EXPLORING THE EVOLUTION PROCESS AND REQUIREMENTS FOR MOLECULAR RECOGNITION *VIA* HIGHLY FUNCTIONALIZED APTAMERS

MATINA MOVAHEDI

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

September 2024

ABSTRACT

Expanding the chemical diversity of oligonucleotide libraries has allowed the evolution of synthetic nucleic acid polymers with enhanced molecular recognition and catalytic abilities. Thus, synthetic methods that enable the sequence-defined incorporation of diverse chemical modifications in developing novel or improved nucleic acid polymers and selection methods that facilitate efficient enrichment of high-quality aptamers are of particular importance in identifying novel or improved nucleic acid polymers for diagnostics and therapeutics. In this thesis, the advancement of ligase-catalyzed oligonucleotide polymerization (LOOPER) is discussed as a method to increase the chemical diversity of oligonucleotide libraries and its application towards the evolution of modified aptamers. An evaluation of the use of different ligases, scope and number of modifications, sequence space, and evolutionary outcomes from in vitro selections is provided, along with a critical lens on challenges to be addressed for the method to mature into a more widely adapted technology. Further to this, a variety of selection methods are discussed striving towards efficient single-round aptamer selection. The successful single-round aptamer selection against Thrombin is an inspiring outcome for future selections.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Ryan Hili, for his invaluable guidance, support, and encouragement throughout this journey. His expertise and dedication have been instrumental in shaping this dissertation. Dr. Hili's commitment to fostering an environment of intellectual curiosity and rigorous scholarship has profoundly influenced my academic growth. His patience, thoughtful feedback, and unwavering belief in my abilities have been a constant source of motivation. I am truly fortunate to have had the opportunity to learn from and work with such an inspiring mentor.

I am also profoundly thankful to my committee members, Dr. Sergey Krylov and Dr. Philip Johnson, for their insightful feedback and advice, which greatly contributed to the development of my research. A special thank you to the Krylov lab for their generous assistance, providing access to the qPCR and the Ion Gene Studio S5 DNA sequencer, which were crucial to my experiments. I would also like to extend my gratitude to An T. H. Le for the fruitful collaborations that enriched this research.

A heartfelt thank you goes to my fellow lab members, Dr. Yasaman Mahdavi-Amiri, Natalie Khamissi, Kimberley Chung, Molly Hu, Morgan Cauley-Le Fevre, Ian Frankel, Emily Anacleto, and Luca Di Fabio, and Nicole Frias whose support and friendships made this experience both rewarding and enjoyable. I am grateful to my undergraduate mentees, Dickson Chen, Tristan Tiburcio, and Kristina Gremi, for their enthusiasm and hard work, which have been a source of inspiration. I am especially appreciative of Natalie Khamissi for our RNA LOOPER collaboration and Christopher Korfmann for his time and dedication to computational analysis of sequencing data.

I would like to acknowledge the incredible academic experience I've had and express my appreciation for the opportunity to engage in such a meaningful and transformative journey. The challenges and learning experiences encountered along the way have been invaluable, and I am deeply thankful for the support that has enabled my growth both academically and personally.

Lastly, to my supportive family and friends, I would like to extent my utmost appreciation to my parents: Maria Etminan and Abdolreza Movahedi, my siblings: Mobina and Sina, and my dearest friends: Nargol Ghazian, Rojin Moghaddas, Sanaz Sakhapour, Lauralee Edwards, and Neushaw Moradi, and Niki Ghomashchi, your unwavering support and encouragement have been a constant source of strength. I am truly blessed to have you in my life.

DEDICATION

To my parents, Maria and Abdolreza, whose boundless love, sacrifice, and encouragement have been the foundation of my academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION TO OLIGONUCLEOTIDE SYNTHESIS FOR THERAPEUTICS

1.1. Nucleic acid polymers

Nucleic acids, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are essential macromolecules that store and translate genetic information in all living organisms.[1,2] Each consists of a phosphate group, a five-carbon sugar (DNA does not have the 2′OH on the ribose sugar), and a nitrogenous base. They can form polymers of nucleotides *via* phosphodiester bonds through the phosphate backbone. DNA encodes the genetic blueprint in a double-stranded form hybridized into a B-helix structure, notwithstanding some viral DNA, while RNA is typically single-stranded enabling protein synthesis and gene regulation.[3,4] The stability of the double-stranded DNA molecule is due to the base-stacking in a helical structure upon complementary base pairing *via* hydrogen bonds between the nitrogenous bases (Figure 1-1).

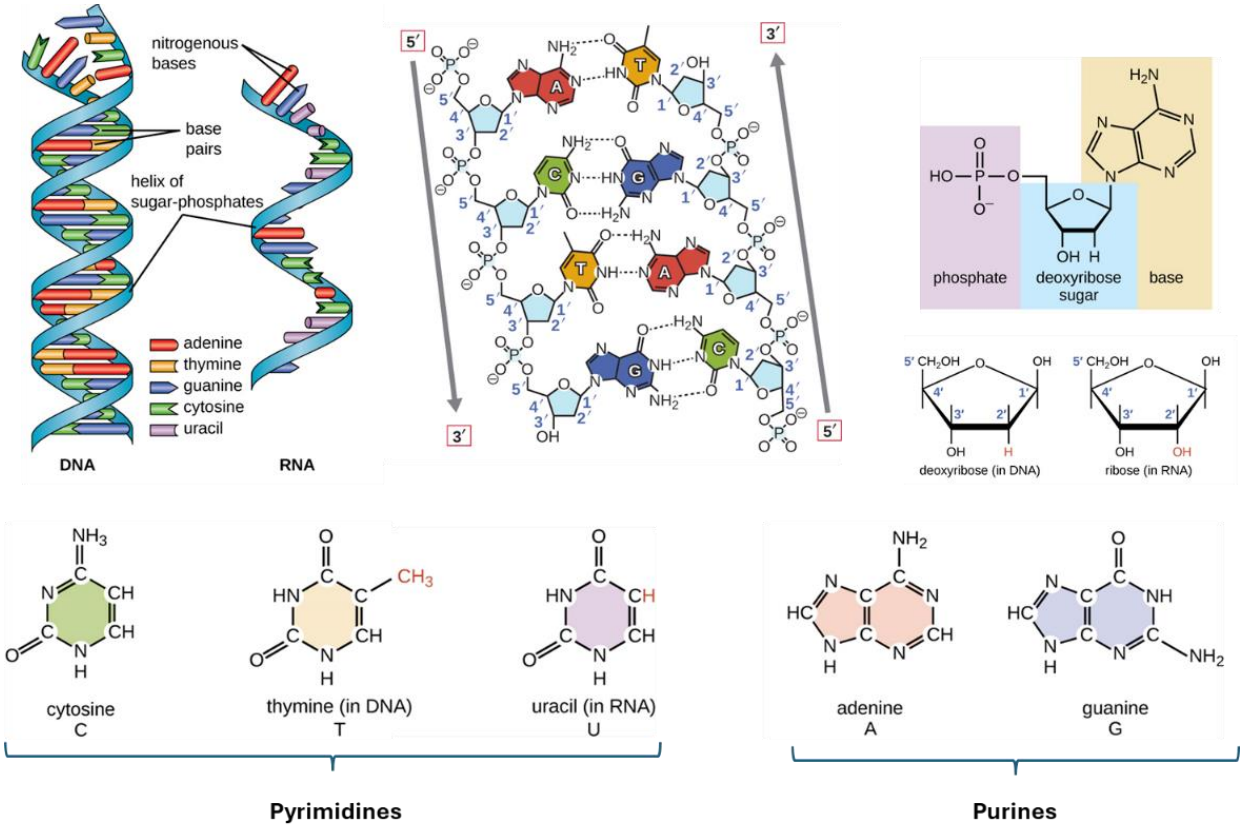


Figure 1-1: Schematic of DNA and RNA nucleobases, Watson Crick base pairing in a double helix. Figure adapted from: Parker et al. Microbiology (2016) from Openstax.

Purines form two hydrogen bonds, adenine (A) pairs with thymine (T), and RNA uses uracil (U) instead of thymine.[4] Pyrimidines form three hydrogen bonds, guanine (G) pairs with cytosine (C). This complementary pairing ensures fidelity during DNA replication, where each strand serves as a template for the synthesis of a new complementary strand. In nature, the DNA sequence is transcribed into mRNA, which is translated into proteins with the help of tRNA (transfer RNA) and rRNA (ribosomal RNA). However, synthetic oligonucleotides are pivotal in biotechnology and biopharma, enabling techniques like PCR[5] (Polymerase Chain Reaction), NGS (Next Generation) DNA sequencing[6], and CRISPR[7,8] (Clustered Regularly Interspaced Short Palindromic Repeats) gene editing, and facilitating the development of therapeutics and diagnostics.

1.1.1. Therapeutic nucleic acids

Within nature's hierarchy of molecular function, proteins have evolved to represent its apex. Equipped with a wide array of functionalities, proteins can achieve exceptionally tight binding, diffusion-controlled catalysis. On the other hand, nucleic acid polymers, which have evolved to display few functionalities as their overarching role remains largely within the domain of information storage and transfer.[9–12] This notwithstanding, vestiges of a potential functional past remain extant in the form of RNAs with binding and catalytic properties.[13,14] These properties prompted groundbreaking work on the laboratory evolution of nucleic acid polymers known as aptamers with binding and catalytic functions.[15–17] Despite decades of research, natural nucleic acids have failed to reach the level of functionality of proteins. Accordingly, bridging the functional gap between proteins and nucleic acids has been a long-standing challenge within the nucleic acids community. The axiom of "diversity begets fitness" is universally accepted in the field of molecular evolution, and certainly, greater chemical space has facilitated the evolution of

functional nucleic acid polymers. However, chemical diversity is the crux of the functional dichotomy between proteins and nucleic acids.

Modified synthetic polymers are generated using creative methodologies, where the functional groups are incorporated into the aptamer sequence either pre- or post-selection. Modifications can be introduced on the base (*e.g.,* 5-methyluridine, pseudouridine, and dihydrouridine), on the sugar (*e.g.,* 2′-O-methyl, 2′-fluoro, 2′-FANA (2′-fluoroarabino nucleic acid), locked nucleic acid (LNA)), or on the phosphate backbone (*e.g.,* morpholino, peptide nucleic acid, phosphorothioate and boranophosphate).[18–20] Some examples include hydrophobic Slow Off-rate (low K_D) SOMAmers[21,22] incorporating 2′-deoxyuridine modified at C5, Thioaptamers[23], X-Aptamers[24], Xeno-nucleic acids or chemically modified unnatural nucleotides on antisense oligonucleotides (ASOs)[25,26] and small interfering RNA (siRNA),[27,28] and Spiegelmers[29–32] which are the enantiomers of natural oligonucleotide built from L-ribose sugars.

The growing field of modified functional aptamers is gradually showing promise for therapy, drug delivery, diagnosis, functional genomics, and bio-sensing. More modified functional aptamers are entering clinical trials. The FDA has approved two RNA aptamers; Macugen (2004) is a modified functional RNA aptamer containing 2′-fluoro and 2′-O-methyl substitutions (Figure 1-2), and Zimura (2023) is a 39-mer aptamer conjugated to a polyethylene glycol group. Both aptamers gained FDA approval for treating age-related macular degeneration. Indeed, expanding the chemical diversity of nucleic acids has been instrumental to improving molecular recognition and catalysis of nucleic acids.[33] While generating modified nucleic acid polymers using polymerases and terminal transferases has enabled the incorporation of up to four different modifications, attaining the level of diversity seen in proteins requires alternative approaches.[34,35]

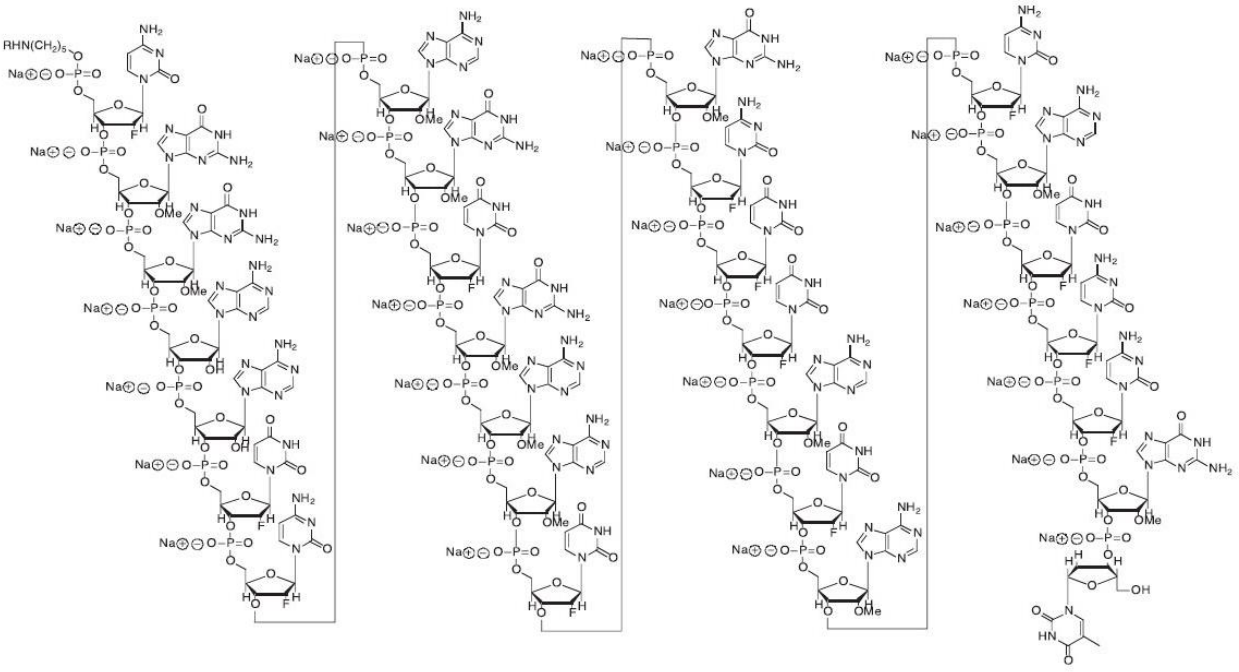


Figure 1-2: Chemical structure of pegaptanib sodium injection in the treatment of neovascular AMD. Figure from: FDA Dermatologic and Ophthalmic Drugs Advisory Committee, NDA 21-756, 2004.

Artificially synthesized DNA and RNA sequences are integral to biopharmaceutical innovation. Recently, their demand has surged with the rise of CRISPR technology and the development of oligonucleotide drugs like ASOs, RNA interference therapies, and aptamers.[36,37] To meet these diverse needs, manufacturers produce synthetic oligonucleotides on scales ranging from micrograms for numerous sequences to kilogram-scale batches for drug products.[38,39] Two primary methods are used to synthesize oligonucleotides: chemical and enzymatic synthesis.[40–42]

1.2. Overview of solid phase oligonucleotide synthesis

The global oligonucleotide market has an estimated net worth of $8.8 billion in 2024.[43] The demand for oligonucleotides is growing, with solid-phase synthesis being the currently predominant method. The solid phase method was first introduced by Robert L. Letsinger and Robert Bruce Merrifield in the 1960s.[44,45] Later the method was improved for oligonucleotide

synthesis *via* phosphoramidites by Marvin Caruthers and earned him the patent titled "Process for Preparing Polynucleotides" issued in 1984.[40] Oligonucleotides can be synthesized on scales ranging from nanomoles for laboratory use to multi-kilogram quantities for industry. The key advantages of solid-phase synthesis include its automation and ability to drive reactions to completion quickly by using excess reagents and washing away impurities between each step. The phosphoramidite method developed by Marvin Caruthers, fundamental to this process, involves the sequential addition of nucleotides in the 3′- to 5′-direction on a solid support.

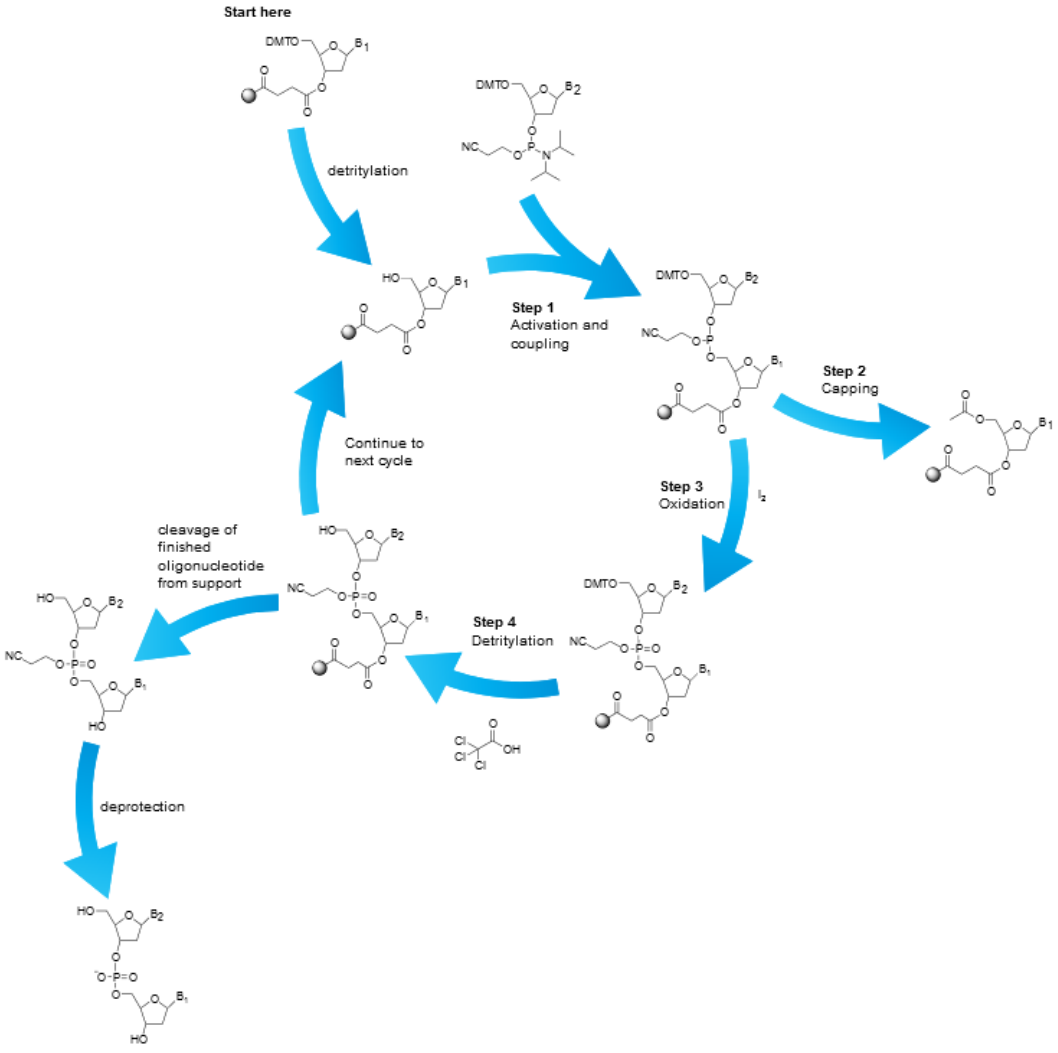


Figure 1-3: The solid-phase phosphoramidite oligonucleotide synthesis cycle. Figure from: Atdbio, nucleic acids book, chapter 5.

The choice of solid support is critical, between controlled pore glass (CPG) and polystyrene (PS) yields may vary based on the required length and scale of oligonucleotide synthesis. CPG, with pore sizes ranging from 75 Å to 4000 Å, enables increased lengths of oligonucleotides to be synthesized based on larger pore sizes.[46] Polystyrene beads, on the other hand, offer excellent moisture exclusion and are effective for short oligonucleotide synthesis on large scales due to their high loading capacity. Each cycle of the synthesis includes several critical steps (Figure 1-3): starting with detritylation to remove the 5′-DMT protecting group, coupling of the phosphoramidite monomer, capping of unreacted 5′-hydroxyl groups to prevent deletion mutations, and oxidation to stabilize the newly formed phosphite triester. The real-time efficiency of the process is monitored by using trityl absorbance at each cycle to estimate coupling yields. Achieving high stepwise yields is essential to minimize cumulative errors over multiple cycles, especially for long oligonucleotides. The golden standard for nucleotide incorporation yields is 99% or higher while ensuring the reagents are pure and the solvents are anhydrous.[38] The final steps following full-length synthesis include cleavage of the linker from the solid support and deprotection. Treatment with concentrated ammonium hydroxide readily cleaves the commonly used succinyl linker, followed by heating to remove the protecting groups from the heterocyclic bases and phosphodiester backbone. This step must be carefully managed to prevent side reactions and ensure the integrity of the synthesized oligonucleotide. Innovations in reagent purity, solid support design, and deprotection strategies have significantly improved the efficiency and reliability of synthesizing complex and modified oligonucleotides.[47] While chemical synthesis excels in producing modified oligonucleotides with high purity, the challenges remain with scalability (<10 kg batches) and sustainability (over 4000 kg of waste from organic solvents and

reagents per kg for a 20 nt long oligonucleotide synthesis).[41] The large volumes of acetonitrile utilized in the multiple washing steps and as a solvent for phosphoramidites raises concerns for supply, costs, and the environmental impact of disposal of bulk volumes.[48]

1.3. Overview of enzymatic oligonucleotide synthesis

Enzymatic methods enable large scale oligonucleotide synthesis in aqueous solutions, offering advantages such as reduced hazardous waste by avoiding toxic organic solvents and the ability to synthesize longer DNA sequences (>1000 nt).[41] This approach leverages enzymes mimicking cellular DNA synthesis processes; naturally occurring terminal transferases (TdT) and polymerases enable template-independent extension of oligonucleotide products by sequentially adding bases from the 5′-end (Figure 1-4).[42,49] The wild-type enzymes are often limited to unmodified bases, while mutated versions of the enzymes allow the incorporation of modified oligonucleotides. However, developing new enzymes and substrates is often necessary to expand enzyme functionality.

