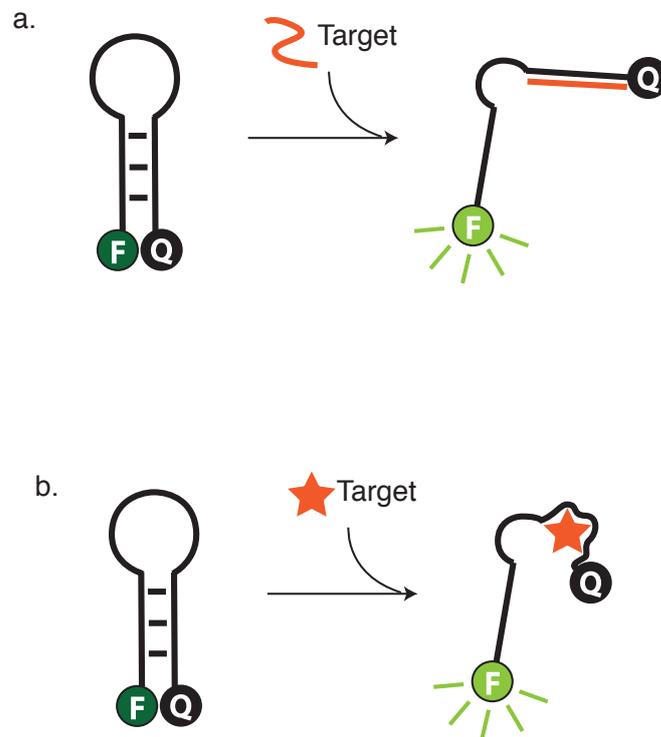


Figure 1-5. Molecular Beacon and Aptamer Beacon. a). a molecular beacon is a ssDNA with partial paring regions on its 3' and 5' ends, respectively, which are also labeled with fluorophores and quenchers. One strand in the paring region acts as the sensing strand, which is complementary with the nucleic acid target, as shown in a; or is the aptamer for a molecular target, as shown in b. The other strand in the paring region is often referred to as the “locking” strand. The binding affinity between the sensing strand and the target is designed to be stronger than that of the paring region, so that upon the presence of the target, the paring region dissociates and lead to the release of fluorescence signal.

Secondly, instead of labelling of chemical moiety, one end of the beacon can be modified into a DNAzyme, and they are sometimes referred to as “aptazymes”. For example, G-quadruplexes can be engineered to play as the signal transducer (139,140). Willner and co-workers used a G-quadruplex named PW17 in the design of a number of DNAzyme-based sensors for various targets including ions, small molecules, DNA, and proteins (141–144). In one of these examples (Figure 1-6), PW17 G-quadruplex was engineered into a catalytic beacon, which forms a hairpin in the absence of target DNA. Added target DNA hybridizes to one strand of the hairpin stem thus unlocking

the beacon, allowing the formation of PW17, which can then catalyze the oxidation of ABTS by H_2O_2 , and lead to a downstream color change from colorless to blue (145).

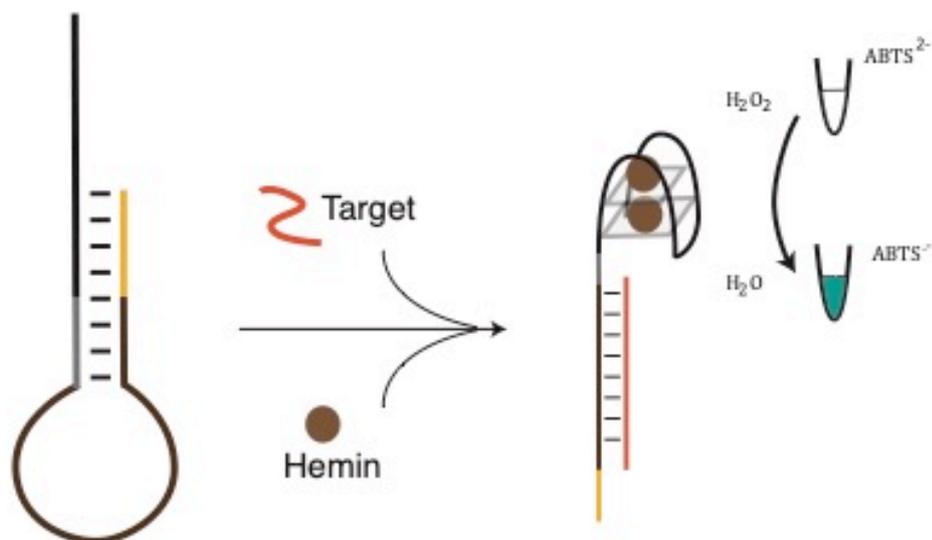


Figure 1-6. Catalytic Beacon Constructed using PW17 G-quadruplex. The brown region is the probing region and is complimentary to the target. The grey region is the “locking” region for a portion of the probing region. The black region is the PW17 region; and the yellow is a C-rich region that “locks” a portion of the PW17. Once the target sequence hybridizes to the probing region, it also releases the yellow locker off the PW17 region. Simultaneously, a G-quadruplex structure is formed by the PW17. It can then associate with hemin and catalyze the colorimetric change of ABTS.

Alternatively, the beacon can be designed such that the non-probe strand is a substrate for an RNA-cleaving DNzyme. Our lab has created an aptzyme biosensor for ATP detection by combining the RNA-cleaving DNzyme MgZ, with the well-known ATP binding DNA aptamer (146). As illustrated in Figure 1-7. The substrate for an RNA-cleaving DNzyme is hybridized to a “lock”, and is therefore inaccessible to the DNzyme, MgZ. The “lock” region is embedded within the sequence of an ATP aptamer, which undergoes conformational change upon addition of ATP. The substrate region is then released and the RNA-DNA linkage can be cleaved by MgZ. In this case, a fluorescent signal is carried out through a unique design of our group. As illustrated in Figure 1-4d, the substrate sequence is internally modified with a fluorophore and

quencher pair that directly sandwiches the single ribonucleotide (119), so that fluorescent enhancement is synchronized with the cleavage. This design is an advancement of several previously designed fluorescent RNA-cleaving reporter systems reported by other groups, which all placed fluorophore and quenchers on the ends of oligonucleotides (Figure 1-4). This unique design has demonstrated its versatility in many studies (36,119,154–157,146–153). We characterized fluorescence enhancement with respect to different fluorophore and quencher pairs, along with different distances relative to the ribonucleotide (148)

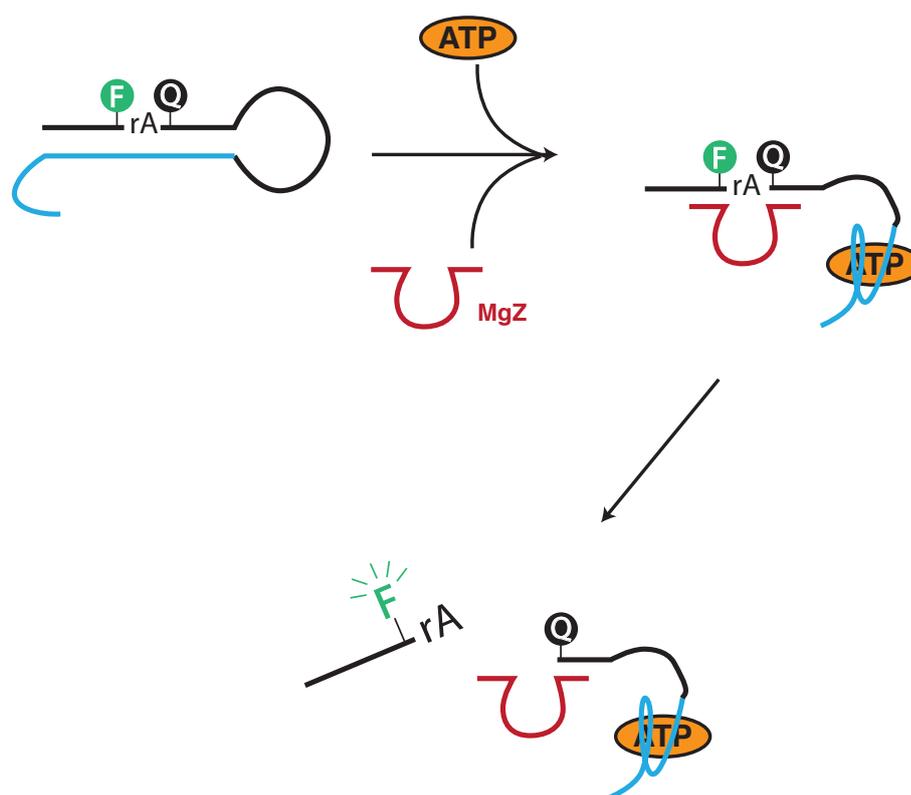


Figure 1-7. Substrate of RNA-cleaving DNAzyme Incorporated into a Molecular Beacon. Black: fluorogenic substrate containing a single ribonucleotide adenosine, sandwiched by a fluorophore and quencher; its sequence is partially complimentary to the ATP binding aptamer, thus acting as a “lock” to the ATP binding aptamer. Blue: ATP binding aptamer; Red: MgZ RNA-cleaving DNAzyme with binding arms that are complimentary to the substrate. Once ATP is added, the ATP binding aptamer changes its conformation and associate with ATP, thereby “unlocking” the substrate, which becomes accessible to the MgZ Upon RNA cleavage, a simultaneous fluorescent signal is generated.

Furthermore, more complex designs of nucleic acid sensors generate amplified signals through either nucleic acid logic gates, or cascades of reactions utilizing protein enzymes (113). For example, our group has developed a system that takes advantage of an aptamer, a structure-switching design, and the enzyme activity of $\phi 29$ DNA polymerase. As illustrated in Figure 1-8, the system uses a pre-primer, an aptamer, and a circular template, which are designed to form a DNA assembly in the absence of the aptamer's cognate target. Once the target is added, the aptamer leaves the assembly to bind to the target. This event leads to the exposure of 3' end of the pre-primer that can be digested by $\phi 29$ DP nucleotide by nucleotide, until $\phi 29$ DP reaches the duplex region. At that point, $\phi 29$ DP starts copying the sequence of circular template (158).

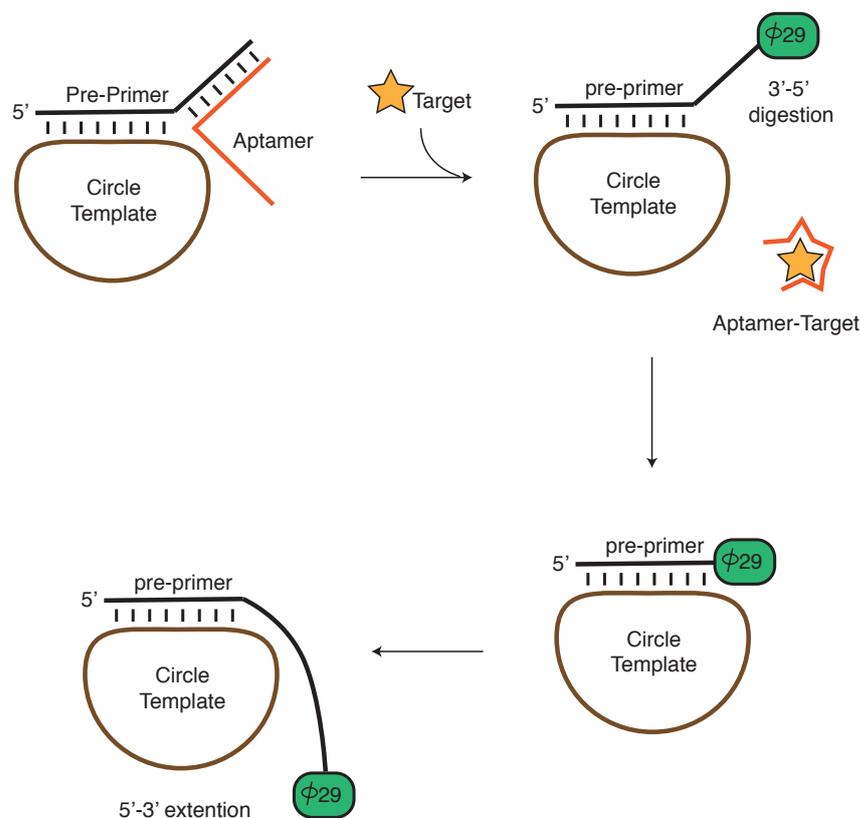


Figure 1-8. A Complex Biosensor Constructed with Nucleic Acids and Protein Enzyme. Black: pre-primer; the 5' half of it is complementary to the circle template and the 3' half is complementary to the 5' end of the aptamer of interest. Red: aptamer of interest; Brown: circle template. With the addition of the target of the aptamer, the aptamer leaves the original complex and binds with the target. Then the 3' end

of the pre-primer becomes susceptible to the digestion of ϕ 29 DNA polymerase. Digestion continues until it reaches the region that is hybridized to the circle template, and it will reverse and start to elongate using the circle template as the template.

Last but not least, an aptazyme biosensor can also be directly selected from a random-sequence DNA pool. To push RNA-cleaving aptazymes a step closer to real world applications, our lab initiated several studies towards using this system for the detection of bacteria. Bacterial infection outbreaks are often sudden, and the emergence of new epidemic strains or antibiotic-resistant strains has always caught the world unprepared. Therefore, to control the spread of infection, there is a need to develop a time-efficient method to generate aptazymes that can provide both high sensitivity and specificity. To design an aptazyme for a bacterium, we would first choose a species-specific target, generate an aptamer for that using in vitro selection, then fuse it with a well-studied DNAzyme, followed by further optimization. Each step of this process can be time consuming and difficult. Therefore, innovatively, we designed a one-step method, which is to directly conduct in vitro selection using the crude extracellular mixture (CEM) of the bacterium of interest. The ideal outcome is to generate one or more aptazymes that can recognize one or more species-specific targets. Successfully, our lab has developed an aptazyme targeting CEM of *Escherichia coli* (150), and another targeting a specific pathogenic strain of *Clostridium difficile* (156).

1.5 Considerations for In Vitro Selection Process

Attributed to the technique of in vitro selection, it seems there is unlimited possibilities of functional nucleic acids that can be generated. However, the reality is proven otherwise. In vitro selection experiment can be laborious and difficult. Over the past few decades, lots of research efforts have been put into studying and optimizing in vitro selection. Our lab also looked into the progress of sequence evolution in terms of

sequence diversity, throughout the selection process (159). With better access to deep sequencing, we start to gain better ideas about population dynamics through the rounds of in vitro selection (155,157). Herein, I discuss some of the considerations worth taking when designing an in vitro selection experiment.

1.5.1 Library Design

Population diversity is the first parameter to consider when designing an in vitro selection. In vitro selection methods assume that among a large randomized population, at least one molecule carries the functional domain of interest. Given that each nucleotide in a sequence can be one of only four possibilities, the number of different sequences in the population can be simply estimated by the formula 4^N , where N is the number of randomized nucleotides. Certainly, longer randomized library would dictate higher probability for finding the ideal sequence, however randomized regions ranging from 40 to 80 nucleotides are generally used.

The first factor that limits the library length is sampling. DNAzymes and aptamers often perform well at low micromolar and nanomolar ranges, and generally, the initial round of in vitro selection would use no higher than 10^{16} molecule. When the length of randomized region passes 25, the actual pool used will undersample the diversity at such concentrations. On the other hand, it has been shown that a long library is required if one is seeking a complex structure or a structure that can perform a rare activity (160). As Wedel suggested in 1996, a 50-nt library was probably too short for the evolution of a self-acylating ribozyme, as the primer binding region was involved in the functional structure (161). In these cases, we would have to use a much larger number of molecules in the initial pool, which may lead to complications in the reactions. Fortunately, many aptamers and nucleic acid enzymes are no longer than 20-

nt. The most studied thrombin-binding aptamer and ATP-binding aptamer are 15 and 25 nucleotides long, respectively (10,108). Both existing RNA-cleaving ribozymes and DNAzymes are typically less than 20-nt in length. For instance, the most monopolized RNA-cleaving DNAzyme, 8-17, only requires a 14-nt long catalytic core, even though it was isolated from a pool with 50-nt randomized regions (14). Such short and simple motifs are also very likely to emerge from the pool, since they are likely to have higher copy numbers (162). Overall, information collected to date indicates that a library between 40 to 80 nucleotides is sufficient.

1.5.2 Population Diversity

Population diversity is a crucial consideration for in vitro selection processes. Besides the use of an appropriate library to create diversity, it is also important to carefully preserve it during experiments. This is especially important in the first few rounds of in vitro selection. During these rounds, active sequences may only constitute a very low percentage of the pool, and each sequence may have only a few copies. Two precautions can be taken to minimize the loss of these valuable sequences. First, we carefully maximize recovery from the step of separating active and inactive species. For example, when isolating RNA-cleaving DNAzymes, we often use denaturing polyacrylamide gel electrophoresis (dPAGE) to separate cleaved sequences from uncleaved ones on the basis size. We tend to excise a large enough gel piece in attempt to include all cleaved sequences in that area, and we elute these sequences under conditions that maximize the elution yield, including longer time of incubation, slightly elevated temperature, and gel crushed into fine pieces. Second, all eluted sample should be used as the template for following PCR amplification, and the number of PCR reactions and number of cycles are set to ensure that enough product is generated for

the next round of selection. This precaution can be less attended as the selection progresses, after the active species rise to occupy a high portion of the pool.

1.5.3 Selective-Pressure Manipulation

The manipulation of selective-pressures is the most important factor in an in vitro selection process. While substantial pressure may allow the survival of less active species, too stringent conditions may cause the loss of important species or generate non-specific species. In the in vitro selection process of generating a target-responsive FNA, generally, there are two parameters to control the amount of selective pressure, concentration of the target and time of reaction. In earlier rounds, typically only a light pressure is put on the population, such as high concentration of targets and a long reaction time. In later rounds, especially after the emergence of a clear active population, time of reaction can be shortened gradually, to select for the most efficient species. Also, the target concentration can be gradually decreased, to select for the most sensitive species.

Conversely, stringent negative selection and counter selection are also necessary to ensure the specificity of evolved sequences. While negative selection is designed to eliminate species active towards the reaction buffer, counter-selection refers to the elimination of species that respond to molecules commonly found in the presence of, or highly similar to, the actual selection target. In our previous experience of generating aptazymes towards bacterial CEM, the library is typically first subjected to a reaction containing buffer only and/or a reaction with other bacterial CEM, so the molecules that are non-specific can be eliminated early enough. Specifically, when isolating the aptazyme specific for *E. coli*, the library was first subjected to reaction buffer and CEM

of *Bacillus subtilis* before exposure to *E. coli* CEM. Experiences in our lab show that non-specific populations are hard to eliminate once a non-specific signal appears.

One of the distinctive advantages provided by in vitro selection is that sensitivity and specificity can be fine-tuned according to the goal of a particular study. For example, our lab has demonstrated this advantage by an in vitro selection study. A buffer at pH 4.0 was used for the first five rounds, then the pool was diverged to selection under pH 3.0, 4.0, 5.0, 6.0, 7.0, respectively. As a result, we generated several pH-specific DNAzymes (36). This method can be potentially generalized to any in vitro experiment to discriminate similar targets and optimize specificity.

1.6 Thesis Objective and Outline

With interests in both understanding the biochemical fundamentals of nucleic acids and propelling the biotechnological applications of functional nucleic acids, my PhD thesis has focused on advancing the in vitro selection method, exploring novel functions of nucleic acid structures, and engineering biosensors using nucleic acids. This thesis will focus on the following three projects, briefly stated in section 1.6.1-1.6.3.

1.6.1 Development of A Homologue-Specific Substrate for Bacterial RNase H2

The utility of in vitro selection has gone beyond the field of functional nucleic acids. In many cases, it has improved our understanding about the functions of many biologically important substances, particularly of nucleic acid interacting proteins (163–165). For example, bacterial ribonuclease P was known to cleave several distinctive substrates that were discovered through the studies of model substrates (166). Through in vitro selection, novel substrates which were cleaved by RNase P as effectively as the intrinsic substrates were discovered (165).

Ribonuclease H2 is a ubiquitous protein enzyme that intrinsically cleaves the RNA-DNA linkage in RNA/DNA hybrids, including RNA/DNA, RNA-DNA/DNA, and DNA-RNA_{1-few}-DNA/DNA (167). RNase H2 is unique among all RNases H, because it can cleave the RNA-DNA linkage embedded in a dsDNA molecule. In comparison, RNase H1 requires four RNA nucleotides in a dsDNA. In this thesis, we propose to develop a novel nucleic acid substrate that: (1) can be more effectively cleaved than known substrates; (2) has specificity towards one homologue of RNase H2, specifically, RNase H2 from the bacterium *Clostridium difficile*. We would like to explore whether this approach can allow us to generate highly sensitive and specific biosensors using nucleic acids. This study aims to demonstrate the capability of in vitro selection to: (a) discover substrates for protein enzymes that are better fit than those discovered through rational tests of model substrates; (b) evolve substrates that can differentiate between homologue proteins.

Gram-positive bacterium *Clostridium difficile* is the leading cause of nosocomial infections. It may reside in 5% of the population without causing any diseases. However, via its resistance to antibiotics, it thrives when a large portion of gut microbiota is killed by broad-scheme antibiotic treatment. *C. difficile* releases harmful toxins that disrupt the epithelial cells of the host intestinal tract, causing severe diarrhea that is often bloody, last for days, and can even be life-threatening. We chose *C. difficile* as the model organism because its detection is crucial for diagnosis, surveillance of the environment, and decision making in treatment.

1.6.2 The Accidental Discovery of a Highly Stable G-Quadruplex with Resistance to 7M-Urea Gel Electrophoresis

During the process of selecting RNA-cleaving DNA sequences from a random-sequence DNA library, we enriched a group of guanine-rich sequences. Later, we found

these sequences to be unusual in that they remain folded in denaturing gel containing 7 M urea, which is used widely to dissolve nucleic acid structures. Examination and analysis of this class (named UD) of sequences indicate that it forms a unique G-quadruplex structure. We reported and characterized this structure, and propose that it will lead to unconventional applications.

1.6.3 Utilizing highly stable UD G-quadruplexes for DNA sensing and highly specific detection of single nucleotide polymorphism

G-quadruplexes are widely incorporated into nucleic acid sensors as a signal transducer, attributed to its ability to produce colorimetric, fluorescent, and chemiluminescent signals. The above-mentioned UD G-quadruplex has several unique features, allowing us to engineer this G-quadruplex into a probe for ssDNA, with the ability of distinguishing a single mismatch of very long DNA sequences.

1.7 References

1. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*. 1983;35(3 PART 2):849–57.
2. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*. 1982;31(1):147–57.
3. Ellington a D, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature*. 1990;346(6287):818–22.
4. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* (80-). 1990;249(4968):505–10.
5. Robertson, Debra L; Joyce GF, Robertson DL, Joyce GF. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* 1990 p. 467–8.
6. Mills DR, Peterson RL, Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci*. 1967;58(1):217–24.
7. Liu J, Cao Z, Lu Y. Functional Nucleic Acid Sensors. *Chem Rev*. 2009;109(5):1948–98.
8. Perreault J-P, Altman S. Important 2'-hydroxyl groups in model substrates for M1 RNA, the catalytic RNA subunit of RNase P from *Escherichia coli*. *J Mol Biol*. 1992;226(2):399–409.
9. Cech TR. The Chemistry of Self-Splicing RNA and RNA Enzymes. *Source Sci New Ser*. 1987;236(19):1532–9.
10. Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*. 1992;355(6):564–6.
11. Ellington AD, Szostak JW. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*. 1992;355(27):850–2.
12. Breaker RR, Joyce GF. A DNA enzyme that cleaves RNA. *Chem Biol*. 1994;1(4):223–9.
13. Carmi N, Shultz LA, Breaker RR. In vitro selection of self-cleaving DNAs. *Chem Biol*. 1996;3(12):1039–46.
14. Santoro SW, Joyce GF. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci*. 1997;94(9):4262–6.
15. Cowperthwaite MC, Ellington AD. Bioinformatic Analysis of the Contribution of Primer Sequences to Aptamer Structures. *J Mol Evol*. 2008;67(1):95–102.
16. Zhou J, Rossi J. Aptamers as targeted therapeutics: Current potential and challenges. *Nature Reviews Drug Discovery*. 2017.
17. Tabarzag M, Jafari M. Trends in the Design and Development of Specific Aptamers Against Peptides and Proteins. *Protein J*. 2016;35(2):81–99.
18. McKeague M, McConnell EM, Cruz-Toledo J, Bernard ED, Pach A, Mastronardi E, et al. Analysis of In Vitro Aptamer Selection Parameters. *J Mol Evol*. 2015;81(5–6):150–61.
19. Lee JF, Hesselberth JR, Meyers LA, Ellington AD. Aptamer Database.
20. Aptagen. Apta-Index (Aptamer Database) [Internet]. 2017 [cited 2018 Sep 21].

- Available from: <https://www.aptagen.com/aptamer-index>
21. Cruz-Toledo J, Mckeague M, Zhang X, Giamberardino A, Mcconnell E, Francis T, et al. Aptamer base: a collaborative knowledge base to describe aptamers and SELEX experiments. Database. 2012;2012:bas006.
 22. Schlosser K, Gu J, Lam JCF, Li Y. In vitro selection of small RNA-cleaving deoxyribozymes that cleave pyrimidine-pyrimidine junctions. Nucleic Acids Res. 2008;36(14):4768–77.
 23. Chiuman W, Li Y. Evolution of High-Branching Deoxyribozymes from a Catalytic DNA with a Three-Way Junction. Chem Biol. 2006;13(10):1061–9.
 24. Breaker RR, Joyce GF. A DNA enzyme with Mg²⁺-dependent RNA phosphoesterase activity. Chem Biol. 1995;2(10):655–60.
 25. Geyer R, Sen D. Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. Chem Biol. 1997;4(8):579–93.
 26. Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CFI. RNA Cleavage by a DNA Enzyme with Extended Chemical Functionality. J Am Chem Soc. 2000;122(11):2433–9.
 27. Perrin DM, Garestier T, Hélè C. Bridging the Gap between Proteins and Nucleic Acids: A Metal-Independent RNaseA Mimic with Two Protein-Like Functionalities. J Am Chem Soc. 2001;123(8):1556–63.
 28. Sidorov A V, Grasby JA, Williams DM. Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and amino functionalities. Nucleic Acids Res. 2004;32(4):1591–601.
 29. Saran R, Chen Q, Liu J. Searching for a DNAzyme Version of the Leadzyme. J Mol Evol. 2015;81(5–6):235–44.
 30. Li J, Zheng W, Kwon AH, Lu Y. In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme. Nucleic Acids Res. 2000;28(2):481–8.
 31. Torabi S-F, Wu P, McGhee CE, Chen L, Hwang K, Zheng N, et al. In vitro selection of a sodium-specific DNAzyme and its application in intracellular sensing. Proc Natl Acad Sci U S A. 2015;112(19):5903–8.
 32. Faulhammer D, Famulok M. Characterization and Divalent Metal-ion Dependence of in Vitro Selected Deoxyribozymes which Cleave DNA/RNA Chimeric Oligonucleotides. J Mol Biol. 1997;269(2):188–202.
 33. Torabi S-F, Lu Y. Identification of the Same Na⁺-Specific DNAzyme Motif from Two In Vitro Selections Under Different Conditions. J Mol Evol. 2015;81(5–6):225–34.
 34. Huang P-JJ, Vazin M, Liu J. Desulfurization Activated Phosphorothioate DNAzyme for the Detection of Thallium. Anal Chem. 2015;87(20):10443–9.
 35. Lam JCF, Withers JB, Li Y. A Complex RNA-Cleaving DNAzyme That Can Efficiently Cleave a Pyrimidine-Pyrimidine Junction. J Mol Biol. 2010;400:689–701.
 36. Liu Z, Mei SHJ, Brennan JD, Li Y. Assemblage of signaling DNA enzymes with intriguing metal-ion specificities and pH dependences. J Am Chem Soc. 2003;125(25):7539–45.
 37. Nelson KE, Bruesehoff PJ, Lu Y. In Vitro Selection of High Temperature Zn²⁺-Dependent DNAzymes. J Mol Evol. 2005;61(2):216–25.
 38. Chandra M, Sachdeva A, Silverman SK. DNA-catalyzed sequence-specific hydrolysis of DNA. Nat Chem Biol. 2009;5(10):718–20.
 39. Fekry MI, Gates KS. DNA-catalyzed hydrolysis of DNA phosphodiester. Nat

- Chem Biol. 2009;5(10):710–1.
40. Gu H, Furukawa K, Weinberg Z, Berenson DF, Breaker RR. Small, Highly Active DNAs That Hydrolyze DNA. *J Am Chem Soc.* 2013;135(24):9121–9.
 41. Parker DJ, Xiao Y, Aguilar JM, Silverman SK. DNA Catalysis of a Normally Disfavored RNA Hydrolysis Reaction. *J Am Chem Soc.* 2013;135(23):8472–5.
 42. Flynn-Charlebois A, Wang Y, Prior TK, Rashid I, Hoadley KA, Coppins RL, et al. Deoxyribozymes with 2'-5' RNA Ligase Activity. *J Am Chem Soc.* 2003;125(1):2444–54.
 43. Hoadley KA, Purtha WE, Wolf AC, Flynn-Charlebois A, Silverman SK. Zn²⁺-Dependent Deoxyribozymes That Form Natural and Unnatural RNA Linkages. *Biochemistry.* 2005;44(25):9217–31.
 44. Purtha WE, Coppins RL, Smalley MK, Silverman SK. General deoxyribozyme-catalyzed synthesis of native 3'-5' RNA linkages. *J AM CHEM SOC.* 2005;127(38):13124–5.
 45. Wang Y, Silverman SK. Deoxyribozymes That Synthesize Branched and Lariat RNA. *J AM CHEM SOC.* 2003;125(23):6880–1.
 46. Coppins RL, Silverman SK. A Deoxyribozyme that Forms a Three-Helix-Junction Complex with its RNA Substrates and has General RNA Branch-Forming Activity. *J Am Chem Soc.* 2005;127(9):2900–7.
 47. Wang Y, Silverman SK. Efficient One-Step Synthesis of Biologically Related Lariat RNAs by a Deoxyribozyme. *Angew Chem Int Ed.* 2005;44(36):5863–6.
 48. Wang W, Billen LP, Li Y. Sequence Diversity, Metal Specificity, and Catalytic Proficiency of Metal-Dependent Phosphorylating DNA Enzymes. *Chem Biol.* 2002;9(4):507–17.
 49. Li Y, Liu Y, Breaker RR. Capping DNA with DNA. *Biochemistry.* 2000;39(11):3106–14.
 50. Cuenoud B, Szostak JW. A DNA metalloenzyme with DNA ligase activity. *Nature.* 1995;375(6532):611–4.
 51. Sreedhara A, Li Y, Breaker RR. Ligating DNA with DNA. *J Am Chem Soc.* 2004;126(11):3454–60.
 52. Carmi N, Balkhi SR, Breaker RR. Cleaving DNA with DNA. *Proc Natl Acad Sci U S A.* 1998;95(5):2233–7.
 53. Sheppard TL, Ordoukhanian P, Joyce GF. A DNA Enzyme with N-glycosylase Activity. *Proc Natl Acad Sci.* 2000;97(14):7802–7.
 54. Chinnapen, Daniel J-F, Sen D. A deoxyribozyme that harnesses light to repair thymine dimers in DNA. Vol. 101, *Proceedings of the National Academy of Sciences.* 2004.
 55. Burmeister J, von Kiedrowski G, Ellington AD. Cofactor-Assisted Self-Cleavage in DNA Libraries with a 3'-5'-Phosphoramidate Bond. *Angew Chemie Int Ed English.* 1997 Jul 4;36(12):1321–4.
 56. Li Y, Sen D. A catalytic DNA for porphyrin metallation. *Nat Struct Biol.* 1996;3(9):743–7.
 57. Sun G, Fan J, Wang Z, Li Y. A Novel DNA-Catalyzed Aldol Reaction. *Synlett.* 2008 Sep 12;2008(16):2491–4.
 58. Höbartner C, Pradeepkumar PI, Silverman SK. Site-selective depurination by a periodate-dependent deoxyribozyme. *Chem Commun.* 2007;14(22):2255–7.
 59. Brandsen BM, Hesser AR, Castner MA, Chandra M, Silverman SK. DNA-Catalyzed Hydrolysis of Esters and Aromatic Amides. *J Am Chem Soc.* 2013;135(43):16014–7.