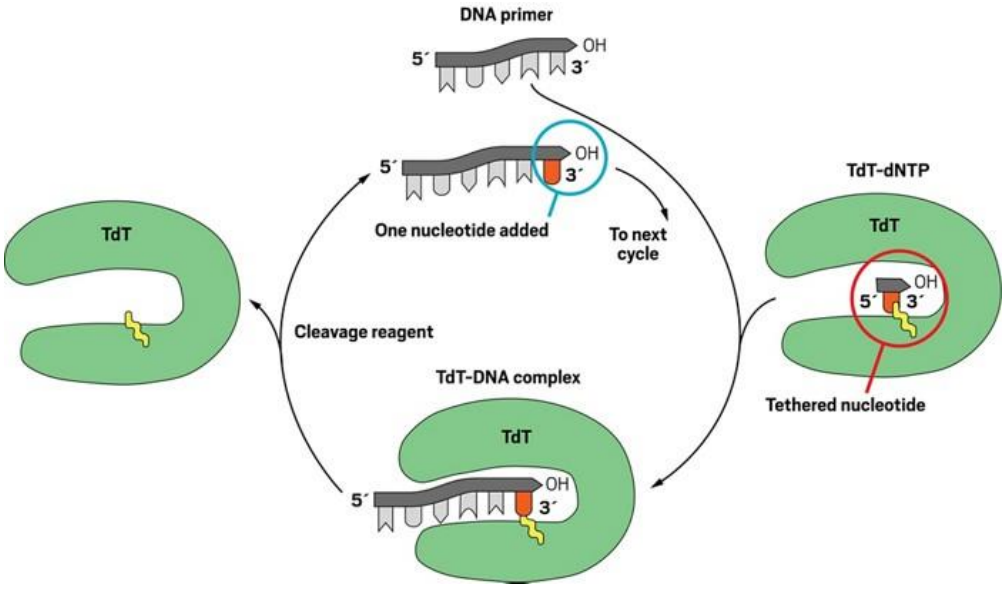


Figure 1-4: Enzymatic synthesis of DNA oligonucleotide *via* TdT; the complex protects the elongating strand. Figure adapted from: Nature biotechnology.[49]

The future of oligonucleotide manufacturing may involve a hybrid approach, combining the strengths of both chemical and enzymatic methods to enhance efficiency, affordability, and precision in oligonucleotide synthesis. However, the challenge remains in synthesizing highly modified diverse oligonucleotide libraries. The number of unique modifications that can be incorporated into aptamer libraries using polymerase-based approaches is theoretically limited to the number of nucleobases that define the genetic code, which canonically is four (A, C, G, and T/U). While up to four modifications have been used in primer extensions, a maximum of three have been successfully implemented during *in vitro* selections.[34,35,50,51] Expanding beyond this limitation requires an alternative approach.

1.4. Concept of Ligase catalyzed OligOnucleotide PolymERization

LOOPER[52–56] (Ligase-Catalyzed Oligonucleotide Polymerization), pioneered in the Hili lab, is a hybrid method for oligonucleotide synthesis, utilizing templates and short-modified oligonucleotides that are synthesized *via* solid-phase chemical synthesis to carry out the enzymatic ligation of highly modified aptamer libraries. To achieve the sequence-defined templated copolymerization of modified anticodon oligonucleotide fragments into diverse libraries of heteromultivalent polymers, LOOPER was developed. LOOPER begins with a library of templates that contain *i*) a reading frame comprising a random combinatorial sequence of codons, whereby each codon is the same length, and *ii*) flanking initiation and termination primers, which define the reading frame and are required for amplification during *in vitro* evolution (Figure 1-5). During LOOPER, a 5′-phosphorylated primer anneals to the initiation primer site, which is then adenylated by a DNA ligase and functionalized anticodons are ligated in a DNA-templated manner along the reading frame. Subsequently, the duplex DNA can be strand separated to isolate the diversely modified ssDNA, which can be ported directly into traditional *in vitro* selection platforms.

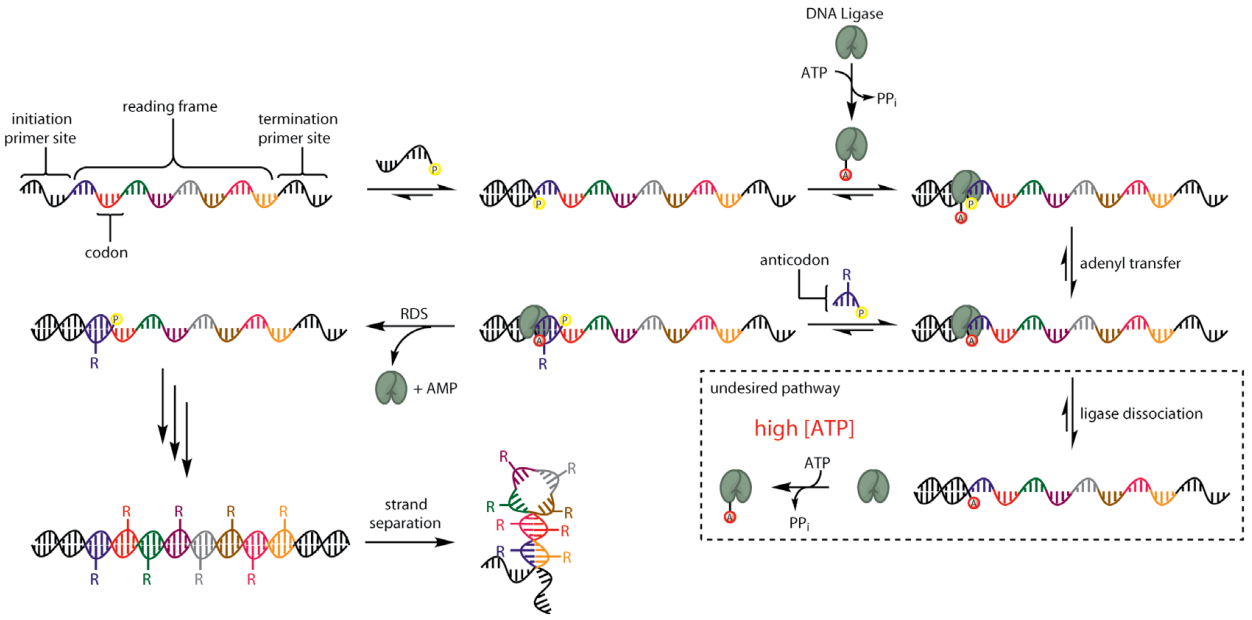


Figure 1-5: General process for ligase Catalyzed Oligonucleotide Polymerization (LOOPER). A DNA template library contains two primer binding sites, and a reading frame comprising multiple instances of codons are a specific length. Upon annealing of a primer, a DNA ligase catalyzes the phosphodiester bond formation of cognate anticodons with encoded modifications in a DNA-templated manner. Following LOOPER, modified dsDNA libraries can be strand separated to provide modified ssDNA libraries for in vitro selection. An undesired pathway (dotted line box) occurs with the ATP concentration is too high, resulting in truncated products.[56]

Thus, LOOPER enables access to highly functionalized aptamer libraries through sequence-defined heteromultivalent display of diverse modifications. The number of modifications increases according to the size of the oligonucleotide fragment, such that the number of unique modifications equals 4^n, where n is the length of the oligonucleotide fragment. Thus, a trinucleotide can theoretically encode 64 modifications, whereas a pentanucleotide can accommodate up to 1024, which occurs at the expense of modification density. Due to the critical role played by the DNA ligase in LOOPER, different ligases have been used, including *E. coli* DNA ligase and T3, T4, and T7 DNA ligases, each demonstrating their unique strengths and weaknesses for the process.[52,57–59]

1.4.1. LOOPER *via* T4 DNA ligase

T4 DNA ligase has been the standard ligase used for the LOOPER method due to its commercial availability, high fidelity and specificity, compatibility with short oligonucleotides, and tolerance for oligonucleotide modifications. At the outset of the development of LOOPER, a crystal structure of T4 DNA ligase had yet to be reported, so initial efforts were devoted to systematically understanding the anticodon preferences of T4 DNA ligase during LOOPER.[60] Several key early findings were critical to subsequent developments and implementation of LOOPER to *in vitro* selection. First, while T4 DNA ligase exhibited some activity with very short anticodons, such as trinucleotides, pentanucleotides were, by and large, the optimal length, with longer anticodons being prone to annealing and ligating out of frame, resulting in a low yield of desired full-length products.[52,61] Second, in a pentanucleotide anticodon, modifications on the Hoogsteen face of adenine, typically *via* position 8 (Figure 1-6), were tolerated at any position except the 3′-end. However, modifications of other nucleobases were poorly tolerated, including modifications *via* the 5′-position of cytidine or thymine, or the N^2-position of guanine.[62]

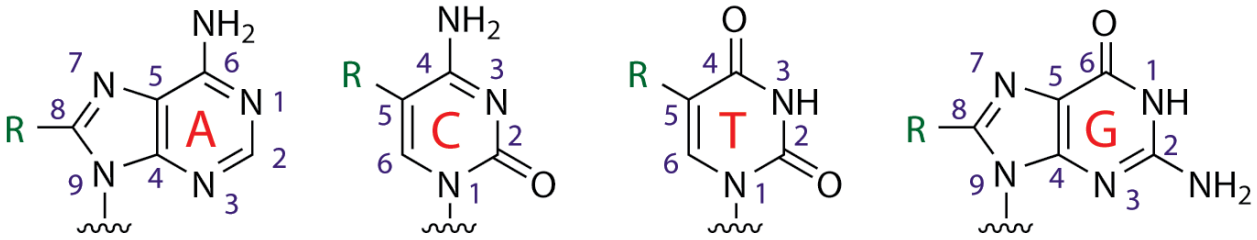


Figure 1-6: Common modification sites on nucleobases tested in LOOPER. "R" denotes modification site.[56]

Third, high ATP concentrations resulted in a precipitous drop in polymerization efficiency, particularly with larger modifications relative to the linker alone. This is hypothesized to result from dissociation of T4 DNA ligase from the adenylated duplex and subsequent readenylation of T4 DNA ligase by excess ATP. Since T4 DNA ligase must be unadenylated to catalyze rate-limiting

phosphodiester bond formation, high ATP concentrations push the equilibrium to an over-adenylated state. To offset this issue, the ATP concentration had to be carefully controlled, with the optimal concentration found to be 25 μM.[62] Fourth, while LOOPER can proceed bidirectionally, as opposed to the unidirectional nature of polymerases, T4 DNA ligase exhibited considerably faster kinetics extending from the 3′-OH.[58] Fifth, in pentanucleotide anticodons, T4 DNA ligase tolerates a large scope of functional groups and can accommodate large modifications such as octapeptides with a broad range of charge and hydrophobicity.[63] Sixth, PEG 6000 was essential as a molecular crowding reagent to increase the effective molarity and thus the rate of LOOPER polymerization reactions.[52] These early findings were essential for the rational design of codon sets to be used in LOOPER and for porting the method into in vitro selections.

1.4.2. LOOPER *via* T3 DNA ligase

While T4 DNA ligase has been successful as a leading ligase in LOOPER, it suffers from several shortcomings, including: *i*) LOOPER efficiency with anticodons less than five nucleotides in length is low, which results in modest density of modifications; and *ii*) modifications are mostly limited to position-8 of adenine, limiting the sequence space explored during evolutions. The use of T3 DNA ligase has partially addressed these issues. T3 DNA ligase was found to be more efficient than T4 and T7 DNA ligases at LOOPER with modified trinucleotides, with full-length products observed for up to 45-nt reading frames (15 codons).[59] T3 DNA ligase was also effective at polymerizing unmodified NNN, NNNN, and NNNNN anticodons with yields of 60 % (13 codons), 89 % (10 codons), and 65 % (8 codons), respectively.[61] Similar to T4 DNA ligase, modifications were optimal on the 5′-end of the anticodon. The accommodation of various small polar and hydrophobic groups has been shown; larger groups have yet to be explored.[59] It is important to note that since T3 DNA ligase can only tolerate small modifications in trinucleotide

systems, custom-modified phosphoramidites are required, albeit this may not be necessary for longer codon sets.[59] Interestingly, T3 DNA ligase has been shown to only tolerate modification *via* the 5-position of cytosine and thymine, thus trinucleotide codon sets have been restricted to a library size of 32 (NNY, where Y = C or T). This codon set was converted into an anticodon set with eight modifications, whereby each modification was represented by four anticodons. This codon/anticodon set has been successfully used for in vitro selections of aptamers. Using Sanger sequencing on a single template, it was qualitatively shown that T3 DNA ligase has good fidelity with a modified 32-membered anticodon library; using duplex DNA sequencing, T3 DNA ligase was later shown to have 97 % fidelity with either unmodified or modified YNN anticodons.[59,61] T3 DNA ligase can handle larger anticodon sets, such as modified YNNN with 92 % fidelity; however, it struggles with pentamer anticodon sets, exhibiting only 74 % fidelity with modified YNNNN, which is too low for in vitro selection.

1.5. Thesis project

The search for high-quality aptamers remains a challenge in the oligonucleotide field, due to the limited diversity of aptamer libraries and the low success rate of *in vitro* selection methods. This thesis will discuss the development, implementations, pitfalls, and outlook for using ligases to achieve greater diversity in aptamer libraries and improved methods for *in vitro* aptamer selections. First the synthesis of highly modified functionalized libraries of aptamers *via* LOOPER is explored for RNA libraries, followed by *in vitro* selections against Thrombin using LOOPER-SELEX and LOOPER-CE, and finally structured aptamer libraries are used to carry out single-round selections. The LOOPER-SELEX and LOOPER-CE selections demonstrate challenges as discussed in Chapter 3 resulting in failed selection, however single-round selections introduced in Chapter 4 demonstrate success *via* a G-quadruplex DNA library selection against Thrombin.

1.6. Contributions

Chapter 2 was a collaboration with Natalie Khamissi from the Hili lab. The work towards expanding LOOPER for the synthesis of modified RNA aptamer libraries was divided evenly; every optimization was carries out with an even breakdown of experiments.

Chapter 3.3 was a collaboration with Dr. An T. H. Le from the Krylov lab. The work towards developing LOOPER-CE for selection *via* modified aptamer libraries was divided based on our specialized expertise. The synthesis of TBL1, synthesis of the starting LOOPER library, qPCR amplifications, PCR amplifications, LOOPER in between rounds of selection, and sequencing preparations were carried out in the Hili lab by me. While Dr. Le carried out the initial CE optimizations with TBL1, selections for three rounds, and affinity analysis of outputs on CE.

CHAPTER 2. HIGHLY MODIFIED LOOPER APTAMER LIBRARIES

2.1. Introduction: LOOPER libraries

LOOPER libraries are assembled *via* ligation of the anticodon sublibraries upon annealing to the corresponding sequence along the reading frame of the template (Figure 2-1). In designing a codon set for LOOPER, there are four essential requirements: *i*) tolerance of modifications on polymerized oligonucleotides; *ii*) high yield of full-length product; *iii*) broad coverage of sequence space; and *iv*) high-fidelity and low codon bias of DNA-templated polymerization. For the first two requirements, gel analysis can demonstrate yields of full-length products. With the development of a sequencing method that allows for template and polymer strands to be directly compared, the fidelity of LOOPER could be accurately calculated for various codon sets and anticodons, allowing for the optimization of the method ahead of its use for *in vitro* selections.[58] The location of a modification on an anticodon influences more than fidelity, it also has a large impact on codon bias during LOOPER. Codon bias is concerned with the frequency of a codon in the template compared to its anticodon frequency in the product strand. Low codon bias is desirable and requires near equal frequencies of codon/anticodon in the template and corresponding product. High codon bias can negatively impact selections as it convolutes the enrichment of sequence space. Generally, codon sets that result in a high yield of full-length products have low codon bias. When using a library of modifications, the choice of how to encode the modifications becomes important. The directionality of the ligase, to extend from the 3′-end versus the 5′-end can explain the location of errors relative to the ligation site. Utilizing insight about errors during *in vitro* evolutions, encoding of anticodon modifications is designed accordingly to fall at the 3′-end since the commonly occurring errors at the 3′-end enable missense mutations.[55,64] These mutations allow the functionally important R-groups to readily evolve, while maintaining a balanced anticodon bearing selection pressure.

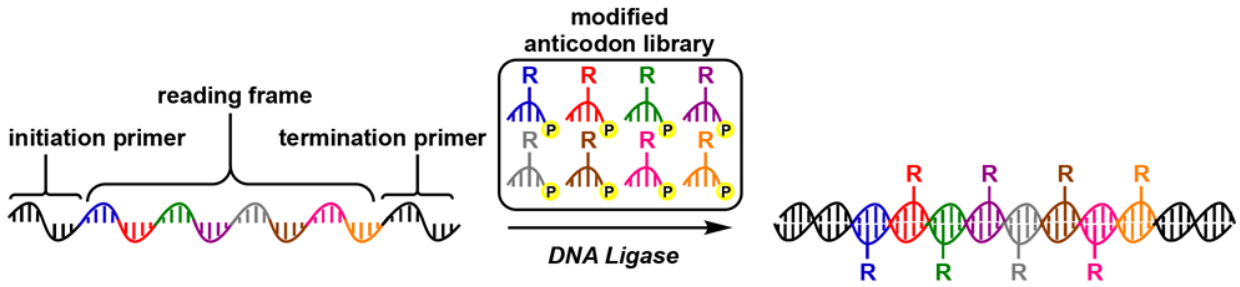


Figure 2-1: LOOPER library synthesis *via* modified anticodon sublibraries.[53]

2.2. Developing LOOPER for RNA oligonucleotides

RNA-based therapeutics are paramount in the future of the pharmaceutical industry with their potential for strategic innovation.[36,37,65,66] Advancements in the field such as incorporation of modifications for increased stability, and delivery methods *via* liposome carriers efficiently protect RNA drugs from degradation by nucleases, improving their longevity.[36,66–68] The potentially long-lasting effectiveness of therapeutic RNA drugs is represented by the recently developed drug Inclisiran, a synthetic siRNA capable of targeting proprotein convertase by inhibiting the translation of the PCSK9 protein.[69,70] Inclisiran was approved by the FDA in December 2021 for the treatment of high low-density lipoprotein (LDL) cholesterol and atherosclerotic cardiovascular disease (ASCVD).[71] The inhibition effect of Inclisiran can last up to six months post-injection, accommodating less frequent administration intervals which can enhance patient experience.[70] The continuing success of RNA therapeutics relies upon curated methods to overcome challenges with low stability and susceptibility to degradation.[72] Chemical modifications introduced along the RNA sequence facilitate lower immunogenicity and enhanced stability, affinity, and specificity toward a target.

The LOOPER method, previously developed for DNA libraries of highly modified functionalized aptamers is herein explored, in collaboration with Natalie Khamisi, to enable the synthesis of modified RNA therapeutics for downstream epitranscriptomic targets. This collaboration was an

equal contribution toward all experimental optimizations and sequencing preparations. The aim is to challenge the synthetic boundaries of various ligases to facilitate the polymerization of modified RNA libraries. Development of the RNA LOOPER method enables the incorporation of multiple modification sites to create a diverse array of chemically modified sequences, improving the fragile nature of RNA while expanding the druggable landscape of RNA therapeutics.

2.3. Modified RNA aptamer libraries

There are over 160 known epigenetic modifications in mammalian cells, specifically posttranscriptional chemical alterations to the RNA transcriptome, significantly impact the biogenesis, structure, and function of RNA.[73] Among these, methyl modifications such as 5-methylcytosine (m^5C), N^1-methyladenosine (m^1A), and N^6-methyladenosine (m^6A) are commonly studied for their diverse cellular functions (Figure 2-2).[74]

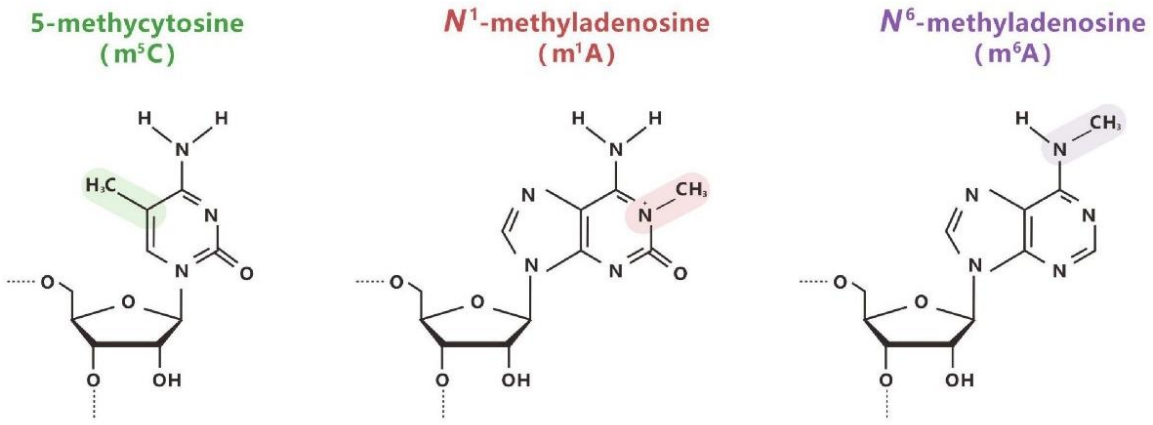


Figure 2-2: Structures of common posttranscriptional RNA modifications. Figure adapted from: Frontiers in cellular neuroscience.[75]

The m^6A modification specifically stands out based on frequency and reversibility. Incorporation of the methyl group is assigned to writers, methyltransferase-like proteins such as METTL3 and METTL14 complex with Wilms tumor 1-associated protein (WTAP).[76,77] The reverse is achieved

by erasers, demethylases such as fat mass and obesity-associated protein (FTO), and AlkB homolog5 (ALKBH5) that target m^{6}A sites and convert them back to adenosine.[78,79] Simultaneously, readers such as the YTH family proteins are key players in epigenetic regulation; YTHDF1-3 and YTHDC1-2 are known readers of m^{6}A-containing transcripts.[80] They participate in RNA splicing, export, translation, and decay *via* recognition of *N*^{6}-methyladenosine found within the DRACH consensus motif (D=G/A/U, R=G/A, H=A/C/U) enriched along coding sequences (CDS) and 3′untranslated region (3′UTR) of genes.[81–83]

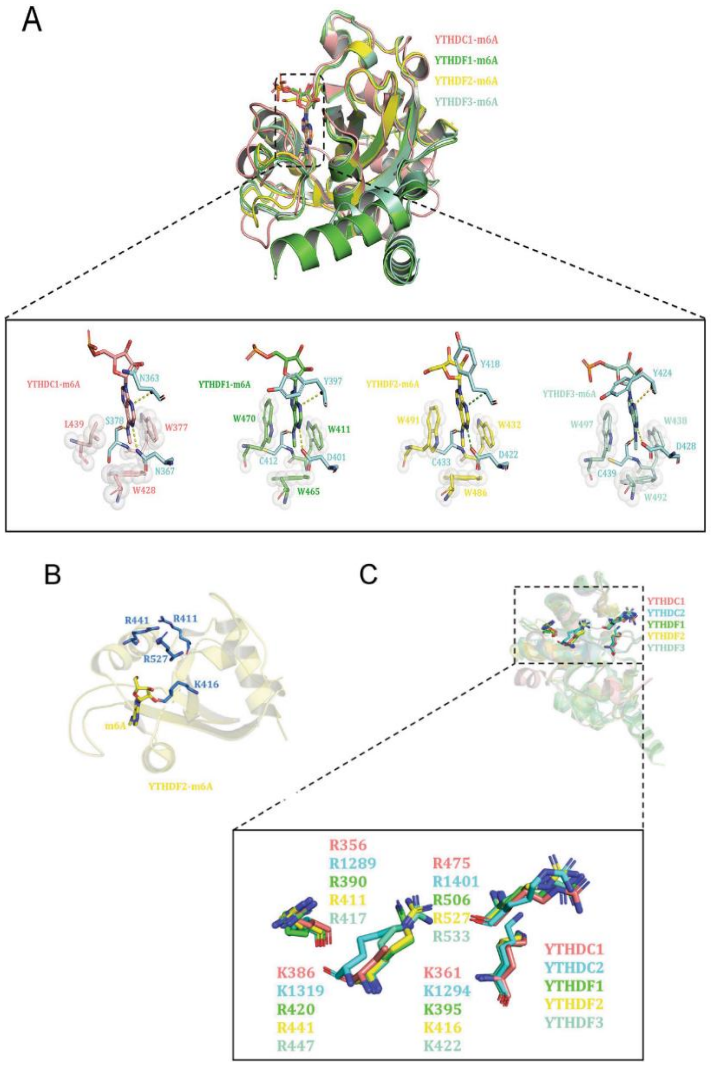


Figure 2-3: Superposition of the conserved YTH domain for m^{6}A recognition. Figure adapted from: Cell death and disease.[84]

The YTH family proteins contain a conserved domain for m^6A recognition. This YTH domain maintains a characteristic β-barrel hydrophobic core surrounded by positively charged basic residues forming a concave surface of α-helices (Figure 2-3).[85,86] The combination of π-π interactions with the adenosine base and hydrophobic interactions with *N*-methyl moiety *via* the aromatic cage, along with stabilizing hydrogen bond formation through neighbouring side-chain residues enable selective binding interactions with N^6-methyladenosine modified RNA.[87] Recent studies highlight the role of YTHDF1 and YTHDF2 in disease implications and various cancers.[88–90] YTHDF1 can favour the translation of certain genes and alter the proteome in cancer cells, promoting cell proliferation and tumourigenesis, making it a promising target for cancer therapy.[91,92] Incorporating methyl modifications in synthetic RNAs can reduce Toll-like receptor (TLR) activation, avoid an immune response, and play a critical role in the translation of these molecules for clinical applications.[93]

2.4. Materials and methods

2.4.1. General Information

Unless otherwise noted, water was purified with the Direct-Q® 3 UV Water Purification System. All degenerate RNA anticodon libraries (rN3, rN4, rN5) and RNA primers were purchased from Horizon Discovery. All DNA templates and primers and m6A modified RNA anticodons (m6ArNrNrNrN, rNm6ArNrNrN, rNrNm6ArNrN, rNrNrNm6ArN, rNrNrNrNm6A) were purchased from Integrated DNA Technologies.

2.4.2. Primers and template sequences

Poly_pr1-PEG

5′-phos/TGCGACGGCAGGCGAATC/iSp18/AACAACAACAACAA3′

3P-F

5′/56-FAM/AGGATCCGAGCTCCACGTG3′

5′P-rPrime1

5′-P-UGCGACGGCAGGCGAAUC3′

FL-rPrime2

5′/56-FAM/GGAUCCGAGCUCCACGUG3′

N39temp_peg

5′GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA

N40temp_peg

5′GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNNCACGTGGAGCTCGGATCC/iSp18/AACAACAACAACAA

PEG-CONT-39MER-TEMP

GATTCGCCTGCCGTCGCAGGTACTGTCGTTTTAGCAGTTTTTGTGCAATGGGTTAGGC

ACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA-3′

PEG-CONT-40MER-TEMP

GATTCGCCTGCCGTCGCAGGTACTGTCGTTTTAGCAGTTTTTGTGCAATGGGTTAGGT

CACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA-3′

CONT-39MER-RNA

5′-phos/CCUAACCCAUUGCACAAAAACUGCUAAAACGACAGUACC-3′

CONT-40MER-RNA

5′-phos/ACCUAACCCAUUGCACAAAAACUGCUAAAACGACAGUACC-3′

N40-5′pr

5′NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCG

GATCCT3′

N40-3′pr

5′GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

NNNNT3′

N40-TNNNN

5′GATTCGCCTGCCGTCGCATNNNNTNNNNTNNNNTNNNNTNNNNTNNNNTNNNNTN

NNNCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA3′

N40-NTNNN

5′GATTCGCCTGCCGTCGCANTNNNNTNNNNTNNNNTNNNNTNNNNTNNNNTNNNNT

NNNCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA3′

N40-NNTNN

5′GATTCGCCTGCCGTCGCANNTNNNNTNNNNTNNNNTNNNNTNNNNTNNNNTNNNN

TNNCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA3′

N40-NNNTN

5′GATTCGCCTGCCGTCGCANNNTNNNNTNNNNTNNNNTNNNNTNNNNTNNNNTNNN

NTNCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA3′

N40-NNNNT

5′GATTCGCCTGCCGTCGCANNNNTNNNNTNNNNTNNNNTNNNNTNNNNTNNNNTNN

NNTCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA3′

3P-F_RNAlooper

5′/56-FAM/AGGCACGGCGAGCGATCTrGrGrArUrCrCrGrArGrCrUrCrCrArCrGrUrG3′

RNA_5HP-N20

5′/phos/rUrGrCrGrArCrGrGrCrArGrGrCrGrArArUrCCAATGATTCGCCTGCCGTCGCANNN

NNNNNNNNNNNNNNNN/6FAM/ 3′

RNA_3HP-N20

5′/6FAM/NNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCCCAATrGrGrArTrCr

CrGrArGrCrTrCrCrArCrGrTrG 3′

5HP_RNAlooper_N40

5′/phos/TGCGACGGCAGGCGAATCCAATGATTCGCCTGCCGTCGCANNNNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCCGCGTGCGGA

CTCCAGCTA3′

5HP_RNAlooper_N40-CONT

5′/phos/TGCGACGGCAGGCGAATCCAATGATTCGCCTGCCGTCGCAGGTACTGTCGTTT

TAGCAGTTTTTGTGCAATGGGTTAGGTCACGTGGAGCTCGGATCCGCGTGCGGACTC

CAGCTA3′

2.4.3. RNA LOOPER via T4 DNA ligase

In a PCR tube, a mixture of 5 µL ligation reaction buffer 4× (40 mM $MgCl_2$, 24% w/v PEG6000, 40 mM DTT, 264 mM Tris pH 7.6), 2.25 µL primer1 (5′P-rprime1 or Poly_Pr1, 10 µM in H_2O),

2.25 µL primer2 (FL-rprime2, 10 µM in H_2O), 1.5 µL template (PEG-CONT-39/40MER-TEMP, 10 µM in H_2O), 1.95/1.5/1.2 µL of rN3/rN4/rN5 anticodons accordingly (7.5 equivalents at 1 mM in H_2O) or 2 µL template (PEG-CONT-39/40MER-TEMP, 10 µM in H_2O) for the control reaction and 2 µL ATP (0.25 mM in H_2O) was brought to 18 µL total volume by the addition of H_2O and heated to 65 °C for 3 minutes then 0 °C for 2 minutes to anneal. This was followed by the addition of 1 µL BSA (2 mg/mL in H_2O) and 1 µL 400 U T4 DNA ligase (New England Biolabs, M0202L). The mixture was allowed to polymerize for 18 hours at 25 °C. Purification was carried out using the Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

2.4.4. RNA LOOPER via T3 DNA ligase

In a PCR tube, a mixture of 5 µL T3 DNA ligase buffer (2X), 0.5 µL primer1 (5′P-rprime1 or Poly_Pr1, 30 µM in H_2O), 0.5 µL primer2 (3PF, 30 µM in H_2O), 1 µL template (PEG-CONT-39/40MER-TEMP, 10 µM in H_2O), 1.3/1/0.8 µL of rN3/rN4/rN5 anticodons accordingly (1 mM in H_2O) or 0.5 µL of RNA (CONT-39/40MER-RNA, 30 µM in H_2O) for the control reaction was brought to 9 µL total volume by the addition of H_2O and heated to 65 °C for 3 minutes then 0 °C for 2 minutes to anneal. This was followed by the addition of 1 µL 3000 U T3 DNA ligase (New England Biolabs, M0317L). The mixture was allowed to polymerize for 18 hours at 25 °C. Purification was carried out using the Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

2.4.5. RNA LOOPER via T4 RNA Ligase II

The control reaction was successful without optimizations. In a PCR tube, a mixture of 2 µL T4 Rnl2 Reaction Buffer (10X), 3 µL primer1 (5′P-rprime1 or Poly_Pr1 or Poly_Pr1_PEG, 10 µM in H_2O), 3 µL primer2 (FL-rprime2, 10 µM in H_2O), 2 µL template (PEG-CONT-39/40MER-TEMP, 10 µM in H_2O), 1.6 µL $MgCl_2$ (100 mM in H_2O), 4 µL 50% PEG 8000, 0.4 µL H_2O and 3 µL RNA

(CONT-39/40MER-RNA, 10 μM in H_2O) was heated to 65 °C for 3 minutes then 0 °C for 2 minutes to anneal. This was followed by the addition of 1 μL 10 U T4 RNA ligase II (New England Biolabs, M0239L). The mixture was allowed to polymerize for 18 hours at 25 °C. Purification was carried out using the Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

The RNA LOOPER reaction with multiple ligations was optimized using the rN5 anticodon system. In a PCR tube, a mixture of 2 μL T4 RNA ligase buffer (10X), 3 μL primer1 (5′P-rprime1, 10 μM in H_2O), 3 μL primer2 (FL-rprime2, 10 μM in H_2O), 2 μL template (template, 10 μM in H_2O), 2.4 μL of rN5 anticodon (15 equivalents, 1 mM in H_2O), 6 μL of PEG6000 (50%), and 0.6 H_2O was heated to 65 °C for 3 minutes then 0 °C for 2 minutes to anneal. This was followed by the addition of 2 μL T4 RNA ligase II (New England Biolabs, M0239L). The mixture was allowed to polymerize for 6 hours, starting at 33 °C for 2 h and then 30 °C for 4 h. The products were then purified into 20 μL of H_2O with a Monarch® RNA Cleanup Kit (#T2030L).

The LOOPER products were then DNase I digested to remove all template sequences. In a PCR tube, 20 μL of sample was added, along with 10 μL of DNAse I Reaction Buffer (10X, New England Biolabs, M0303), 2 μL of DNAse I (New England Biolabs, M0303), and 68 μL of nuclease-free water. The mixture was incubated at 37 °C for 10 minutes. It was then purified with Monarch® RNA Cleanup Kit (#T2030L) and eluted in 7 μL of water before PAGE analysis.

2.4.6. RNA LOOPER sequencing preparation

RNA LOOPER was carried out using the optimized T4 RNA ligase II method, without DNase I digestion, using an HP-seq hairpin template. The purified product was confirmed by 10 % denaturing PAGE gel (150 V, 55 °C, 45 min), and the full-length product was gel extracted. The extracted band was crushed in a microcentrifuge tube and incubated overnight in 100 μL of 0.3 M NaCl at 37 °C. The elution was purified using Centri-sep spin columns, loaded with 60 mg of

Sephadex® G-50 beads (S5897-25G), and dried on the speedvac. The purified RNA HP-seq products were diluted in 12 µL of H_2O, 2 µL of 3Prev-Seq primer (50 µM in H_2O), 1 µL of dNTPs (10 mM in H_2O), 4 µL 5× Induro RT reaction buffer, 1 µL of Induro RT (200 U/mL) and incubated at 55 °C for 30 min followed by 95 °C for 1 min. The product was purified in 30 µL of H_2O using the Omega Kit. The sample was dried to 10 µL on the speedvac to allow quantification using the Qubit fluorometer.

The cDNA was barcoded by PCR amplification using Q5 polymerase. In a microcentrifuge tube a mixture of 10 µL cDNA (10 pM in H_2O), 10 µL H_2O, 2.5 µL of 3Prev-Seq primer (10 µM in H_2O), 2.5 µL of ioncode barcoded primer (10 µM in H_2O), and 25 µL of Q5® high-fidelity 2× master mix (NEB# M0494) was PCR amplified for 16 cycles (98 °C for 2 min, 98 °C for 10 sec, 55 °C for 30 sec, 72 °C for 30 sec, go to step 2 for 1×, 98 °C 10 sec, 71 °C for 30 sec, 72 °C for 30 sec, go to step 6 for 14X). The PCR product was purified in 30 µL of H_2O using the Omega Kit and dried on the speedvac. The double-stranded product was confirmed by 10 % native PAGE gel (150 V, 100 min), and the full-length product was gel extracted. The extracted band was crushed in a microcentrifuge tube and incubated overnight in 100 µL of 0.3 M NaCl at 37 °C. The elution was purified using Centri-sep spin columns, loaded with 60 mg of Sephadex® G-50 beads (S5897-25G), and dried on the speedvac.

Finally, the samples were diluted to 5 pM in H_2O after quantification using the Qubit fluorometer. For each chip, an equal volume of every sample was combined in a microcentrifuge tube, and 25 µL of the mixture was loaded onto the Ion Chef™ instrument.

2.5. Results and discussion

Despite significant advancements in RNA synthesis technology, current methods are limited when it comes to creating highly modified RNA libraries directly. For aptamer selection, which requires

the evolution of RNA sequences to bind specific targets with high affinity and specificity, the ability to introduce various modifications during the synthesis phase would be advantageous.[94] Currently, post-selection modifications are a common approach to enhance the properties of RNA aptamers. These modifications can improve stability, binding affinity, and specificity as new functional groups are introduced.

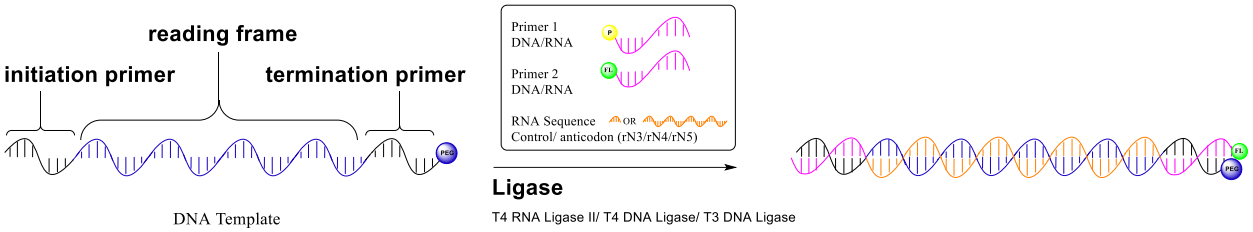


Figure 2-4: Schematic of the RNA LOOPER process; preliminary method *via* control 40-mer RNA, and unmodified RNA anticodons using T4 RNA ligase II, T4 DNA ligase, and T3 DNA ligase.

The LOOPER method enables the incorporations of various modifications expanding DNA into a library of heteromultivalent polymers for downstream evolution of higher molecular function. Herein, the development of the RNA LOOPER method aims to extend the capabilities of the LOOPER technique beyond its application in creating densely modified DNA aptamer libraries. Although the core idea is similar to DNA LOOPER, the RNA LOOPER method presents a novel approach by employing LOOPER templated copolymerization for the enzymatic synthesis of modified RNA libraries, using a DNA template library (Figure 2-4). The development of this new method carefully considers the various factors involved in LOOPER polymerizations.

Preliminary experiments utilized a model system; RNA LOOPER was carried out with a specific PEGylated DNA template and a full-length RNA complement of the reading frame to reduce the ligation sites to the two required at the primer nicks. The control LOOPER experiment facilitated optimizations of ligations between DNA and RNA primer combinations, and selection amongst

three commercially available candidate ligases (T3 DNA ligase, T4 DNA ligase, and T4 RNA ligase II). All primer combinations (5′-DNA and 3′-RNA, 5′-RNA and 3′-DNA, 5′ and 3′-DNA, and both RNA) were examined for all three enzymes. The efficacy of the model RNA LOOPER experiments was examined *via* pre-stained PAGE analysis (Figure 2-5), where the 5′-primer (DNA/RNA) was fluorescently tagged for reference. The results demonstrated three clear bands (when LOOPER was successful) regardless of the enzyme or the primer combination; where the bottom band corresponded to a single primer extended ligation (Figure 2-5, red dotted box), the middle band was full-length product RNA (Figure 2-5, green box), and the top band represented duplex DNA/RNA hybrid that did not denature fully (Figure 2-5, yellow dashed box).