60. Mcmanus SA, Li Y. Multiple Occurrences of an Efficient Self-Phosphorylating Deoxyribozyme Motif. *Biochemistry*. 2007;46(8):2198–204.
61. Sun H, Zu Y. A Highlight of Recent Advances in Aptamer Technology and Its Application. *Molecules*. 2015 Jun 30;20:11959–80.
62. Silverman SK. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem Sci*. 2016;41(7):595–609.
63. Lao Y-H, Phua KKL, Leong KW. Aptamer Nanomedicine for Cancer Therapeutics: Barriers and Potential for Translation. *ACS Nano*. 2015 Mar 24;9(3):2235–54.
64. Xiang D, Shigdar S, Qiao G, Wang T, Kouzani AZ, Zhou SF, et al. Nucleic acid aptamer-guided cancer therapeutics and diagnostics: The next generation of cancer medicine. *Theranostics*. 2015;5(1):23–42.
65. Sun H, Zhu X, Lu PY, Rosato RR, Tan W, Zu Y. Oligonucleotide Aptamers: New Tools for Targeted Cancer Therapy. *Mol Ther Nucleic Acids*. 2014;3:e182.
66. Wu X, Chen J, Wu M, Xiaojun Zhao J. Aptamers: Active Targeting Ligands for Cancer Diagnosis and Therapy. *Theranostics*. 2015;5(4):322–44.
67. Xiang D, Shigdar S, Qiao G, Zhou S-F, Li Y, Wei MQ, et al. Aptamer-Mediated Cancer Gene Therapy. *Curr Gene Ther*. 2015;15:109–19.
68. Sun H, Zu Y. Aptamers and their applications in nanomedicine. *Small*. 2015;11(20):2352–64.
69. Zhou J, Rossi JJ. Cell-type-specific, Aptamer-functionalized Agents for Targeted Disease Therapy. *Mol Ther Nucleic Acids*. 2014;3:e169.
70. Saberian-Borujeni M, Johari-Ahar M, Hamzeiy H, Barar J, Omidi Y. Nanoscaled aptasensors for multi-analyte sensing. *BioImpacts*. 2014;4(4):205–15.
71. Chen A, Yang S. Replacing antibodies with aptamers in lateral flow immunoassay. *Biosens Bioelectron*. 2015;71:230–42.
72. Famulok M, Mayer G, Mayer uNTER. Aptamer Modules as Sensors and Detectors. *Acc Chem Res*. 2011;44(12):1349–58.
73. Wu J, Zhu Y, Xue F, Mei Z, Yao L, Wang X, et al. Recent trends in SELEX technique and its application to food safety monitoring. *Mikrochim Acta*. 2014;181(5–6):479–91.
74. Hayat A, Marty JL, Stoytcheva MS, Yu B. Aptamer based electrochemical sensors for emerging environmental pollutants. 2014;
75. Baum DA, Silverman SK. Deoxyribozyme-Catalyzed Labeling of RNA. *Angew Chem Int Ed*. 2007;46(19):3502–4.
76. Büttner L, Javadi-Zarnaghi F, Höbartner C. Site-Specific Labeling of RNA at Internal Ribose Hydroxyl Groups: Terbium-Assisted Deoxyribozymes at Work. *J Am Chem Soc*. 2014;136(22):8131–7.
77. Brown, III CW, Lakin MR, Stefanovic D, Graves SW. Catalytic Molecular Logic Devices by DNAzyme Displacement. *ChemBioChem*. 2014 May 5;15(7):950–4.
78. Orbach R, Willner B, Willner I. Catalytic nucleic acids (DNAzymes) as functional units for logic gates and computing circuits: from basic principles to practical applications. *Chem Commun*. 2015;51(20):4144–60.
79. Wang F, Liu X, Willner I. DNA Switches: From Principles to Applications. *Angew Chemie - Int Ed*. 2015;54(4):1098–129.
80. Zhang M, Xu S, Minter SD, Baum DA. Investigation of a Deoxyribozyme As

- a Biofuel Cell Catalyst. *J Am Chem Soc.* 2011;133(40):15890–3.
81. Cho E-A, Moloney FJ, Cai H, Au-Yeung A, China C, Scolyer RA, et al. Safety and tolerability of an intratumorally injected DNAzyme, Dz13, in patients with nodular basal-cell carcinoma: a phase 1 first-in-human trial (DISCOVER). *Lancet.* 2013;381(9880):1835–43.
 82. Biologicals SR. Allergen-Induced Asthmatic Responses Modified by a GATA3-Specific DNAzyme. *N Engl J Med.* 2015;372(21):1987–95.
 83. Lerner L, Roupioz Y, Ting R, Perrin DM. Toward an RNaseA Mimic: A DNAzyme with Imidazoles and Cationic Amines. *J AM CHEM SOC.* 2002;124(34):9960–1.
 84. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M^{2+}). *Nucleic Acids Res.* 2009;37(5):1638–49.
 85. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM, Hollenstein M, Hipolito CJ, et al. A DNAzyme with Three Protein-Like Functional Groups: Enhancing Catalytic Efficiency of M^{2+} -Independent RNA Cleavage. *ChemBioChem.* 2009;10(12):1988–92.
 86. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM. Toward the Combinatorial Selection of Chemically Modified DNAzyme RNase A Mimics Active Against all-RNA Substrates. *ACS Comb Sci.* 2013;15(4):174–82.
 87. Fokina AA, Meschaninova MI, Durfort T, Venyaminova AG, François J-C. Targeting Insulin-like Growth Factor I with 10–23 DNAzymes: 2'-O-Methyl Modifications in the Catalytic Core Enhance mRNA Cleavage. *Biochemistry.* 2012;51(11):2181–91.
 88. Du Y, Dong S. Nucleic acid biosensors: Recent advances and perspectives. *Anal Chem.* 2017;89(1):189–215.
 89. Jenison RD, Gill SC, Pardi A, Polisky B. High-Resolution Molecular Discrimination by RNA. *Science (80-).* 1994;263(5152):1425–9.
 90. Cho EJ, Lee J-W, Ellington AD. Applications of Aptamers as Sensors. *Annu Rev Anal Chem.* 2009;2:241–64.
 91. Mok W, Li Y. Recent Progress in Nucleic Acid Aptamer-Based Biosensors and Bioassays. *Sensors.* 2008;8(11):7050–84.
 92. Nimjee SM, Rusconi CP, Sullenger BA. APTAMERS: AN EMERGING CLASS OF THERAPEUTICS. *Annu Rev Med.* 2005;56:555–83.
 93. Renata H, Wang ZJ, Arnold FH, Arnold FH. Expanding the Enzyme Universe: Accessing Non- Natural Reactions by Mechanism-Guided Directed Evolution. *Angew Chem Int Ed.* 2015;54(11):3351–67.
 94. Roiban G-D, Reetz MT. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem Commun.* 2015;51(12):2208–24.
 95. Kiss G, Elebi-Ölâüm NÇ, Moretti R, Baker D, Houk KN. Computational Enzyme Design. *Angew Chem Int Ed.* 2013;52:5700–25.
 96. Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N, et al. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol.* 2013 Jun 9;9(8):494–8.
 97. Tinoco Jr I, Bustamante C. How RNA Folds. *J Mol Biol.* 1999;293(2):271–81.
 98. Blackwell HE, Grubbs RH. Highly Efficient Synthesis of Covalently Cross-Linked Peptide Helices by Ring-Closing Metathesis. *Angew Chem Int Ed.*

- 1998;37(23):3281–4.
99. Hermann T, Patel DJ. Adaptive Recognition by Nucleic Acid Aptamers. *Science* (80-). 2000;287(4):820–5.
 100. Perumal V, Hashim U. Advances in biosensors: principle, architecture and applications. *J Appl Biomed*. 2014;12(1):1–15.
 101. Turner APF. Biosensors: sense and sensibility. *Chem Soc Rev*. 2013;42:3175–648.
 102. Liu A, Wang K, Weng S, Lei Y, Lin L, Chen W, et al. Development of electrochemical DNA biosensors. *Trends Anal Chem*. 2012;37:101–11.
 103. Drummond TG, Hill MG, Barton JK. Electrochemical DNA sensors. *Nat Biotechnol*. 2003;21:1192–9.
 104. Sassolas A, Leca-Bouvier BD, Blum LJ. DNA Biosensors and Microarrays. *Chem Rev*. 2008;108(1):109–39.
 105. Storhoff JJ, Elghanian R, Mucic RC, Mirkin CA, Letsinger RL. One-Pot Colorimetric Differentiation of Polynucleotides with Single Base Imperfections Using Gold Nanoparticle Probes. *J Am Chem Soc*. 1998;120(9):1959–64.
 106. O'malley RP, Mariano TM, Siekierka J, Mathews MB. A Mechanism for the Control of Protein Synthesis by Adenovirus VA RNAI. *Cell*. 1986;44(3):391–400.
 107. Cullen BR, Greene WC. Regulatory pathways governing HIV-1 replication. *Cell*. 1989;11(58):423–6.
 108. Huizenga DE, Szostak JW. A DNA Aptamer That Binds Adenosine and ATP. *Biochemistry*. 1995;34(2):656–65.
 109. Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*. 1992;355(6):564–6.
 110. Saidur MR, Abdul Aziz AR, Basirun WJ. Recent advances in DNA-based electrochemical biosensors for heavy metal ion detection: A review. *Biosens Bioelectron*. 2016;90:125–39.
 111. Willner I, Shlyahovsky B, Zayats M, Willner B. DNAzymes for sensing, nanobiotechnology and logic gate applications. *Chem Soc Rev*. 2008 Jun;37(6):1153–65.
 112. Liu J, Lu Y. Colorimetric biosensors based on DNAzyme-assembled gold nanoparticles. *J Fluoresc*. 2004 Jul;14(4):343–54.
 113. Wang F, Lu C-H, Willner I. From Cascaded Catalytic Nucleic Acids to Enzyme–DNA Nanostructures: Controlling Reactivity, Sensing, Logic Operations, and Assembly of Complex Structures. *Chem Rev*. 2014 Mar 12;114(5):2881–941.
 114. Gong L, Zhao Z, Lv Y-F, Huan S-Y, Fu T, Zhang X-B, et al. DNAzyme-based Biosensors and Nanodevices. *Chem Commun*. 2015;51(6):979–95.
 115. Lan T, Furuya K, Lu Y. A highly selective lead sensor based on a classic lead DNAzyme. *Chem Commun (Camb)*. 2010;46(22):3896–8.
 116. Li J, Lu Y. A Highly Sensitive and Selective Catalytic DNA Biosensor for Lead Ions. *J Am Chem Soc*. 2000 Oct;122(42):10466–7.
 117. Zhou W, Zhang Y, Ding J, Liu J. In Vitro Selection in Serum: RNA-Cleaving DNAzymes for Measuring Ca²⁺ and Mg²⁺. *ACS Sensors*. 2016 May 27;1(5):600–6.
 118. Luo Y, Zhang Y, Xu L, Wang L, Wen G, Liang A, et al. Colorimetric sensing of trace UO₂²⁺ by using nanogold-seeded nucleation amplification and label-

- free DNAzyme cleavage reaction. *Analyst*. 2012 Mar 19;137(8):1866–71.
119. Mei SHJ, Liu Z, Brennan JD, Li Y. An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. *J Am Chem Soc*. 2003;125(2):412–20.
 120. Chen Y, Chen L, Ou Y, Wang Z, Fu F, Guo L. DNAzyme-based biosensor for Cu²⁺ ion by combining hybridization chain reaction with fluorescence resonance energy transfer technique. *Talanta*. 2016 Aug 1;155:245–9.
 121. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*. 2007;35(2):406–13.
 122. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*. 2012;13(11):770–80.
 123. Li Y, Li X, Ji X, Li X. Formation of G-quadruplex-hemin DNAzyme based on human telomere elongation and its application in telomerase activity detection. *Biosens Bioelectron*. 2011;26(10):4095–8.
 124. Sen D, GILBERT W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature*. 1988;334(6180):364–6.
 125. Miyoshi D, Nakao A, Sugimoto N. Structural transition from antiparallel to parallel G-quadruplex of d(G 4 T 4 G 4) induced by Ca²⁺. *Nucleic Acids Res*. 2003;31(4):1156–63.
 126. Qu K, Zhao C, Ren J, Qu X. Human telomeric G-quadruplex formation and highly selective fluorescence detection of toxic strontium ions. *Mol Biosyst*. 2012;8(3):779–82.
 127. Chen FM. Strontium(2+) facilitates intermolecular G-quadruplex formation of telomeric sequences. *Biochemistry*. 1992 Apr 21;31(15):3769–76.
 128. Ono A, Togashi H. Highly Selective Oligonucleotide-Based Sensor for Mercury(II) in Aqueous Solutions. *Angew Chemie Int Ed*. 2004;43(33):4300–2.
 129. Li T, Dong S, Wang E. A Lead(II)-Driven DNA Molecular Device for Turn-On Fluorescence Detection of Lead(II) Ion with High Selectivity and Sensitivity. *J Am Chem Soc*. 2010;132(38):13156–7.
 130. Kotch FW, Fettinger JC, Davis JT. A lead-filled G-quadruplex: insight into the G-Quartet's selectivity for Pb(2+) over K(+). *Org Lett*. 2000 Oct;2(21):3277–80.
 131. Wang DY, Lai BHY, Sen D. A General Strategy for Effector-mediated Control of RNA-cleaving Ribozymes and DNA Enzymes. *J Mol Biol*. 2002;318(1):33–43.
 132. Achenbach JC, Nutiu R, Li Y. Structure-switching allosteric deoxyribozymes. *Anal Chim Acta*. 2005;534(1):41–51.
 133. Li Y, Geyer CR, Sen D. Recognition of anionic porphyrins by DNA aptamers. *Biochemistry*. 1996;35(21):6911–22.
 134. Travascio P, Li Y, Sen D. DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chem Biol*. 1998;5(9):505–17.
 135. Li T, Wang E, Dong S. Parallel G-quadruplex-specific fluorescent probe for monitoring DNA structural changes and label-free detection of potassium ion. *Anal Chem*. 2010;82(18):7576–80.
 136. Yurke B, Turberfield AJ, Mills Jr AP, Simmel FC, Neumann JL. A DNA-fuelled molecular machine made of DNA. *Nature*. 2000;406(6796):605–8.
 137. Yu Zhang D, Seelig G. Dynamic DNA nanotechnology using strand-

- displacement reactions. 2011;
138. Tyagi S, Kramer FR. Molecular Beacons: Probes that Fluoresce Upon Hybridization. *Nat Biotechnol.* 1996;14(3):303–8.
 139. He H-ZZ, Chan DSH, Leung C-HH, Ma D-LL, Shiu-Hin Chan D, Leung C-HH, et al. G-quadruplexes for luminescent sensing and logic gates. *Nucleic Acids Res.* 2013;41(8):4345–59.
 140. Kong D-M. Factors influencing the performance of G-quadruplex DNAzyme-based sensors. *Methods.* 2013 Dec 15;64(3):199–204.
 141. Li D, Wieckowska A, Willner I. Optical Analysis of Hg 2+ Ions by Oligonucleotide–Gold-Nanoparticle Hybrids and DNA-Based Machines. *Angew Chemie Int Ed English.* 2008;47(21):3927–31.
 142. Li D, Shlyahovsky B, Elbaz J, Willner I. Amplified Analysis of Low-Molecular-Weight Substrates or Proteins by the Self-Assembly of DNAzyme-Aptamer Conjugates. *J AM CHEM SOC.* 2007;129(18):5804–5.
 143. Xiao Y, Pavlov V, Niazov T, Dishon A, Kotler M, Willner I. Catalytic Beacons for the Detection of DNA and Telomerase Activity Scheme 1. (A) Analysis of DNA by Opening of a Beacon Nucleic Acid and the Generation of a DNAzyme. (B) Analyzing Telomerase Activity by a Functional DNA Beacon that Self-Generates a DNAzym. *J AM CHEM SOC.* 2004;126:7430–1.
 144. Niazov T, Pavlov V, Xiao Y, Gill R, Willner I. DNAzyme-Functionalized Au Nanoparticles for the Amplified Detection of DNA or Telomerase Activity.
 145. Xiao Y, Pavlov V, Niazov T, Dishon A, Kotler M, Willner I. Catalytic Beacons for the Detection of DNA and Telomerase Activity. *J Am Chem Soc.* 2004;126(24):7430–1.
 146. Chiuman W, Li Y. Simple Fluorescent Sensors Engineered with Catalytic DNA “MgZ” Based on a Non-Classic Allosteric Design. *PLoS One.* 2007;2(11):1224.
 147. Shen Y, Brennan JD, Li Y. Characterizing the Secondary Structure and Identifying Functionally Essential Nucleotides of pH6DZ1, a Fluorescence-Signaling and RNA-Cleaving Deoxyribozyme †. *Biochemistry.* 2005;44.
 148. Chiuman W, Li Y. Efficient signaling platforms built from a small catalytic DNA and doubly labeled fluorogenic substrates. *Nucleic Acids Res.* 2007;35(2):401–5.
 149. Shen Y, Mackey G, Rucpich N, Gloster D, Chiuman W, Li Y, et al. Entrapment of fluorescence-signaling DNA enzymes in sol-gel derived materials for metal ion sensing. *Ballesteros, A Trends Biotechnol.* 2006;1(1):6.
 150. Monsur Ali M, Aguirre SD, Lazim H, Li Y, Ali MM, Aguirre SD, et al. Fluorogenic DNAzyme probes as bacterial indicators. *Angew Chemie - Int Ed.* 2011;50(16):3751–4.
 151. Aguirre SD, Ali MM, Salena BJ, Li Y. A Sensitive DNA Enzyme-Based Fluorescent Assay for Bacterial Detection. *Biomolecules.* 2013;3(3):563–77.
 152. Qu L, Ali MM, Aguirre SD, Liu H, Jiang Y, Li Y. Examination of Bacterial Inhibition Using a Catalytic DNA. *PLoS One.* 2014;9(12):e115640.
 153. He S, Qu L, Shen Z, Tan Y, Zeng M, Liu F, et al. Highly Specific Recognition of Breast Tumors by an RNA-Cleaving Fluorogenic DNAzyme Probe. *Anal Chem.* 2015;87(1):569–77.
 154. Tram K, Xia J, Gysbers R, Li Y. An Efficient Catalytic DNA that Cleaves L-RNA. *PLoS One.* 2015;10(5):1–14.
 155. Gysbers R, Tram K, Gu J, Li Y. Evolution of an Enzyme from a Noncatalytic

- Nucleic Acid Sequence. *Sci Rep*. 2015;5:11405.
156. Shen Z, Wu Z, Chang D, Zhang W, Tram K, Lee C, et al. A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen. *Angew Chemie - Int Ed*. 2016;55(7):2431–4.
 157. Chan L, Tram K, Gysbers R, Gu J, Li Y. Sequence Mutation and Structural Alteration Transform a Noncatalytic DNA Sequence into an Efficient RNA-Cleaving DNAzyme. *J Mol Evol*. 2015 Dec 3;81(5–6):245–53.
 158. Liu M, Zhang W, Zhang Q, Ohn J, Brennan D, Li Y. DNA Amplification Biosensing by Tandem ReactionsofS structure Switching, Nucleolytic Digestion, and DNAAmplification of DNA Assembly. *Angew Chem Int Ed*. 2015;54(33):9637–41.
 159. Schlosser K, Li Y. Tracing Sequence Diversity Change of RNA-Cleaving Deoxyribozymes under Increasing Selection Pressure during in Vitro Selection. *Biochemistry*. 2004;43(30):9695–707.
 160. Sabeti PC, Unrau PJ, Bartel DP. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem Biol*. 1997;4(10):767–74.
 161. Wedel AB. Fishing the best pool for novel ribozymes. *Trends Biotechnol*. 1996;14(12):459–65.
 162. Salehi-Ashtiani K, Szostak JW. In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature*. 2001;414(6859):82–4.
 163. Kim S, Shi H, Lee DK, Lis JT. Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res*. 2003;13(7):1955–61.
 164. Nazarenko IA, Uhlenbeck OC. Defining a smaller RNA substrate for elongation factor Tu. *Biochemistry*. 1995;34(8):2545–52.
 165. Pan T. Novel RNA Substrates for the Ribozyme from *Bacillus subtilis* Ribonuclease P Identified by in Vitro Selection. *Biochemistry*. 1995;34(26):8458–64.
 166. Altman S. Ribonuclease P. *Philos Trans R Soc B*. 2011;366(1580):2936–41.
 167. Rychlik MP, Chon H, Cerritelli SM, Klimek P, Crouch RJ, Nowotny M. Crystal structures of rnae h2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. *Mol Cell*. 2010;40(4):658–70.

Chapter 2

In vitro Selection of a Fluorogenic DNA/RNA molecule as the Substrate for Ribonuclease H2 from *Clostridium difficile*

2.1 Author's Preface

This chapter describes an *in vitro* selection that was in fact the second attempt with the same target, RNase H2 from *C. difficile*. During my first attempt, as a junior graduate student, I initiated the selection in the conventional manner. Although it generated promising results after a steep learning curve, I came to realize that I could have designed the *in vitro* selection experiments more suitable for the goal being pursued. Therefore, I took the courage to restart the selection with critical modifications to the protocol. The results of the recent *in vitro* selection were not only promising, but also demonstrated the ability of FNA sensors to distinguish similar target, empowered by *in vitro* selection.

2.1 Introduction

Biosensors often include a recognition element that binds to a target of interest, and a signal transducer so that the recognition event can be observed. Sensitivity and specificity are the two key parameters used to evaluate biosensing methods (1,2). To enhance the sensitivity, many methods take the advantage of amplification strategies to either amplify the target or the signal. For example, many nucleic acid amplification-based tests, such as polymerase chain reaction (PCR) and loop-mediated isothermal amplification (LAMP) (3), detect targets by producing many amplicons from the original sequence. In contrast, protein targets cannot be readily amplified, therefore antibody-based protein detection methods often use a coupled enzyme to amplify the recognition signal. For example, the popular enzyme-linked immunosorbent assay (ELISA) utilizes an enzyme linked to the secondary antibody to generate an amplified signal upon target binding by the primary antibody (3).

Functional nucleic acids, particularly DNA aptamers and DNAzymes, have attracted a great deal of attention for biosensor research as they can be used as the key building blocks to set up biosensors (2,4–22). A variety of DNA aptamers have been produced by *in vitro* selection for recognizing a wide range of targets, ranging from metal ions to whole cells, with impressive sensitivity and specificity (4,7,8,10,14,18,19,22–30). In addition, many DNAzymes generated by *in vitro* selection have been employed to design biosensors, either as the sensing or the signalling component (11,12,16,21,31–40).

Our group has a long-standing interest in developing DNA aptamer and DNAzyme based biosensors (11,41–49). For example, we have developed a number of RNA-cleaving DNAzymes that are capable of detecting metal ions, pH change and bacterial pathogens (31,36,53–56,41,43–45,49–52). These systems are designed to

function using the signaling mechanism illustrated in Figure 2-1. Specifically, when the target of interest binds the DNAzyme, the cleavage occurs at the RNA location of a fluorogenic DNA/RNA (FDR) chimeric substrate. Because the cleavage site is surrounded by a pair of nucleotides modified with a fluorophore and a matching quencher, the cleavage leads to the generation of a real-time fluorescent signal (57). The successful design of these DNAzyme systems offer multiple turnovers and are thus able to provide signal amplification. The fluorogenic RNA-cleaving DNAzymes are typically derived from a random-sequence DNA pool by *in vitro* selection in which the DNA pool is subjected to a selection process that favor the enrichment of DNAzymes over non-catalytic DNA sequences round by round until the most active DNAzymes dominate the DNA pool (58,59).

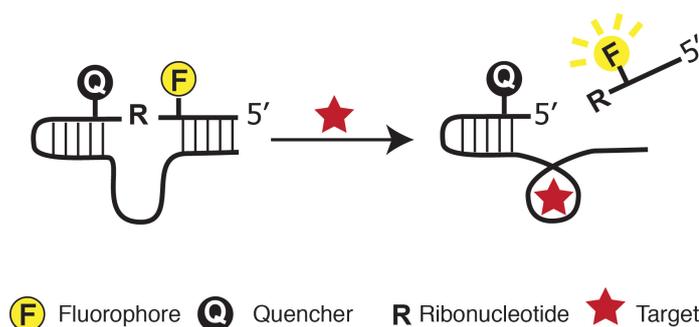


Figure 2-1. RNA cleavage by a target-induced RNA-cleaving DNAzyme. The RNA-cleaving reaction is coupled with fluorescent signal generation. The introduction of the target activates the DNAzyme, leading to the cleavage of the phosphodiester bond between RNA and DNA. The cleavage event causes the separation between fluorophore and quencher, resulting in increase in fluorescence.

Another conceivable way of developing fluorogenic RNA-cleaving biosensors with potentially great sensitivity is to develop effective FDR molecules for specific ribonucleases (RNases), as we will demonstrate using RNase H2 from *Clostridium difficile* (*C. difficile*) in this chapter. RNase H2 is an ubiquitously expressed enzyme that can cleave an RNA linkage in a hetero RNA/DNA duplex (60). Because they are highly efficient enzymes with great turnovers, RNase H2-FDR system can function as

highly effective sensors that take advantage of the turnover capability of the RNase H2. Earlier studies have shown that *in vitro* selection can be used to identify novel nucleic acid sequences or structures of nucleic acid-binding proteins (61–64). Therefore, it is feasible to engineer the proposed RNase H2/FDR systems.