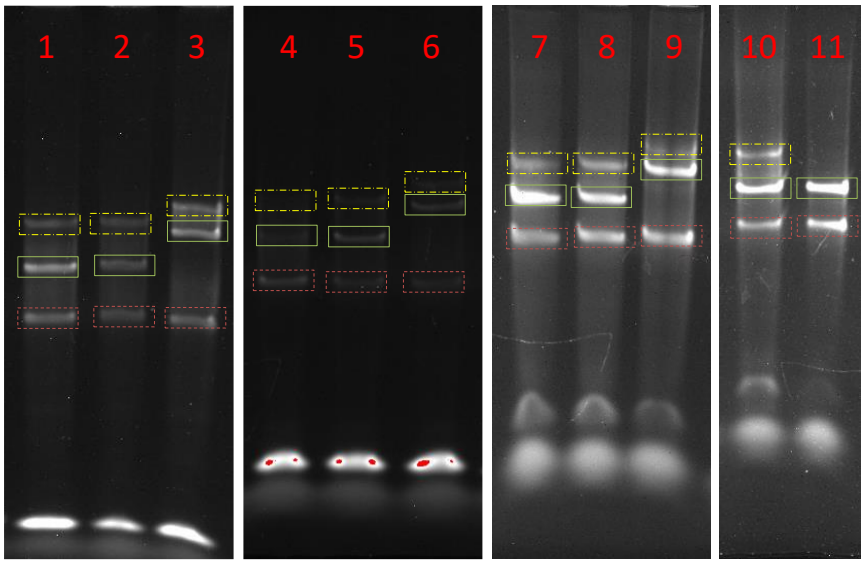


Figure 2-5: 10 % Denaturing PAGE gel of primer sampling control ligation *via* T3/T4 DNA ligase and T4 RNA ligase II; lanes 1-3 correspond to T3 DNA ligase (all contain 5′-FL DNA primer, 1: 3′-RNA primer, 2: 3′-DNA primer, 3: 3′-PEGylated DNA primer), lanes 4-6 correspond to T4 DNA ligase, lanes 7-9 correspond to T4 RNA ligase II (all contain 5′-FL RNA primer, 4/7: 3′-RNA primer, 5/8: 3′-DNA primer, 6/9: 3′-PEGylated DNA primer), lane 10 and 11 both represent T4 RNA ligase II products (lane 11 was DNase I digested, and missing the duplex band seen in lane 10). Note, that bands outlined in yellow represent the duplex, green boxes represent the full-length product, and red boxes represent single primer ligation.

Primer preference at the 5′-end is specific to each ligase. T3 DNA ligase showed preference for a 5′-DNA primer, while T4 DNA ligase and T4 RNA ligase II both demonstrated a preferred primer arrangement with a 5′-RNA primer. The overall best control system results were observed with T4 RNA ligase II, ligating RNA primers annealed at both 5′- and 3′-primer binding sites. This overall best setup allowed for further analysis of the top band on the gel *via* a paired parallel reaction where one sample was DNase I digested prior to PAGE analysis. The top band disappeared upon DNase digestion as seen in lane 11 in Figure 2-5, confirming it to result from a DNA/RNA hybrid not fully denatured due to its strong hybridization. Due to the high T_m of the DNA/RNA hybrid, denaturation of the full-length product required temperatures as high as 98 °C and remnant duplex was still present in some cases.

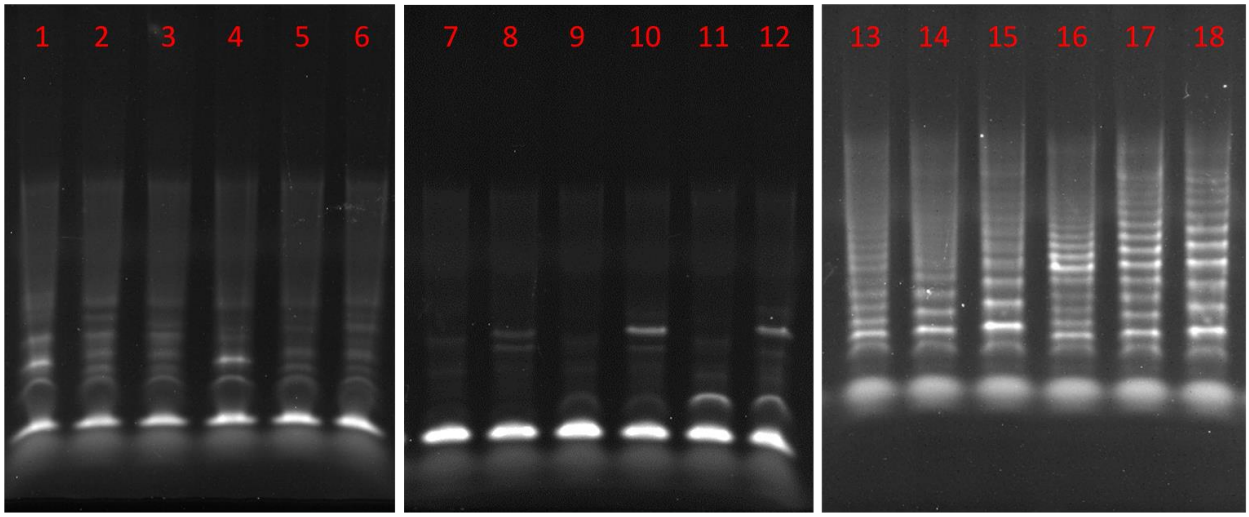


Figure 2-6: 10 % Denaturing PAGE gel of RNA LOOPER with T3/T4 DNA ligase and T4 RNA ligase II; lanes 1-6 correspond to T3 DNA ligase (left to right rN3/rN4/rN5 in order, odd number lanes: RNA/DNA primer combination and even number lanes: both DNA primers), lanes 7-12 correspond to T4 DNA ligase (left to right rN3/rN4/rN5 in order, 7-9: DNA/RNA primer combination and 10-12: both RNA primers), lanes 13-18 correspond to T4 RNA ligase II (left to right rN3/rN4/rN5 in order, 13-15: DNA/RNA primer combination and 16-18: both RNA primers). All samples were DNase I digested.

Optimization efforts primarily focused on qualitative relative yields of full-length product bands. It is crucial to differentiate between trimer, tetramer, and pentamer anticodon systems, as they impact ligation efficiency according to the number of nicks as well as annealing behaviour. The preferred primer combinations for each enzyme were tested with the unmodified RNA anticodon systems to start (degenerate trimer, tetramer, and pentamer sequences). The anticodon experiments utilized degenerate PEGylated template libraries and successful primer combinations from the control experiments. Along the reading frame of a 76mer library spanning 39 or 40 bases (the 39-mer reading frame corresponds to trimers and the primer binding sites adjust the total length to a 76-mer), trimers required 13 incorporations *via* 14 ligations, tetramers required 10 incorporations *via* 11 ligations, and pentamers required 8 incorporations *via* 9 ligations. Partial product bands were observed as laddering based on the number of incorporations elongating the RNA strand. However, due to multiple purifications before PAGE analysis, the density of these laddering byproduct bands was not sufficiently useful for assessing full-length product yield. The pentamer system of RNA LOOPER was most successful in synthesizing full-length products, possibly due to fewer ligation sites. Conclusive to Figure 2-6, T4 RNA ligase II was consistently effective *via* the method utilizing both RNA primers and the pentamer RNA anticodons.

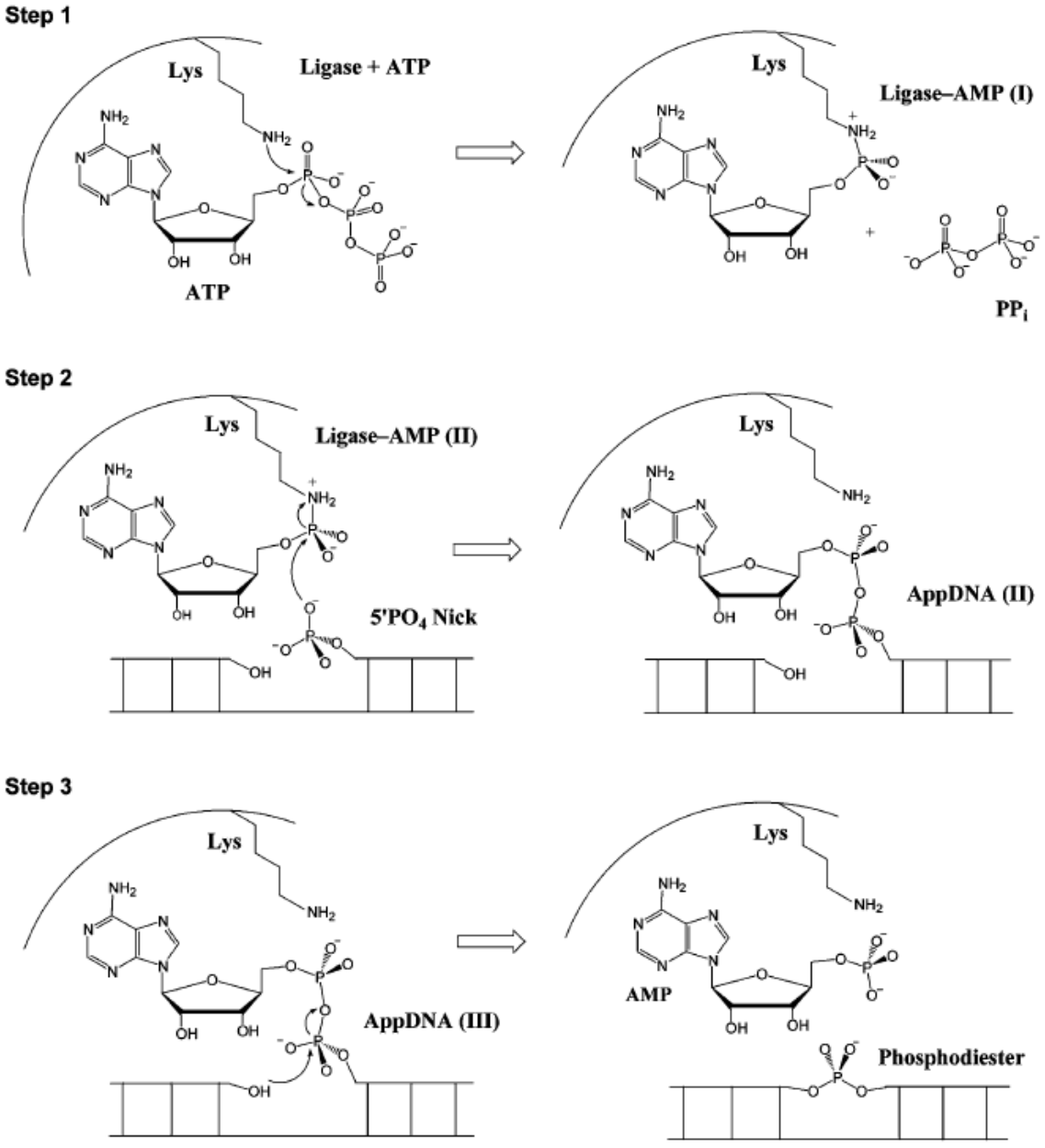


Figure 2-7: Pathway of nick ligation by ATP-mediated T4 RNA ligase II; step 1 is the adenylation of the ligase *via* active-site lysine attack on ATP to form a covalent ligase-AMP, releasing pyrophosphate (PPi). Step 2 ligase-AMP is attacked to transfer AMP to the nick 5′ phosphate. In step 3, the nick 3′OH attacks the adenylated 5′ phosphorus to form a phosphodiester bond, releasing AMP. Figure adapted from: Cell.[95]

The bacteriophage T4 RNA ligase II repairs nicks in duplex oligonucleotide polymers where the 3′OH is RNA.[95] The mechanism of ligation occurs in three steps (Figure 2-7): i) ATP-dependent

adenylation of the Rnl2 (RNA ligase 2) enzyme, an intermediate to ligation, ii) transfer of the AMP (Adenosine Mono-Phosphate) group to the 5′PO_4 of the annealed strand at the nick to activate the anticodon for ligation, and iii) Rn12 directed attack of the 3′OH on the activated 5′Phosphoanhydride linkage to seal the nick *via* a phosphodiester bond formation, releasing AMP in the process.[95] The 2′OH group of the 3′-end RNA base at the nick is not chemically involved in the phosphodiester bond formation, however it is required as a major contributor to the RNA-like folding at the ligation site particularly for step 3.[95] The ribose ring directly affects the helical form of the duplex, where the 2′OH provides stability for the duplex *via* the A-form helix with increased melting temperature (Tm) compared to the standard B-form helix form of DNA. Structural analysis on helical parameters of hybrid duplexes in literature suggest more inclination of the DNA/RNA duplex towards an A-form helix.[96,97]

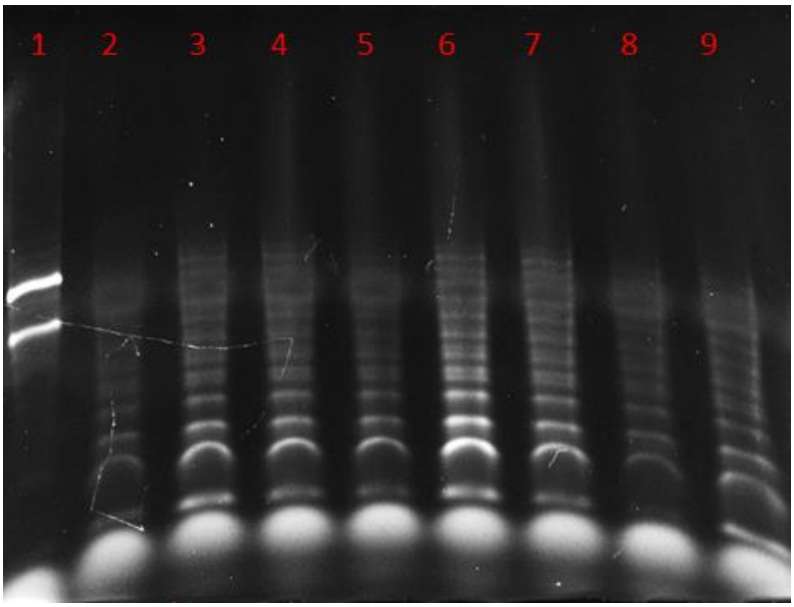


Figure 2-8: 10 % Denaturing PAGE gel of time and temperature optimizations for rN5 RNA LOOPER *via* T4 RNA ligase II; lane 1 represents a positive control experiment, lanes 2-4 correspond to 30 °C for 2/4/6 hours respectively, lanes 5-7 correspond to 33 °C for 2/4/6 hours respectively, lanes 8-9 correspond to 37 °C for 2/4 hours respectively. All samples were DNase I digested and denatured at 98 °C for 2 min before loading.

The unmodified RNA pentamer anticodon system facilitated further optimizations of the RNA LOOPER method *via* T4 RNA ligase II. Multiple experiments examined a wide array of ligation conditions and additives; exploring parameters including ligation temperature, time, the equivalency of each component, secondary ligase additions, and more to maximize full-length product yields. Preliminary optimizations of time and temperature challenged the ligase to identify the most effective window of ligation, between 2 to 24 hours at temperatures ranging from 4 °C to 37 °C. The final PAGE analysis results at varying time lengths and temperatures demonstrated optimal full-length product within 4 hours of ligation at 33 °C (Figure 2-8, lane 6). These optimizations were followed by additional cycling temperatures for improved full-length product yields.

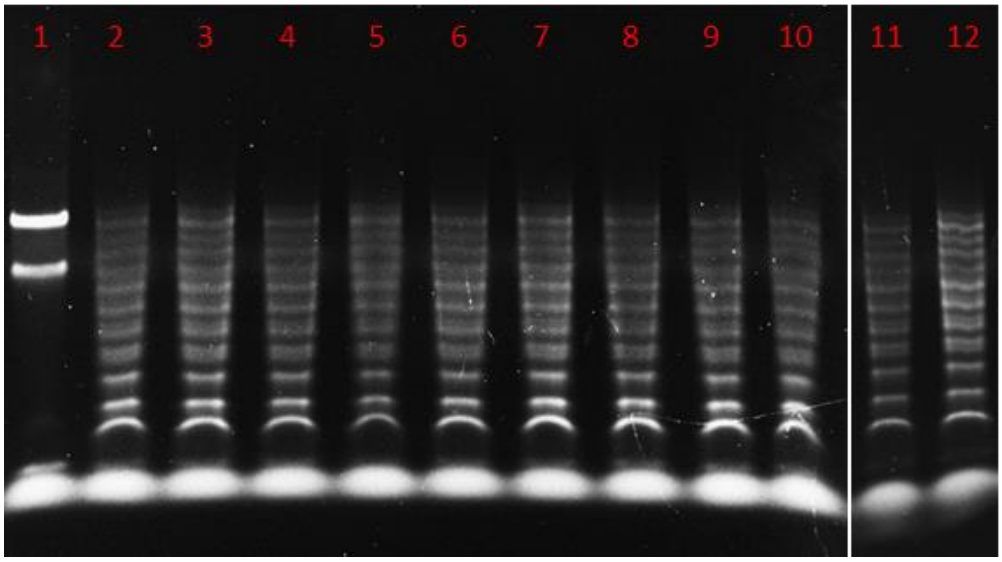


Figure 2-9: 10 % Denaturing PAGE gel ligase addition optimizations for rN5 RNA LOOPER *via* T4 RNA ligase II; lane 1 represents a positive control experiment, lanes 2-7 correspond to 30 °C for 2 hours followed by 10 U of ligase addition and 30/33/37 °C for 2/4 hours respectively, lanes 8-10 correspond to 33 °C for 4 hours followed by 10 U of ligase addition and 30/33/37 °C for 2 hours respectively, and lanes 11-12 correspond to 33 °C for 2 hours followed by 30 °C for 4 hours with a and additional unit of ligase in between incubation steps versus at the beginning respectively. All samples were DNase I digested and denatured at 98 °C for 2 min before loading.

Enzymatic reactions are often carried out on a small scale and are limited by the amount of enzyme available for synthesis. Thus far the RNA LOOPER method was successful while using only 10 U of T4 RNA ligase II per reaction, which was comparably much lower than the T4 DNA ligase LOOPER using 400 U and the T3 DNA Ligase LOOPER using 3000 U of each ligase per reaction. The efficiency of RNA LOOPER was improved by a secondary addition of T4 RNA ligase II; this was tested *via* an intermediate addition in between cycling temperatures (Figure 2-9, lanes 1-11), and was found to be more effective when 2 units of ligase were added in full before incubation (Figure 2-9, lane 12). The yields were noticeably higher upon cycling ligation at 33 °C for 2 hours followed by 30 °C for 4 hours. These optimizations reduced ligation time drastically, avoiding overnight incubation while improving RNA LOOPER product yields.

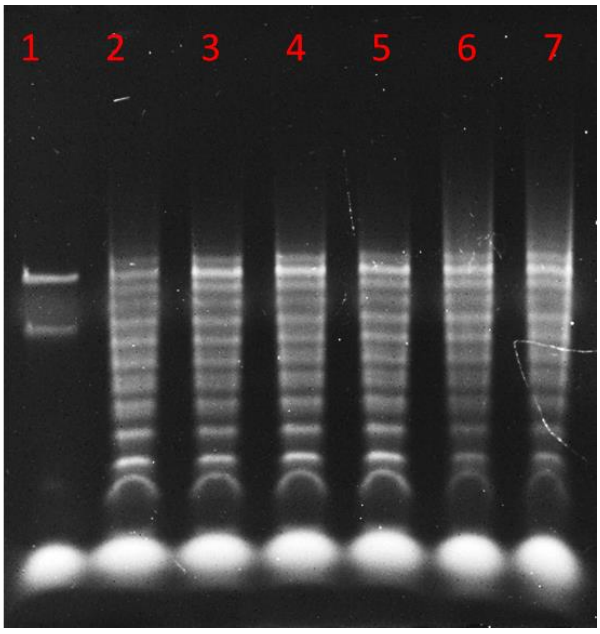


Figure 2-10: 10 % Denaturing PAGE gel of anticodon equivalency and volume optimizations for rN5 RNA LOOPER *via* T4 RNA ligase II; lane 1 represents a positive control experiment, lane 2 represents 10 equivalent rN5 as a control thus far, lanes 3-5 represent 10/15/20 equivalent rN5 reactions respectively without added $MgCl_2$, lanes 6-7 correspond to 10/15 equivalent repeats respectively with no added $MgCl_2$ at a reduced overall reaction volume of 13 µL.