The key objective of this work is to use *in vitro* selection to identify FDR molecules that can be efficiently cleaved by CDH2, RNase H2 from *C. difficile* (Figure 2-2). *C. difficile* is a rod-shaped, gram-positive bacterium that is resistant to many broad-scheme antibiotics. It is the leading cause of nosocomial infections. *C. difficile* infection (CDI) is estimated to affect more than 300,000 hospitalized patients annually in the United States (65). *C. difficile* causes infection when the gut flora is largely affected by broad scheme antibiotic treatment. *C. difficile* infection is characterized by constant and bloody diarrhea, fever, and pseudomembrane formation in the colon. It leads to death in severe cases (66–70). Data collected showed that it caused 14,000 deaths in the US in 2007 (71). CDI costs U.S. acute care facilities in excess of \$4.8 billion annually (72–74). As suggested in the latest report by Canadian Nosocomial Infection Surveillance Program, a rise in CDI cases has been seen, and likely attributed to community-acquired and recurrent cases (65). Therefore, the surveillance of *C. difficile* is important for diagnosis and the control of its infection. As a long-term goal of our group, we wish to develop a highly selective fluorogenic substrate for CDH2 and use it to detect a biosensor to detect *C. difficile*.

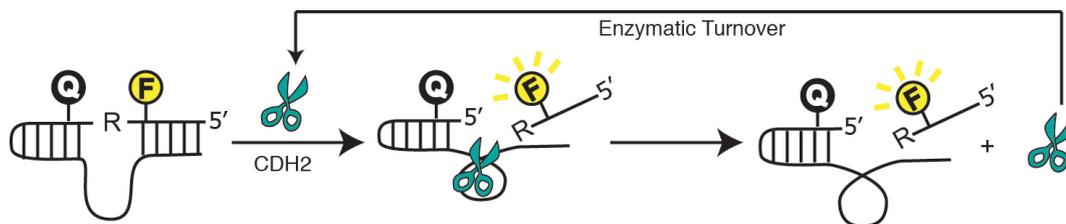


Figure 2-2. Fluorogenic DNA/RNA (FDR) Substrate for *C. difficile* RNase H2 (CDH2). CDH2 cleaves the RNA-DNA junction within the substrate, leads to fluorescence signal, and then cleaves more substrates due to enzymatic turnover.

For the motivation stated above, we want to determine if we can derive FDR molecules with significant selectivity for CDH2 over RNase H2 from other bacterium using an *in vitro* selection method that incorporates a counter-selection step prior to a positive selection step. For this purpose, we choose to use RNase H2 from *Escherichia coli* (ECH2) and *Salmonella typhimurium* (STH2) as the counter-selection targets.

E. coli and *S. typhimurium* are gram-negative bacteria, which are very different from gram-positive *C. difficile*. However, the sequence alignment of RNase H2 protein from these three organisms shows >50% similarity (Figure 2-3). They all include highly conserved catalytic domains and substrate binding sites. These shared properties present a great challenge to select for highly specific FDR molecules. We would rely on non-conserved domains to provide unique binding between the FDR and RNase H2 from *C. difficile*. This may be feasible, given that several published *in vitro* selection studies have successfully used the combination of counter-selection and positive-selection to drive for the recognition specificity (75–78).

```

EC -----
ST -----
CD MQDKSVREIKEI IETLEVEKYMEYIELLRVDERKSVQGLAIKLAKKLDNIRKEEERLETI

EC ---MIEFVYPHTQLVAGVDEVGRGPLVGAVVTAAVILDPARPIAGLNDSKKLSEKRRLLAL
ST ---MIEFVYPHTLHVAGVDEVGRGPLVGAVVTAAVILDPARPIVGLNDSKKLSEKRRLSL
CD NIFENEGYDKGYLYIGGIDEAGRGPLAGPVVASVVVFKKDTKIEGVNDSKKLSEAKRDEL
      *          :. :* .***** .* .** :. : * * :***** : * *

EC YEEIKEKALSWSLGRAEPHEIDELNILHATMLAMQRAVAGLHIAPEYVLIIGNRCPKLP
ST YDEIKEKALSWSLGRAEAHEIDELNILHATMLAMQRAVAGLHIAPEYVLIIGNRCPPELPV
CD FEVIKKEALDYGIGIVNNEEIDFNILNATYAMKKA INCLKKAPDYLLVLAATI PGIDI
      :: ** :** :. : * . : .***** :** :** :** : * : ** :** :* . * : :

EC PAMAVVKGDSRVPEISAASILAKVTRDAEMAALDIVFPQYGF AQHKGYP TAFHLEKLAEH
ST PSMAVVKGDSRVAEISAASILAKVTRDAEMAALDIVFPQYGF AQHKGYP TAFHLEKLAQY
CD SQNP I V K G D S K S I S I A A A S I L A K V T R D S I M Y Q Y D R V Y P E Y G F K S H K G Y G T K E H Y E A I E K Y
      . . :***** : . :***** : * * * * :** :** .***** * * * : :

EC GATEHRRSF GPVKRALGLAS
ST GATAHRRSF APVKRALGLVS
CD GITPIHRKSF LKNIL-----
      * * ** :**

```

Yellow: amino acids involved in active sites
Cyan: amino acids involved in binding domains

Figure 2-3. Sequence Alignment of RNase H2 of *E. coli* (EC), *S. typhimurium* (ST), and *C. difficile* (CD), and conserved binding sites and catalytic sites. *Completely conserved amino acids; /. Non-conserved amino acids.

2.2 Results and Discussion

2.2.1 In vitro Selection

The selection scheme is illustrated in Figure 2-4A. The DNA library used for the selection, DL1, contains a 40-nt (nucleotide) random-sequence domain flanked by a 15-nt fixed-sequence on each side (Figure 2-4B). The library is ligated to the 3' end of 30-nt FDR1 in the presence of LT1, the ligation template (Step 1). The ligated construct was purified in Step 2 using denaturing (7 M urea) polyacrylamide gel electrophoresis (dPAGE), followed by incubation with ECH2 and STH2 in Step 3. The uncleaved fraction of the DNA pool in this counter-selection (CS) step was purified using dPAGE (Step 4), and then reacted with CDH2 (Step 5). The DNA molecules cleaved in this positive selection (PS) step were isolated by dPAGE (Step 6), and then

subjected to two consecutive polymerase chain reactions (PCRs) in Steps 7 and 8. The first PCR used the standard DNA primers FP1 and RP1, but the second PCR used FP1 and a modified reverse primer, RP2, that contained the A20 at the 5'-end and a non-amplifiable Spacer 9 carbon linker (L), and therefore, the antisense DNA strand of the PCR product in this step is 20-nt longer than the sense DNA strand. Following the second PCR step, the sense DNA strand was separated from the antisense strand by dPAGE (Step 9). The purified sense strand was ligated to FDR1 and used for the next round of selection (repeating Steps 2-10). A total of 10 rounds of selection were carried out using these procedures. Table 2-1 provides the concentrations of DNA pool, CDH2, ECH2 and STH2 as well as the reaction time used for both CS and PS steps in each round of selection, along with the observed percent cleavage in both CS and PS.

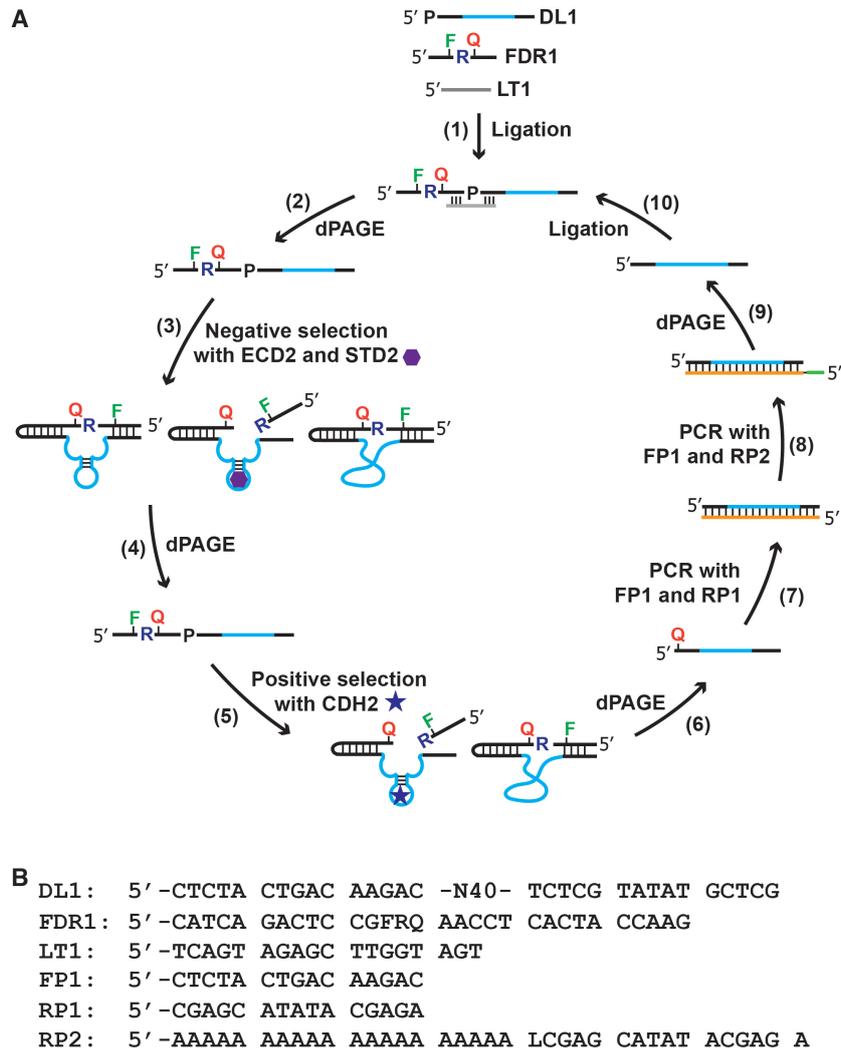


Figure 2-4. In vitro Selection. (A) Steps in the in vitro selection strategy. (B) Sequences of DNA molecules used in panel A. L: Spacer 9 carbon linker.

Selection Round	[ECH2] for CS (nM)	[STH2] for CS (nM)	[DNA] for CS (nM)	NS reaction time (min)	% Clv	[CDH2] for PS (nM)	[DNA] for PS (nM)	PS reaction time (min)	% Clv
1	5	5	1000	180	3.4	20	100	180	1.2
2	5	5	100	O/N	9.1	10	100	150	3.5
3	5	5	100	O/N	38	5	100	30	5.3
4	5	5	100	O/N	42.5	1	100	15	1.9
5	5	5	100	O/N	17	1	100	15	2.1
6	5	5	100	O/N	70	1	100	5	3.5
7	10	10	100	O/N	79	1	100	1	1
8	10	10	100	O/N	71.6	1	100	1	2.3
9	10	10	100	O/N	76	1	100	1	2.6
10	10	10	100	O/N	65	1	100	1	3.4

Table 2-1. In vitro selection conditions and outcomes. CS: Counter selection; PS: Positive selection; O/N: overnight (16-20 h).

It should be noted that the first round of selection used ~1 nmole of the ligated FDR1-DL1 construct, corresponding to $\sim 10^{15}$ individual molecules. Several strategies were implemented to drive the selection of FDR molecules that have high activity for CDH2 and high selectivity over ECH2 and STH2. First, the concentration of CDH2 was progressively reduced: from 20 nM in round 1, to 10 nM in round 2, 5 nM in round 3 and 1 nM in rounds 4 and after. Second, the reaction time in the PS step was progressively reduced, from 180 min in round 1 to 150 min in round 2, to 30 min in round 3, to 5 min in rounds 4 and 5, and to 1 min rounds 6-10. Third, the concentrations of EC2 and STH2 was increased from 5 nM each in rounds 1-6 to 10 nM each in rounds 7-10. Fourth, the reaction time used for the CS step was increased from 180 min in round 1 to overnight incubation (16-20 h) in rounds 2-10.

Because the changes in reaction conditions from round to round, we decided to use a fluorogenic chimeric DNA/RNA duplex molecule, named FDR1D (Figure 2-5A), as the reference substrate to judge the progress of in vitro selection. FDR1D consisted of FDR1 and its full complementary sequence, which was designed to mimic the intrinsic substrate for RNase H2. We determined the comparative cleavage activity (CCA) for each round of selection, which was calculated as the ratio of the percent cleavage of the pool by CDH2 by the percent cleavage of FDR1D by CDH2 at the same concentration. As shown in Figure 2-5B, the CCA increased significantly with the progress of the selection: at round 2, the relative activity was merely 0.2; by round 10, it increased to 8.6.

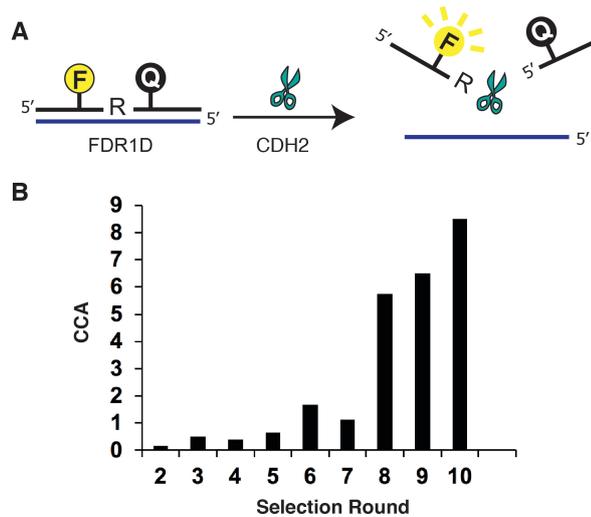


Figure 2-5. Monitoring selection progress. (A) FDR1D, a reference duplex to measure the progress of selection. (B) Selection progress measured by comparative cleavage activity (CCA) defined as the ratio of percent cleavage of each pool by CDH2 divided by percent cleavage of FDR1D by CDH2 at the same concentration.

2.2.2 Activity Assessment of Enriched Sequences

We carried out next-generation sequencing to obtain enriched sequences in G10, the selected pool in round 10. The top 10 sequences are provided in Table 2-2. These sequences account for 31% of the total sequence reads, with 11% for the most abundant sequence, which is named FDRC1.

Class	Copy No.	% in G10	N ₄₀ Random-Sequence Domain*
FDRC1	42507	10.6	CCTAA CATGT GACGT AGTGT GCGGT TGAGG TTCGT ACGTG
FDRC2	18554	4.6	GCAAA AGGAA GATGT GTGCG GTTGA GGTTA TAGTA CGTGG
FDRC3	14529	3.6	CATGC CGGAA CTTAC CTATG TGACG GAGAG GTTAT TCGGG
FDRC4	13030	3.2	CAAGC AGAAA GGACA GTGTG TGC GG TATGG GTTGT ATGGG
FDRC5	8756	2.2	CATGG ACTCA GGTAG TGTGA GGTG AGTGT TCGTA TTGGG
FDRC6	7422	1.8	CAACA GCAAG CACAG TGTGT TGGAG GGAGG TTATA ACGTG
FDRC7	5075	1.3	CACCG GGCAA ACAGT GTGTG TCGGT TGAGG TTGTA TGGGG

FDRC8	4825	1.2	CCAGG ACGGA CGGTG GTGAG AGGTT GAGGG TTGTA TTGGG
FDRC9	4253	1.1	CAAAT CAGGG CAGTG TGGTG CAGAG TGGGT TATAA CGCCG
FDRC10	4105	1.0	CAAGG ATAAG GACGT AGTGT GAGGG GTAGG TTGTA CGTCG

Table 2-2. Top 10 sequences in G10 pool

*The full library sequence is 5'- CATCA GACTC CGT (F) rAT (Dab) AACCT CACTA CCAAG CTCTA CTGAC AAGAC-N40-TCTCG TATAT GCTCG-3'

Next, we measured the activity of all top 10 sequences in the presence of CDH2, ECH2 or STH2 (all of which was used at 1 nM), with an incubation time of 15 min. To simplify the comparison, we calculated the relative activity (RA) using the following normalization:

$$\text{RA} = (\text{Percent Cleavage of a FDRC by an RNase H2}) / (\text{percent cleavage of FDRC1 by CDH2}) \times 100.$$

The data is plotted as Figure 2-6A. Without exception, all top 10 sequences exhibited the highest activity with CDH2 (green bar in each set) and significant reduced activities with ECH2 (red bar) and STH2 (blue bar). The most dominant class, FDRC1, demonstrates the highest activity towards CDH2 (RA of 100), followed by FDRC8 (RA of 91), and followed by FDRC6 (70), FDRC4 (67) and FDRC5 (60).

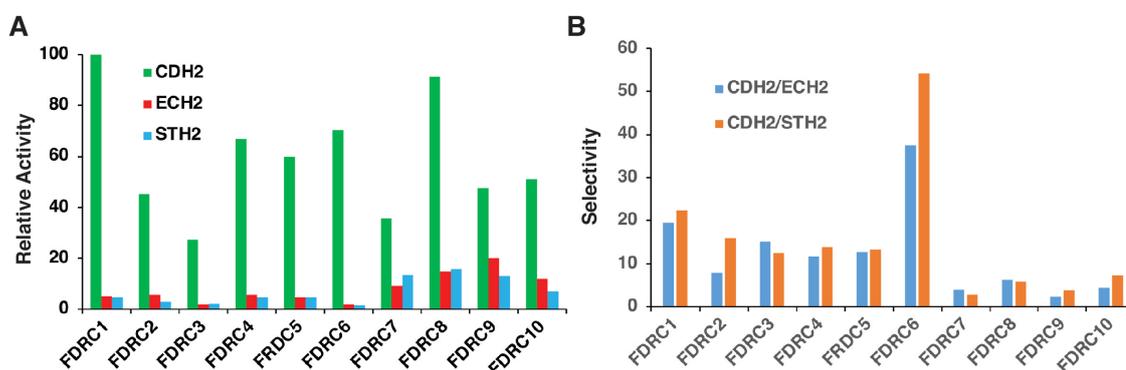


Figure 2-6. (A) Relative activity and (B) selectivity of top 10 sequences. The top 10 sequences were ligated to FDR1 and purified using dPAGE. Purified sequences were diluted using 1x SB to the concentration of 100 nM. RNase H2 of *C. difficile*, *E. coli*, or *S. typhimurium* was added to a final concentration of 1 nM. The relative activity was obtained by taking the cleavage activity of FDRC1 by CDH2 as 100. The selectivity was derived the ratio of the cleavage activity of an FDRC by CDH2 over the cleavage activity of the same FDRC by ECH2 or STH2.

We also calculated selectivity activity (SA) using the following normalization:

$$\text{SA} = (\% \text{ Cleavage of a FDRC by CDH2}) / (\% \text{ cleavage of the same FDRC by ECH2 or STH2})$$

The data is plotted as Figure 2-6B. Interestingly, FDRC6 showed the best selectivity – it registered an SA of 54 and 37 against STH2 and ECH2, respectively. FDRC1 also exhibited high levels of selectivity, scored an SA of ~20 against both STH2 and ECH2. FDR2-5 also demonstrated excellent levels of selectivity, with SA values varying between 8 and 16. FDR7-FDR10, however, produced reduced levels of selectivity, with SA values ranging from 3-7.

The data presented above clearly indicate that our selection strategy has led to the isolation of FDR molecules with the desired properties: high activity for CDH2 and high selectivity for CDH2 over ECH2 and STH2.

2.2.3 Secondary Structure Prediction

A structural model predicted for the top rank sequence FDRC1 using Mfold is shown in Figure 7A. The structure contains 7 short Watson-Crick base-pairing regions (named P1-P7), 6 internal bulges (named B1-B6) and one loop (named L1). Interestingly, the structural model predicted for FDRC6 using mfold (Figure 7B) shares considerable similarity to the structural model of FDRC1. Specifically, the structural model of FDRC6 also contains 7 pairing regions (P1-P7), 6 internal bulges (named B1-B6) and one loop (named L1).

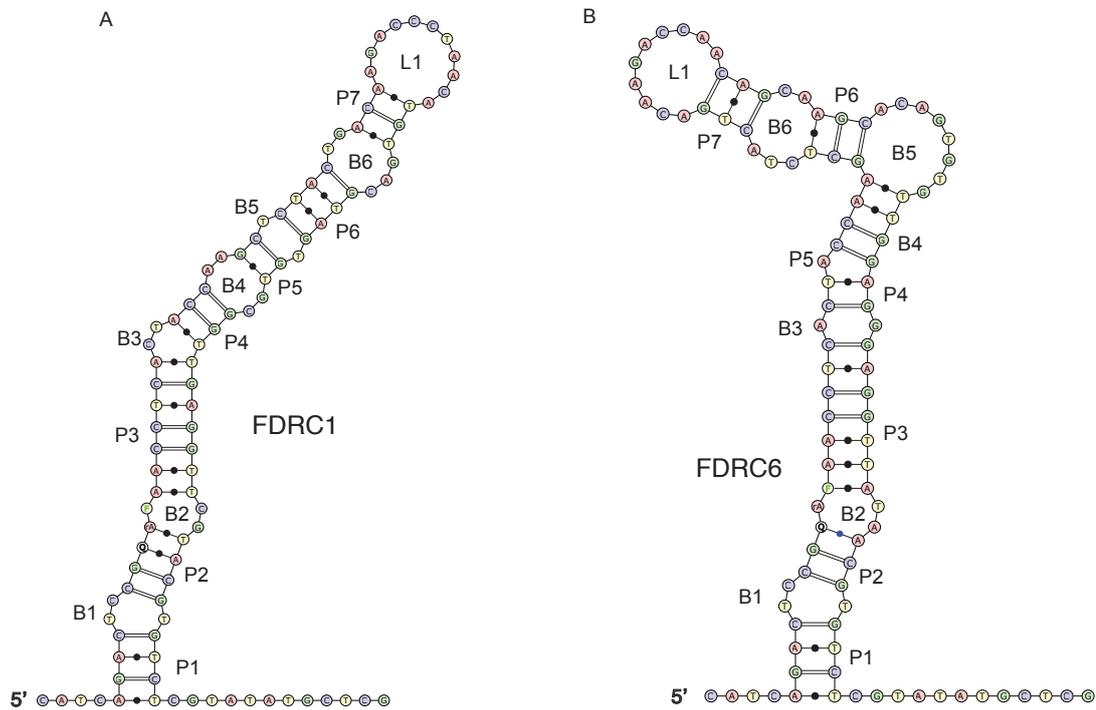


Figure 2-7. Structural models predicted with m-fold for (A) FDRC1, and (B) FDRC6. Letter P in the figure represent pairing regions; and letter B represent bulges.

FDRC1 and FDRC6 share great similarity at the bottom half of their secondary structures (Figure 2-8). Both of them contain the identical GAGGTT sequence (sequence highlighted in red in Figure 2-8) that forms part of P3. This may explain the high activity of both molecules towards CDH2 and high selectivity against ECH2 and STH2 (Figure 2-6). There are noticeable differences in the pairing arrangement concerning F, R and R in the structural context of P2, B2 and bottom part of P3. This may also explain that the higher activity of FDRC1 but better selectivity of FDRC6.

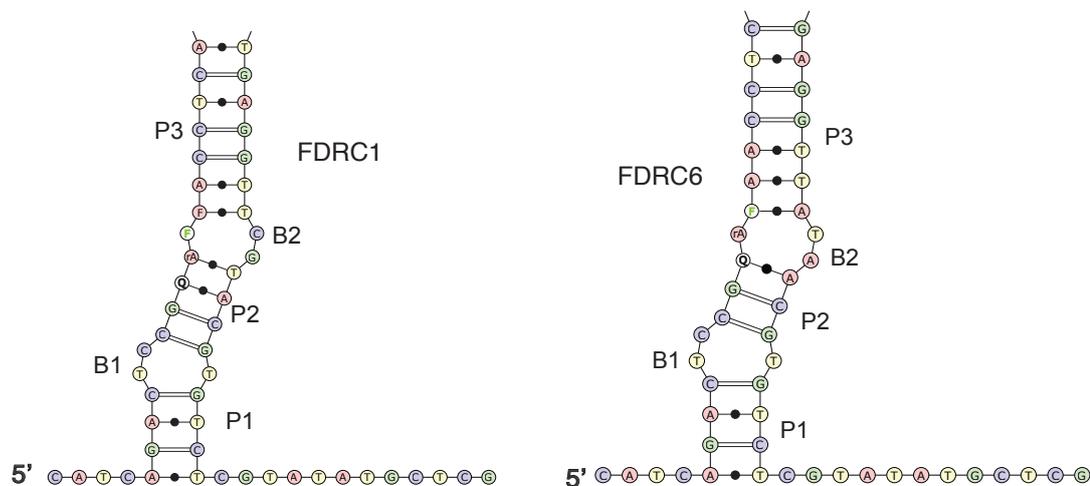


Figure 2-8. Comparing the similarity of the bottom half of the secondary structures predicted for FDRC1 and FDRC6. They both appear to have three pairing regions and two bulges at similar regions. In particular, both of them have a bulge near the cleavage site (rA ribonucleotide).

Secondary structures for the other top 10 classes were also predicted with Mfold, which are provided in Figure 2-9. The common features observed with these structures are that all of them contain multiple pairing regions and bulges, features shared by FDRC1 and FDRC6. However, few bulges are found at the bottom half of the secondary structures predicted for lowly ranked top 10 sequences. For example, FDRC7 contains 0 bulge in that area; both FDRC8 and FDRC10 contain a single bulge therein. These observations seem to suggest the bulge elements in the top rank sequences might be responsible for the observed high activity and selectivity.

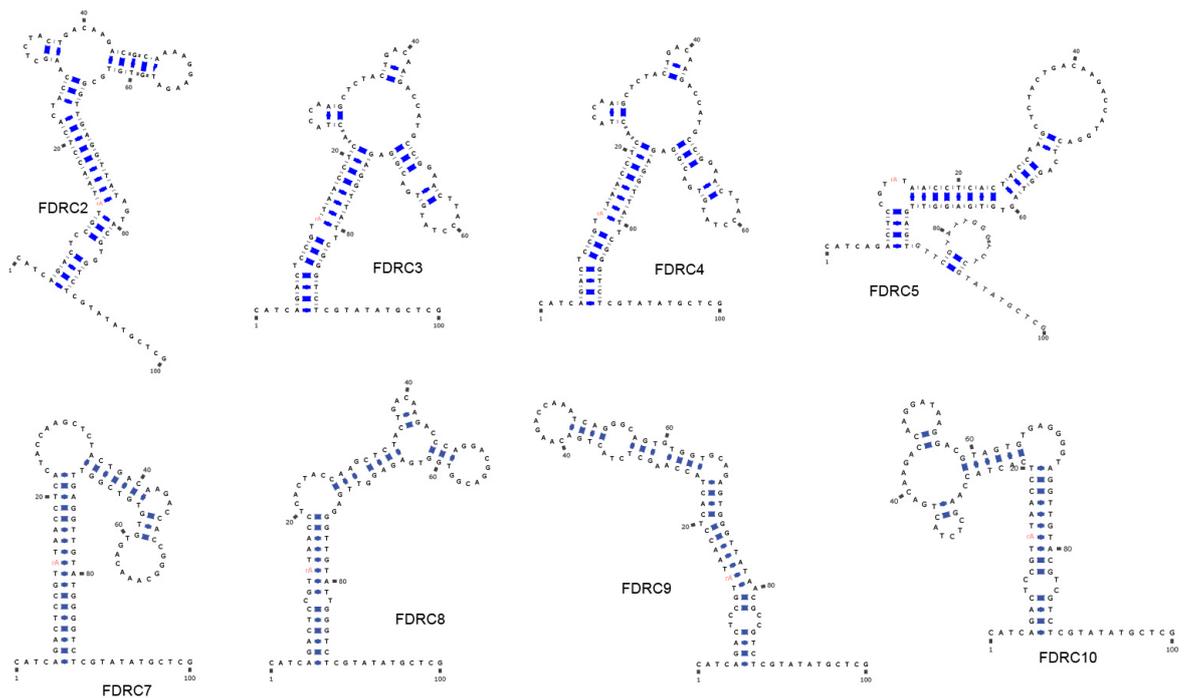


Figure 2-9. Structural models predicted with Mfold for the remaining 8 top 10 sequences

2.2.4 Secondary Structure Characterization of FDR1

As shown in Figure 2-7A, FDR1 is predicted to have multiple duplex regions sandwiched between several loops and bulges. We investigated the significance of these sequence and structural elements to obtain better understanding of the sequence requirement by CDH2. It is worth noting that in this experiment, FDR1 was tested in *trans*, using a series of constructs shown in Figure 2-10. This is because constructing FDR molecules in *trans* is less laborious and resulted in better activities. Shortening L1 from 11-nt loop to 5-nt loop (FDR1-1), removing P7 and L1 (FDR1-2) and removing L1, P7, and B6 altogether (FDR1-3) did not impose a detrimental effect on both the activity towards CDH2 and the specificity over ECH2 and STH2. In fact, the truncated construct FDR1-3 exhibited enhanced activity (35% higher than FDR1). However, further truncation by removing B5 and P6 (FDR1-4) resulted in the complete loss of activity.

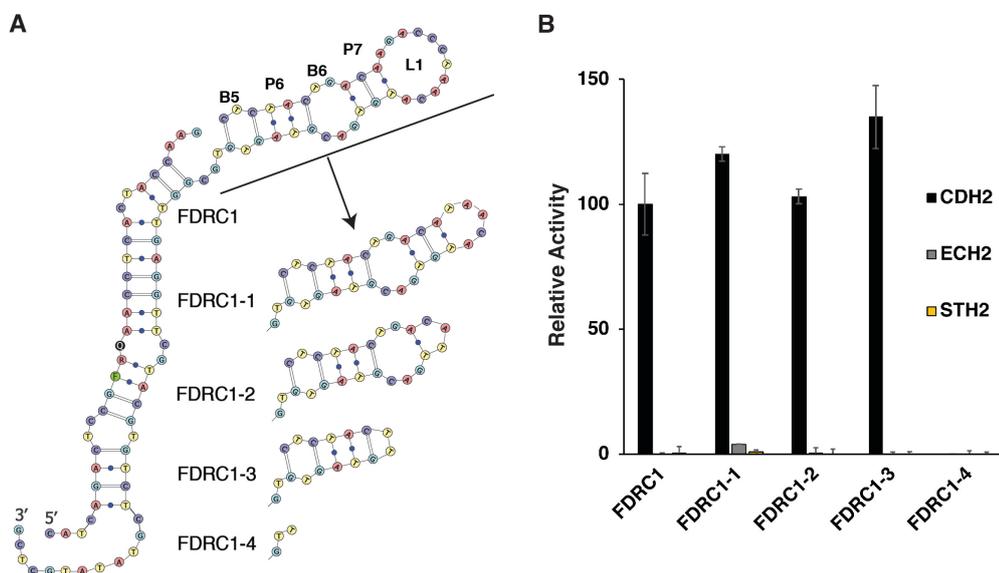


Figure 2-10. Assessing the significance of pairing regions (P6 and P7), loop (L1), and bulges (B5 and B6). Each test used 100 nM of FDR1 or one of its variants, 100 nM of FDR1, and 1 nM of CDH2, ECH2, or STH2. The reaction time was 15 minutes. The resultant reaction mixture was loaded onto dPAGE for analysis. The cleavage activity of FDR1 by CDH2 was taken as 100.

We also made an attempt to truncate nucleotides from the 3'-end of FDR1-3. The removal of 6 (FDR1-3A) and 12 (FDR1-3B) nucleotides from the 3'-end led to shortened sequences with slightly enhanced activity. However, further truncation by removing the next 5 (FDR1-3C) and 9 (FDR1-3D) resulted in the complete loss of activity (Figure 2-11). These observations suggest that P1, B1 and P2 elements play highly important functional role for FDR1.

We next investigated the functional significance of bulges within FDR1-3A. Five new constructs were tested that is based on the secondary structure of FDR1-3A and each construct was designed to eliminate each of the five bulges, B1-B5, by creating a pairing element. Eliminating B1, B3, B4 and B5 led to significant reduction in reactivity towards CDH2. For example, the construct FDR1-3A-B4 in which the bulge B4 was converted into a perfectly pairing region, only exhibited 45% activity of

FDRC1-3A. Interestingly, the elimination of B2 resulted in a construct (FDRC1-3A-B2) that had similar CDH2 reactivity to FDRC1-3A, much increased reactivity to the other two RNases, especially to ECH2. This observation seems to suggest that the B2 element plays a very important role in selectivity.

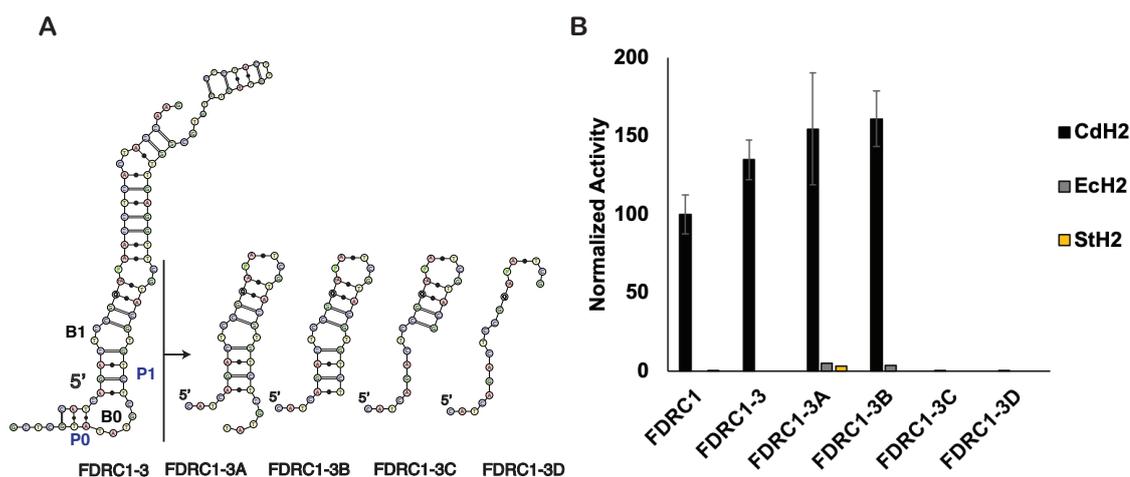


Figure 2-11. FDRC1-3 Constructs with Nucleotides Truncated from the 3'-End of the Sequence. Each test used 100 nM of FDRC1 or one of its variants, and 1 nM of CDH2, ECH2, or STH2. The reaction time was 15 minutes. The resultant reaction mixture was loaded onto dPAGE for analysis. The cleavage activity of FDRC1 by CDH2 was taken as 100.

The observation that certain bulges are required to achieve the selectivity for CDH2 over ECH2 and STH2 is a very interesting discovery. Prior to our work, all RNase H2 enzymes are known to cleave fully complementary hetero RNA/DNA duplexes. Our work shows that RNase H2 enzymes are also capable of cleaving duplex-like molecules that contain unpaired nucleotides. More importantly, there are bulge-containing heteroduplex structures that can be selectively cleaved by a specific RNase H2 but exhibit significantly reduced activity towards other RNase H2 homologues. This demonstrates the advantages of in vitro selection in the search of unique structures that can serve as binding partners or substrates for nucleic acid processing enzymes, as rational designs are hard to be applied in such cases.

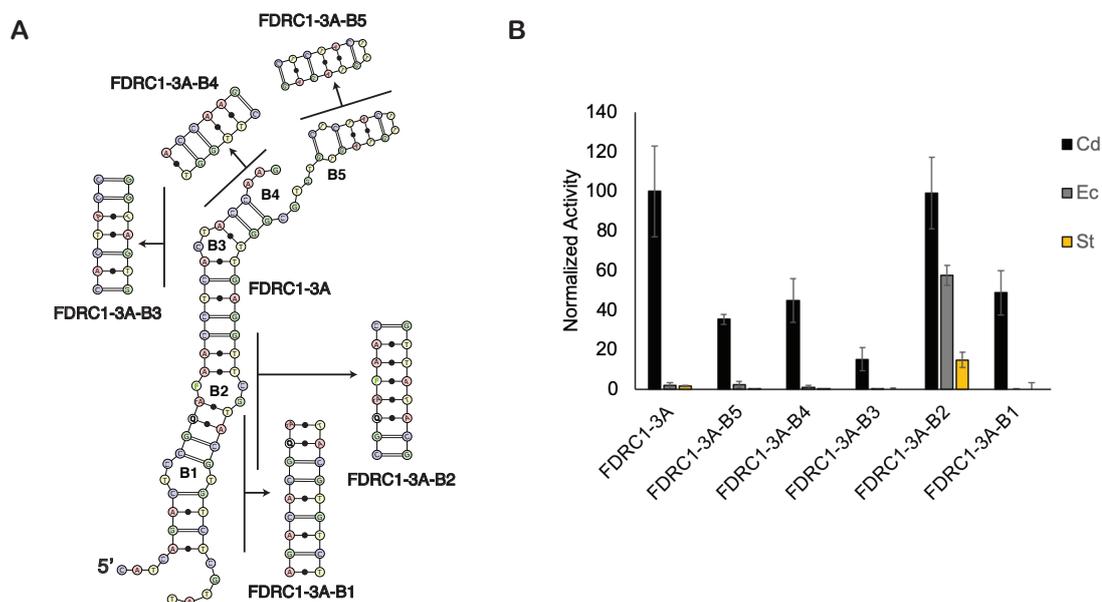


Figure 2-12. Assessment of the Significance of Bulges within FDRC1-3A. 100 nM of each FDRC1 variant was reacted towards 1 nM of CDH2, ECH2, and STH2, respectively, and incubated for 15 minutes. The resultant reaction mixture was loaded onto dPAGE for analysis.

2.2.5 Requirement of F and Q

Our in vitro selection experiment used a fluorogenic chimeric DNA/RNA substrate that contains a single ribonucleotide as the cleavage site, which are surrounded by a pair of nucleotides modified with fluorescein (F) and DABCYL (Q). To assess the essentiality of the F and Q moieties, we assessed the activity of DRC1-3B, the version of FDRC1-3B with the removal of the F and Q (Figure 2-13). For this experiment, the substrate strand was labeled with a 5'-³²P. DRC1-3B was a better substrate than FDRC1-3B towards CDH2 (400% more active). At the same time, it maintained a high level of selectivity over ECH2 and STH2. Based on this observation, it appears that F and Q do not contribute to the selectivity.

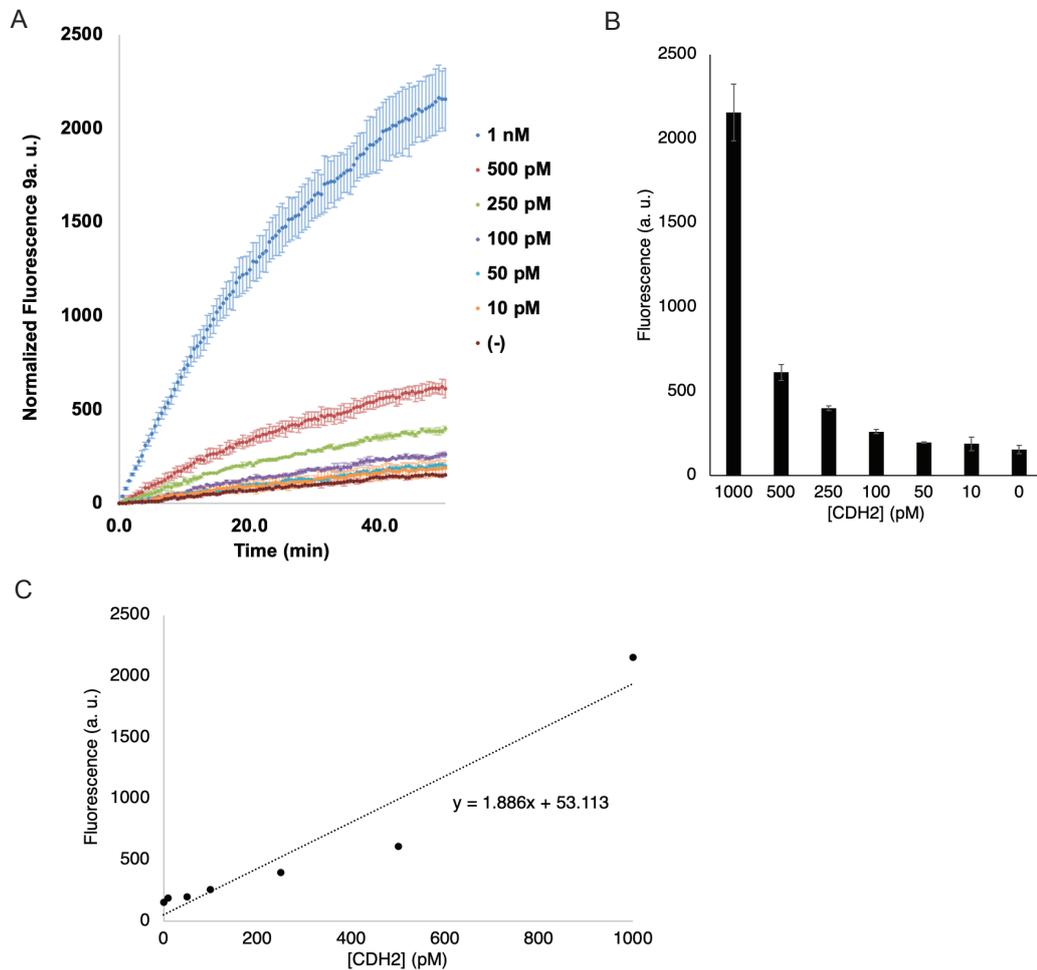


Figure 2-14. Assessing the Sensitivity of FDRC1-3B Using Fluorescence Spectrometer. A). RNase H2 of *C. difficile* at the concentration of as low as 50 pM (electric blue) can be detected using fluorescence spectrometer by the FDRC1-3B complex after 50 minutes. The fluorescence reading was started ~10 minutes after the addition of RNase H2 at various concentrations to the mixture of 100 nM of FDRC1-3B. B). The fluorescence at 50 minutes versus the concentration of *c. difficile* RNase H2. C). The LOD was determined using formula $LOD = 3 \cdot (SD \text{ of } y\text{-intercept}) / \text{Slope}$

2.3 Concluding Remarks and Future Directions

We have shown for the first time that in vitro selection can be used to derive chimeric DNA/RNA molecules, from a random-sequence pool, with two intriguing properties: high activity for a targeted RNase H2 and significantly reduced activity towards non-targeted RNase H2. This finding was made using three RNase H2

homologues: CDH2 as the intended enzyme, and a combination of ECH2 and STH2 as the unintended enzymes. We have isolated many sequences and assessed the activity the top 10 sequences towards these RNases. Without exception, all the top 10 sequences exhibit strong cleavage activity towards CDH2 and much reduced activity towards the two control enzymes. It is clear that the successful isolation of these sequences benefits from the incorporation of a counter selection step with ECH2/STH2 and a positive selection step with CDH2. Our work clearly demonstrates that it is feasible to isolate chimeric DNA/RNA molecules that are highly selective for a particular RNase H2 homologue within the large RNase H2 family.

An interesting observation is that all selected DNA molecules exhibit a duplex-like secondary structure with multiple bulges, unpaired regions flanked by short duplex elements. This observation can be explained from two aspects. First, all RNase H2 enzymes recognize hetero DNA/RNA duplexes as their native substrates (79), which may be the key reason for the duplex-like structure seen with all the selected chimeric FDR molecules as the key determinant for the affinity by RNase H2 enzymes towards DNA is the duplex recognition. Second, the design of the selection method pushed for the selection of FDR molecules from the random-sequence DNA pool that can be recognized and cleaved by CDH2 but not by ECH2 and STH2. Because all three enzymes belong to same family and exhibit similar biochemical capability, they are largely similar in their active sites for their substrate binding and cleavage. However, there are notable differences in the sequences of the three RNase H2 enzymes, and expectedly, significant differences in the active site. The counter-selection and positive-selection strategy we employed has helped to promote the selection of DNA molecules capable of recognizing the differences. In particular, we suspect that some unpaired nucleotides within the duplex-like structures play critical roles in recognizing the

differences. For example, the B2 bulge in FDRC1 seems to play an important role in defining its high selectivity for CDH2 over ECH2 and STH2.

Our work points to the possibility of expanding the same approach to search for unique DNA or RNA sequences as highly specific substrates for a specific homologue of the other RNase enzyme families or and even DNase enzyme families.

A key question remains to be answered is whether the selected FRD molecules, such as FDR1, can be used to develop effective biosensors for the selective detection of *C. difficile* in the complex biological samples such as human stools and use it to diagnose *C. difficile* infections. This constitutes an important future direction of this project. FDR1 is set up well for a fluorescence-based bioassay as it is capable of generating a robust fluorescence signal upon RNA-cleavage upon RNA cleavage. It should also be compatible with other signaling mechanisms that have been demonstrated for aptamers and DNAzymes, such as colorimetric and electrochemical detection of analytes (8,14,18,22,29).

2.4 Materials and Methods

2.4.1 Enzymes, chemicals, and other materials.

T4 polynucleotide kinase (PNK) and T4 DNA ligase were purchased from Thermo Scientific (Ottawa, ON, Canada). [γ - ^{32}P]-ATP was obtained from Perkin Elmer (Woodbridge, ON, Canada). Urea (ultrapure) and 40% polyacrylamide solution (29:1) were acquired from BioShop Canada (Burlington, ON, Canada). The water used was purified via Milli-Q Synthesis A10 water purifier. All other chemicals were purchased from Sigma-Aldrich (Oakville, ON, Canada). 10 \times TBE (Tris-borate EDTA) (1 L) was made of 108 g Tris-base (0.89 M), 55 g boric acid (0.89 M), 20 mL of 0.5 M ethylenediaminetetraacetic acid (EDTA; pH 8.0; 10 mM). 2 \times native gel-loading buffer (per 100 mL) was made with 20 g sucrose, 10 mL of 10 \times TBE, 1 mL of 10% (w/v) SDS (sodium dodecyl sulphate), 25 mg bromophenol blue, and 25 mg xylene cyanol FF. 2 \times denaturing gel-loading buffer was made with the same recipe, with the addition of 84.1 g urea (14 M). Note that the concentration of Tris-boric acid in the 2 \times gel-loading buffer was 89 mM. 10% denaturing polyacrylamide gel stock was made of 250 mL of 40% polyacrylamide solution (29:1), 100 mL of 10 \times TBE, and 425 g urea.

2.4.2 Synthesis and purification of oligonucleotides.

DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA), prepared using automated DNA synthesis using standard phosphoramidite chemistry. All oligonucleotides were purified using 10% dPAGE.

2.4.3 Molecular Cloning and Expression of RNase H2.

Candidate genes encoding various RNase H2 (*rnhb* of *C. difficile*, *E. coli*, *S. typhimurium*) were cloned into pET-15b vector (Novagen) with His-tag incorporated.

NdeI and XhoI restriction sites were used for *C. difficile* *rnhb* gene; and NdeI and BamHI were used for *rnhb* of *E. coli* and *S. typhimurium*. PCR of each candidate gene was performed with genomic DNA extracted from *C. difficile* BI/027-H strain, *E. coli* K12 strain, *S. typhimurium*, respectively. High fidelity polymerase was used to amplify target genes. Amplified PCR product was purified using the Qiagen PCR purification kit. Purified PCR product and vector were digested with restriction enzymes (NEB) and gel-extracted using the Qiagen QIAquick Gel Extraction kit. Ligation reaction (20 μ L) was performed overnight at room temperature, using 70 ng of digested PCR product and 70 ng of digested plasmid, in 1 \times T4 DNA ligase buffer, and 2.5 units of T4 DNA ligase. Ligated DNA was transformed into *E. coli* competent cells by electroporation. Cells were then plated onto agar plates containing 25 g/L LB media, 20 g/L agar, and 100 μ g/mL ampicillin, and incubated over night at 37 °C. Positive clones were picked and grown in LB media, and plasmids were extracted using PureYield™ Plasmid Miniprep System. Purified vectors were verified by PCR and sequencing (MOBIX Lab, McMaster University), and then transformed into expression cell lines.

C. difficile RNase H2 was expressed in the *E. coli* BL21 (DE3)-pRARE cell line, and RNase H2 of *E. coli*, and *S. typhimurium* were expressed in the *E. coli* BL21 (DE3) cell line. *E. coli* cells with desired construct were grown at 37°C for overnight in LB with 100 μ g/mL ampicillin. The next morning, a 1:50 dilution of the overnight culture into fresh liquid media was performed. Cells were grown at 37°C until an OD₆₀₀ reached 0.6-0.7, and isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce overexpression. RNase H2 of *C. difficile*, *E. coli*, and *S. typhimurium* expressed well under the condition of 37 °C for 3 hours. Cells were centrifuged and stored at -80 °C.

2.4.4 Protein Purification.

Purification of RNase H2 of *C. difficile* was performed using fast protein liquid chromatography (FPLC) (Amersham Biosciences AKTA). Cells re-suspended from a 1 L culture was lysed with sonication, following the addition of mini EDTA-free protease inhibitor tablet (Roche). The crude cell lysate was loaded onto a 5 mL General Electric (GE) Healthcare Ni²⁺ column. A stepwise gradient of imidazole was used to elute the protein. Fractions containing the target protein was pooled, and loaded onto an ion exchange column, Mono Q (GE Healthcare) that was equilibrated with 20 mM Tris pH 8, 5 mM DTT, 5% glycerol, 0.05 M NaCl. A NaCl gradient from 0.05 M to 0.5 M was used to elute the protein of interest. Purified protein was stored in 14 mM Tris pH8, 3.5 mM DTT, 150 mM NaCl, 30% glycerol at -80 °C. The His-tag of *C. difficile* RNase H2 was removed using 0.25 U (Sigma-Aldrich) of thrombin at room temperature for 1 h prior to Mono Q purification. Sodium dodecyl sulfate polyacrylamide (SDS) gel electrophoresis was used for the assessment of protein expression and purification in above described experiments. RNase H2 of *E. coli* and *S. typhimurium* were purified using Qiagen QIAexpress Ni-NTA Fast Start Kit, following the protocol provided with the kit. Purified proteins were stored in 70% elution buffer, 30% glycerol, at -80 °C.

2.4.5 In vitro selection.

First round of selection was initiated by ligating the random sequence DNA library DL1 to FDR1. 2000 pmol of DL1 was first phosphorylated with 1 mM of ATP at the 5'-end in the presence of 10 units of T4 PNK, and 1× T4 PNK buffer A for 45 min at 37°C in a reaction volume of 100 µL. The reaction was stopped by heating at 90°C for 5 min. After cooling to room temperature, 2200 pmol of FDR1 and 2200 pmol

of LT1 were added. The reaction mixture was then heated to 90°C for 1 min and cooled to RT. 10 units of T4 DNA ligase and 25 µL of 10× T4 DNA ligase buffer were added to the reaction mixture (total reaction volume: 250 µL). The ligation reaction was carried out at room temperature for 2 h. The DNA in the mixture was precipitated by ethanol and the ligated product was purified by 10% dPAGE.

The purified FDR1-DL1 was quantified using NanoVue and resuspended in 1× selection buffer (1× SB; 45 mM HEPES pH 7.5, 300 mM NaCl, and 10 mM MgCl₂) to a final concentration of 100 nM. The pool was heated to 90°C for 1 min and cooled to RT, then ECH2 and STH2 were added to a final concentration of 5 nM each. The reaction mixture was left at RT for 3 h, and terminated by the addition of EDTA to a final concentration of 100 mM. The mixture was precipitated by adding 1/10 volume of 3 M NaOAc pH 5.2, and 2.5× 100% ethanol. The precipitated DNA was purified using dPAGE. The uncleaved sequences were excised from dPAGE, and eluted in 1× elution buffer (200 mM NaCl, 10 mM Tris pH 7.5, and 1 mM EDTA) at room temperature by vigorous shaking on vortex.

Eluted DNA was precipitated, and resuspended to a final concentration of 100 nM, and heated to 90°C for 1 min and cooled to RT. Then CDH2 was added to a final concentration of 20 nM, and reacted for 3 h. The reaction mixture was subjected to dPAGE using the protocol similar to the one used for the counter-selection, except that cleaved DNA band was excised, and elution was performed at RT overnight.

The isolated DNA above was amplified by polymerase chain reaction (PCR) in 50 µL of 1× PCR buffer (diluted from 10× PCR buffer supplied by the vendor), 0.2 mM of each standard dNTPs, 1.25 U of Taq DNA polymerase, 0.5 µM of FP1 (forward primer), and 0.5 µM RP1 (reverse primer) were added. 12 cycles of PCR were performed with

the following parameters: 94°C, 30 s (2 min for the first cycle); 52°C, 40 s; 72°C, 45 s. A 1/100-fold dilution of the first PCR product was used for the second PCR using the same condition described above with the exception that RP2 was used instead of RP1. Also, a minimum of 8 reactions were performed to ensure that enough DNA was generated for the next round of selection. This was especially important for this selection because each round of selection involved both counter-selection and positive-selection steps. The desired DNA molecules were purified by 10% dPAGE.

A total of 10 rounds of selection were conducted. Table 2-1 provides all the reaction conditions for each round of selection. The cleavage product from round 10 was amplified, cloned and sequenced using a protocol we published previously (11).

2.4.6 Assessment of the activity of top 10 classes.

Each of the top 10 classes (500 pmol) was first phosphorylated with 1 mM ATP at the 5'-end in the presence of 5 units of T4 PNK, and 1× T4 PNK buffer A for 45 min at 37 °C in a reaction volume of 50 µL. The reaction was stopped by heating at 90°C for 5 min. After cooling to room temperature, 550 pmol of FDR1 and 550 pmol of LT1 were added. The reaction mixture was then heated to 90°C for 1 min and cooled to RT. 5 units of T4 DNA ligase and 10 µL of 10× T4 DNA ligase buffer were added to the reaction mixture (total reaction volume: 100 µL). The ligation reaction was carried out at room temperature for 2 h. The DNA in the mixture was precipitated and purified by 10% dPAGE. The purified construct was quantified using NanoVue and resuspended in 1× selection buffer to a final concentration of approximately 110 nM. 9 µL of each class was heated to 90°C for 1 min and cooled to RT, and 1 µL of CDH2, ECH2 or STH2 diluted to 10 nM in 1× selection buffer was added to corresponding reactions.

The reaction was stopped after 15 min by addition of equal volume of 2× denaturing gel loading dye. 8 μL of each reaction was analyzed using dPAGE.

2.4.7 Assessment of Activity of FDRC1 and its variants.

These experiments were performed using FDR1 and a relevant DNA molecule at the ratio of 1:50 (to ensure all FDR1 molecules were hybridized to R10C1 variants). 4.5 μL of 220 nM FDR1 (in 1× SB) was mixed with 4.5 μL of 11 μM candidate DNA (also in 1× SB). The mixture was heated to 90°C for 1 min and cooled to RT, and subsequent reaction with CDH2, ECH2 or STH2 was carried out in the same way as described in 2.4.6.

2.4.8. Activity Assessment of DRC1-3B.

DR1 was used in this experiment to replace FDR1 – DR1 is identical to FDR1 except for the removal of F and Q. 10 pmol DR1 was labeled with 1 μL of [γ -³²P]ATP (10 μCi) in 20 μL of 1× PNK buffer A containing 2.5 units of PNK. After incubation at 37°C for 45 min, PNK was inactivated by heating at 90°C for 5 min. The phosphorylated DNA was purified by 10% dPAGE. The purified DR1 was mixed with 100 pmol of unlabeled DR1 and used for activity assessment of DRC1-3B using the same protocol described in 2.4.6.