Follow-up experiments targeted buffer components such as $MgCl_2$ (magnesium chloride), DTT (Dithiothreitol), and ATP (Adenosine Triphosphate). Out of these buffer components, $MgCl_2$ and ATP can affect ligation efficiency as they work cooperatively to activate the ligase *via* adenylation to form Rnl2-AMP. The divalent Mg^{2+} metal ion coordinates the triphosphate moiety providing stabilization. On the contrary, DTT may not affect ligation yields during RNA LOOPER as the reducing agent mainly contributes to enzyme stability over long incubation times. The commercially available NEB buffer (20x) paired with T4 RNA ligase II contains 2 mM $MgCl_2$, 1 mM DTT, and 400 µM ATP at 1x concentration in solution. All components were individually tested at varying concentrations; $MgCl_2$ concentrations from 0 to 7.5 mM, DTT concentrations from 0 to 1.5 mM, and ATP concentrations from 0.2 to 1 mM were experimentally evaluated for optimal LOOPER efficiency. The best overall conditions were consistent with the NEB buffer concentrations of each component.

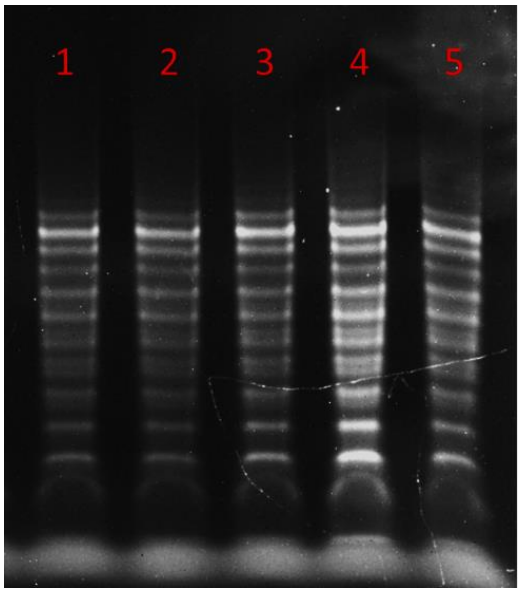


Figure 2-11: 10% Denaturing PAGE gel of PEG optimizations for rN5 RNA LOOPER *via* T4 RNA ligase II; lane 1 corresponds to 15 % PEG 8000 (6 µL of 50 % PEG 8000), lane 2 corresponds to 16.5 % PEG 8000, lane 3 corresponds to 12.5 % PEG 6000, lane 4 corresponds to 15 % PEG 6000, and lane 5 corresponds to 16.5 % PEG 6000.

LOOPER, akin to other polymerizations, relies upon the building blocks to synthesize a complementary strand. The enzyme activity was examined in the presence of increased anticodon equivalencies. Interestingly, up to 15 equivalents of the anticodon improved ligation yields; however, higher concentrations did not provide any added benefits (Figure 2-10). In parallel, the reaction was carried out at a reduced volume to accommodate a more concentrated environment for the enzyme (Figure 2-10, lanes 6-7). The lower volume resulted in slightly better yields to previous reaction conditions likely due to the closer proximity interactions of anticodons with the template promoting efficient ligation of nicks; however, was not explored further due to smeary band from higher salt concentrations relative to the volume. A similar effect was achieved with PEG (Polyethylene Glycol), a crowding agent, that helps increase ligation yields by occupying empty spaces in the reaction. Thus, PEG is an important component during ligation; optimizations of PEG 8000 were carried out from 2.5 % to 16.5 %, and PEG 6000 from 12.5 % to 16.5 % (Figure 2-11). Ligase efficiency was drastically higher when using 15 % PEG 6000 in RNA LOOPER.

Subsequently, the optimized RNA LOOPER method facilitated the synthesis of modified RNA libraries *via* the pentamer anticodon system. Commercially available m^6A modified pentamer libraries allowed the incorporation of up to eight m^6A sites along the 40-mer library reading frame. The pentamer anticodon sublibraries were purchased with a single m^6A at every possible position shifting from the 5′- to 3′-end of the sequence. Each modified anticodon sublibrary was paired with a corresponding template to carry out LOOPER. The optimized method demonstrated the successful synthesis of full-length modified RNA libraries regardless of the location of the m^6A site. Considering the results, T4 RNA ligase II showed higher efficiency for LOOPER polymerization of pentamer anticodons containing m^6A sites at the second and third position from the 5′-end (Figure 2-12, lanes 2-3) and at the 3′-end (Figure 2-12, lane 5). Thus far, the RNA

LOOPER method had promising results for the synthesis of diverse libraries of modified RNA oligonucleotides; however, T4 RNA ligase II fidelity was uncharted territory both for the RNA LOOPER method and has never been previously investigated in literature.

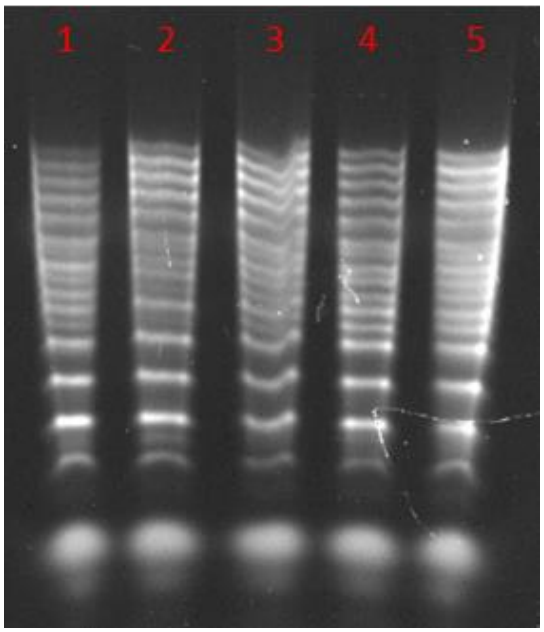


Figure 2-12: 10% Denaturing PAGE gel of m6A modified RNA library synthesized *via* the fully optimized RNA LOOPER method using T4 RNA ligase II and corresponding templates; lane 1 corresponds to m6ANNNN anticodon, lane 2 corresponds to Nm6ANNN anticodon, lane 3 corresponds to NNm6ANN anticodon, lane 4 corresponds to NNNm6AN anticodon, lane 5 corresponds to NNNNm6A anticodon.

Prior to any further optimizations on the RNA LOOPER system, NGS (Next-Generation Sequencing) was used to study the capacity of discrimination by T4 RNA ligase II between anticodon substrates containing paired and mispaired nucleotides near the ligation sites during LOOPER. The correct anticodon annealing and ligation in correspondence to the respective template sequence during LOOPER is a critical requirement for the effectiveness of the system. Ion Torrent[98] sequencing combined with a newly developed sequencing preparation method for fidelity analysis, Hairpin Sequencing (HP-seq) pioneered by Natalie Khamissi, provided the opportunity to analyse the fidelity of T4 RNA ligase II during RNA LOOPER. Sequencing analysis

included a degenerate trimer anticodon DNA LOOPER control with T3 DNA ligase, the RNA LOOPER products from the degenerate anticodon systems (trimer, tetramer, and pentamer), and the five m^6A containing libraries mentioned in Figure 2-12.

Figure 2-13: HP-seq assembly for RNA-LOOPER sequencing. Figure from: Natalie Khamissi.

Notwithstanding the significance of ligase efficiency, the ability of the enzyme to synthesize full-length LOOPER products with high fidelity particularly important for RNA LOOPER aptamer regeneration during *in vitro* evolutions. The HP-seq method facilitated fidelity analysis *via* a hairpin template, enabling LOOPER as an extension of the hairpin primer on the 5′-end and an overhang primer on the 3′-end (Figure 2-13). Once LOOPER was completed using the HP-seq templates the full-length products were confirmed by denaturing PAGE and gel extracted for follow-up sequencing preparations. The overhang allowed for barcode assembly and PCR amplification of samples for sequencing. Computational analysis of the sequencing data provided quantitative information about codon misincorporations. The overall ligation fidelities of each anticodon system were calculated using the ratio of total anticodon errors over the total number of anticodons occurrences. The resulting fidelities ranged from 5.60 to 16.6 % (Table 2-1) for T4 RNA ligase II. The highest fidelity of 16.60 % corresponded to the trimer anticodon system (rN3), while pentamer anticodon system (rN5) demonstrated a lower fidelity of 11 %. The m^6A modified anticodons dropped the fidelity further below 10 %. However, the overall extremely low fidelities of T4 RNA ligase II for the RNA LOOPER anticodon systems, compared to the acceptable fidelity for LOOPER aptamer libraries demonstrated by the T3 DNA ligase control LOOPER, makes the

RNA LOOPER method susceptible to undesirable errors during ligation. This inconsistency of the ligation method undermines its effectiveness for synthesizing modified RNA libraries for aptamer selections, making it difficult to regenerate the aptamer libraries with accuracy as the method is not robust.

Table 2-1: LOOPER fidelities; DNA LOOPER *via* T3 DNA ligase and RNA LOOPER *via* T4 RNA ligase II, where fidelities are based on the anticodon errors.

Library	Fidelity
NNN (DNA)	88.5%
rN3	16.6%
rN4	8.60%
rN5	11.0%
(m^6A)NNNN	9.80%
N(m^6A)NNN	5.60%
NN(m^6A)NN	6.60%
NNN(m^6A)N	6.40%
NNNN(m^6A)	8.40%

First, considering the variability between enzyme activity and structure, T4 RNA ligase II (329 amino acids) contains two conserved domains. Namely the *N*-terminal domain (aa 1–234) which is active only during step 1 and 3 of ligation (defective in binding to a nicked duplex substrate as an isolate) and near identical to adenyltransferase domains of DNA ligases, on the contrary the *C*-terminal domain (aa 244–329) is unique to RNA ligases enabling the recognition of the 3′-sequence as RNA which is critical for step 2 of nick ligation. The T4 RNA ligase II enzyme interacts with the nicked duplex, spanning 13 base pairs, where the adenyltransferase domain mainly interacts with the last base on the 5′-PO_4 side *via* the AMP moiety while most enzyme/nucleotide contacts are on the 3′-OH side. On the contrary, the ligase points of contact with the template backbone are limited and located 6 or 7 nucleotides away from the nick on the 5′-side.[95] It can be speculated that the T4 RNA ligase II enzyme cannot discriminate between canonical and mismatched base pairs

due to the lack of overlapping contacts with the nucleotides along the anticodon and the corresponding portion of the template strands near the nick. A second potential factor contributing to the low fidelities observed with RNA LOOPER using T4 RNA ligase II is the stability of RNA/DNA duplexes. Sequences that are stable, forming multiple hydrogen bonds and normal helical diameter, such as GU base pairs and purine-purine stacking, may bypass the fidelity requirement of A-form helical structure and be favoured for ligation.[99] The tighter annealing and increased rigidity of these duplex structures compared to DNA duplexes may result in non-specifically annealed anticodons being ligated before hybridization can be reversed to the correct base-pairing. Further analysis of the RNA anticodon annealing to the DNA template can concur these speculations.

2.6. Conclusions

Expanding the scope of LOOPER relies upon new and improved anticodon systems and ligases for templated copolymerization. Considering the efficacy of RNA aptamer modifications post-selection, as previously seen in FDA approved RNA aptamers, and the lack of method availability for the synthesis of modified RNA aptamer libraries, herein the LOOPER method was explored for RNA anticodon sublibraries with commercially available T4 RNA ligase II. Although extensive optimizations significantly improved the efficiency of RNA LOOPER yields with T4 RNA ligase II, sequencing analysis revealed critically low fidelities, highlighting the limited capacity of the LOOPER method using commercially available ligases. To achieve higher fidelities, future research could explore evolving high fidelity RNA ligases *via* mutated versions of the T4 RNA ligase II enzyme for the LOOPER derived synthesis of modified RNA libraries. Despite this limitation, the overall concept of LOOPER remains a powerful tool, facilitating the

templated copolymerization of diverse libraries of highly modified oligonucleotides through sublibraries of functionalized anticodons.

CHAPTER 3. *IN VITRO* EVOLUTION OF LOOPER-DERIVED APTAMERS

3.1. Modified DNA aptamer libraries

Anticodon modifications play a significant role in target engagement during molecular recognition and catalysis and influence the tertiary structures of aptamers. Thus, the choice of modifications is paramount and may be governed by the molecular target or type of catalysis. A variety of Brønsted acids and bases, polar and hydrophobic groups, and metal chelators have been successfully implemented in DNA-LOOPER during *in vitro* selections. Fidelities ranging from 87-98% (average 94%) have been used during successful selections.[53,100] Anticodons modified with aldehydes, and peptide fragments have also been shown to be compatible with LOOPER. Peptides of up to eight amino acids in length, comprising charged, polar, or hydrophobic residues, are accommodated by T4 DNA ligase.[62]

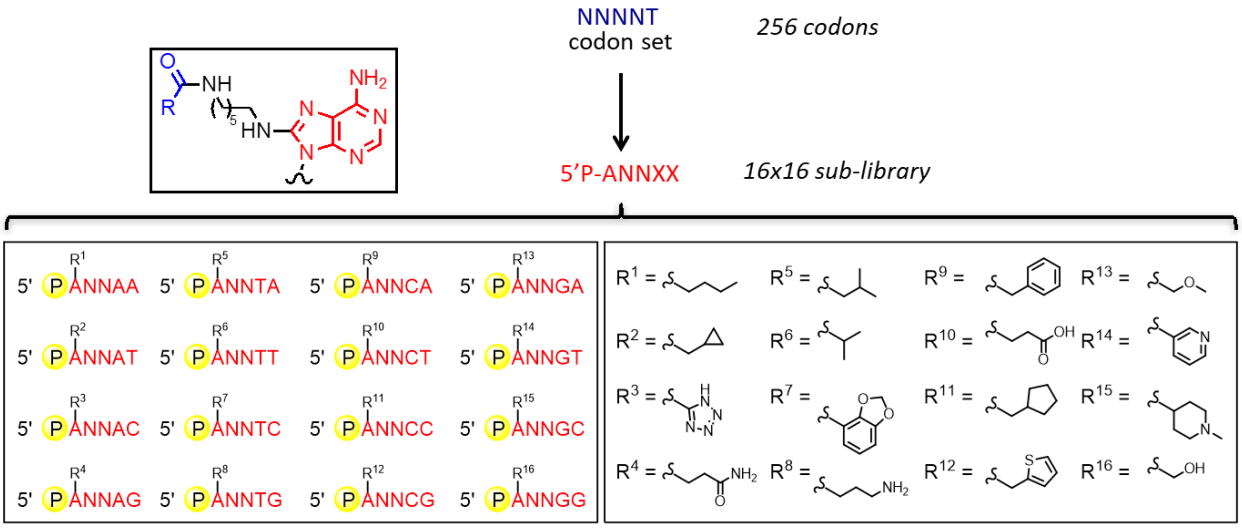


Figure 3-1: Synthesis of modified DNA by ligase-catalyzed oligonucleotide polymerization (LOOPER); library design encoding of modifications on anticodons. Note R8 is just HMDA, linker only.[53,55,56]

An increase in fidelity can benefit in vitro applications by increasing the enrichment factors and allowing access to larger reading frames. To date, the largest codon set that has been used in LOOPER is defined as XXNNT (anticodon defined as ANNXX), where XX is a specific dinucleotide that encodes for one of 16 possible modifications on the cognate anticodon. In

contrast, NN is a variable dinucleotide sequence that does not encode for the modification but provides sequence diversity (Figure 3-1). Together, this codon set consists of 256 codons, such that each of the 16 modifications is encoded by 16 unique codons with the highest chemical diversity available for the evolution of nucleic acid polymers. This original LOOPER library has been successfully used for aptamer evolutions against human α-Thrombin resulting in TBL1[53,55] (Figure 3-2), the first Thrombin-binding aptamer to deviate form a G-quadruplex structure, discussed in further detail throughout this thesis.

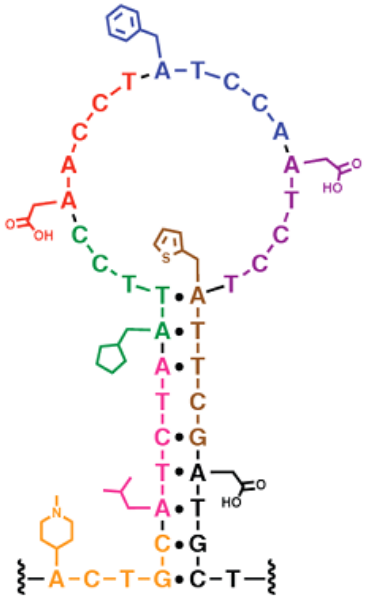


Figure 3-2: Schematic of TBL1; a highly modified functional LOOPER aptamer selected against human α-Thrombin.[53,55,56]

3.1.1. Swap-codon libraries

LOOPER-derived aptamers utilize the secondary structures formed by their specific oligonucleotide sequences to display incorporated functional groups along the sequence. When properly oriented, these LOOPER modifications can interact effectively with the target, enhancing binding quality. Inspired by a previously successful LOOPER aptamer evolution, swap-codon libraries are developed to evaluate the target-binding interactions of select modifications across

diverse anticodon sublibraries. The original LOOPER library was utilized as a point of reference to design and synthesize three swap-codon libraries, encoding the same modifications as the original LOOPER library used to evolve TBL1 *via* unique anticodons in each library (Figure 3-3). The same XXNNT codon set was used along a 40-mer reading frame flanked by 18-mer primer binding sites. Additionally, the same 16 small-molecule modifications enabled the synthesis of 256-member sublibraries such that every library is unique; the same pentamer anticodon never encodes the same functional group twice (especially the combinations present in the original LOOPER library used to evolve TBL1). For instance, R^1 is encoded by ANNAA in the original library, ANNCA in library 1, ANNCG in library 2, and ANNCC in library 3. The swap-codon libraries provide an opportunity to conduct an in-depth analysis of the LOOPER aptamer evolution process based on sequence homology versus chemical homology in selection outputs, comparing the outcome for libraries with the same functional groups along a diverse scope of sequence space.

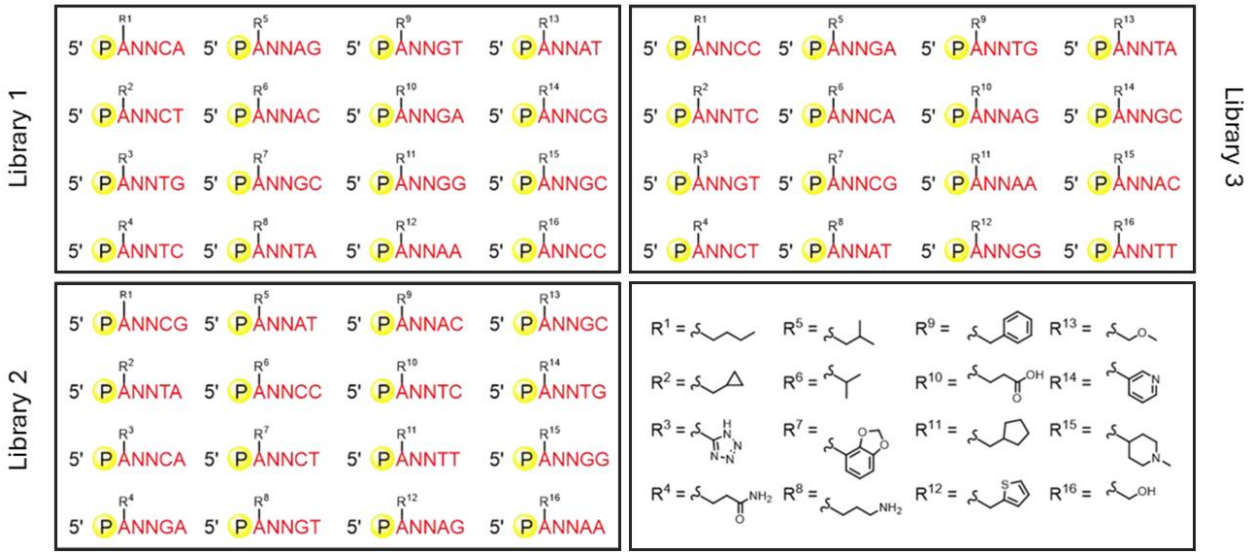


Figure 3-3: Three swap-codon libraries and corresponding modifications; the carboxylic acids are numbered for simplicity.

3.1.2. Materials and methods

3.1.2.1. General information

Unless otherwise noted, water was purified with the Direct-Q® 3 UV Water Purification System. Aminomodified pentanucleotide libraries (ANNXX anticodon libraries) were purchased from Integrated DNA Technologies or synthesized on the Applied Biosystems DNA/RNA synthesizer model 394 as mentioned below. All materials and reagents used for oligo-synthesis were purchased from Glen Research™. All materials and reagents used for pentamer functionalization were purchased from Sigma Millipore. Functional pentamer products were purified by reverse-phase high-performance liquid chromatography (HPLC, 1260 Infinity II LC System) using a C18 stationary phase (Eclipse-XDB C18, 5 μm, 4.6 × 150 mm) and an acetonitrile/ 0.1 M triethylammonium acetate (TEAA buffer) gradient. Purified products were lyophilized using a FreeZone 2.5 L -84 °C Benchtop Freeze Dryer. Coupled product concentrations were quantified from A_{260} values using a Nanodrop One UV-Vis Spectrophotometer. Mass spectrometry of coupled products was carried out using the Velos Pro Mass spectrometer (Orbitrap Elite MS). The mass accuracy of Velos Pro Orbitrap ranges from 0-6.62 ppm.