2.5 Reference

1. Perumal V, Hashim U. Advances in biosensors: principle, architecture and applications. *J Appl Biomed*. 2014;12(1):1–15.
2. Turner APF. Biosensors: sense and sensibility. *Chem Soc Rev*. 2013;42:3175–648.
3. Lequin RM. Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clin Chem*. 2005;51(12):2415–8.
4. Wu J, Zhu Y, Xue F, Mei Z, Yao L, Wang X, et al. Recent trends in SELEX technique and its application to food safety monitoring. *Mikrochim Acta*. 2014;181(5–6):479–91.
5. Zhan S, Wu Y, Wang L, Zhan X, Zhou P. A mini-review on functional nucleic acids-based heavy metal ion detection. *Biosens Bioelectron*. 2016;86:353–68.
6. Torabi S-F, Lu Y. Functional DNA nanomaterials for sensing and imaging in living cells. *Curr Opin Biotechnol*. 2014;28:88–95.
7. Tabarzad M, Jafari M. Trends in the Design and Development of Specific Aptamers Against Peptides and Proteins. *Protein J*. 2016;35(2):81–99.
8. Famulok M, Mayer G, Mayer U. Aptamer Modules as Sensors and Detectors. *Acc Chem Res*. 2011;44(12):1349–58.
9. Kaur H, Bruno JG, Kumar A, Sharma TK. Aptamers in the Therapeutics and Diagnostics Pipelines. *Theranostics*. 2018;8(15):4016–32.
10. Sun H, Zu Y. A Highlight of Recent Advances in Aptamer Technology and Its Application. *Molecules*. 2015 Jun 30;20:11959–80.
11. Zhang W, Feng Q, Chang D, Tram K, Li Y. In vitro selection of RNA-cleaving DNazymes for bacterial detection. *Methods*. 2016;106:66–75.
12. Willner I, Shlyahovsky B, Zayats M, Willner B. DNazymes for sensing, nanobiotechnology and logic gate applications. *Chem Soc Rev*. 2008 Jun;37(6):1153–65.
13. Drummond TG, Hill MG, Barton JK. Electrochemical DNA sensors. *Nat Biotechnol*. 2003;21:1192–9.
14. Liu J, Cao Z, Lu Y. Functional Nucleic Acid Sensors. *Chem Rev*. 2009;109(5):1948–98.
15. Yoshida W, Abe K, Ikebukuro K. Emerging techniques employed in aptamer-based diagnostic tests. *Expert Rev Mol Diagn*. 2014;14(2):143–51.
16. Wang F, Liu X, Willner I. DNA Switches: From Principles to Applications. *Angew Chemie - Int Ed*. 2015;54(4):1098–129.
17. Zhou W, Ding J, Liu J. Theranostic DNazymes. *Theranostics*. 2017;7(74):1010–25.
18. Cho EJ, Lee J-W, Ellington AD. Applications of Aptamers as Sensors. *Annu Rev Anal Chem*. 2009;2:241–64.
19. Wu X, Chen J, Wu M, Xiaojun Zhao J. Aptamers: Active Targeting Ligands for Cancer Diagnosis and Therapy. *Theranostics*. 2015;5(4):322–44.
20. Schlosser K, Li Y. Biologically Inspired Synthetic Enzymes Made from DNA. *Chem Biol*. 2009;16(3):311–22.
21. Huang J, Su X, Li Z. Metal ion detection using functional nucleic acids and nanomaterials. *Biosens Bioelectron*. 2017;96:127–39.
22. Du Y, Dong S. Nucleic acid biosensors: Recent advances and perspectives. *Anal Chem*. 2017;89(1):189–215.
23. Hayat A, Marty JL, Stoytcheva MS, Yu B. Aptamer based electrochemical

- sensors for emerging environmental pollutants. 2014;
24. Wang T, Chen C, Larcher LM, Barrero RA, Veedu RN. Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnol Adv.* 2019;37(1):28–50.
 25. Sun H, Zhu X, Lu PY, Rosato RR, Tan W, Zu Y. Oligonucleotide Aptamers: New Tools for Targeted Cancer Therapy. *Mol Ther Nucleic Acids.* 2014;3:e182.
 26. Nutiu R, Li Y. Aptamers with fluorescence-signaling properties. *Methods.* 2005;37(1):16–25.
 27. Navani NK, Li Y. Nucleic acid aptamers and enzymes as sensors. *Curr Opin Chem Biol.* 2006;10(3):272–81.
 28. Breaker RR, Joyce GF. The expanding view of RNA and DNA function. *Chem Biol.* 2014;21(9):1059–65.
 29. Mok W, Li Y. Recent Progress in Nucleic Acid Aptamer-Based Biosensors and Bioassays. *Sensors.* 2008;8(11):7050–84.
 30. Zhuo Z, Yu Y, Wang M, Li J, Zhang Z, Liu J, et al. Recent Advances in SELEX Technology and Aptamer Applications in Biomedicine. *Int J Mol Sci.* 2017;18(10):e2142.
 31. Shen Y, Mackey G, Rucpich N, Gloster D, Chiuman W, Li Y, et al. Entrapment of Fluorescence Signaling DNA Enzymes in Sol-Gel-Derived Materials for Metal Ion Sensing. *Anal Chem.* 2007;79(9):3494–503.
 32. Kosman J, Juskowiak B. Peroxidase-mimicking DNAzymes for biosensing applications: A review. *Anal Chim Acta.* 2011;707(1–2):7–17.
 33. Mcghee CE, Loh KY, Lu Y. DNAzyme sensors for detection of metal ions in the environment and imaging them in living cells. *Curr Opin Biotechnol.* 2017;45:191–201.
 34. Xu H, Wu Q, Shen H. A DNAzyme sensor based on target-catalyzed hairpin assembly for enzyme-free and non-label single nucleotide polymorphism genotyping. *Talanta.* 2017;167:630–7.
 35. Saberian-Borujeni M, Johari-Ahar M, Hamzeiy H, Barar J, Omidi Y. Nanoscaled aptasensors for multi-analyte sensing. *BioImpacts.* 2014;4(4):205–15.
 36. Chiuman W, Li Y. Simple Fluorescent Sensors Engineered with Catalytic DNA “MgZ” Based on a Non-Classic Allosteric Design. *PLoS One.* 2007;2(11):1224.
 37. Wang DY, Lai BHY, Sen D. A General Strategy for Effector-mediated Control of RNA-cleaving Ribozymes and DNA Enzymes. *J Mol Biol.* 2002;318(1):33–43.
 38. Li Y. As with development of any new technology, what we have demonstrated represents only a small, albeit key, step towards developing a practically useful bacterial detection method. ”. *Future Microbiol.* 2011;973–6.
 39. Tram K. Exploring Synthetic Functional Dna Molecules for Biosensor Development. 2015.
 40. Tram K, Kanda P, Li Y. Lighting Up RNA-cleaving DNAzymes for biosensing. *J Nucleic Acids.* 2012;2012(Figure 1).
 41. Liu Z, Mei SHJ, Brennan JD, Li Y. Assemblage of signaling DNA enzymes with intriguing metal-ion specificities and pH dependences. *J Am Chem Soc.* 2003;125(25):7539–45.
 42. Achenbach JC, Nutiu R, Li Y. Structure-switching allosteric deoxyribozymes.

- Anal Chim Acta. 2005;534(1):41–51.
43. Chiuman W, Li Y. Efficient signaling platforms built from a small catalytic DNA and doubly labeled fluorogenic substrates. *Nucleic Acids Res.* 2007;35(2):401–5.
 44. Aguirre SD, Ali MM, Salena BJ, Li Y. A Sensitive DNA Enzyme-Based Fluorescent Assay for Bacterial Detection. *Biomolecules.* 2013;3(3):563–77.
 45. Qu L, Ali MM, Aguirre SD, Liu H, Jiang Y, Li Y. Examination of Bacterial Inhibition Using a Catalytic DNA. *PLoS One.* 2014;9(12):e115640.
 46. Li Y, Geyer CR, Sen D. Recognition of anionic porphyrins by DNA aptamers. *Biochemistry.* 1996;35(21):6911–22.
 47. Wang W, Billen LP, Li Y. Sequence Diversity, Metal Specificity, and Catalytic Proficiency of Metal-Dependent Phosphorylating DNA Enzymes. *Chem Biol.* 2002;9(4):507–17.
 48. Liu M, Zhang Q, Chang D, Gu J, Brennan JD, Li Y. A DNAzyme Feedback Amplification Strategy for Biosensing. *Angew Chemie - Int Ed.* 2017;56(22):6142–6.
 49. Monsur Ali M, Aguirre SD, Lazim H, Li Y, Ali MM, Aguirre SD, et al. Fluorogenic DNAzyme probes as bacterial indicators. *Angew Chemie - Int Ed.* 2011;50(16):3751–4.
 50. Shen Y, Brennan JD, Li Y. Characterizing the Secondary Structure and Identifying Functionally Essential Nucleotides of pH6DZ1, a Fluorescence-Signaling and RNA-Cleaving Deoxyribozyme. *Biochemistry.* 2005;44.
 51. Kandadai SA, Mok WWK, Ali MM, Li Y. Characterization of an RNA-Cleaving Deoxyribozyme with Optimal Activity at pH 5. *Biochemistry.* 2009;48(31):7383–91.
 52. He S, Qu L, Shen Z, Tan Y, Zeng M, Liu F, et al. Highly Specific Recognition of Breast Tumors by an RNA-Cleaving Fluorogenic DNAzyme Probe. *Anal Chem.* 2015;87(1):569–77.
 53. Tram K, Xia J, Gysbers R, Li Y. An Efficient Catalytic DNA that Cleaves L-RNA. *PLoS One.* 2015;10(5):1–14.
 54. Gysbers R, Tram K, Gu J, Li Y. Evolution of an Enzyme from a Noncatalytic Nucleic Acid Sequence. *Sci Rep.* 2015;5:11405.
 55. Shen Z, Wu Z, Chang D, Zhang W, Tram K, Lee C, et al. A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen. *Angew Chemie - Int Ed.* 2016;55(7):2431–4.
 56. Chan L, Tram K, Gysbers R, Gu J, Li Y. Sequence Mutation and Structural Alteration Transform a Noncatalytic DNA Sequence into an Efficient RNA-Cleaving DNAzyme. *J Mol Evol.* 2015;81(5–6):245–53.
 57. Mei SHJ, Liu Z, Brennan JD, Li Y. An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. *J Am Chem Soc.* 2003;125(2):412–20.
 58. Ellington a D, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature.* 1990;346(6287):818–22.
 59. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (80-).* 1990;249(4968):505–10.
 60. Rychlik MP, Chon H, Cerritelli SM, Klimek P, Crouch RJ, Nowotny M. Crystal structures of rnaase h2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. *Mol Cell.*

- 2010;40(4):658–70.
61. Kim S, Shi H, Lee DK, Lis JT. Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res.* 2003;13(7):1955–61.
 62. Nazarenko IA, Uhlenbeck OC. Defining a smaller RNA substrate for elongation factor Tu. *Biochemistry.* 1995;34(8):2545–52.
 63. Altman S. Ribonuclease P. *Philos Trans R Soc B.* 2011;366(1580):2936–41.
 64. Pan T. Novel RNA Substrates for the Ribozyme from *Bacillus subtilis* Ribonuclease P Identified by in Vitro Selection. *Biochemistry.* 1995;34(26):8458–64.
 65. Katz K, Thampi N, Taylor G, Frenette C, Kibsey P, Suh K, et al. Canadian Nosocomial Infection Surveillance Program (CNISP) 2018 Surveillance for *Clostridium difficile* infection (CDI). 2017.
 66. Dethlefsen L, Huse S, Sogin ML, Relman DA. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol.* 6(11):280.
 67. Larson HE, Borriello SP. Quantitative Study of Antibiotic-Induced Susceptibility to *Clostridium difficile* Enterocolitis in Hamsters. *Antimicrob Agents Chemother.* 1990;34(7):1348–53.
 68. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, et al. Decreased Diversity of the Fecal Microbiome in Recurrent *Clostridium difficile*-Associated Diarrhea. *J Infect Dis.* 2008;197(3):435–8.
 69. Merrigan M, Sambol S, Johnson S, Gerding DN. Susceptibility of hamsters to human pathogenic *Clostridium difficile* strain B1 following clindamycin, ampicillin or ceftriaxone administration. *Anaerobe.* 2003;9(2):91–5.
 70. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol.* 2009;7(7):526–36.
 71. Hall AJ, Curns AT, McDonald LC, Parashar UD, Lopman BA. The roles of *Clostridium difficile* and norovirus among gastroenteritis-associated deaths in the United States, 1999-2007. 2012;
 72. Dubberke ER, Olsen MA. Burden of *Clostridium difficile* on the Healthcare System (Suppl 2) • Dubberke and Olsen. *Clin Infect Dis.* 2012;55(S2):S88-92.
 73. Magill SS, Edwards JR, Stat M, Bamberg W, Beldavs ZG, Dumyati G, et al. Multistate Point-Prevalence Survey of Health Care-Associated Infections. *N Engl J Med.* 2014;370(13):1198–208.
 74. Miller BA, Chen LF, Sexton DJ, Anderson DJ. Comparison of the Burdens of Hospital-Onset, Healthcare Facility-Associated *Clostridium difficile* Infection and of Healthcare-Associated Infection due to Methicillin-Resistant *Staphylococcus aureus* in Community Hospitals. *Infect Control Hosp Epidemiol.* 2011;32(4):387–90.
 75. Wang L, Wang R, Chen F, Jiang T, Wang H, Slavik M, et al. QCM-based aptamer selection and detection of *Salmonella typhimurium*. *Food Chem.* 2017;221:776–82.
 76. Setlem K, Mondal B, Ramlal S, Kingston J. Immuno Affinity SELEX for Simple, Rapid, and Cost-Effective Aptamer Enrichment and Identification against Aflatoxin B1. *Front Microbiol.* 2016;7:1909.
 77. Almasi F, Latif S, Gargari M, Bitaraf F, Rasoulinejad S. Development of a Single Stranded DNA Aptamer as a Molecular Probe for LNCap Cells Using

- Cell-SELEX. *Avicenna J Med Biotechnol.* 2016;8(3):104–11.
78. Jenison RD, Gill SC, Pardi A, Polisky B. High-Resolution Molecular Discrimination by RNA. *Science.* 1994;263(5152):1425–9.
 79. Arraiano CM, Andrade JM, Domingues S, Guinote IB, Malecki M, Matos RG, et al. The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev.* 2010;34(5):883–923.

Chapter 3

Serendipitous Discovery of a Guanine-rich DNA Molecule with a Highly Stable Structure in Urea

3.1 Author's Preface

This study reports the serendipitous discovery of an ultra-stable G-rich structure, which resists the denaturation by denaturing polyacrylamide gel containing 7 M urea, a widely used denaturant for DNA and RNA structures. This finding suggests that our knowledge about the structural and functional boundaries of nucleic acids may be still limited. We are excited about this discovery and looking forward to future unconventional discoveries of functional nucleic acids. We expect such eye-widening discoveries will not only expand our knowledge, but also lead to many superlative applications of nucleic acids.

This work has been published and is presented in the published format (full citation listed below). I am the first author of this publication. I took the leading role in project design, experimentation and data analysis. Dr. Meng Liu has provided significant insights and suggestions during the process of this work. I conducted all the experiments and summarized all the results. The manuscript was written with assistance from Dr. Yingfu Li and other co-authors. Its supplementary information is included in Appendix 1.

Zhang, W., Liu, M., Lee, C., Salena, B. J. & Li, Y. (2018). Serendipitous Discovery of a Guanine-rich DNA Molecule with Highly Stable Structure in Urea. *Scientific Report*. 8(1):1935. DOI:10.1038/s41598-018-20248-w

3.2 Abstract

We have made an accidental discovery of an unusual, single-stranded, guanine-rich DNA molecule that is capable of adopting a folded structure in 7 M urea (7MU) known to denature nucleic acid structures. The folding of this molecule requires Na^+ and Mg^{2+} and the folded structure remains stable when subjected to denaturing (7MU) polyacrylamide gel electrophoresis. Results from sequence mutagenesis, DNA methylation, and circular dichroism spectroscopy studies suggest that this molecule adopts an intramolecular guanine-quadruplex structure with 5 layers of guanine tetrads. Our finding indicates that DNA has the ability to create extremely stable structural folds despite its limited chemical repertoire, making it possible to develop DNA-based systems for unconventional applications.

3.3 Introduction

DNA is a rather simple polymer from the chemical functionality perspective, especially when compared to proteins. This is because DNA is built with four chemically similar nucleobases distributed over a backbone of sugar-phosphodiester. And yet, many man-made DNA molecules have been produced, through the technique of *in vitro* selection (1, 2), to recognize a broad range of molecular targets or catalyze various chemical transformations (3-12). However, the majority of functional DNA molecules have been derived to perform tasks under conditions conducive to nucleic acid structure folding, such as neutral pH and amiable temperature. Previous examples of functional DNA molecules that are active under challenging reaction conditions include RNA-cleaving DNazymes working at high acidity (21-24) or high temperature (25), although these DNazymes exhibit much

reduced catalytic efficiency than their counterparts derived to work at neutral pH and normal temperature (25). Nevertheless, it is largely unknown if single-stranded DNA can create intricate structures under structure-disrupting conditions.

Urea at high concentrations represents a disruptive condition for nucleic acid structures (26, 17). Because urea can act as hydrogen-bond donor and acceptor, it can easily denature structures of nucleic acids. In fact, 6–8 M urea is the key component for denaturing polyacrylamide gel electrophoresis (dPAGE) widely used to separate DNA oligonucleotides by size (28, 29). In this technique, DNA molecules are completely denatured and migrate as unfolded linear polymers through the polyacrylamide gel matrix. To our great surprise, we came across a DNA molecule that can form a folded structure in 7 M urea (7MU), which remains stable during the process of 7MU-dPAGE.

3.4 Results and Discussion

The original goal of this study was to use *in vitro* selection technique to isolate RNA-cleaving DNazymes to function under the condition of 20 mM Tris-HCl, pH 8.5, 50 mM NaCl, and 10 mM MgCl₂ (denoted 1× Selection Buffer, or 1× SB), from a random-sequence synthetic DNA pool. The DNA pool was made of ~10¹⁴ 98-nt DNA molecules (nt: nucleotide) with 43-nt constant sequence at the 5'-end, followed by 40-nt random sequence and 15-nt constant sequence (Fig. 3-1a). Note that the 5'-sequence contains an adenosine ribonucleotide at 14th position (R in Fig. 3-1a) as the cleavage site, which is surrounded by a pair of thymine deoxyribonucleotides modified with a fluorescein (F) and the dabcyI quencher (Q)—a design intended for making DNazymes for biosensing application (30, 31). The selection strategy, depicted in Fig. S1, is straightforward: RNA-cleaving DNA sequences produce two

products, 14-nt P1 and 84-nt P2, which is the desired product because it contains the catalytic DNA sequence. Because of large differences in size, P2 can easily be separated from both P1 and the uncleaved molecules using 7MU-dPAGE. Purified P2 is first multiplied via DNA amplification and then subjected to the next cycle of cleavage, product separation and DNA amplification. This procedure is repeated until a strong cleavage activity is observed.

The described selection strategy has been successfully used to derive RNA-cleaving DNazymes in our previous studies (21, 30-33). As expected, a strong P2 DNA band on 7MU-dPAGE gel was observed after 16 cycles of selection. The DNA pool was then subjected to cloning and sequencing. 33 sequences were obtained, which can be grouped into 3 sequence classes (Table [S1](#)). A representative sequence from each class was chosen for further characterization (Fig. 3-[1b](#)). These sequences are named UD1, UD2, and UD3.

The first evidence suggesting that these DNA molecules were not RNA-cleaving DNazymes came from the experiment depicted in Fig. 3-[1c](#). For this experiment, each UD was made to have a ^{32}P -labeled phosphodiester linkage between the 28th and 29th nucleotides in addition to the fluorescein label at the 13th position. Therefore, the RNA cleavage reaction should produce a fluorescent P1 and a radioactive P2, as illustrated in the inserted reaction scheme in Fig. 3-[1c](#). The uncleaved UD should be detectable by both fluorimaging (gel image on the left) and radioimaging (image on the right), while P1 and P2 should be detectable only by fluorimaging and radioimaging, respectively. However, P1 was not detected at all whereas P2-like DNA bands appeared in both images. These observations strongly suggest that we obtained novel DNA sequences that are able to fold into stable structures in 7MU. The folded DNA molecules are more compact than unfolded DNA

sequences, and thus migrate faster on 7MU-dPAGE. These molecules were selected simply because the folded structure (albeit 98-nt in length) has a gel mobility that is similar to the linear 84-nt DNA molecule used as the gel excision marker.

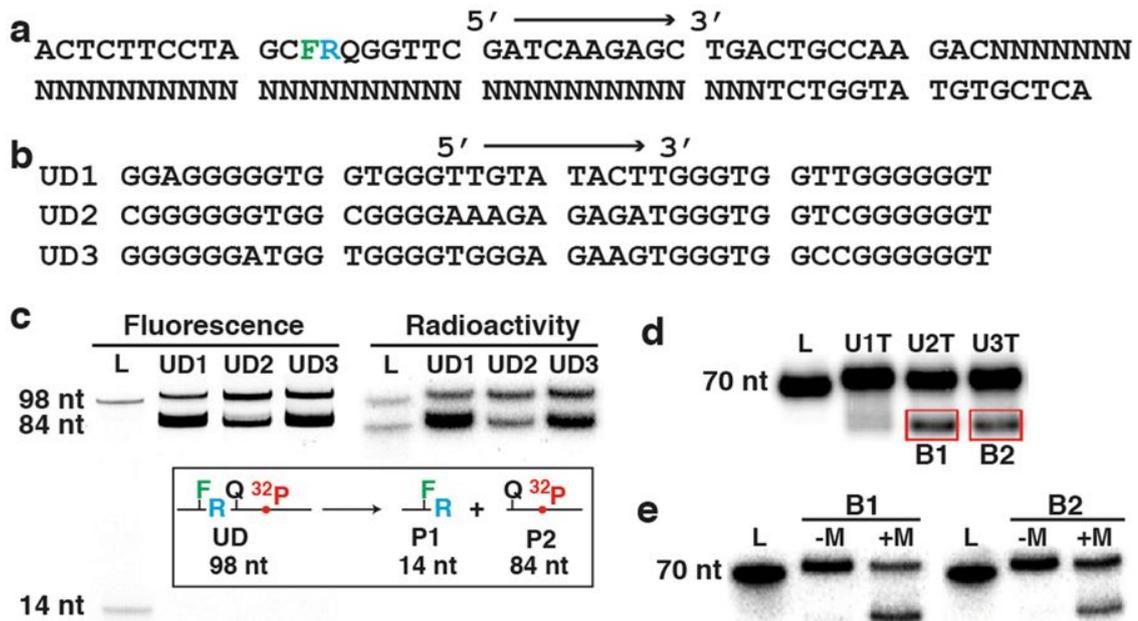


Figure 3-1. Discovery of DNA sequences with urea-resistant structures. (a) The sequence of the DNA library used for in vitro selection. R: adenosine ribonucleotide; F: fluorescein-labeled dT; Q: dabcyl-labeled dT; N: mixture of ACGT (25% each). (b) The random-sequence regions of UD1, UD2 and UD3. (c) Fluorimager and phosphorimager analysis of a 7MU-dPAGE gel conducted to analyze the cleavage reaction of UD1–3. Insert: The cleavage reaction with 14-nt P1 and 84-nt P2 as the cleavage product. L: DNA ladder lane. (d) 7MU-dPAGE analysis of U1T, U2T and U3T (truncated UD1, UD2 and UD3 with the removal of the first 28 nucleotides). (e) 7MU-dPAGE analysis of the boxed DNA bands from panel d. +M and -M: with and without 50 mM NaCl and 5 mM MgCl₂, respectively. Experimental details: Panel c: 150 pmol of synthetic oligonucleotides, UD1, UD2 and UD3 were phosphorylated with 10 μCi of [γ-³²P]ATP by PNK. The reactions were incubated at 37°C for 15 min, followed by adding 100 nmol of nonradioactive ATP and another 20 min of incubation. They were then ligated to 150 pmol S1 in the presence of 150 pmol T1 and 5 units of T4 ligase. Ligated products were then purified by dPAGE. 5 pmol of full-length UD1, UD2 and UD3 for were used for the experiment. For each DNA, 5 pmol in 5 μL were mixed with 5 μL of 2' SB, and then incubated at RT for 15 min. 10 μL of 2' dGLB were then introduced. The resultant mixture was heated to 90°C and cooled at RT for 15 min. 10 μL (out of 20 μL) were taken for dPAGE analysis. Panel d: 5 pmol of U1T, U2T and U3T in 5 μL were mixed with 5 μL of 2' SB, and then incubated at RT for 15 min. 10 μL of 2' dGLB were then introduced. The resultant mixture was heated to 90°C and cooled at RT for 15 min. 10 μL (out of 20 μL) were taken for dPAGE analysis. Panel e: the boxed DNA bands from panel c were excised and the DNA eluted. The DNA was dissolved in 10 μL of water. 5 μL was subjected to the same procedure as described for panel d. +M: the reaction mixture contained NaCl and MgCl₂; -M: the two salts were omitted.

To provide additional evidence, we performed an experiment using altered UD1–3 (named U1T, U2T and U3T; Fig. 3-1d) where the first 28 nucleotides (including the RNA cleavage site) were removed. A faster migrating DNA band appeared for each shortened DNA, particularly for U2T and U3T.

The folded DNA band from U2T and U3T, labeled B1 and B2, were excised from the gel, eluded, denatured by heat, and refolded under the *in vitro* selection condition. When they were subjected to dPAGE analysis, two DNA bands were observed again. This experiment further verified that UD1, UD2 and UD3 are not RNA-cleaving DNazymes. More importantly, the appearance of the fast-migrating DNA band was dependent on the presence of 50 mM NaCl and 5 mM MgCl₂ (“+M” lanes, Fig. 3-[1e](#)).

We realized by this time that the experimental procedure used for *in vitro* selection in this study had two key deviations from the protocol previously used by us for isolating RNA-cleaving DNazymes. Our standard practice has a step of quenching the cleavage reaction by adding sufficient amount of EDTA to chelate divalent metal ions, followed by DNA recovery using ethanol precipitation. The DNA is then dissolved in 1× dPAGE gel loading buffer containing 7MU (1× dGLB), heated at 90 °C for 5 min and cooled at room temperature for 15 min. The mixture is then subjected to 7MU-dPAGE to purify the cleavage product. However, in the current study, the cleavage reaction mixture was directly combined with 2× dGLB, followed by heating, cooling and 7MU-dPAGE purification as usual. In other words, the altered protocol omitted EDTA-chelation and ethanol-precipitation steps. These omissions resulted in a DNA mixture containing 50 mM NaCl and 5 mM MgCl₂ in addition to 7MU prior to gel loading. We believe that the presence of metal ions in the loading mixture provided a chance for UD sequences to fold into a tight structure in the presence of 7MU and get selected via isolation on 7MU-dPAGE.

To probe into this hypothesis, we carried out an experiment to examine the concentration effect of Na⁺ and Mg²⁺ on the folding of U2T. Sodium itself at high concentrations (375 mM or above; Fig. 3-[2a,d](#)) can induce a high level (~85%) of

structural formation of U2T. Magnesium was less effective: only ~20% U2T were folded when 100 mM MgCl₂ was provided (Fig. 3-2b,d). However, a synergistic effect was seen when Na⁺ and Mg²⁺ were used simultaneously: in the presence of 5 MgCl₂, ~85% U2T adopted folded structure when the sodium concentration reached 150 mM. K⁺, Li⁺ and NH₄⁺ were also found to support the structural formation (Fig. S2), and the effectiveness of the tested monovalent cations follows the order of Na⁺ > K⁺ > NH₄⁺ > Li⁺.

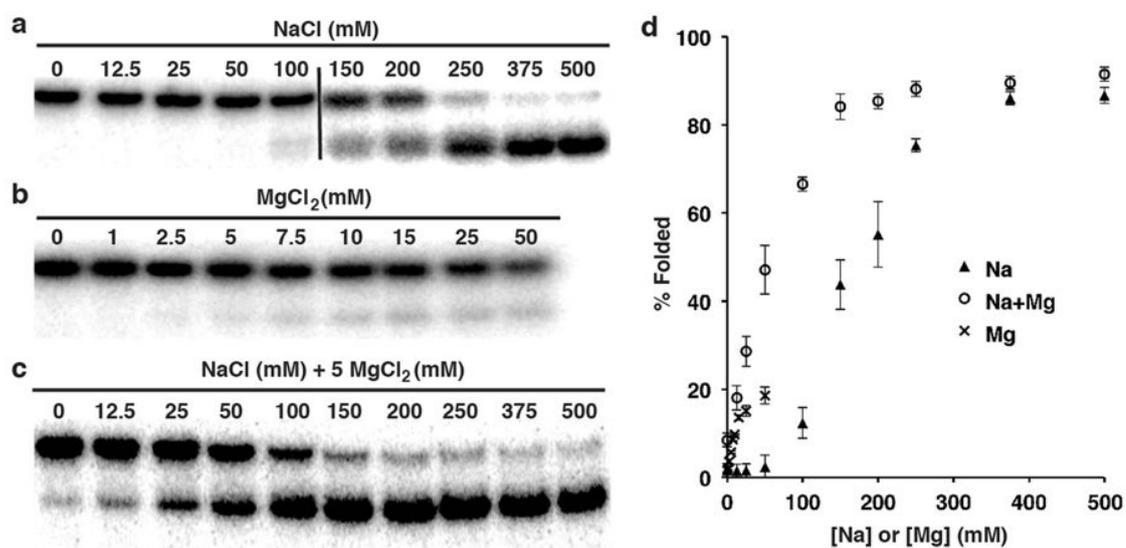


Figure 3-2. Metal ion requirements for the folding of U2T. (a) Effect by Na⁺ (in the absence of Mg²⁺). (b) Effect by Mg²⁺ (in the absence of Na⁺). (c) Effect by Mg²⁺ and Na⁺. Each reaction mixture contained 44.5 mM Tris-borate, 10 mM HEPES, 7 M urea, and the specified metal ions. (d) % Folded structure vs. NaCl concentration. The mixture was heated at 90 °C for 5 min and left standing at room temperature for 15 min before loading. Experimental details: 5 pmol of U2T in 5 μL were mixed with 5 μL of various 4' folding buffers containing 40 mM HEPES (pH 7.5), and 4' concentrations of metal ions listed for each lane. Then 10 μL of 2' dGLB were introduced and the resultant mixture was heated to 90°C and cooled at RT for 15 min. 10 μL (out of 20 μL) were taken for dPAGE analysis.

We next investigated the time for folding. U2T samples containing 150 mM NaCl, 5 mM MgCl₂, and 7MU were heated at 90 °C for 5 min and then incubated at room temperature for times specified in Fig. 3-3a prior to dPAGE analysis. Note that each folding reaction was started at a different starting time so that all the reactions reached the same endpoint for gel loading. The graph presented in Fig. 3-3b indicates

that ~40% U2T adopted the folded structure in 5 min, 70% in 60 min, and additional 10% molecules were able to fold in the next 3 h. The last lane was a control where a sample was loaded immediately following the heating step, and the lack of fast-migrating DNA band in this lane indicates that U2T was unable to fold within the gel, most probably due to the absence of metal ions in the gel. The folded structure, however, remained stable within the gel.

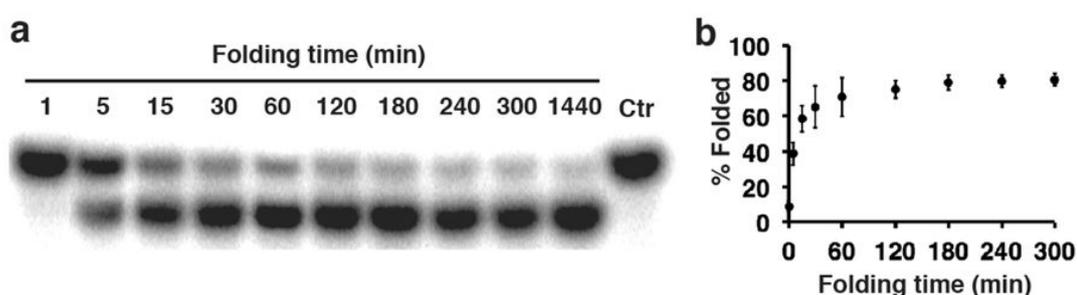


Figure 3-3. Time for U2T folding. (a) 7MU-dPAGE analysis. (b) % Folded structure vs. folding time. U2T was heated at 90 °C for 5 min and then incubated at room temperature for the specified time before 7MU-dPAGE analysis. Ctr: the reaction mixture was immediately loaded following heat denaturation. All the reaction mixtures contained 7MU, 5 mM MgCl₂, and 150 NaCl. Experimental details: 5 pmol of U2T in 5 μ L were mixed with 5 μ L of 4' FB, and 10 μ L of 2' dGLB, then heated at 90°C for 5 min and cooled at RT for the specified time before dPAGE analysis. The Ctr lane was an exception where the reaction mixture was immediately loaded following the heat denaturation step. Each reaction was started at different time points so that all the reactions ended at the same time. For example, the 4 h reaction was started one hour after the start of 5 h reaction. 5 μ L of each reaction mixture were used for the dPAGE analysis.

Sequence truncation experiment provided in Fig. [S3](#) indicates that nucleotides outside of the random region can be completely removed. The 40-nt UD2 is named U2R. It is highly rich in guanine (G; 65%) and such sequences tend to form G-quadruplex structures comprising stacks of G-quartets, planar arrangement of four guanines bonded with 8 hydrogen bonds (34-40).

Footprinting experiment with dimethyl sulfate (DMS) was then performed with U2T and the data is provided in Fig. [S4](#). This technique can detect the sensitivity of N7 atoms of guanines in G-quartets to methylation (so called “methylation interference”). It was found that 20 of 26 guanines exhibited significant methylation interference (green G residues in Fig. [3-4a](#)), indicating that these guanine bases play

important roles in UD2's structural folding. Interestingly, these guanines are conserved in U1R and U3R (Fig. 3-4a) and distributed into six G-stretches in the following arrangement: G₅N₁₋₂G₂N₁G₃N₁₀₋₁₁G₃N₁G₂N₂G₅ (N stands for A, C or T). This finding suggests that these 3 selected DNA molecules are sequence relatives.

The folding property of several U2R mutants (Fig. 3-4b) were examined next. Removing 3 internal, non-conserved guanine nucleotides (U2R1) had no effect on folding. However, deleting (U2R2) or mutating (U2R3) conserved guanine nucleotides was totally disruptive. As expected, the 5' C and 3' T (U2R4) nucleotides had no effect.

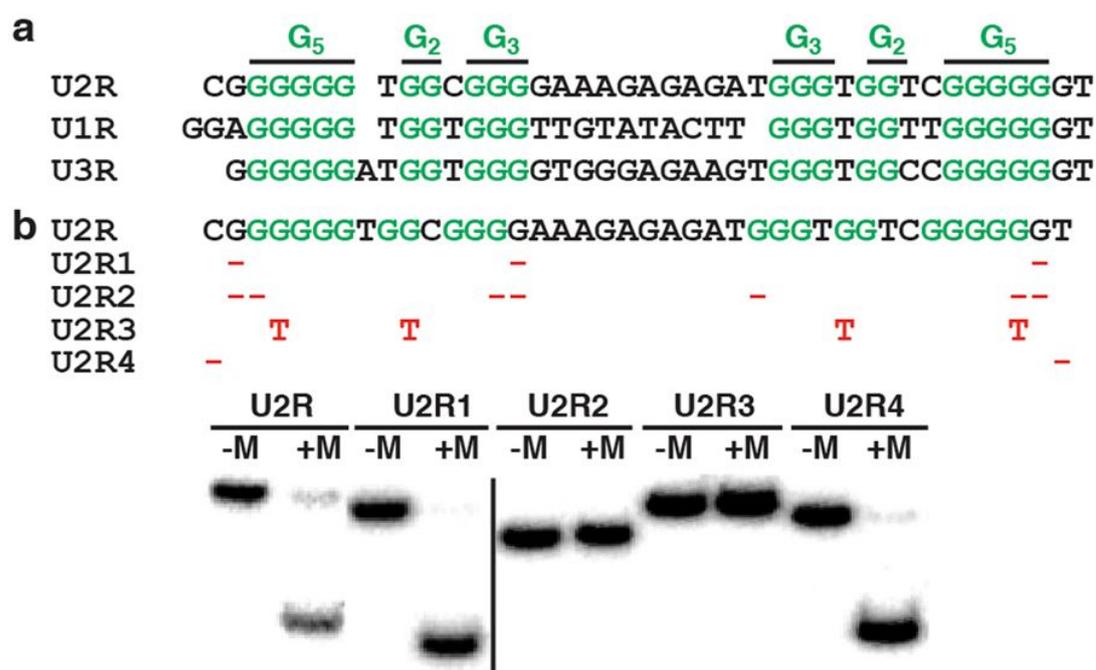


Figure 3-4. Testing variant sequences of U2R. (a) Sequence alignment of U2R, U1R and U3R. Guanines with significant methylation interference are shown in green. Only altered nucleotides are shown. -: nucleotide deleted; red T: guanine mutated to thymine. (b) Analysis of folding property of U2R mutants in 7MU with (+) and without (-) of 5 mM MgCl₂, and 150 NaCl. Two different gels, separated by the black dividing line, were grouped. Experimental details: 5 pmol of each of listed U2R mutants in 5 μL were mixed with 5 μL of 4' FB and 10 μL of 2' dGLB, and heated at 90°C for 5 min, and cooled at RT for 15 min. 5 μL were taken for dPAGE analysis.

The conserved sequence motif has two G₅ elements and two G₂+G₃ elements each with a pyrimidine nucleotide insertion (pink C and T, Fig. 3-5a). This observation prompted us to examine whether these two nucleotides are essential to

the structural folding, particularly considering their removal would produce two additional G₅ elements to allow the DNA sequence to form a quadruplex structure with 5 layers of G-tetrads. Surprisingly, these two nucleotides are indispensable to the structural folding (U2R5, Fig. 3-[5a](#))—this point will be further discussed below.

Circular dichroism spectroscopy (CD) has been widely used to examine guanine quadruplex structures, particularly for identifying different G-quadruplex folds (36, 41-45). The CD spectrum of U2R1 in the presence of urea (unfilled circles, Fig. 3-[5b](#)) contains a positive peak at 260 nm and a negative peak at 240 nm, both of which were not seen in the spectrum of U2R5 (unfilled triangles). The CD spectrum of U2R1 is indicative of a parallel G-quadruplex fold extensively reported in the literature for other G-quadruplex forming DNA and RNA sequences, strongly suggesting that U2R1 has G-quadruplex made of exclusively parallel DNA strands (46-48). The parallel chain directionality of U2R1 was further tested with protoporphyrin (PPIX), which has been shown to selectively bind to parallel G-quadruplexes and produce fluorescence enhancement (49). Indeed, strong fluorescence increases were observed when PPIX was incubated with U2R1 both in the absence and presence of 7MU (Fig. [S5a](#)).

Based on the above information, we propose a structural model, which is depicted in Fig. 3-[5c](#). The proposed structure consists of two continuous stacks of tetrads (top and bottom two-tiers) sandwiching a middle stack with two backbone openings. The model not only places all 20 conserved guanines into 5-layer G-quartets, but the nature of all parallel strands is also consistent with the CD data (Fig. 3-[5b](#)) and the results from the PPIX binding experiment (Fig. [S5a](#)). Furthermore, the model explains the indispensability of the aforementioned pyrimidine nucleotides because they are needed for creating all-parallel strand

orientations. The structure also has a long linking loop (10 nt) connecting two distal guanines located on the top and bottom tetrads, respectively. We found that nucleotide identities are not important as they can be changed to thymidines (U2R7, Fig. 3-5d). However, the number of nucleotides in this loop has pronounced impact on formation of the urea-resistant structure: while 8 and more nucleotides (U2R6, 11 thymidines; U2R7, 7 thymidines) support the structure formation, shortened loops (5 nt in U2R8 and 2 nt in U2R9) are completely ineffective. These findings are consistent with the proposed structural model.

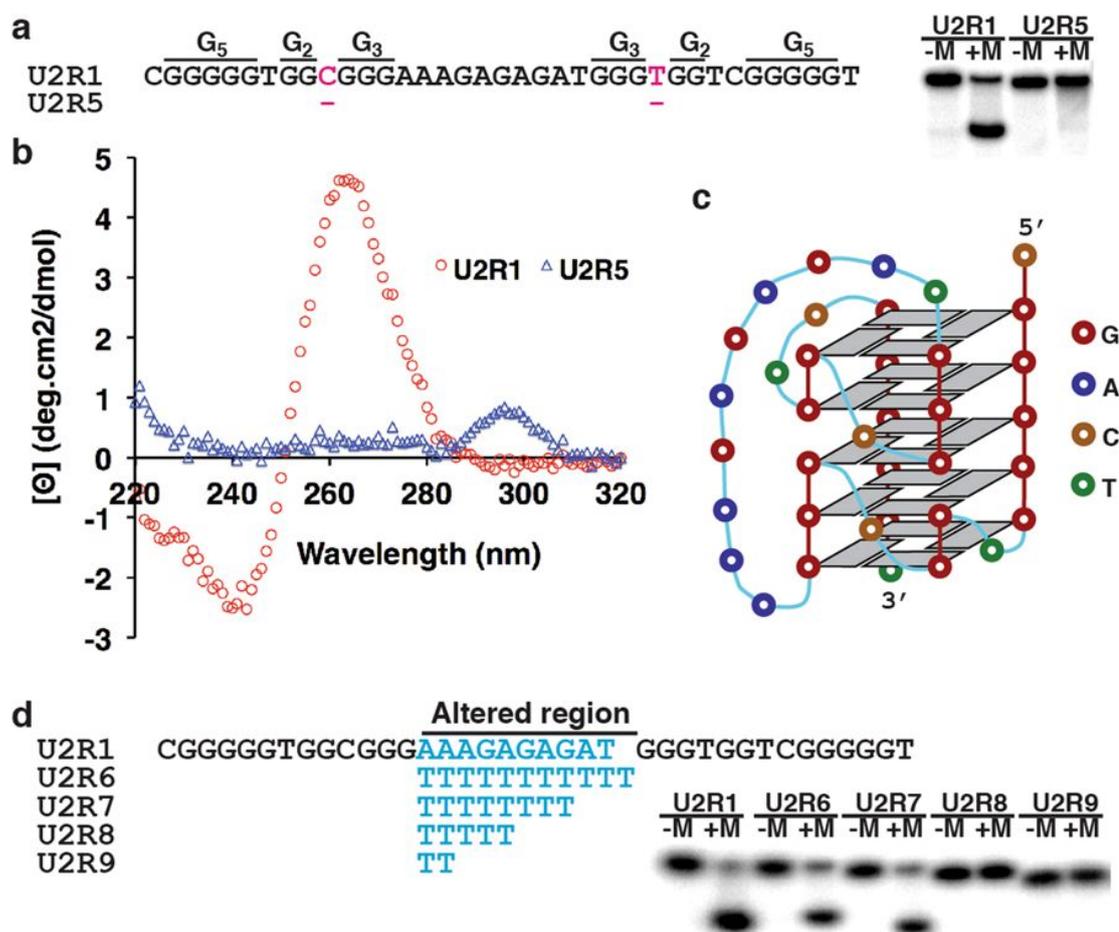


Figure 3-5. A structure model for U2R1. (a) Analysis of folding property of U2R1 and its mutant U2R5 in the presence of 7M urea. (b) Circular dichroism spectra obtained with U2R1 and U2R5 in 7M urea, 150 mM NaCl and 5 mM MgCl₂. (c) Proposed structure model. (d) Analysis of folding property of U2R1 mutants altered at the long linking loop. +M and -M in panels (a,d): with and without 150 mM NaCl and 5 mM MgCl₂. Experimental details: Panel b: 5 pmol of U2R1 or U2R5 in 5 μL were mixed with 5 μL of 4' FB and 10 μL of 2' dGLB, heated at 90°C for 5 min, and cooled at RT for 15 min. 5 μL were taken for dPAGE analysis. Panel c: the same procedure for panel b was used, however, the dPAGE was substituted with native PAGE. Panel d: All samples (500 μL each) contained 4 μM U2R1 or U2R5, 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM MgCl₂, and 7 M urea. Each sample was heated to 90°C for 5 min and

incubated at RT for 20 h. 350 μ L was placed in a 0.1 cm quartz cuvette. CD studies were carried out using an AVIV model 410 CD spectrometer (Lakewood, NJ, USA) as per the manufacturer's instructions. The samples were scanned from 320 to 220 nm. All shown spectra are the average of 2 individual scans.

We examined the thermal stability of U2R1 by measuring their melting points (T_m values) in the absence and presence of 7MU and the data are shown in Fig. S5b. U2R1 has a melting point of 68 $^{\circ}$ C and 59 $^{\circ}$ C in the absence and presence of 7MU, and more importantly, its T_m decreases only by 9 $^{\circ}$ C in the presence of 7MU. These results are consistent with a highly stable structure associated with U2R1 even in the presence of 7MU.

Finally, we tested whether the RNA version of U2R1 (named RNA-U2R1) could fold into a similar urea-resistant structure. Interestingly, RNA-U2R1 behaved very differently from U2R1 as it did not produce a defined fast-migrating band on the 7MU-dPAGE (Fig. S5c). This particular observation suggests that the stable structure adopted in 7MU by U2R1 does not straightforwardly translate into a similar stable structure by the RNA with the same sequence. However, we believe RNA molecules with similar properties do exist and can be isolated from random-sequence RNA pools.

3.5 Conclusion

In summary, we have accidentally discovered a remarkable DNA molecule that can adopt a folded structure in the presence of 7 M urea, which is known to be a strong denaturant for nucleic acid structures. This DNA molecule is highly rich in guanine and several lines of evidence suggest that this molecule forms an intricate 5-tiered guanine-quadruplex with all-parallel strand orientations. To our knowledge, such a structural arrangement has never been described before. The formation of the structure is contingent on the presence of metal ions. Overall, our work demonstrates

that DNA, despite its chemical simplicity, is able to create highly stable structural folds in a structurally disruptive environment. This finding opens up the possibility of developing DNA-based systems for unconventional applications. For example, it is conceivable that DNA aptamers or ligand-responsive DNazymes for biosensing applications can be isolated to function under denaturing conditions designed to overcome issues such as non-specific binding or degradation by nucleases in biological samples. Our finding may also have implications for origin of life. It has been hypothesized that life on Earth may have emerged from an “RNA world” where RNA played the dual roles of being the hereditary material and carrying enzymatic reactions (50-53). It is possible that the early RNA catalysts may have to deal with structure-disrupting conditions imposed by urea-like denaturants. Furthermore, these un-supportive conditions might also facilitate the replication of nucleic acid enzymes as they can help denature the replicated duplexes at elevated temperature and allow the regeneration of functional enzymes under denaturing conditions.

3.6. Methods

3.6.1 Enzymes, chemicals, and other materials

T4 DNA ligase and T4 polynucleotide kinase (PNK) were purchased from Thermo Scientific (Ottawa, ON, Canada). [γ - ^{32}P]dATP was obtained from Perkin Elmer (Woodbridge, ON, Canada). Urea (ultrapure) and 40% polyacrylamide solution (29:1) were acquired from BioShop Canada (Burlington, ON, Canada). The water used was purified via Milli-Q Synthesis A10 water purifier. All other chemicals were purchased from Sigma-Aldrich (Oakville, ON, Canada). 10 \times TBE (Tris-borate EDTA) (1 L) was made of 108 g Tris-base (0.89 M), 55 g boric acid (0.89 M), 20 mL of 0.5 M ethylenediaminetetraacetic acid (EDTA; pH 8.0; 10 mM). 2 \times native gel-

loading buffer (per 100 mL) was made with 20 g sucrose, 10 mL of 10× TBE, 1 mL of 10% (w/v) SDS (sodium dodecyl sulphate), 25 mg bromophenol blue, and 25 mg xylene cyanol FF. 2× denaturing gel-loading buffer was made with the same recipe, with the addition of 84.1 g urea (14 M). Note that the concentration of Tris-boric acid in the 2× gel-loading buffer was 89 mM. 10% denaturing polyacrylamide gel stock was made of 250 mL of 40% polyacrylamide solution (29:1), 100 mL of 10× TBE, and 425 g urea.

3.6.2 Synthesis and purification of oligonucleotides

DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA), prepared using automated synthesis using standard phosphoramidite chemistry. Each oligonucleotide was purified using 10% denaturing polyacrylamide gel electrophoresis (dPAGE) containing 7 M urea. Note that for the purification of urea-resistant DNA molecules featured in this study, DNA samples in 1× denaturing gel loading buffer at 90 °C for 5 min and then immediately loaded onto dPAGE gel. This procedure was adopted to minimize the structure formation.

3.6.3 ³²P-labelling of oligonucleotides

1 μL of [γ -³²P]ATP (10 μCi) was used to label ~150 pmol of DNA or RNA in 50 μL of 1× PNK buffer A containing 5 units of PNK. After incubation at 37 °C for 30 min, PNK was inactivated by heating at 90 °C for 5 min. The phosphorylated DNA was first concentrated by ethanol precipitation and then purified by 10% dPAGE purification.

3.6.4 In vitro selection procedures

In vitro selection was conducted using a similar protocol that we previously described (54). The *in vitro* selection schematic and all the DNA sequences are provided in Fig. [S1](#). 1000 pmol of DL1 were used as the initial library. The DNA molecules in this pool were first labeled with ^{32}P at the 5'-end in the presence of 10 units (U) of T4 PNK, 10 μCi of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$, and 1 \times T4 PNK buffer A (using the 10 \times buffer supplied by the vendor) for 20 min at 37 $^{\circ}\text{C}$ in a reaction volume of 100 μL . This was followed by the addition of non-radioactive ATP to a final concentration of 1 mM and further incubation at 37 $^{\circ}\text{C}$ to ensure complete phosphorylation. The reaction was stopped by heating at 90 $^{\circ}\text{C}$ for 5 min. Upon cooling to room temperature (~ 23 $^{\circ}\text{C}$), 1000 pmol of S1 and 1000 pmol of T1 were added. The reaction mixture was then heated to 90 $^{\circ}\text{C}$ for 1 min and cooled to RT. 10 units of T4 DNA ligase and 25 μL of 10 \times T4 DNA ligase buffer (supplied by the vendor) were added to the reaction mixture (total reaction volume: 250 μL). The ligation reaction was carried out at room temperature for 2 h. The DNA in the mixture was precipitated by ethanol and the ligated product was purified by 10% dPAGE (54).

The purified DNA above was suspended in 25 μL of H_2O , heated to 90 $^{\circ}\text{C}$ for 30 s and cooled to room temperature over 15 min. The cleavage reaction was initiated by the addition of 25 μL of 2 \times selection buffer (40 mM Tris-HCl, pH 8.5, 100 mM NaCl, and 20 mM MgCl_2), and the reaction mixture was incubated at room temperature for 1 h. The reaction solution was then mixed with 50 μL of 2 \times denaturing gel loading buffer, which was heated at 90 $^{\circ}\text{C}$ for 5 min and cooled to RT for 15 min (54). Note that the metal ion concentrations were: $[\text{Na}^+] = 25$ mM and $[\text{Mg}^{2+}] = 5$ mM; the buffering agent Tris concentration were 32.2 mM [(40 mM from 2 \times SB + 89 mM from 2 \times dGLB)/4]. This mixture was then subjected for 10% dPAGE

purification. A DNA marker that has the identical size to the cleavage product was used to guide the excision of the desired DNA band.

The isolated DNA above was amplified by polymerase chain reaction (PCR) in a volume of 50 μ L containing 1 \times PCR buffer (supplied by the vendor as the 10 \times buffer), 0.2 mM each of the standard dNTPs, 1.25 U of Tth DNA polymerase, 0.5 μ M FP1 (forward primer) and 0.5 μ M RP1 (reverse primer). Twelve thermocycles were carried out with the following parameters: 94 $^{\circ}$ C, 30 s (2 min for the first cycle); 52 $^{\circ}$ C, 40 s; 72 $^{\circ}$ C, 45 s. A 1/100-fold dilution of the first PCR product was used for the second PCR using the same condition described above with the exception that RP2 was used instead of RP1. Another PCR2 was performed for internal labeling with 32 P. This was achieved by following the PCR2 protocol except that 10 μ Ci of [α - 32 P]dGTP and 0.02 mM non-radioactive dGTP were used to substitute 0.2 mM dGTP. The non-radiolabeled and 32 P-labeled PCR solutions were combined and the DNA in the mixture was precipitated by ethanol. The desired DNA molecules were purified by 10% dPAGE (54).

The purified DNA was used to carry out the second selection using the same procedure. 16 selection rounds were conducted. The cleavage product from round 16 was amplified, cloned and sequenced using a protocol we published previously (54, 55). 33 DNA sequences were revealed, which are provided in Table [S1](#).

3.6.5 General procedure for dPAGE-based folding analysis of urea-resistant DNA molecules

The folding buffer (4 \times FB) contained 40 mM HEPES, pH 7.5, 600 mM NaCl, and 20 mM MgCl₂. The general procedure was as follows: a DNA sequence of interest (radioactively labeled; 5 pmol in 5 μ L) was first mixed with 5 μ L of 4 \times FB, and then combined with 10 μ L of 2 \times dGLB. The final concentration for each

component in the mixture was: [DNA] = 12.5 nM, [Na⁺] = 150 mM, [Mg⁺] = 5 mM, [HEPES] = 10 mM, [Tris-boric acid] = 44.5 mM and [urea] = 7 M. The final volume was 20 μ L and the final pH was 8.4 (directly measured). This mixture was heated to 90 °C for 5 min, incubated at room temperature (RT) for 15 min, and then loaded onto a 10% dPAGE gel, which contained 10% polyacrylamide, 7 M urea, 89 mM Tris-boric acid and 1 mM EDTA. The unfolded and folded DNA bands separated by dPAGE gel was imaged with a Typhoon Trio + Imager (GE Healthcare) and the radioactivity of each DNA band was quantified with ImageQuant software (Molecular Dynamics). The percentage of folded structure was then calculated using Microsoft Excel (54).

3.7. References

1. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
2. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
3. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564-566 (1992).
4. Ellington, A. D. & Szostak, J. W. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **355**, 850-852 (1992).
5. Huizenga, D. E. & Szostak, J. W. A DNA aptamer that binds adenosine and ATP. *Biochemistry* **34**, 656-665 (1995).
6. Shangguan, D. *et al.* Aptamers evolved from live cells as effective molecular probes for cancer study. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11838-11843 (2006).
7. Zhang, N. *et al.* Intercellular connections related to cell-cell crosstalk specifically recognized by an aptamer. *Angew. Chem. Int. Ed. Engl.* **55**, 3914-3918 (2016).
8. Breaker, R. R. & Joyce, G. F. A DNA enzyme that cleaves RNA. *Chem. Biol.* **1**, 223-229 (1994).
9. Cuenoud, B. & Szostak, J. W. A DNA metalloenzyme with DNA ligase activity. *Nature*. **375**, 611-614 (1995).
10. Chinnapen, D. J. F. & Sen, D. A deoxyribozymes that harnesses light to repair thymine dimers in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 65-69 (2004).
11. Cruz, R. P., Wiithers, J. B. & Li, Y. Dinucleotide junction cleavage versatility of 8-17 deoxyribozyme. *Chem. Biol.* **11**, 57-67 (2004).
12. Torabi, S-F. *et al.* In vitro selection of a sodium-specific DNAzyme and its application in intracellular sensing. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5903-5908 (2015).
13. Huang, P-J. J., Vazin, M. & Liu, J. In vitro selection of a DNAzyme cooperatively binding two Lanthanide ions for RNA cleavage. *Biochemistry* **55**, 2518-2525 (2016).
14. Lee, Y. *et al.* DNA-catalyzed DNA cleavage by a radical pathway with well-defined product. *J. Am. Chem. Soc.* **139**, 255-261 (2017).
15. Wilson, D. S. & Szostak, J. W. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611-647 (1999).
16. Breaker, R. R. DNA enzymes. *Nature Biotechnology* **15**, 427-431 (1997).
17. Schlosser, K. & Li, Y. Biologically inspired synthetic enzymes made from DNA. *Chem. Biol* **16**, 311-322 (2009).
18. Mayer, G. The chemical biology of aptamers. *Angew. Chem. Int. Ed. Engl.* **48**, 2672-2689 (2009).
19. Silverman, S. K. DNA as a versatile chemical component for catalysis, encoding, and stereocontrol. *Angew. Chem. Int. Ed. Engl.* **49**, 7180-7201 (2010).
20. Liu, J., Cao, Z. & Lu, Y. Functional nucleic acid sensors. *Chem. Rev.* **109**, 1948-1998 (2009).
21. Liu, Z., Mei, S. H. J., Brennan, J. D. & Li, Y. Assemblage of signaling DNA enzymes with intriguing metal-ion specificities and pH dependences. *J. Am. Chem. Soc.* **125**, 7539-7545 (2003).

22. Kandadai, S. A. & Li, Y. Characterization of a catalytically efficient acidic RNA-cleaving deoxyribozyme. *Nucleic Acids Res.* **33**, 7164-7175 (2005).
23. Ali, M. M., Kandadai, S. & Li, Y. Characterization of pH3DZ1-an RNA-cleaving deoxyribozyme with optimal activity at pH 3. *Can. J. Chem.* **85**, 261-273 (2007).
24. Kandadai, S. A., Mok, W. W., Ali, M. M. & Li, Y. Characterization of an RNA-cleaving deoxyribozyme with optimal activity at pH 5. *Biochemistry*, **48**, 7383-7391 (2009).
25. Nelson, K. E., Bruesehoff, P. J. & Lu, Y. In vitro selection of high temperature Zn²⁺-dependent DNAzymes. *J. Mol. Evol.* **61**, 216-225 (2005).
26. Kourilsky, P., Manteuil, S., Zamansky, M. H. & Gros, F. DNA-RNA hybridization at low temperature in the presence of urea. *Biochem Biophys Res Commun.* **41**, 1080-1087 (1970).
27. Kourilsky, P., Leidner, J. & Tremblay, L. G. DNA-DNA hybridization on filters at low temperature in the presence of formamide or urea. *Biochimie* **53**, 1111-1114 (1971).
28. Maniatis, T., Jeffrey, A. & van deSande, H. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry* **14**, 3787-3794 (1975).
29. Summer, S. H., Grämer, R. & Dröge, P. Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *J. Vis. Exp.* e1485 (2009).
30. Ali, M. M., Aguirre, S. D., Lazim, H. & Li, Y. Fluorogenic DNAzyme probes as bacterial indicators. *Angew. Chem. Int. Ed. Engl.* **50**, 3751-3754 (2011).
31. Shen, Z. *et al.* A catalytic DNA activated by a specific strain of bacterial pathogen. *Angew. Chem. Int. Ed. Engl.* **55**, 2431-2434 (2016).
32. Mei, S. H. J., Liu, Z., Brennan, J. D. & Li, Y. An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. *J. Am. Chem. Soc.* **125**, 412-420 (2003).
33. Chiuman, W. & Li, Y. Simple fluorescent sensors engineered with catalytic DNA 'MgZ' based on a non-classic allosteric design. *PLoS One.* **2**, e1224 (2007).
34. Sen, D. & Gilvert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364-366 (1988).
35. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871-880 (1989).
36. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770-780 (2012).
37. Hänsel-Hertsch, R., Antonio, M. D. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell. Biol.* **18**, 279-284 (2017).
38. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **19**, 414-422 (2009).
39. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 2013-2018 (1962).
40. Ralph, R. K., Connors, W. J. & Khorana, H. G. Secondary structure and aggregation in deoxyguanosine oligonucleotides. *J. Am. Chem. Soc.* **84**, 2265-2266 (1962).
41. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482-5515 (2008).

42. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402-5415 (2006).
43. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908-2916 (2005).
44. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406-413 (2007).
45. Cech, T. R. Beginning to understand the end of the chromosome. *Cell.* **116**, 273-279 (2004).
46. Kypr, J., Kejnovská, I., Renčiuk, D. & Vorlíčková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713-1725 (2009).
47. Paramasivan, S., Rujan, I. & Bolton, P. H. Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* **43**, 324-331(2007).
48. Vorlíčková, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64-75 (2012).
49. Gilvert, W. Origin of life: the RNA world. *Nature* **319**, 618 (1986).
50. Joyce, G. F. RNA evolution and the origins of life. *Nature* **338**, 217-224 (1989).
51. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214-221 (2002).
52. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* **114**, 285-366 (2014).
53. Wang, W., Billen, L. P. & Li, Y. Sequence diversity, metal specificity, and catalytic proficiency of metal-dependent phosphorylating DNA enzymes. *Chem. Biol.* **9**, 507-517 (2002).