3.1.2.2. Sequences and carboxylic acids

Aminomod-AA 5′- /5Phos/ANNAA

Aminomod-AT 5′- /5Phos/ANNAT

Aminomod-AC 5′- /5Phos/ANNAC

Aminomod-AG 5′- /5Phos/ANNAG

Aminomod-TA 5′- /5Phos/ANNTA

Aminomod-TT 5′- /5Phos/ANNTT

Aminomod-TC 5′- /5Phos/ANNTC

Aminomod-TG 5′- /5Phos/ANNTG

Aminomod-CA 5′- /5Phos/ANNCA

Aminomod-CT	5′- /5Phos/ANNCT
Aminomod-CC	5′- /5Phos/ANNCC
Aminomod-CG	5′- /5Phos/ANNCG
Aminomod-GA	5′- /5Phos/ANNGA
Aminomod-GT	5′- /5Phos/ANNGT
Aminomod-GC	5′- /5Phos/ANNGC
Aminomod-GG	5′- /5Phos/ANNGG
Carboxylic acid 1	Valeric acid
Carboxylic acid 2	Cyclopropylacetic acid
Carboxylic acid 3	1H-Tetrazole-5-acetic acid
Carboxylic acid 4	Succinamic acid
Carboxylic acid 5	Isovaleric acid
Carboxylic acid 6	Isobutyric acid
Carboxylic acid 7	3,4-(Methylenedioxy)phenylacetic acid
Carboxylic acid 8	no R-group
Carboxylic acid 9	Phenylacetic acid
Carboxylic acid 10	succinic anhydride
Carboxylic acid 11	Cyclopentylacetic acid
Carboxylic acid 12	3-Thiopheneacetic acid
Carboxylic acid 13	Methoxyacetic acid
Carboxylic acid 14	Niacin (Nicotinic Acid)
Carboxylic acid 15	1-methyl-piperidine-4-carboxylic acid
Carboxylic acid 16	Glycolic acid

3.1.2.3. Solid-phase synthesis of anticodons

Pentanucleotides were synthesized on an ABI 394 synthesizer using a (DiMethoxyTrityl) DMT-ON protocol on a 1 μmol scale (1000 Å CPG column). Amine-modified C6 dA (Glen Research 10-1089), dA+dC+dG+dT-CE Phosphoramidite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (CPRII, 10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 25 °C in 400 μL of a 1:1 mixture of ammonium hydroxide and methylamine for 25 minutes. The cleaved resin was filtered away, and the elution was incubated at 60 °C for 30 minutes to remove the protecting groups on the phosphoramidites. The oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 100 μL of H_2O, and purified using reverse-phase HPLC purification using a DMT-ON method (Table 3-1) on the preparative column (ZORBAX Eclipse XDB-C18, P.N. 990967-202) *via* acetonitrile in TEAA (0.1 M, pH 7) solvent gradient, column temperature of 40 °C. The full-length product containing the DMT group was collected as an elution at 15 minutes, and the truncated products were discarded.

Table 3-1: HPLC DMT-ON (collect at 15 min) and DMT-OFF (collect at 3 min) protocol.

Time (min)	Solvent A (%)	Solvent B (%)	Flow (mL/min)
0	90	10	4
10	80	20	4
20	20	80	4
25	20	80	4
28	90	10	4
30	90	10	4

The purified oligonucleotide was then incubated at room temperature in 1 mL of 40 % aqueous acetic acid for 1 hour to cleave the DMT group, and then frozen using liquid nitrogen to lyophilize.

The dry oligonucleotide was incubated in 500 µL 30 % ammonium hydroxide at room temperature for 15 minutes to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using a speedvac. The dried product was dissolved into 100 µL H_2O and subjected to reverse-phase HPLC purification using the DMT-OFF method on the preparative column *via* acetonitrile in TEAA (0.1 M, pH 7) solvent gradient, column temperature of 40 °C. The deprotected product was collected as an elution at 3 minutes. The purified oligonucleotide was dissolved in water. The pentamer concentrations were quantitated by UV spectroscopy at 260 nm.

3.1.2.4. Amide coupling of carboxylic acid modifications

A mixture of 25 µL carboxylic acid (100 mM in DMSO), 25 µL N-hydroxysuccinimide (NHS, 100 mM in 1:1 mixture of DMSO and H_2O), 5 µL 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC, 100 mM in DMSO) and 7.5 µL DMSO was incubated at room temperature for 30 minutes. The coupling reaction is demonstrated in Figure 3-4. This was followed by the addition of 20 µL of amino-modified pentanucleotides (1 mM in H_2O) and 30 µL Na_2CO_3 buffer (500 mM in H_2O, pH 9). The mixture was incubated at room temperature overnight on a vortex. The reaction was then quenched by the addition of 15 µL Tris-HCl buffer (500 mM, pH 8 in H_2O) at room temperature for one hour.

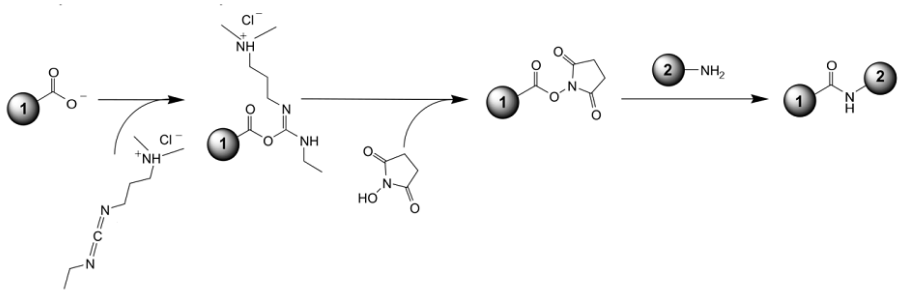


Figure 3-4: Functionalization of C6-Aminomodified dA pentanucleotides; (1) Carboxylic acid side chain (R-groups), (2) aminomodified pentanucleotide.

Acid #10 (succinic anhydride) does not require the activation step as it is not a carboxylic acid. A mixture of 25 μL succinic anhydride (100 mM in DMSO), 25 μL 1 M $NaHCO_3$, 163 μL DMSO, 17 μL water, and 25 μL of aminomodified pentanucleotides (1 mM in H_2O) was incubated at room temperature overnight with vortex. The reaction was then quenched by the addition of 50 μL Tris-HCl buffer (500 mM, pH 8 in H_2O) at room temperature for one hour.

Functionalized pentanucleotides were lyophilized and dissolved in 100 μL of H_2O for HPLC purification (injection volume: 100 μL) using the analytical column (ZORBAX Eclipse XDB, P.N. 993967-902). The HPLC quaternary pump flow was set to 1.00 mL/min with a pressure limit range of 0-400 bar and a maximum flow gradient of 100.00 mL/min^2. Injection (volume of 100 μL, at 40 °C) speed was set to draw at 200.00 μL/min and eject at 400.00 μL/min. The solvent composition of TEAA buffer (0.1 M, pH 7) and acetonitrile was programmed based on the method. The faction collection was set to start at 2 min and stop at 18 min. The peak detection mode (at 260 nm, margin for negative absorbance: 100 mAU) was set to the threshold (15.000 mAU–3000.00 mAU); up/down slope set to 5 mAU. Analog attenuation was set to 1000 mAU. Functionalized pentamer concentrations were quantitated by UV spectroscopy at 260 nm.

Table 3-2: HPLC protocol; DMT-OFF-earlyseq for anticodons with polar modifications.

Time (min)	Solvent A (%)	Solvent B (%)	Flow (mL/min)
0	95	5	1
5	80	15	1
15	55	45	1
16	20	80	1
18	90	10	1
20	90	10	1

The anticodons containing polar modifications were purified *via* the earlyseq method (Table 3-2) and the anticodons containing non-polar modifications were purified using the lateseq method (Table 3-3).

Table 3-3: HPLC protocol; DMT-OFF-lateseq for anticodons with non-polar modifications.

Time (min)	Solvent A (%)	Solvent B (%)	Flow (mL/min)
0	80	20	1
10	55	45	1
16	20	80	1
18	90	10	1
20	90	10	1

Negative mode mass spectrometry of 5 µL sample (10 µM in H_2O) injections was set up through the HPLC connection into the Heated Electrospray Ionization (HESI-II) Probe (Heated Temperature: 300 °C, Sheath Gas flow rate: 10, Aux Gas flow rate: 10, Sweep gas flow rate: 5, Spray voltage: 3.5 kV, Capillary temperature: 350 °C, S-Lens RF Level: 67.1 %). The mass range was defined between 150–2000 m/z (normal, resolution: 120000, analyzer: FTMS). The LCMS method time was set to 1 min/sample. Deconvolution of results was done using Freestyle (Xcalibur). Finally, the 16 modified pentanucleotide sublibraries were dissolved to 480 µM in H_2O for follow-up LOOPER.

3.1.3. Results and discussion

Three unique swap-codon libraries, each containing 256-member LOOPER pentamer sublibraries, were successfully synthesized and purified. Functionalization of C6 AdA pentanucleotides was carried out using the EDC/NHS coupling method *via* an NHS-ester intermediate. The purification of the 16 functionalized aminomodified pentanucleotide products for each swap-codon library was

optimized on HPLC. The Analytical column demonstrated higher peak resolution and better signal detection of coupled products; however, it did compromise the product yields in the process. To counter the lower yields due to increased column absorption, the initial amount of aminomodified pentanucleotide was adjusted for the coupling reaction, from 10 nmol to 20 nmol. Additionally, intermittent wash methods were implemented between each sample purification to enhance product yield, as subsequent samples in the sequence without washes showed lower yields.

Table 3-4: Swap-codon library #1; functionalized C6 AdA pentanucleotides (256 ANNXX anticodons divided into 16 sublibraries based on unique functional R-groups, coupled on *via* their carboxylic acid derivatives). The percentage yield of the coupled products ranged from 17-90 %.

Pentamer	**Lowest MW (Da)**	**Highest MW (Da)**	**Acid#**	**Coupling prod. Lowest MW (Da)**	**Coupling prod. Highest MW (Da)**	**HPLC retention time (min)**	**Yield (%)**	**Coupling prod. MS (Da)**
A'NNCA	1625	1705	1	1710	1790	6.9-9.5	32	1709-1789
A'NNCT	1616	1696	2	1699	1779	5.6-8.7	43	1698-1778
A'NNTG	1656	1736	3	1767	1847	2.0-5.0	40	1766-1846
A'NNTC	1616	1696	4	1716	1796	2.0-4.2	85	1715-1795
A'NNAG	1665	1745	5	1750	1830	6.2-10	90	1749-1829
A'NNAC	1625	1705	6	1696	1776	5.4-7.6	59	1695-1775
A'NNGC	1641	1721	7	1804	1884	7.9-9.2	17	1803-1883
A'NNTA	1640	1720	8	1641	1721			
A'NNGT	1656	1736	9	1775	1855	7.9-9.8	61	1774 -1854
A'NNGA	1665	1745	10	1766	1846	1.8-3.9	83	1765-1845
A'NNGG	1681	1761	11	1792	1872	9.1-11	83	1791-1871
A'NNAA	1649	1729	12	1774	1854	7.3-10	55	1773-1853
A'NNAT	1640	1720	13	1713	1793	4.2-6.5	81	1712-1792
A'NNCG	1641	1721	14	1747	1827	4.5-6.7	58	1746-1826
A'NNTT	1631	1711	15	1757	1837	2.0-5.0	38	1756-1836
A'NNCC	1601	1681	16	1660	1740	2.0-4.3	70	1659-1739

The retention time of each functionalized anticodon library (ANNXX) directly depended on the polarity of the coupled modification, with the codon size and HMDA (hexamethylenediamine) linker remaining constant throughout each library. Longer retention times indicate stronger attraction of functional groups to the stationary phase. Hydrophobic R-groups (non-polar) demonstrated longer retention times, while heterocyclic and acidic R-groups (polar) were collected earlier during purification.

Table 3-5: Swap-codon library #2; functionalized C6 AdA pentanucleotides (256 ANNXX anticodons divided into 16 sublibraries based on unique functional R-groups, coupled on *via* their carboxylic acid derivatives). The percentage yield of the coupled products ranged from 21-82 %.

Pentamer	**Lowest MW (Da)**	**Highest MW (Da)**	**Acid #**	**Coupling prod. Lowest MW (Da)**	**Coupling prod. Highest MW (Da)**	**HPLC retention time (min)**	**Yield (%)**	**Coupling prod. MS (Da)**
A'NNCG	1641	1721	1	1726	1806	6.8-9.1	31	1725-1805
A'NNTA	1640	1720	2	1723	1803	5.5-8.5	36	1722-1802
A'NNCA	1625	1705	3	1736	1816	2.0-5.0	39	1735-1815
A'NNGA	166	1745	4	1765	1845	2.0-4.8	33	1764-1844
A'NNAT	1640	1720	5	1725	1805	6.6-9.6	78	1724-1804
A'NNCC	1601	1681	6	1672	1752	4.8-6.6	63	1671-17751
A'NNCT	1616	1696	7	1779	1859	7.4-10.	49	1778-1858
A'NNGT	1656	1736	8	1657	1737			
A'NNAC	1625	1705	9	1744	1824	7.5-9.6	63	1743-1823
A'NNTC	1616	1696	10	1717	1797	1.8-3.6	73	1716-1796
A'NNTT	1631	1711	11	1742	1822	8.4-11	80	1741-1821
A'NNAG	1665	1745	12	1790	1870	7.4-10	55	1789-1869
A'NNGC	1641	1721	13	1714	1794	3.5-5.4	82	1713-1793
A'NNTG	1656	1736	14	1762	1842	4.7-6.7	58	1761-1841
A'NNGG	1681	176	15	1807	1887	2.0-5.3	21	1806-1886
A'NNAA	1649	1729	16	1708	1788	2.0-4.5	58	1707-1787

The HPLC method was optimized for polar modifications due to overlap with salts (charged ions); collection times coincided closely with a strong peak at 2 minutes. Adjustment of the solvent gradient for acids 3, 4, 10, 15, and 16 enabled higher column retention for better purification. Generally, compared to the expected retention times following the same solvent composition as a previous LOOPER library synthesis using the preparative column, the retention time of the same R-groups (on coupled pentamer anticodons) was decreased by 2 minutes. This trend was consistent among the three swap-codon libraries. After HPLC purification, the synthesized sublibraries underwent mass characterization *via* LCMS (Liquid-Chromatography Mass-Spectrometry). The LCMS data were deconvoluted to confirm the presence of 16 sublibrary members (amino-modified ANNXX products) by analyzing their specific mass-to-charge ratios. The range of molecular weights for each sublibrary was calculated and tabulated for comparison with experimental data; Table 3-4 represents swap-codon library1, Table 3-5 represents swap-codon library2, and Table 3-6 represents swap-codon library3. Negative mode LCMS data displayed peaks at a charge of -1 confirming the expected molecular weight range for these libraries; for each mass peak, the m/z (mass/charge) value plus 1 represented the actual molecular weight of the product. Please note that the anticodon with HMDA (acid #8) as the modification was purchased in this case. The percentage yields displayed in the tables above for each library are representative of varying coupling efficiencies as well as varying purification loss. The range of yields observed was between 17 to 90 %. The lowest yield of 17 % was sufficient for follow-up LOOPER and selection based on the large size of the anticodon library.

Table 3-6: Swap-codon library #3; functionalized C6 AdA pentanucleotides (256 ANNXX anticodons divided into 16 sublibraries based on unique functional R-groups, coupled on *via* their carboxylic acid derivatives). The percentage yield of the coupled products ranged from 18-80 %.

Pentamer	Lowest MW (Da)	Highest MW (Da)	Acid#	Coupling prod. Lowest MW (Da)	Coupling prod. Highest MW (Da)	HPLC retention time (min)	Yield (%)	Coupling prod. MS (Da)
A'NNCC	1601	1681	1	1686	1766	6.6-8.7	55	1685-1765
A'NNTC	1616	1696	2	1699	1779	4.8-7.9	48	1698-1778
A'NNGT	1656	1736	3	1767	1847	2.0-4.3	36	1766-1846
A'NNCT	1616	1696	4	1716	1796	2.0-4.8	35	1715-1795
A'NNGA	1665	1745	5	1750	1830	6.5-8.6	51	1749-1829
A'NNCA	1625	1705	6	1696	1776	4.9-7.3	49	1695-1775
A'NNCG	1641	1721	7	1804	1884	7.5-10	42	1803-1883
A'NNAT	1640	1720	8	1641	1721			
A'NNTG	1656	1736	9	1775	1855	7.9-10	63	1774-1854
A'NNAG	1665	1745	10	1766	1846	1.8-3.8	80	1765-1845
A'NNAA	1649	1729	11	1760	1840	9.5-11	48	1759-1839
A'NNGG	1681	1761	12	1806	1886	7.1-10	54	1805-1885
A'NNTA	1640	1720	13	1713	1793	4.6-6.3	18	1712-1792
A'NNGC	1641	1721	14	1747	1827	4.2-6.3	67	1746-1826
A'NNAC	1625	1705	15	1751	1831	2.0-4.7	42	1750-1830
A'NNTT	1631	1711	16	1690	1770	2.0-4.9	72	1689-1769

3.2. Introduction to LOOPER-SELEX

In an effort to mimic the surface of proteins for molecular recognition, the LOOPER method enables a sequence-defined display of functional groups along a library of ssDNA at high densities. Combining the LOOPER process with the Systematic Evolution of Ligands by EXponential enrichment (SELEX)[16] provides access to high-quality functionalized aptamers with potential improvements of affinity and selectivity against molecular targets (Figure 3-5).

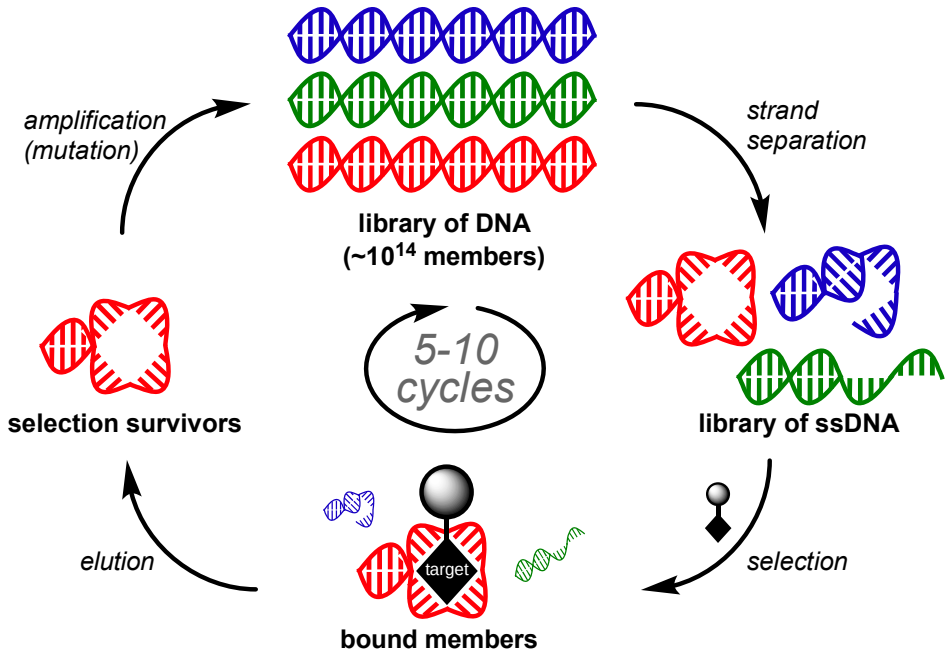


Figure 3-5: Systematic Evolution of Ligands by EXponential enrichment (SELEX).

Provided that fidelity, codon bias, and polymerization yield of LOOPER are well optimized for DNA libraries, porting LOOPER-derived libraries into SELEX has been straightforward. Upon synthesis of the LOOPER aptamer library, *in vitro* evolution begins with selection against a target of choice followed by elution and amplification of the output library. Minor differences in LOOPER-SELEX cycles are largely related to the nature of highly modified DNA, which requires polymerases with larger active sites, such as family B polymerases (*e.g.*, KOD DNA polymerase),[101] to amplify during PCR steps. Thus far, three protein targets have been successfully selected against using LOOPER-derived functionalized oligonucleotide libraries, which have provided insights into the impact of dense and diverse modifications on the evolution of aptamers.