3.8 Supplementary Tables and Figures

Table S1. DNA sequences isolated after 16 rounds of in vitro selection.

Sequence ID	Nucleotides in random-sequence region ^a	Sequence class
SQ01	GGAGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	UD1
SQ02	A GAGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ03	G T AGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ04	G T AGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ05	G T AGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ06	G T AGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ07	GGAGG T GGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ08	GGAGG T GGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ09	GGAGGGGGTGGTGGG C TGTATACTTGGGTGGTTGGGGGGT	
SQ10	GGAGGGGGTGGTGGG C TGTATACTTGGGTGGTTGGGGGGT	
SQ11	GGAGGGGGTGGTGGG C TGTATACTTGGGTGGTTGGGGGGT	
SQ12	GGAGGGGGTGGTGGG C TGTATACTTGGGTGGTTGGGGGGT	
SQ13	GGAGGGGGTGGTGGG TG CATACTTGGGTGGTTGGGGGGT	
SQ14	GGAGGGGGTGGTGGG TGT GTA C TGGGTGGTTGGGGGGT	
SQ15	GGAGGGGGTGGTGGG TGTAT G C TTGGGTGGTTGGGGGGT	
SQ16	GGAGGGGGTGGTGGG TGTATACT C G GGTGGTTGGGGGGT	
SQ17	GGAGGGGGTGGTGGG TGTATACTTGGG C G TTGGGGGGT	
SQ18	GGAGGGGGTGGTGGG TGTATACTTGGGTGGTTGGGGGG G	
SQ19	G T AGG T GGTGGTGGGTTGTATACTTGGGTGGTTGGGGGGT	
SQ20	G T AGGGGGTGGTGGG T C G TATACTTGGGTGGTTGGGGGGT	
SQ21	G T AGGGGGTGGTGGG TG C A TACTTGGGTGGTTGGGGGGT	
SQ22	GGAGGGGGTGGTGGGTTGTATACTTGGGTGGTTGGGGG TC	
SQ23		
SQ24	CGGGGGGTGGCGGGGAAAGAGAGATGGGTGGT C GGGGGGT	UD2
SQ25	CGGGGGGTGG T GGGG A AGAGAGATGGGTGGT C GGGGGGT	
SQ26	CGGGGGGTGGCGGG TG AAGAGAGATGGGTGGT C GGGGGGT	
SQ27	CGGGGGGTGGCGGG AG AAGAGAGATGGGTGGT C GGGGGGT	
SQ28	CGGGGGGTGGCGGG AG AAGAGAGATGGGTGGT C GGGGGGT	
SQ29	CGGGGGGTGG T GGG AG AAGAGAGATGGGTGGT C GGGGGGT	
SQ30	GGGGGGATGGTGGGGTGGGAGAAGTGGGTGGCCGGGGGGT	UD3
SQ31	GGGGGGATGGTGGGGTGGGAGAAGTGGGTGGCCGGGGGGT	
SQ32	G T GGGGATGGTGGGGTGGGAGAAGTGGGTGGCCGGGGGGT	
SQ33	GGGGGGATGGTGGGGTGGGAGAAGTGGGTGGCCGGGGG TC	
SQ33		

^a: Nucleotide highlighted in yellow denotes a mutation in comparison to the first sequence listed in each class at the same position.

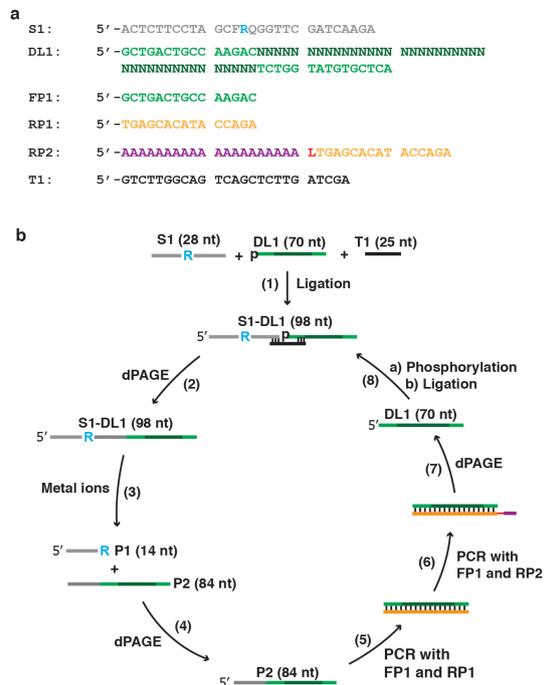


Figure S3-1. In vitro selection. (a) The sequences of the DNA molecules used for the study. DNA molecules are color-coded to assist the understanding of panel b. (b) In vitro selection scheme with the following 8 steps. (1) Ligation of phosphorylated DL1 (DNA library) to S1 (fluorogenic substrate) templated by T1. (2) Purification of ligated S1-DL1 by dPAGE. (3) incubation of S1-DL1 in the selection buffer (40 mM HEPES, pH 7.5, 100 mM NaCl, and 10 mM MgCl₂). (4) Purification of cleaved products by dPAGE. (5) PCR using FP1 and RP1 as primers. (6) PCR with FP1 and RP2 as primers. Note: RP2 contains a triethylene glycol spacer (L, panel a) and A20 tail at the 5' end; the spacer prevents the poly-A tail from being amplified, making the non-DNAzyme-coding strand 20 nucleotides longer than the coding strand. (7) Purification of DL1 strand by dPAGE. (8) Phosphorylating DL1 and ligating it to S1. The cycle of steps 2-10 was repeated for 16 times in this study. The experimental details are described in Section A above under "In vitro selection procedures".

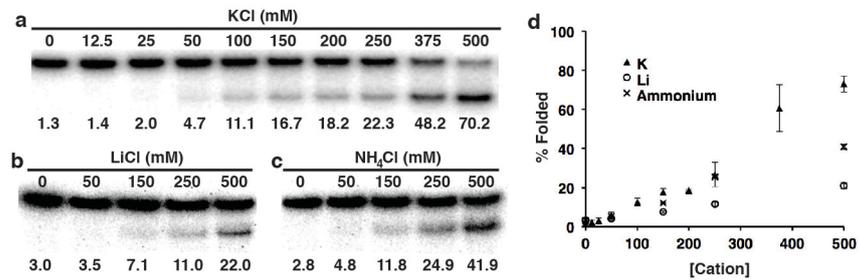


Figure S3-2. Effect of varying concentrations of K^+ , Li^+ and NH_4^+ on the folding of U2T. Each reaction mixture contained 10 mM HEPES (pH 7.5), 7 M urea, and specified concentrations of KCl (a), LiCl (b) or NH_4Cl (c). (d) % Folded structure vs. cation concentration. The mixture was heated at $90^\circ C$ for 5 min and left standing at room temperature for 15 min prior to dPAGE analysis.

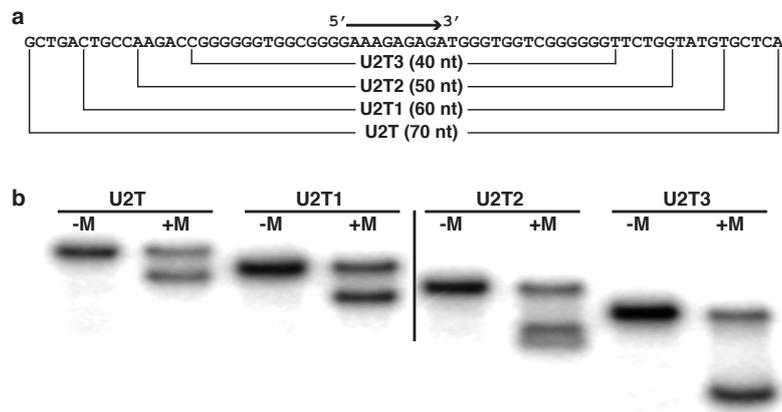


Figure S3-3. Examination of sequence truncation on the structure folding of UD2. (a) Four progressively shortened UD2 sequences tested for this study. U2T3 is subsequently named U2R as it contains only 40 nucleotides located in the random-sequence region. (b) dPAGE analysis of each truncated sequence in the presence of $1\times$ FB1 (10 mM HEPES, pH 7.5, 150 mM NaCl and 5 mM $MgCl_2$ (+M) or 10 mM HEPES only (-M).

was heated to 90°C and cooled at RT for 15 min. The test sample was subjected to dPAGE, to exclude any unfolded molecules in the test sample. The control sample was also purified by dPAGE. The DNA recovered from the gel were then treated with 100 μ L of 10% piperidine at 90°C for 30 min, dried by a vacuum concentrator. Each sample was then dissolved in 10 μ L of 1' dPAGE loading buffer, heated at 90°C for 5 min, and immediately loaded onto a dPAGE gel.

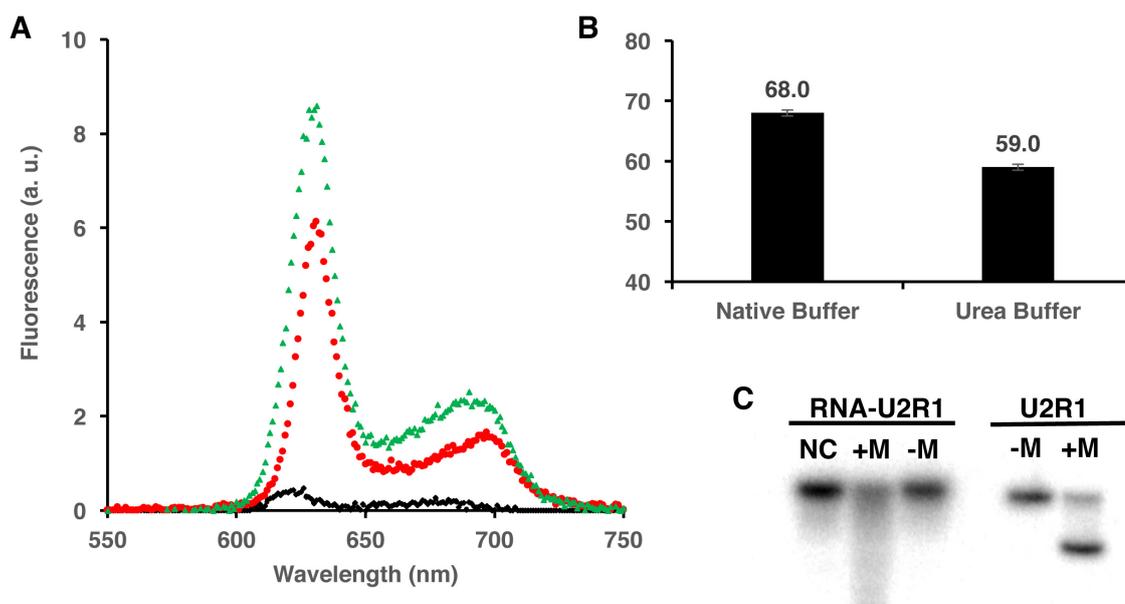


Figure S3-5. (a) Testing the presence of parallel G-quadruplex in the structure of U2R1 with protoporphyrin IX (PPIX). PPIX is known to produce enhanced fluorescence upon binding to parallel G-quadruplex. Three conditions were tested: PPIX alone (data in black), PPIX with U2R1 (data in red) and PPIX with U2R1 in the presence of 7M urea. (b) UV- melting points of U2R1 obtained under native and urea-containing conditions. (c) dPAGE analysis of RNA-U2R1 in the presence of 10 mM HEPES, pH 7.5, 150 mM NaCl and 5 mM MgCl₂ (+M) or 10 mM HEPES only (-M). Experimental details: (a) Samples with volumes of 60 μ L contained 1 μ M U2R1, 10 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM MgCl₂, in the presence of absence of 7 M urea. Corresponding negative control mixtures (with no DNA) were also made. The mixtures were heated to 90°C for 5 min and cooled to RT. 1.2 μ L of 100 μ M Protoporphyrin IX (PPIX) was added to each sample. After 10 minutes of incubation, samples were scanned with CARY Eclipse Fluorescence Spectrophotometer, with the excitation wavelength of 410 nm, and emission between 550 nm – 750 nm was recorded. (b) The UV measurements were run using Cary 300UV/Vis spectrophotometer. Samples with volumes of 400 μ L were prepared to contain 5 μ M U2R1, 10 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM MgCl₂, in the presence of absence of 7 M urea, along with corresponding blank mixtures (with no DNA). Each sample and blank was heated at 90°C for 5 min and cool to RT, then loaded into quartz cuvettes. The cuvettes were then loaded into the spectrophotometer. The block temperature was set to 23°C and the samples were incubated for 15 min before the melting experiment was initiated. Absorbance at 295 nm was measured between 23°C to 95°C. The ramp was set at 1°C/min. Melting points were calculated using the derivative of absorbance vs. temperature graphs. (c) 5 pmol of U2R1 or RNA-U2R1 in 5 μ L were mixed with 5 μ L of 4' FB and 10 μ L of 2' dGLB, heated at 90°C for 5 min, and cooled at RT for 15 min. 5 μ L was taken for dPAGE analysis. +M: the reaction mixture contained NaCl and MgCl₂; -M: the two salts were omitted. NC: 2' dGLB was mixed with 10 μ L of RNA-U2R1 dissolved in water.

Chapter 4

A DNA Switch for Detecting Single Nucleotide Polymorphism within a Long DNA Sequence Under Denaturing Conditions

4.1 Author's Preface

This study describes the potential application in biosensing of the aforementioned urea-resistant G-quadruplex. Using this structure, we have designed a catalytic beacon can detect single nucleotide mutation within a long target sequence. We demonstrate that it can discriminate a mutation-bearing target from one with 98.9% similarity.

This work is presented in the manuscript format, which has been submitted to *Chemistry, A European Journal* and has been accepted. I am the first author of this publication. I took the leading role in project design, experimentation and data analysis. Dr. Jiuxing Li has provided significant insights and suggestions during the process of this work. I conducted all the experiments and summarized all the results. The manuscript was written with assistance from Dr. Yingfu Li and other co-authors.

4.2 Manuscript: A DNA Switch for Detecting Single Nucleotide Polymorphism within a Long DNA Sequence Under Denaturing Conditions

Wenqing Zhang, Jiuxing Li, Bruno Salena and Yingfu Li*

Abstract: DNA detection is usually conducted under nondenaturing conditions to favor the formation of Watson-Crick base-pairing interactions. However, although such a setting is excellent for distinguishing a single-nucleotide polymorphism (SNP) within short DNA sequences (15-25 nucleotides), it does not offer a good solution to SNP detection within much longer sequences. Here we report on a new detection method capable of detecting SNP in a DNA sequence containing 35-90 nucleotides. This is achieved through incorporating into the recognition DNA sequence a previously discovered DNA molecule that forms a stable G-quadruplex in the presence of 7 molar urea, a known condition for denaturing DNA structures. The systems are configured to produce both colorimetric and fluorescent signals upon target binding.

Using synthetic oligonucleotides as recognition elements to achieve detection of a specific DNA and RNA sequence is a widely accepted conception and practice in molecular biology and medical diagnostics. Commonly used technologies include FISH (fluorescence in situ hybridization), molecular beacon and microarray.^[1-3] A key goal of these DNA detection technologies is to achieve specific detection of DNA sequences that may only differ by a single nucleotide polymorphism (SNP) because SNP represents important determinants in many genetic diseases.^[4-13] Although techniques such as FISH and molecular beacon can detect SNP within a short DNA sequence (25-nt or shorter; nt: nucleotide), they are unable to detect SNP in much longer DNA sequences. Certain DNA detection technologies, particularly microarray, use very long DNA recognition sequences (60-80 nt in length) as recognition probes because these long

DNA probes are able to achieve excellent detection sensitivity.^[14,15] However, such long DNA sequences as the detection probe are unable to distinguish DNA sequences differing by 30% nucleotide identity,^[15-17] let alone detect a SNP. Performing detection at an elevated temperature under precise temperature control can improve detection specificity, but such a measure is still unable to achieve SNP detection.^[6,14] More recently, it has been shown that the method of strand displacement is capable of achieving SNP detection in long DNA sequences,^[16] but such a method typically uses fluorescence to report target recognition, limiting their range of applications. In this work, we report a simple, "mix-and-read" approach capable of achieving both fluorescent and colorimetric detection of a SNP within DNA sequences containing as many as 90 nt.

The key component of our system is a guanine-rich (G-rich) DNA molecule that can form a stable G-quadruplex structure (G4 structure) in the presence of 7 molar urea (7MU). Another important component is a hairpin design that acts as a toehold for strand displacement.^[18] As we will explain in greater detail below, the combined system places the G-rich element "locked" in a long hairpin structure and unable to form the G4 structure under 7MU, unless the sequence to be detected (the target sequence) is present in the solution.

We have recently discovered a family of G-rich DNA sequences, that are capable of folding into highly stable anti-parallel G4 structures even in the presence of 7MU.^[19] The folding requires the assistance of Na⁺ and Mg²⁺ ions. In addition to the G-rich feature, this sequence family has a central loop region that needs to be longer than 7 nt but does not require a specific sequence (Figure 1A).^[19] In this

W. Zhang, Dr. J. Li, Prof. Dr. Y. Li
M.G. DeGrootte Institute for Infectious Disease Research
Department of Biochemistry and Biomedical Sciences
DeGrootte School of Medicine
McMaster University
1280 Main Street West, Hamilton, ON, L8S 4K1 (Canada)
*E-mail: living@mcmaster.ca

Prof. Dr. B. Salena
Department of Medicine
DeGrootte School of Medicine
McMaster University
1280 Main Street West, Hamilton, ON L8S 4K1 (Canada)

Supporting information for this article is given via a link at the end of the document.

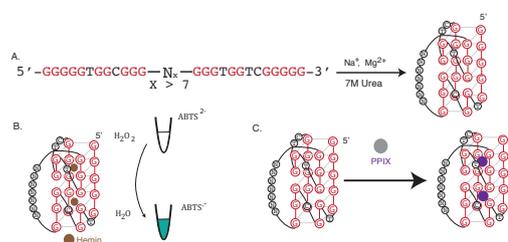


Figure 4-1. Reporting systems constructed with ULX. (A) Anti-parallel G-quadruplex (G4) structure of ULX in 7 M urea (7MU) containing Na⁺ and Mg²⁺. (B) Oxidation of ABTS in the presence of a ULX, hemin and H₂O₂. (C) Fluorescence enhancement of protoporphyrin IX (PPIX) upon its binding to the G4 structure of a ULX.

work, we rename this family of sequences as ULX in

which X stands for the number of nucleotides in the loop (L).

Anti-parallel G4 structures are known to be excellent signal transducers and are capable of

UL10, UL15 and UL20 were found to produce a fast-moving band when they were subjected to 10% denaturing (7MU) polyacrylamide gel electrophoresis (Figure 2B). In contrast, this band was absent with

UL5. These observations are consistent with our previous finding that the minimal length is 7 nt.^[19] We tested UL5, UL10, UL15 and UL20 for peroxidase activity. UL10, UL15 and UL20 were capable of oxidizing ABTS in the presence of hemin and H₂O₂ to produce the blue color both in 0MU (i.e. the reaction solution containing no urea) and 7MU; in contrast, UL5 did not induce the color change (Figure 2C). These results are consistent with the structural folding data in Figure 2B. Color intensity analysis shows that UL10 produced a stronger color change in 7MU than UL15 and

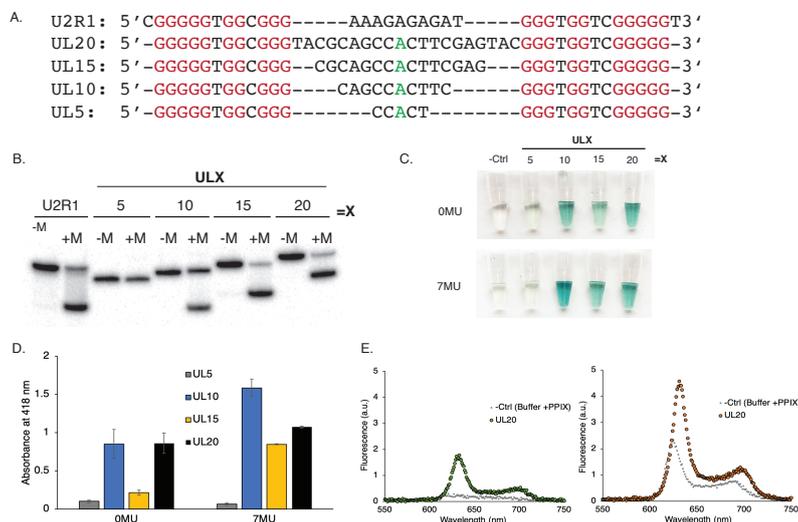


Figure 4-2. G4 structure formation by ULX. (A) The sequences of DNA molecules tested. U2R1 is the reference sequence previously reported. (B) Detection of G4 structures of ULX by 7MU-dPAGE. -M/+M: in the absence and presence of Na⁺ and Mg²⁺. For each ³²P-labeled DNA molecule, the appearance of the fast-moving DNA band (the bottom band) in the +M lane indicates the successful G4 structure formation. (C) The peroxidase activity of the four ULX constructs in 0MU and 7MU. (D) Absorbance readings at 418 nm of the reactions in C. -Ctrl: no DNA. (E) Emission spectra of UL20-PPIX mixture in 0 MU (left) and 7MU (right). -Ctrl: no DNA.

producing both fluorescence and colorimetric signals.^[20-23] They can catalyze the oxidation of dyes such as 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) in the presence of hemin and hydrogen peroxide, which produces a color change (Figure 1B). This activity is known as peroxidase activity, and therefore, these types of DNA molecules are also called peroxidase DNAzymes.^[24-26] They can also bind to certain fluorogenic molecules, such as protoporphyrin IX (PPIX), to produce a strong fluorescence signal (Figure 1C).^[27-29] Therefore, ULX can function as an excellent colorimetric or fluorescent signaling module to build a biosensing system to achieve target detection in 7MU.

We tested 4 ULX molecules with a loop length of 5, 10, 15, and 20 nt (Figure 2A) to determine their folding patterns and the ability to produce either colorimetric signal with ABTS/hemin/H₂O₂ or fluorescence with PPIX. Radioactively labeled (³²P)

UL20 (Figure 2D); however, we decided to choose UL20 for further biosensor work because of the need of a stronger "lock", a requirement for the sensor

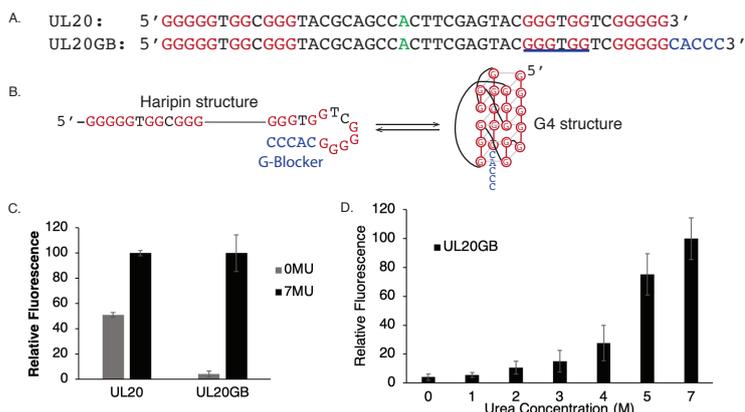


Figure 4-3. Design of a harpin structure with a G-blocker element. (A) The sequences of UL20 and UL20GB. (B) Alternative structures of UL20GB. The G-blocker creates a hairpin structure as shown. (C) Fluorescence readings of the mixtures of UL20 and UL20GB with PPIX in 0MU and 7MU. (D) Relative fluorescence of the reaction solutions containing UL20GB and 0-7MU.

design to be discussed below. For this reason, we also examined the ability of UL20 to enhance the fluorescence of PPIX. Interestingly, the fluorescence intensity of UL20 in 7MU was twice as high as in 0MU (Figure 2E). These findings suggest that UL20 can be

used both as colorimetric and fluorescence reporter for a biosensor to function in 7MU.

To design a biosensor, we need to establish a strategy that can keep UL20 from forming the G4 structure before target binding but permit the G4 structure formation upon target binding. We achieved this by incorporating into UL20 two additional sequence elements: a target-binding DNA sequence (named “probe”) and a G4 structure blocking sequence (named “G-blocker”). Together, these two sequence elements create a target-responsive system that is capable of distinguishing DNA targets differing by a SNP, as we will progressively demonstrate below.

We first appended the 3'-end of UL20 with CACCC (blue nucleotides in Figure 3A). The new sequence is named UL20GB because the CACCC motif functions as a “G-blocker”: this motif is designed to pair with the GGGTG element (underlined in Figure 3A) to create a hairpin structure to compete with the G4 structure, as illustrated in Figure 3B.

We compared the fluorescence intensity of UL20 and UL20GB when they were combined with PPIX in 0MU and 7MU (Figure 3C). For UL20, high levels of fluorescence were observed in both 0MU and 7MU; however, UL20GB produced much stronger fluorescence in 7MU than 0MU. The fluorescence intensity of the UL20GB-PPIX mixture was also found to increase with the increasing urea concentrations (Figure 3D). These results support the following statements: (1) UL20 adopts the G4 structure in both 0MU and 7MU; (2) the hairpin structure of UL20GB is much more stable than its G4 structure in 0MU, but the stability reverses in 7MU. Consequently, if the hairpin is the starting structure, it will give rise to the G4 structure in solution containing 7MU. This transition is denoted “H2G transition”.

Another sequence element appended to UL20 is the probe sequence, the green nucleotides added right after the G-blocker (Figure 4A). We chose a 50-nt probe sequence (P50) as the first model probe and the resultant sensor DNA construct is thus named UL20GBP50. P50 was designed to achieve two functions. The first 20 nucleotides were designed to be a “lock” sequence, with the purpose of forming base pairs with the orange nucleotides in UL20GB, the long connecting loop in the G4 structure. This portion of the probe and the G-blocker work together

to create a very strong stem to inhibit the H2G transition even in 7MU. The remaining 30-nt sequence in P50, which is in the single-stranded form (labeled “tail” in Figure 4A), was designed to function

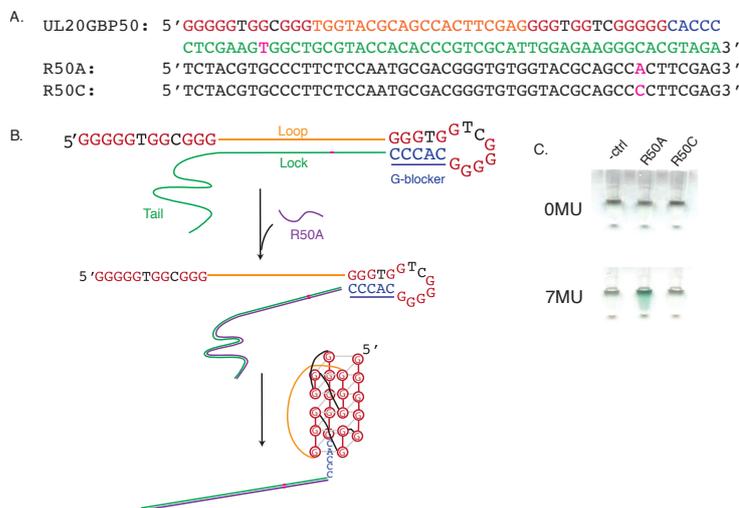


Figure 4-4. SNP detection using a hairpin/G4 structure switch. (A) The sequences of UL20GBP50 (sensor), R50A (DNA target) and R50C (SNP control). (B) Graphic illustration of the H2G transition of the sensor target binding. (C) Peroxidase reactions before and after target binding in 0MU and 7MU.

as a toehold to initiate the binding to a 50-nt target sequence that is fully complementary to P50. By this design, the toehold-initiated probe-target binding event is expected to produce a relatively weak hairpin containing only the 5-base pair stem (the G-blocker stem), which should undergo the H2G transition for signal generation (Figure 4B).

The ability of the aforementioned sensing system in distinguishing a SNP was tested with two DNA targets named R50A and R50C, respectively (their sequences are provided in Figure 4A). R50A is the intended target, which represents a sequence segment of the human rhodopsin gene that carries the crucial C-to-A point mutation responsible for the genetic eye disease retinitis pigmentosa,^[30] and R50C (representing the wildtype sequence) is the SNP control. Note that: (1) the tail and the lock portions of UL20GBP50 are designed to be perfectly complementary to R50A; (2) the point mutation was placed at the 8th position of the probe sequence (the corresponding nucleotides in R50A and R50C (which are labeled pink in Figure 4A), which was chosen for the consideration that it was located near the center of the lock and thus should favor the H2G transition by the match SNP but not by the mismatch target.

The lock and the G-blocker create a very strong stem with the G4 loop, which was found to prevent

R35A, as the peroxidase activity was observed for both targets.

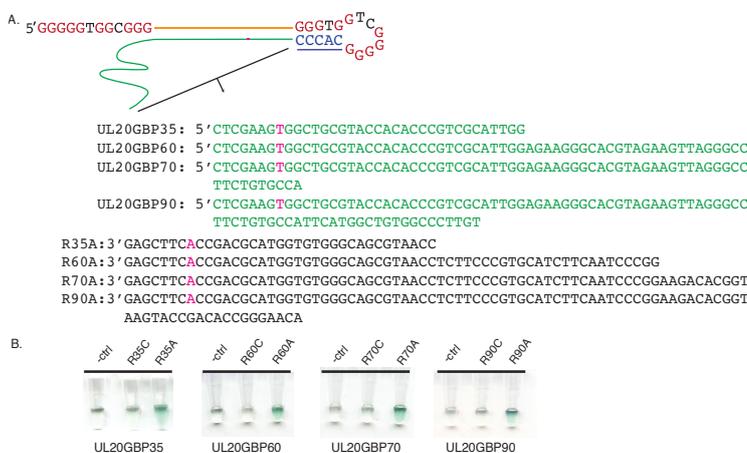


Figure 4-5. Four additional sensors with lengthened probe sequences. (A) The sequences of additional sensors and targets. The four new sensors were named UL20GBP35, 60 and 90. Their perfectly matched DNA targets were named R35A, R60A, R70A, and R90A, respectively. The corresponding SNP controls were named R35C, R60C, R70C, and R90C (their sequences are not shown, as the only difference for each SNP control is the replacement of the pink A with C). (B) Peroxidase reactions of the four sensors before and after target binding in OMU and 7MU.

the G4 structure formation in 7MU, based on the lack of peroxidase activity (the tube labeled “Ctrl” in Figure 4C). When R50A was added to the same solution (the test with the matching DNA target), a blue color was produced, signifying that the formation of the G4 structure. However, no color change was detected when R50C was added (the SNP control). Furthermore, no color change was seen in all three reactions in OMU.

We also tested the target-promoted structure transition using PPIX (Figure S1) and the results were consistent with the colorimetric test: high level of fluorescence signal was only seen with R50A in the presence of 7MU but not in the two controls.

We also tested four additional constructs with a probe sequence of 35, 60, 70 and 90 nt in length (Figure 5A). All these sequences were designed to be fully complementary to the human rhodopsin gene that carries the same C-to-A point mutation but vary in the number of nucleotides in the 5'-end (Figure 5A). All these probes were found to be able to distinguish the mutant gene from the wildtype in the colorimetric peroxidase assay. The fact that the sensing system can identify a SNP within a 90-nt sequence indicates that it is capable of differentiating non-specific targets with 98.9% identity to the true target (89/90).

The G-blocker element is crucial for SNP detection capability, as the data in Figure S2 shows. In this experiment, the probe length was chosen to be 35 nt. With the removal of the G-blocker (the sensor DNA is now named UL20P35 instead of UL20GBP35), the system was incapable of distinguishing R35C from

To show that the detection method is capable of detecting SNP in different sequence context, we designed two additional UL20GBP50 systems, targeting a SNP within a human mitochondria DNA fragment^[16a] and a SNP within an *E. coli. rpo gene* fragment carrying mutations responsible for antibiotic resistance^[16b], respectively. Each system was capable of producing a color in the peroxidase assay in the presence of the intended target but not in the SNP control (Figure S3). Note that the two new targets were purposely designed to contain G-rich and GC-rich sequence elements to assess whether such elements might disrupt the correct folding of the G4 structure. The ability

to detect these targets indicates that the biosensor design is broadly applicable to diverse DNA sequences.

In summary, although hybridization of DNA in the presence of denaturing agents, such as urea and formamide, has been investigated before,^[31] the use of very high concentrations of urea as the key solution component to enable a DNA switch to achieve SNP detection has not been attempted until the current study. Carrying out DNA detection under 7 M urea comes with a significant advantage in SNP detection: the capability to detect a SNP in very long DNA sequences. Towards this end, we have shown that our method is capable of detecting a SNP in a ssDNA sequence containing as many as 90 nucleotides.

The successful implementation of SNP detection in a long DNA sequence in a solution containing 7 M urea is attributed to the discovery of a family of G-rich DNA sequences that have two distinct properties. The first property is their capability to form a stable G-quadruplex structure in 7 M urea. The second property is that the G-quadruplex structure of this family has a loop sequence that can be changed to any sequence as long as this sequence is longer than 7 nucleotides. In this work, we have chosen the length of the loop to be 20 nucleotides and the sequence of the loop to be complementary to a chosen 20-nt sequence segment of three different DNA targets. For the detection of a specific DNA sequence, one simply needs to change the loop sequence to be the sequence complementary to the DNA target of interest. We envision that the strategy

demonstrated by this work can be exploited for the design of biosensors to detect many other SNPs in long DNA sequences.

Acknowledgements

Funding was provided by Natural Sciences and Engineering Research Council of Canada (NSERC) via a Discovery Grant to YL. Part of the work was conducted at the McMaster Biointerfaces Institute.

Keywords: DNA detection • single nucleotide polymorphism • G-quadruplex • DNAzyme • denaturing condition

- [1] a) P. R. Langer-Safer, M. Levine, D. C. Ward, *Proc. Natl. Acad. Sci. U. S. A.* **1982**, 4381-4385. b) M. Yoshimoto, O. Ludkovski, J. Good, C. Pereira, R. J. Gooding, J. McGowan-Jordan, *et al.*, *Lab Investig.* **2018**, 98, 403-413. c) H. Frickmann, A. E. Zautner, A. Moter, J. Kikhney, R. M. Hagen, H. Stender, *et al.*, *Crit. Rev. Microbiol.* **2017**, 43, 263-293.
- [2] a) T. Kuang, L. Chang, X. Peng, X. Hu, D. Gallego-Perez, *Trends Biotechnol.* **2017**, 35, 347-359. b) Z. Ma, X. Wu, C. J. Krueger, A. K. Chen, *Genomics Proteomics Bioinforma.* **2017**, 15, 279-286. c) S. Tyagi, F. R. Kramer, *Nat. Biotechnol.* **1996**, 14, 303-308.
- [3] a) M. C. Pirrung, E. M. Southern, *Biochem. Mol. Biol. Educ.* **2014**, 42, 106-113. b) S. E. Wildsmith, F. J. Elcock, *J. Clin. Pathol. Mol. Pathol.* **2001**, 54, 8-16. c) S. W. Cheung, W. Bi, *Expert. Rev. Mol. Diagn.* **2018**, 18, 531-542.
- [4] S. J. Smith, C. R. Nemr. S. O. Kelly, *J. Am. Chem. Soc.* **2017**, 139, 1020-1028.
- [5] B. Mehta, R. Daniel, C. Phillips, D. McNevin, *Int. J. Legal. Med.* **2017**, 131,21-37.
- [6] X. Luo, I. M. Hsing, *Biosens. Bioelectron.* **2009**, 25, 803-808.
- [7] M. T. Hwang, P. B. Landon, J. Lee, D. Choi, A. H. Mo, G. Glinsky, R. Lal, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 7088-7093.
- [8] P. Y. Kwok, X. Chen, *Curr. Issues. Mol. Biol.* **2003**, 5, 43-60.
- [9] H. Deng, W. Shen, Z. Gao, *Biosens. Bioelectron.* **2015**, 68, 310-315.
- [10] H. Xu, Q. Wu, H. Shen, *Talanta* **2017**, 167, 630-637.
- [11] Y. Xiao, V. Pavlov, T. Niazov, A. Dishon, M. Kotler, I. Willner, *J. Am. Chem. Soc.* **2004**, 126, 7430-7431.
- [12] Z. F. Gao, Y. Ling, L. Lu, N. Y. Chen, H. Q. Luo, N. B. Li, *Anal. Chem.* **2014**, 86, 2543-2548.
- [13] S. Bhadra, V. Codrea, A. D. Ellington, *Anal. Biochem.* **2014**, 445, 38-40.
- [14] a) C. C. Chou, C. H. Chen, T. T. Lee, K. Peck, *Nucleic Acids Res.* **2004**, 32, e99. b) M. D. Kane, T. A. Jatkoe, C. R. Stumpf, J. Lu, J. D. Thomas, S. J. Madore, *Nucleic Acids Res.* **2000**, 28, 4552-4557.
- [15] T. R. Hughes, M. Mao, A. R. Jones, J. Burchard, M. J. Marton, K. W. Shannon, *et al.*, *Nat. Biotechnol.* **2001**, 19, 342-347.
- [16] a) S. X. Chen, D. Y. Zhang, G. Seelig, *Nature Chemistry.* **2013**, 5, 782-789. b) D. A. Khodakov, A. S. Khodakova, A. Linacre, A. V. Ellis, *J. Am. Chem. Soc.* **2013**, 135, 5612-5619.
- [17] Z. He, L. Wu, X. Li, M. W. Fields, J. Zhou, *Appl. Environ. Microbiol.* **2005**, 71, 3753-3760.
- [18] D. Y. Zhang, G. Seeling, *Nat. Chem.* **2011**, 3, 103-113.
- [19] W. Zhang, M. Liu, C. Lee, B. J. Salena, Y. Li, *Sci. Rep.* **2018**, 8, 1935.
- [20] J. Kosman, B. Juskowiak, *Anal. Chim. Acta.* **2011**, 707, 7-17.
- [21] S. Bhadra, V. Codrea, A. D. Ellington, *Anal. Biochem.* **2014**, 445, 38-40.
- [22] H. Z. He, D. S. H. Chan, C. H. Leung, D. L. Ma, *Nucleic Acids Res.* **2013**, 41, 4345-4359.
- [23] D. M. Kong, *Methods* **2013**, 64, 199-204.
- [24] J. Ren, T. Wang, E. Wang, J. Wang, *Analyst* **2015**, 140, 2556-2572.
- [25] Y. Li, R. Geyer, D. Sen, *Biochemistry* **1996**, 35, 6911-6922.
- [26] B. Ruttkay-Nedecky, J. Kudr, L. Nejdil, D. Maskova, R. Kizek, V. Adam, *Molecules* **2013**, 14760-14779.
- [27] T. Li, E. Wang, S. Dong, *Anal. Chem.* **2010**, 82, 7576-7580.
- [28] H. Yaku, T. Murashima, D. Miyoshi, N. Sugimoto, *Molecules* **2012**, 17, 10586-10613.
- [29] B. H. Kang, N. Li, S. G. Liu, N. B. Li, H. Q. Luo, *Anal. Sci.* **2016**, 32, 887-892.
- [30] T. P. Dryja, T. L. McGee, E. Reichel, L. B. Hahn, G. S. Cowley, D. W. Yandell *et al.*, *Nature* **1990**, 343,364-366.
- [31] a) Ph. Kourilsky, J. Leidner, G. Y. Tremblay, *Biochimie* **1971**, 53, 1111-1114. b) L. S. Yilmaz, A. Loy, E. S. Wright, M. Wagner, D. R. Noguera, *PLoS One* **2012**, 7, e43862. c) D. Moralli, Z. L. Monaco, *Methods Mol. Biol.* **2010**, 659, 203-218. d) J. Fuchs, D. Dell'Atti, A. Buhot, R. Calemczuk, M. Mascini, T. Livache, *Anal. Biochem.* **2010**, 397, 132-134. e) R. S. Santos, N. Guimarães, P. Madureira, N. F. Azevedo, *J. Biotechnol.* **2014**, 187, 16-24.

4.3 Supplementary Information: A DNA Switch for Detecting Single Nucleotide Polymorphism within a Long DNA Sequence Under Denaturing Conditions

Wenqing Zhang, Jiuxing Li, Bruno Salena and Yingfu Li*

MATERIALS AND METHODS

Enzymes, chemicals, and other materials. [γ - ^{32}P]ATP was obtained from Perkin Elmer (Woodbridge, ON, Canada). Urea (ultrapure) and 40% polyacrylamide solution (29:1) were purchased from BioShop Canada (Burlington, ON, Canada). Water was purified via Milli-Q Synthesis A10 water purifier. All other chemicals were purchased from Sigma-Aldrich (Oakville, ON, Canada).

Synthesis and purification of oligonucleotides. DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA), and synthesized using automated DNA synthesis using standard phosphoramidite chemistry. All oligonucleotides were purified using 10% denaturing polyacrylamide gel electrophoresis (dPAGE) containing 7 M urea. Note that for the purification of urea-resistant DNA molecules featured in this study, DNA samples were resuspended in 1.5 \times denaturing gel loading buffer (dGLB), heated at 90°C for 5 min and then immediately loaded onto dPAGE gel. This procedure was effective in minimizing the structure retention on urea dPAGE.

^{32}P -labelling of oligonucleotides. 1 μL of [γ - ^{32}P]ATP (10 μCi) was used to label ~150 pmol of DNA in 50 μL of 1 \times PNK buffer A containing 5 units of PNK. After incubation at 37°C for 30 min, PNK was inactivated by heating at 90°C for 5 min. The phosphorylated DNA was first concentrated by ethanol precipitation and then purified by 10% dPAGE purification. Similarly, this purification step requires the DNA to be resuspended in 1.5 \times dGLB, heated at 90°C for 5 min and then immediately loaded onto dPAGE gel.

Denaturing-Polyacrylamide Gel Electrophoresis (dPAGE)-based folding analysis of urea-resistant DNA molecules. The folding buffer (4× FB) contained 40 mM HEPES, pH 7.5, 600 mM NaCl, and 20 mM MgCl₂. 10× TBE (Tris-borate EDTA) (1 L) was made of 108 g Tris-base (0.89 M), 55 g boric acid (0.89 M), 20 mL of 0.5 M ethylenediaminetetraacetic acid (EDTA; pH 8.0; 10 mM). 2× dGLB (per 100 mL) was made with 20 g sucrose, 10 mL of 10× TBE, 1 mL of 10% (w/v) SDS (sodium dodecyl sulphate), 25 mg bromophenol blue, and 25 mg xylene cyanol FF, and 84.1 g urea (14 M). Firstly, ³²P-labeled U2R1 and ULX sequences (5 pmol in 5 μL) was each mixed with 5 μL of 4× FB, and then combined with 10 μL of 2× dGLB. This mixture was heated to 90°C for 5 min, incubated at room temperature (RT) for 15 min, and then loaded onto a 10% dPAGE gel, which contained 10% polyacrylamide, 7 M urea, 89 mM Tris-boric acid and 1 mM EDTA; The gel electrophoresis was run in the presence of 1× TBE The unfolded and folded DNA bands separated by dPAGE gel was imaged with a Typhoon Trio+ Imager (GE Healthcare).

Colorimetric Experiments. The peroxidase activity of ULX sequences (UL5, UL10, UL15, and UL20) and UL20GB, and UL20GBPX (UL20GBP50, UL20GBP35, UL20GBP60, UL20GBP70, and UL20GBP90) were tested under both native (0MU) and denaturing conditions (7MU). A folding buffer A (4× FBa) contained 160 mM Tris pH 8.5, 600 mM NaCl, and 20 mM MgCl₂ was used. Specifically, a final volume of 46 μL was made to contain 1.08 μM of DNA sequence of interest, 4.35 μM of hemin, 1.08× FBa for a 0MU test; and additional 7.6 M urea for a 7MU test. This mixture was heated to 90 °C for 3 min and cooled to RT for 15 min. 3 μL of 50 mM ABTS and 1 μL of 50× diluted

H₂O₂ were added. The reaction had a final concentration of 1 μM ULX, 4 μM hemin, 1× FBa, 3 mM of ABTS, and 4 μM of H₂O₂. Pictures were taken approximately 1 min right after.

For colorimetric tests involving UL20GBP_X and R_XA or R_XC (X = 35, 50, 60, 70, or 90), reactions were conducted similarly, with the addition of equimolar of R_XA or R_XC before heating.

Probing G-quadruplex Formation and DNA Detection with Protoporphyrin (PPIX). To test formation of G-quadruplex formation by ULX variants, a mixture of 57 μL was made to have 1.05 μM ULX and 1.05× FBa for a native test, and an additional of 7.37 M of urea for a test under 7MU condition. The mixture was heated to 90 °C for 3 min and cooled to RT. PPIX was added so that the mixture has 1 μM ULX, 1× FBa, and 2 μM PPIX, and an additional 7 M urea if the test is under denaturing condition. The mixture was incubated for approximately one hour, before analyzed by CARY Eclipse Fluorescence Spectrophotometer, with the excitation wavelength of 410 nm, and the emission spectrum between 550 nm - 750 nm was recorded. Alternatively, the emission of 631 nm was recorded using Teacan M1000, and the emissions were corrected against fluorescence generated by buffer alone. For experiments involving R_XA or R_XC detection, DNA targets at various concentrations were added together with UL20GB probes prior to heating.

For Figure 3, relative fluorescence was calculated assuming that UL20 at 7MU produces 100% fluorescence, using the following equation:

$$\text{Relative Fluorescence} = (\text{Sample Fluorescence}/\text{UL20 Fluorescence at 7MU}) * 100\%$$

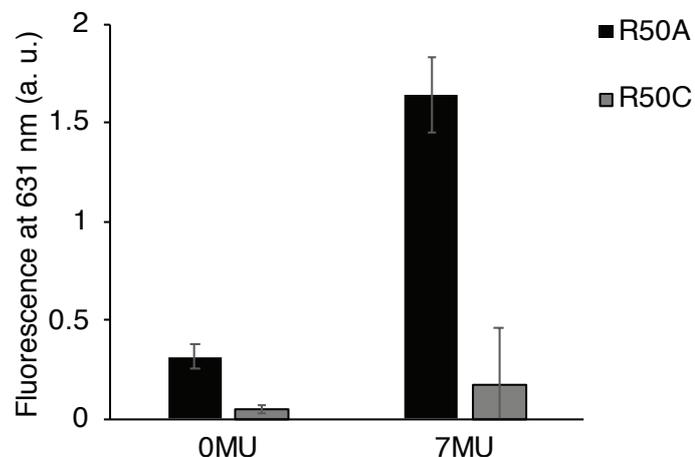


Figure 4-S1. Discrimination of R50A and R50C using UL20GBP50, under 0MU or 7MU. 1 μ M UL20GBP50 and equimolar R50A or R50C, in combination of 2 μ M PPIX were heated at 90 $^{\circ}$ C for 3 minutes, and incubated at RT for one hour, followed by the measurement of fluorescence at 631 nm. The concentrations are final concentration in 50 μ L reaction solution.

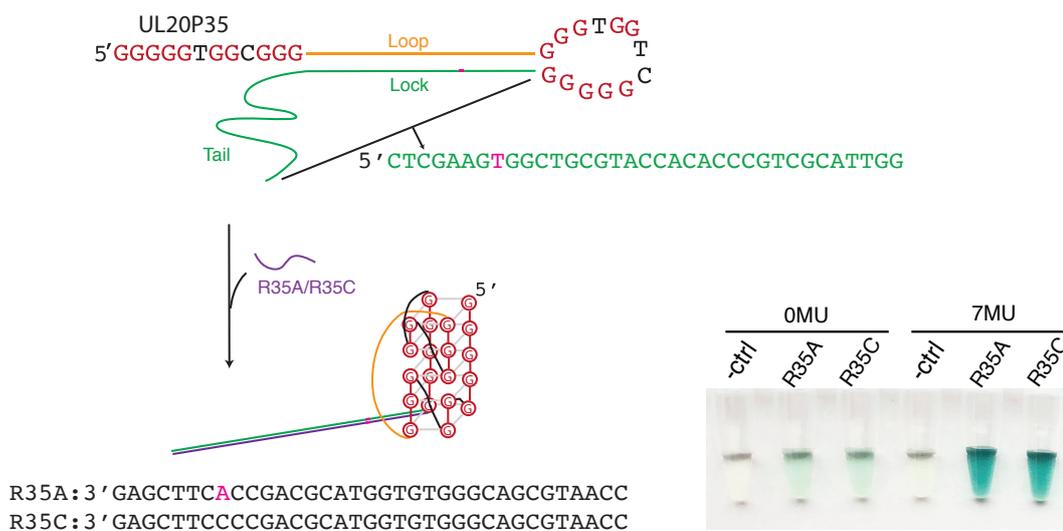


Figure 4-S2. Testing the importance of the G-blocker. 1 μ M UL20P35, equimolar R50A or R50C, and 4 μ M hemin were heated at 90 $^{\circ}$ C for 3 minutes, and the mixture was incubated at RT for 15 min, followed by the addition of ABTS and H₂O₂ to a final concentration of 3 mM and 4 μ M respectively. The concentrations are final concentration in 50 μ L reaction solution. The picture was taken approximately 1 min after the addition of H₂O₂.

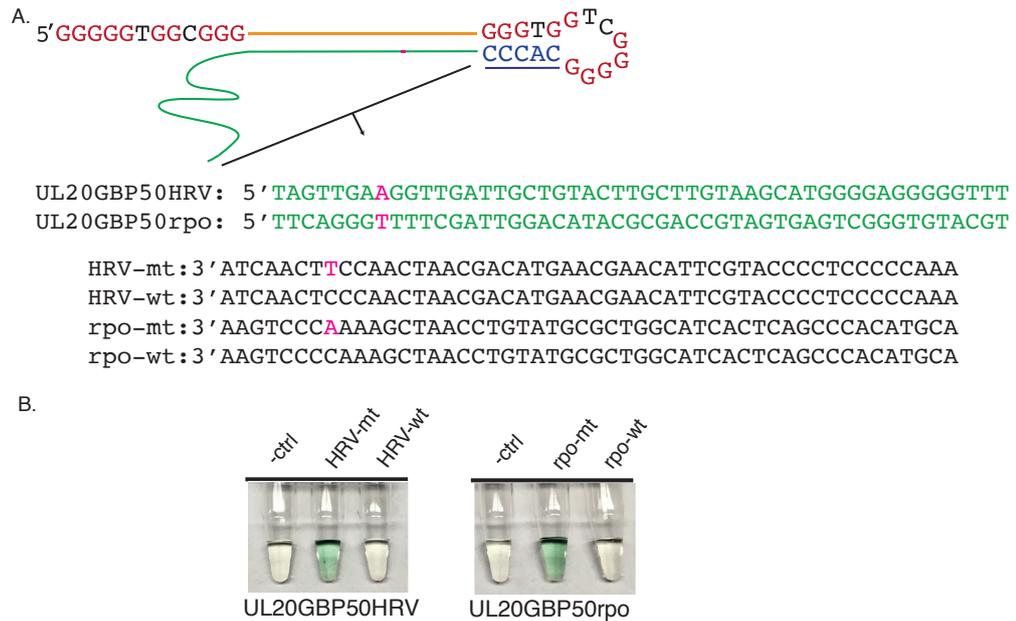


Figure 4-S3. Testing the UL20GB probe using other DNA targets. 1 μ M UL20GBP50HRV or UL20GBP50rpo, equimolar their perfect match or mismatch targets (HRV-mt, HRV-wt, rpo-mt, and rpo-wt, respectively), and 4 μ M hemin were heated at 90 °C for 3 minutes, and the mixture was incubated at RT for 15 min, followed by the addition of ABTS and H₂O₂ to a final concentration of 3 mM and 4 μ M respectively. The concentrations are final concentration in 50 μ L reaction solution. The picture was taken approximately 1 min after the addition of H₂O₂.

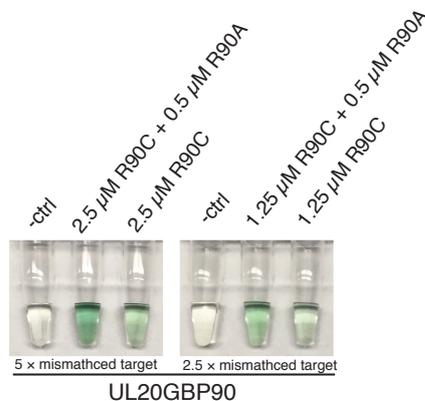


Figure 4-S4. Testing the specificity of UL20GBP90 in the presence of excess mis-matched targets. 500 nM of UL20GBP90, DNA targets (concentrations as indicated) and 4 μ M hemin were heated at 90 °C for 3 minutes, and the mixture was incubated at RT for 15 min, followed by the addition of ABTS and H₂O₂ to a final concentration of 3 mM and 4 μ M respectively. The concentrations are final concentration in 50 μ L reaction solution. The picture was taken approximately 1 min after the addition of H₂O₂.

Chapter 5

Conclusion and Outlooks

In this thesis, I have had the opportunity to apply the *in vitro* selection technique to search from random-sequence DNA libraries for some very unique single-stranded DNA molecules with interesting functions. I started my project with a goal of developing nucleic acid-based sensors that can facilitate the detection of *clostridium difficile*, a bacterial pathogen that causes wide-spread infections in hospitals in North America and Europe. Along the way, I have developed a deep appreciation of functional capability of single-stranded DNA molecules as well as the power of *in vitro* selection in identifying such sequences from DNA pools containing as many as 10^{15} different sequences.

The goal of the work presented in Chapter 2 is to determine whether it is possible to use the strategy of combining a positive-selection step and a counter-selection step to achieve the isolation of chimeric RNA/DNA molecules that can be selectively cleaved by the RNase H2 from *C. difficile* while exhibiting significantly reduced activities towards analogous RNase H2 enzymes from two control bacteria, *E. coli* and *Salmonella*. The idea was proven working as I was able to isolate several chimeric RNA/DNA molecules with high selectivity for the RNase H2 from *C. difficile*. This study has strengthened my appreciation of the power of the *in vitro* selection technique, as it can deliver an outcome that rational designs are unable to do. Two important implications come out of my findings. The first implication is that the high selectivity of the selected chimeric nucleic acid molecules towards the RNase H2 from *C. difficile* can be further utilized for the development of selective biosensors for this important bacterial pathogen. Given the

isolated nucleic acids are programmed to produce a strong fluorescence signal upon RNase H2 mediated cleavage, they can be used immediately as a fluorescence biosensor. Alternatively, they can also be designed into colorimetric biosensors using several strategies discussed in Chapter 1 of this thesis that have been successfully implemented for nucleic acid cleavage based systems (such as DNAzymes). However, given the fact that there are many different bacteria in any given biological sample, as a future direction, these nucleic acid molecules need to be further examined for cleavage activity by RNase H2 enzymes from more control bacteria. This will further evaluate their usefulness as biosensing systems for the detection of *C. difficile*. The second implication, also an important future research direction, is that the same approach can be adopted for isolating nucleic acid substrates highly selective for another nuclease (an endonuclease and another RNase) from *C. difficile* and minimally cross-reactive for the same type of nucleases from control bacteria. With the choice of a proper nuclease as the targeted enzyme and the choice of the same enzyme from control bacteria, it should be possible to obtain a set of nucleic acid molecules that together can provide the needed specificity for the detection of *C. difficile*.

Chapter 3 describes the unexpected discovery of a unique G-rich sequence that still is able to fold into a parallel G-quadruplex structure in the presence of 7 M urea, a condition known to denature nucleic acid structures. It is an accidental discovery due to the alteration of a common experimental procedure in selecting chimeric RNA/DNA molecules with cleavage activities. In a standard procedure, after the cleavage reaction is done, EDTA is added to chelate divalent metal ions and ethanol precipitation is performed to concentrate both the cleavage products and uncleaved starting material. The concentrated DNA is

dissolved in 7 M urea and subjected to denaturing (7 M urea) polyacrylamide gel electrophoresis so that the cleaved products (which are shorter) can be purified from the uncleaved starting material. I skipped the EDTA and ethanol precipitation steps; instead the reaction mixture was directly made into a solution containing 7 M urea. This alteration allowed for the emergence of several G-rich sequences that took advantage of the presence of Mg^{2+} and Na^+ in the urea containing solution to create a folded structure that has the same gel mobility exhibited by the cleaved products in 7 M polyacrylamide gel. In other words, these uncleaved starting chimeric RNA/DNA sequences use structural folding to create a structured DNA with a gel mobility equivalent to that of the cleaved products and get themselves selected. I was quite amazed at what some single-stranded DNA molecules are capable of doing. One would not be able to rationally predict the existence of such miraculous DNA species, but again “irrational” in vitro selection did the trick. As a future research idea, I am profoundly interested in learning whether it is possible to isolate DNA molecules with well-folded structures under other challenging conditions, such as high acidity, high basicity, high temperature, solutions containing high content of organic solvents, to list a few.

New basic scientific discovery often leads to new applications. Chapter 4 provides an example of how the urea-resistant G-quadruplex structures can be utilized as a highly unique nucleic acid detection tool. Specifically, we have re-engineered the G-quadruplex into a molecular switch that is able to detect a single nucleotide polymorphism (SNP) in a targeted DNA sequence containing as many as 90 nucleotides in the presence of 7 M urea. In addition, this sensing method can provide both measurable fluorescent signal and color

change that is visible to the naked eye. With further optimization, we hope this unique molecular switch can find useful application in genetic disease diagnosis.

In conclusion, my thesis work continues to showcase the power of in vitro selection technique. In the process, we have discovered two very interesting classes of functional nucleic acid molecules that hold some promise for the development of unique biosensors for unique applications.