3.2.1. Successful LOOPER-SELEX evolutions

The first successful LOOPER-SELEX aptamer evolution was demonstrated using a 256-membered ANNNN anticodon library encoding 16 unique modifications, generated using T4 DNA

ligase (Figure 3-6).[53] Selections were performed against human α-Thrombin, which after six rounds resulted in convergence of the library onto a C-rich consensus sequence within the eight-codon reading frame (Table 3-7); the highest frequency sequence, TBL1, represented 1.1 % of the total output library.[55]

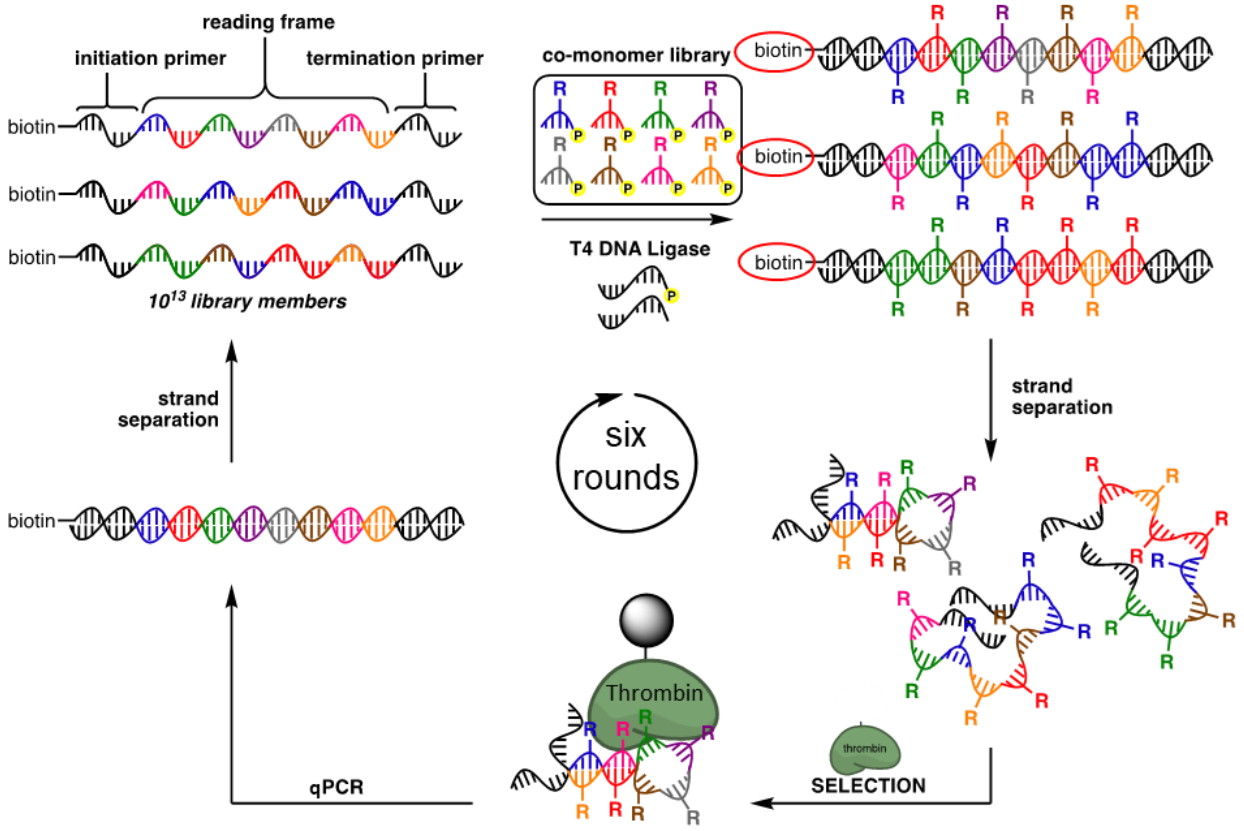
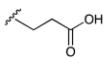
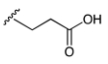
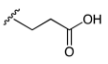
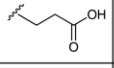
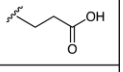
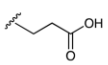
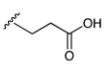
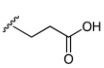
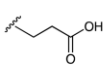
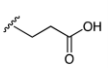
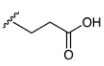
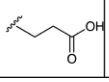
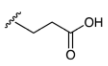


Figure 3-6: *In vitro* selection of aptamers *via* LOOPER-SELEX against human α-Thrombin.[53,55,56]

Further to this, there was strong functional group homology observed from anticodons 3-6 within the reading frame, namely a cyclopentyl group at position 3, a carboxylic acid at positions 4 and 6, and a phenyl ring at position 5. This region within the reading frame was under strong evolutionary pressure during SELEX. While TBL1 was a C-rich sequence, its corresponding template was a G-rich sequence that was found to form a stable G-quadruplex. Since G-quadruplexes are known to bind Thrombin, the affinity of the TBL1 template for Thrombin was measured. Surprisingly the template bound to Thrombin with a K_D of 350 nM.[55] Even more

surprising was that TBL1 also bound to Thrombin, and with much tighter affinity (K_D = 1.6 nM).[53,55] This represented the first known instance where both the genotype and phenotype achieved the same function during molecular evolution.

Table 3-7: Highest frequency read sequences from the first LOOPER-SELEX evolutions.[55]

name	reads[a]	K_d	position within reading frame[b]							
			1	2	3	4	5	6	7	8
TBL1	9525	1.6 nM	ACTGC 	ATCTA 	ATTCC 	AACCT 	ATCCA 	ATCCT 	ATTCG 	ATGCT 
TBL2	5596	16.7 nM	AGTTT 	ACTTA 	AGTCC 	AACCT 	ACCCA 	ACCTT 	ACTTA 	AGTCT 
TBL3	5445	2.3 μM	ATGTC 	ATCGC 	ATTCC 	AACCT 	AATCC 	ATCCT 	ATGCG 	ATAAT 
TBL4	3021	14.5 nM	ATGCT 	AGTGC 	AATCC 	ACCCT 	ACCCA 	AACCT 	ATTGC 	ACCTT 
TBL5	2796	29 nM	AACCT 	AAACT 	ATCCC 	ATCCT 	ATCCA 	ACCTA 	AGTTT 	AGTGT 

The modifications displayed on TBL1 were shown to be essential for binding. Deletion of all modifications fully ablated binding, as did deletions within the loop region of its hairpin structure. Furthermore, single modification deletions resulted in approximately 10-fold loss of binding affinity. Beyond this, the modifications were shown to be critical to the thermal stability of the folded structure of TBL1. TBL1 has a melting temperature of 85 °C, while TBL1 without modification has a melting temperature of between 50-55 °C. This suggests that the modifications are integral to the folding of TBL1, which was supported by molecular modeling studies. Further to this, TBL1 shares functional group similarities with known peptide-based inhibitors, such as hirugen[102,103] and the FDA-approved drug bivalirudin[104] – both are bivalent peptides, rich in acidic and hydrophobic residues similar to TBL1 (Figure 3-7). These peptide inhibitors are known to

bind exosite-I on Thrombin suggesting a potentially similar mode of action for TBL1.[105] Indeed, competitive binding assays demonstrated that TBL1 binds exosite-I without interacting with nearby exosite-II.

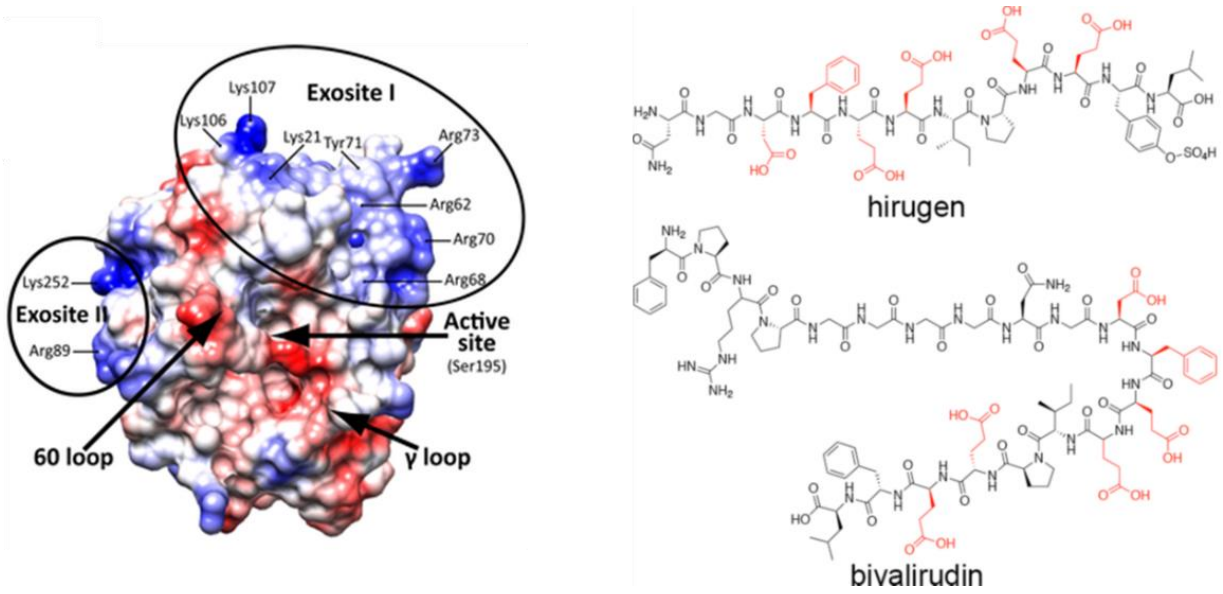


Figure 3-7: Schematic 3-D model of human α-Thrombin (left) and chemical structure of known peptide inhibitors (hirugen and bivalirudin). The 3D structure of human α-Thrombin with the charge density illustrated with blue for basic or positively charged regions and red for negatively charged or acidic regions. The color intensity reflects the charge intensity. Thrombin structure from PDB, code 1DM4 run using UCSF Chimera v1.8 and annotated in Adobe Illustrator CS5. Figure adapted from: PLOS ONE.[55,106]

The target, Thrombin, has two identified positively charged aptamer binding sites: exosite-I which binds TBA (anti-parallel G-quadruplex structure), and exosite-II which binds HD22 (duplex/G-quadruplex mixed structure).[107] Both aptamers contain complex structures; however, they comprise unmodified (native) nucleobases. The G-quadruplex motif (protruding loop residues) enable hydrophobic interactions, and the negatively charged DNA backbone facilitates polar interactions with the protein binding site.[108] TBL1 exhibits considerably tighter binding compared to the original Thrombin-binding aptamer TBA, which exhibits a K_D of ~99 nM.[109] TBL1 was also shown to inhibit Thrombin activity in serum assays with an IC_{50} of 23 nM.[56,110]

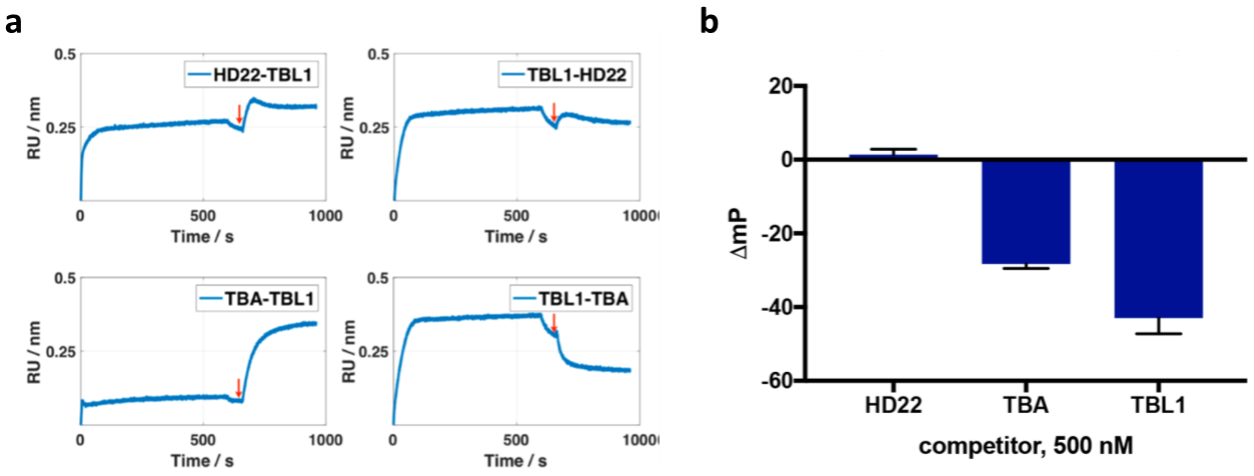


Figure 3-8: Interaction of TBL1 with Thrombin. (a) Binding sensorgrams on biolayer interferometry (BLI). HD22 and TBL1 binding; TBA and TBL1 binding are shown. Red arrow indicated time of injection of competitor. (b) Competitive fluorescence polarization analysis of Thrombin exosite-I binding.[55]

As the first LOOPER-derived aptamer, TBL1 highlights the importance of functional group diversity as it represents the first known DNA aptamer for Thrombin that deviates from a G-quadruplex motif. Epitope binning by competitive BioLayer Interferometry[111] (BLI) and competitive Fluorescence Polarization[112] (FP) suggest selective binding of TBL1 at the allosteric exosite-I on human α-Thrombin, Figure 3-8.[55] I performed the FP assay by displacement of fluorescein-labeled TBA in competition with either unlabeled TBA, HD22, or TBL1 at 500 nM concentration; this data along with the computational modeling I provided using the YASARA[113–115] software (Figure 3-11) further established the binding site of TBL1 to exosite-I on Thrombin.

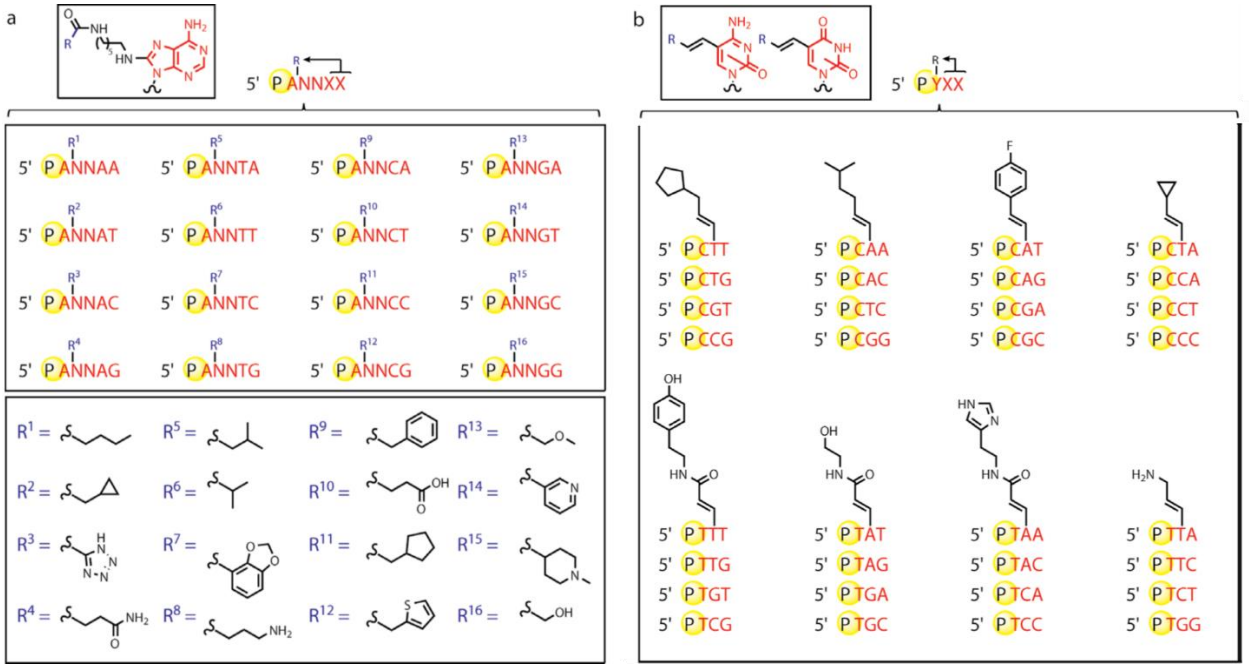


Figure 3-9: Modified anticodon sets used during LOOPER. a) 256-membered pentanucleotide anticodon set encoding 16 unique chemical modifications. b) 32-membered trinucleotide anticodon set encoding eight unique chemical modifications.[56]

More recently, T3 DNA ligase has been used in LOOPER-SELEX with a 32-membered YNN anticodon library encoding eight modifications (Figure 3-9). This library was used to generate aptamers against Proprotein convertase subtilisin/kexin type 9 serine protease (PCSK9) and interleukin-6 (IL- 6).[59] After 9 rounds of selection, aptamer PCSK9-A5 was isolated, demonstrating a K_D of 98 nM. This shows a markedly higher affinity against PCSK9 than a recently evolved unmodified DNA aptamer, known as AP-1, which had a K_D of 294 nM.[116] Deletion analysis showed that three residues were important for binding, namely phenol at position 9 isobutyl at position 11, and cyclopentyl at position 12. PCSK9-A5 was further reselected for another 6 rounds of SELEX to yield PCSK9-evo5, which exhibited a K_D of 3 nM against PCSK9 in its truncated form. Interestingly, this aptamer showed no predicted secondary structure, again highlighting that the dense modifications installed using LOOPER can lead to non-canonical folded structures beyond Watson-Crick-Franklin base-pairing. T3 DNA ligase-mediated

LOOPER-SELEX was also performed against IL-6, which after 7 rounds of selection, generated aptamer IL6-A7 that exhibited high affinity for IL6 with a K_D = 12 nM (Figure 3-10).

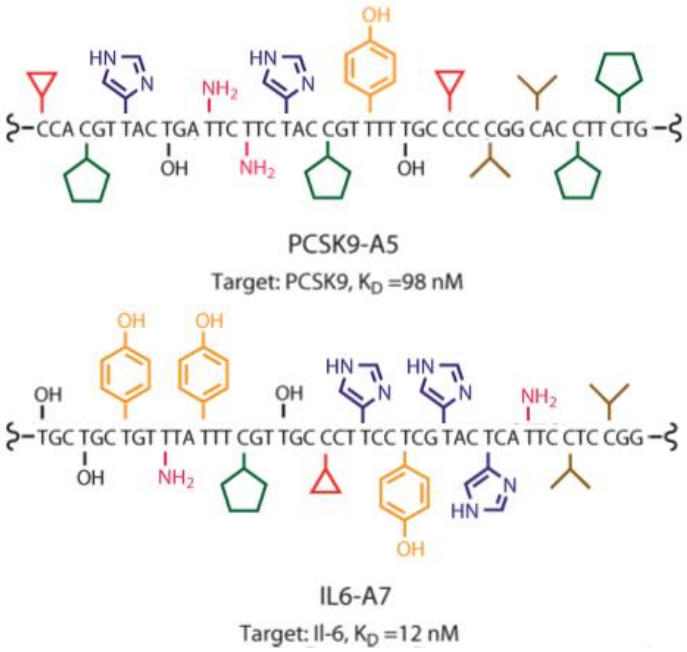


Figure 3-10: Aptamers evolved using T3 DNA ligase driven LOOPER-SELEX *via* modified trinucleotides including their binding and inhibition constants.[56]

This affinity was only marginally tighter than previously reported unmodified DNA aptamers for IL6, such as aptamer 12L, which was isolated after 5 rounds of selection against IL6 with a K_D of 17 nM.[117] IL6-A7 also contained two critical isobutyl groups at positions 14 and 15, which when deleted ablated binding. In a recent follow-up study, anticodon libraries that were unfunctionalized, charged, polar, nonpolar, or fully functionalized were used during selection against PCSK9 and IL-6.[118] The selection results demonstrated that modified aptamer libraries containing non-polar modifications converged more quickly and outperformed libraries lacking non-polar modifications, providing guidelines for the development of future LOOPER-derived aptamer libraries.

3.2.2. Limitations LOOPER-SELEX

The natural evolution of the biological genome utilizes codon sets that are three nucleotides in length, similar to LOOPER-evolved aptamers; there is a variety of codon bias and mutation-selection drift associated with this type of selection. The incorporation of a given modification is affected by the GC content of its codon and compatibility with the ligase. LOOPER exhibits varying levels of codon bias *via* differences in anticodon incorporation efficiencies. LOOPER codon enrichment is calculated as the frequency of a codon in the polymer divided by the frequency of the cognate anticodon in the template. Enrichment >1 means favored codon incorporation during LOOPER while <1 means disfavored codon incorporation.[55] The selection pressure of the target can be quantified using Jensen−Shannon (JS) divergence, which measures the difference between two probability distributions. The enrichment limit of functionalized LOOPER anticodons quickly decreases after each round of *in vitro* selection.

Successful *in vitro* selection requires multiple rounds of SELEX to evolve aptamers with high affinity towards a target. Throughout the evolution process, high-affinity binders (stable and complex structures) may be eliminated due to PCR bias and other limitations; inaccessibility of primer binding sites, complex structural motifs and GC-rich regions, and modified nucleotides.[119–121] While traditional multi-round SELEX can often enrich a single sequence to well above 10% of the library, SELEX coupled with high-throughput sequencing is sufficient to detect lower levels of enrichment with fewer rounds.[55] Furthermore, even with high enrichment, winning sequences may not represent the tightest binders as PCR bias and other factors can adversely affect SELEX at high cycle numbers. This notwithstanding, the theoretical limit to the enrichment of a single sequence within a LOOPER-derived library can be estimated by dividing the amount of unique pentanucleotide anticodon against the amount of unique template. The LOOPER protocol uses 480

pmol of equimolar anticodon library consisting of 256 unique anticodons. This means 1.9 pmol of each unique anticodon. After round 1, template libraries typically hover at about 15 pmol at the beginning of each round. Thus, the estimated limit is 12.7 % (1.9/15) enrichment of a single sequence assuming that the aptamer uses 8 different anticodons.[55] Repeated anticodons within the same aptamer sequence would further lower the estimated limit of enrichment. This comparatively low sequence enrichment was also observed in LOOPER-SELEX against IL-6 protein. This limit to enrichment is an expected result of anticodon scarcity during LOOPER. Nevertheless, future practitioners of LOOPER-SELEX should keep in mind its limits in sequence enrichment when performing and analyzing selections.

3.2.3. LOOPER-SELEX evolutions *via* Swap-codon libraries

Unmodified aptamer libraries rely primarily on folding and stacking of the sequence within itself, forming inherent shapes like the stem and loop hairpins, bulges, symmetric and asymmetric internal loops, pseudo knots, and G quadruplexes.[122,123] On the contrary, introducing modifications augment target interactions through additional binding modalities. Modifications must be oriented and displayed appropriately for the opportunity to interact with a target; hence, modified oligonucleotides are predisposed to complex secondary structures that accommodate binding interactions. Herein, the swap-codon library selections encompass an in-depth evaluation of the LOOPER-SELEX method. Comparing empirical observations from previous success with TBL1 against newly designed LOOPER libraries with randomized codon sets to delineate the preference for secondary structure versus enriched chemical modifications. Notably, the modifications are displayed *via* a hexamethylene diamine (HMDA) linker, they have enough flexibility to engage in binding with the protein target properly, however, the linker alone is not believed to aid in binding

interactions.[55] The TBL1 sequence synthesized with the HMDA linker alone in place of modifications does not show any binding to Thrombin.

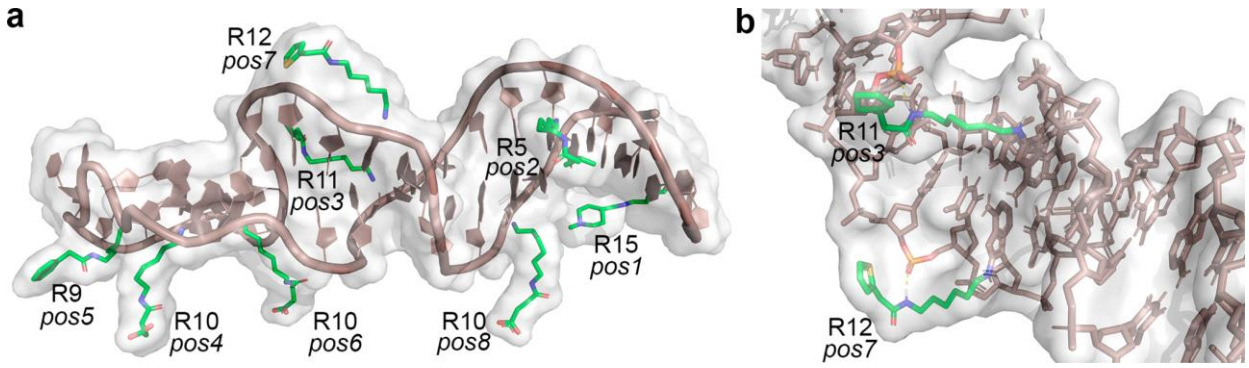


Figure 3-11: Structural models of TBL1. (a) Full aptamer structure. Green residues are the HMDA-linked modifications. (b) Major groove containing modifications that engage in hydrogen bonding (yellow hashed line) between the amide of the linker and the phosphate backbone.[55]

To gain insight into the possible folded structures and display of modifications on TBL1, computational modeling was simulated *via* YASARA structural modeling software (Figure 3-11). The software utilizes molecular dynamics and provides all the functions required to predict and validate macromolecular structures, including highly accurate force fields containing knowledge-based potentials. The simulated model of TBL1 in the presence of monovalent salts enabled the structure to fold into the structure shown in Figure 3-11a. The functional group modifications in the loop region are positioned in close proximity to each other, due to the slight twist along the oligonucleotide, and are displayed on the same face of the aptamer. The carboxylic acid modification in position 8 is also displayed on the same face, positioned *syn* to the loop modifications. As proposed earlier, this enables the hydrophobic and carboxylic acid modifications to interact with the allosteric exosite-I site on Thrombin. Additionally, hydrogen bonding between the amide of the HMDA linker and the phosphate backbone of the aptamer (Figure 3-11b) such as the bridging in the major groove (position 7) may be involved in the observed increased thermal stability in the folded structure.

3.2.4. Materials and methods

3.2.4.1. General information

Unless otherwise noted, water was purified with the Direct-Q® 3 UV Water Purification System. All primers and templates were purchased from Integrated DNA Technologies. Buffer recipes are included below.

Ligation reaction buffer 4X pH 7.6

(40 mM $MgCl_2$, 24% w/v PEG6000, 40 mM DTT, 264 mM Tris)

Strand separation buffer pH 7.5

(40 mM HEPES, 125 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.05% TWEEN- 20)

Selection buffer pH 7 (SELEX buffer)

(20 mM Tris-HCl, 140 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$)

Elution buffer pH 8

(40 mM Tris-HCl, 10 mM EDTA, 3.5 M urea and 0.02% TWEEN20)

3.2.4.2. Primers and templates

LOOPERtemp /5BioTinTEG/GAT TCG CCT GCC GTC GCA NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TCA CGT GGA GCT CGG ATC C

poly_pr1 5′- /5Phos/TGC GAC GGC AGG CGA ATC -3′

poly_pr2 5′- GGA TCC GAG CTC CAC GTG -3′

PrBt_PCR /5BioTinTEG/GATTCGCCTGCCGTCGCA

PEG-prB AACAACAACAACAA/iSp18/GGATCCGAGCTCCACGTG

LOOPERtemp-V2 /5BioTinTEG/CAG GTC CTG CTC GTG CCA NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TCG TCA GAG CTC GCG TCG A

poly_pr1-V2 5′- /5Phos/TGG CAC GAG CAG GAC CTG -3′

poly_pr2-V2 5′- TCG ACG CGA GCT CTG ACG -3′

PrBt_PCR_V2 5′-/5BiotinTEG/CAG GTC CTG CTC GTG CCA

PEG-prB_V2 AACAACAACAACAA/iSp18/TCG ACG CGA GCT CTG ACG

LOOPERtemp_V3 5′-/5BiotinTEG/GGT GCC GTC CTC ATA CGC NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TCG CAC TGC TCA GGA TGC G -3′

LOOPERtemp_V4 5′-/5BiotinTEG/CCG CAC GTA CCG GTT GTC NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TAG TGC CGC TTG CGG CAC A -3′

LOOPERtemp_V5 5′-/5BiotinTEG/TCT TCA GCG CCT GCG ACG NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TGC CAG CAG CTT GAC TGG C -3′

poly_pr1_V3 5′-/5Phos/GCG TAT GAG GAC GGC ACC -3′

poly_pr1_V4 5′-/5Phos/GAC AAC CGG TAC GTG CGG -3′

poly_pr1_V5 5′-/5Phos/CGT CGC AGG CGC TGA AGA -3′

poly_pr2_V3 5′- CGC ATC CTG AGC AGT GCG -3′

poly_pr2_V4 5′- TGT GCC GCA AGC GGC ACT -3′

poly_pr2_V5 5′- GCC AGT CAA GCT GCT GGC -3′

PrBt_PCR_V3 5′-/5BiotinTEG/GGT GCC GTC CTC ATA CGC

PrBt_PCR_V4 5′-/5BiotinTEG/CCG CAC GTA CCG GTT GTC

PrBt_PCR_V5 5′-/5BiotinTEG/TCT TCA GCG CCT GCG ACG

PEG-prB_V3 AACAACAACAACAA/iSp18/CGC ATC CTG AGC AGT GCG

PEG-prB_V4 AACAACAACAACAA/iSp18/TGT GCC GCA AGC GGC ACT

PEG-prB_V5 AACAACAACAACAA/iSp18/GCC AGT CAA GCT GCT GGC

3.2.4.3. LOOPER: Polymerization

The first round of polymerizations was set up at a 120 pmol scale; 8 reactions were set up at 15 pmol scale in parallel. In a PCR tube, a mixture of 1.5 µL LOOPERtemp (10 µM in H_2O), 2.25 µL poly_pr1 (10 µM in H_2O), 2.25 µL poly_pr2 (10 µM in H_2O), 5 µL ligation reaction buffer 4X, 2 µL ATP (0.25 mM in H_2O) and 4 µL H_2O was heated to 90 °C for 2 minutes and then cooled to 25 °C at a rate of 0.1 °C/s. This was followed by the addition of 1 µL functionalized pentamer mixture (480 µM in H_2O), 1 µL BSA (2 mg/mL in H_2O), and 400U (1 µL) T4 DNA ligase (New England Biolabs, M0202L). The mixture was incubated to polymerize for 24 hours at 25 °C. The products were then purified with an OMEGA® PCR Purification Kit before strand separation.

After the first selection round, the polymerization reaction was carried out at a 15 pmol scale every subsequent round. Note, a minimum requirement of 15 pmol of the biotinylated template library from the output after each selection round for follow-up rounds.

3.2.4.4. LOOPER: Strand Separation

Prior to start, 40 µL (5 µL/15 pmol DNA) of streptavidin beads (Dynabeads™ MyOne™ Streptavidin C1) were washed with 400 µL (10× the volume of beads) 1× strand separation buffer, three times. The dry beads were then treated with 200 µL of 2× strand separation buffer before incubating (30 minutes on a rotator at room temperature) with the purified dsDNA (200 µL in

miliQ-water). Please note the beads can crack if left dry or frozen (at -20 °C). Post incubation, the supernatant was extracted, and the streptavidin beads were washed with 400 µL of 1× strand separation buffer three times. Then to elute the non-biotinylated template strand, the beads were treated with 40 µL (320 µL for the first round) of 150 mM NaOH and mixed by pipetting to avoid using a vortex. Since the aptamer library is the non-biotinylated strand, the supernatant was extracted into a new microcentrifuge tube and neutralized using 4 µL (32 µL for the first round) of 1.5 M HCL. The beads were discarded. The neutralized supernatant (containing the LOOPER aptamer library) was purified with Princeton Separation columns and dried on a speed vacuum.

3.2.4.5. LOOPER-SELEX: In Vitro Selection against Thrombin

The dry aptamer libraries were brought to 55 µL volume with 11 µL 5× SELEX buffer and 44 µL purified elution (8-fold everything for the first round, total volume= 440 µL). To renature the aptamers in solution, they were incubated at 90 °C for 4 minutes and snap-cooled on ice for 10 minutes. The solutions were maintained at room temperature for 30 minutes while 50 µL of Thrombin-bound agarose beads were pre-washed (with 50 µL 1× SELEX buffer, three times); prior to the first pre-wash, the initial supernatant was discarded. Please note the larger radius beads, such as sepharose or agarose beads, require a wide bore pipette tip to extract even and accurate volumes from the stock solution. Normal pipette tips can be made wide bore by cutting off the tip of the pipette using a clean razor. After all pre-washes were complete, the supernatant was discarded prior to incubation (15 minutes on rotator) with the 55 µL LOOPER-aptamer library (in 1× SELEX buffer). Please note in the first round of LOOPER-SELEX four selections were set up (2-fold the volume), for each library, washing 100 µL of beads with 100 µL 1× SELEX buffer. The LOOPER aptamer library solution was divided into four samples (i.e., 110 µL of aptamer library in each microcentrifuge tube). After incubation, the beads were spun down at 376 RCF for 1 min; the

supernatant was extracted into a new microcentrifuge tube labeled “Flow Through” (FT). This was followed by three washes, 50 µL 1× SELEX buffer, and every time the supernatant was collected as “Washes” 1-3 (W1, W2, W3). It is important to leave some elution above the beads when pipetting out FT, W1, W2, and W3, but for the final elution the beads were left completely dry as the elution contained binders. Finally, to elute the binders from the solution, the beads were treated with 50 µL of elution buffer and incubated for 4 minutes at 90 °C. Please note low speeds on the centrifuge result in poor separation. High speeds result in the aggregation of agarose beads. The binders were then purified using pre-hydrated Princeton Separation columns. These SELEX winners from each round were then dried on the speed vacuum prior to qPCR amplification.

3.2.4.6. LOOPER-SELEX: qPCR amplification

The dry purified SELEX winners were brought to 81 µL with nuclease-free water to be divided up as eight (10 µL) qPCR samples and one (1 µL in 9 µL nuclease-free water) output sample. The collected supernatants during selection were prepared as FT, W1, W2, and W3 samples (5 µL of each in 45 µL of nuclease-free water, using 10 µL for qPCR). A 10-fold dilution series of the LOOPERtemp library was also prepared as a positive control (5 dilutions, 10 µL used for qPCR, 1 pM-10 nM). Please note the samples along with water only (10 µL) No Template Control (NTC), 45 samples in total, required at least 1800 µL of qPCR master mix (made fresh before the run). To make the master mix scale up the following 1× mixture enough for 1 sample. In a PCR tube combine 2.5 µL of each primer (10 µM in H_2O), 5 µL of dNTP (2 mM from KOD pol. kit), 3 µL of $MgSO_4$ (25 mM from KOD pol. kit), 0.5 µL of Sybr Green (100× in H_2O), 5 µL of KOD buffer (10× from KOD pol. kit), 20.5 µL of H_2O, and finally 1 µL of KOD Hot-start polymerase. 40 µL of the qPCR master mix (please check the supplementary information) was added to each sample.

The loss of diversity during SELEX rounds was monitored by the default melting temperature of the qPCR instrument. Please note all samples were prepared in white qPCR tubes with clear flat caps and the master mix was added last while the tubes were placed on ice. qPCR was carried out using a Bio-Rad instrument, real-time monitoring was enabled by incorporation of Sybr Green in the master mix. Initially, the KOD Hot-start polymerase was activated at 95 °C for 3 minutes, followed by one cycle of denaturation (95 °C for 20 seconds), annealing (54 °C for 10 seconds), and extension (70 °C for 10 minutes). This was followed by 30 cycles of denaturation (95 °C for 20 seconds), annealing (54 °C for 10 seconds), and extension (70 °C for 1 second). Please note the first 10-minute step was necessary for KOD polymerase to read the modified strand. To avoid by-products *via* overamplification, the 8 aliquots of SELEX winners were taken out at the halfway point before the plateau. The winning aptamer aliquots were pooled and purified using the OMEGA PCR cleanup kit.

3.2.4.7. LOOPER-SELEX: PCR amplification

For PCR the samples were prepared with 2.5 µL of each primer, 10 µL sample, 25 µL Q5 polymerase (2× Master Mix), and 10 µL of nuclease-free water. PCR amplification was carried out using Q5 Hot-start polymerase, 10 cycles. The amplified aliquots were pooled and purified using an OMEGA PCR cleanup kit. Note, 15 pmol requirement of template going into each new selection round.

3.2.4.8. LOOPER-SELEX: Biotinylated Strand Recovery

The eight qPCR aliquots were pooled and purified using at least three OMEGA PCR cleanup kits. The purified samples were combined and dried on a speed vacuum to about 10 µL. Samples were then quantified using a Nanodrop spectrophotometer to determine the required amount of streptavidin agarose beads (binding capacity 5 µL/15 pmol of DNA) for strand separation. Please

note, no less than 5 µL should be used if there is less than 15 pmol product. After every round of SELEX, about 1 % of the winners (dsDNA) were saved in qPCR tubes and kept at -20 °C for later use. The streptavidin beads (Dynabeads™ MyOne™ Streptavidin C1), were pre-washed with 1× strand separation buffer (10-fold the volume of beads), three times. The dry beads were then treated with 2× strand separation buffer (5-fold the volume of beads) before incubation with the purified post-selection library (5-fold the volume of beads), for 30 minutes on a rotator at room temperature. Please note the beads can crack if left dry or frozen at -20 °C. Post incubation, the streptavidin beads were washed with 1× strand separation buffer (10-fold the volume of beads), three times. Then to elute the PEGylated strand, the beads were treated with 150 mM NaOH (40 µL/ 5 µL beads) and mixed by pipetting to avoid using a vortex. Since the template strand for the next round was the biotinylated strand, the supernatant was discarded. The beads were resuspended in 10 mM EDTA, pH 8.2 with 95 % formamide (10× the volume of beads) and incubated for 2 minutes at 90 °C to extract the biotinylated LOOPER sense-strand. The supernatant was quickly removed to avoid re-binding to the beads. The biotinylated aptamer template strand was purified using the PCR purification kit and dried on a speed vacuum.

3.2.4.9. PAGE Gel Extraction

A 10% denaturing PAGE gel was prepared to fully isolate the biotinylated template strand from any leftover impurities by gel extraction. The gel was cast and cooked at 55 °C while running for 30 minutes at 150 V to eliminate excess urea. The dry sample was brought to 20 µL (10 µL of nuclease-free water and 10 µL of 2× dye-containing formamide) and denatured prior to loading, at 95 °C for 4 min. The samples were loaded and run at 55 °C for 30 min, at 150 V. To be able to observe the ssDNA bands, the gel was stained in 1× Sybr Safe staining solution for 30 minutes. The stained gel was placed over a blue light transilluminator to visualize and cut out the

biotinylated band; the biotinylated strand was shorter than the PEGylated strand and traveled farther down the gel. The extracted gel was then crushed into a slurry with 100 µL of 0.3 M NaCl before overnight incubation at 37 °C. The eluted ssDNA (biotinylated template strand) was purified from the slurry by Princeton Separation columns (pre-hydrated with water). A Nanodrop spectrophotometer was used to quantify purified DNA (dried with a vacuum centrifugation and brought to 10 µL prior to quantification). This purified DNA was used as the biotinylated template strand for the following round of LOOPER-SELEX until 6 rounds had been performed. Prior to each round, the template was dried on a speed vacuum and resuspended in 1.5 µL nuclease-free water.

3.2.4.10. NGS sequencing sample preparation

The selected output libraries were PCR amplified using Q5 polymerase and Ioncode barcoded primers. The full-length products were gel extracted on a native PAGE gel. The purified product bands were dried and quantified using Qubit. Concentration readings were taken in the Qubit® Fluorometer (ng/µL) and converted to molar units using the DNA copy number and dilution calculator provided by Thermo Fisher. The samples were brought to 50 pM concentrations for sequencing. A mixture of all samples at equimolar concentrations was prepared, same volume of each. Minimum total volume of 25uL of this mixture was the required loading volume onto the Ion Chef sequencing chip. Samples were sequenced using the Ion Gene Studio S5 System, setup to carryout 500 flows on 530 chip type.

3.2.5. Results and discussion

The three swap-codon libraries were synthesized and strand-separated as individual LOOPER libraries in parallel. Each library underwent six iterative rounds of LOOPER-SELEX, where the outputs from each selection round were qPCR amplified for 5-10 cycles to avoid overamplification

with KOD polymerase. Further amplification was carried out post-qPCR *via* Q5 polymerase to reduce PCR bias. The quality of binder enrichment was evaluated from qPCR data after each round, comparing the ratio of eluted output (representative of binders) relative to flowthrough (representative of non-binders) provided a metric to track the progress of library recovery.[119,124] Whereby the cycle at which the amplification curve crosses a set threshold signal (horizontal line) at a concentration represented by the C_T value, which is inversely proportional to the sample concentration such that [DNA] $\propto e^{-C_T}$. Amplification curves of post-selection libraries demonstrated enrichment as the output sample C_T values gradually decreased while the flow-through sample C_T values increased (Figure 3-12). This evaluation considers the high abundance of non-binders in the starting library during the preliminary rounds of selection versus the desirable enrichment of binders in the final selection rounds as non-binders become scarce. The first selection round was exempt from qPCR-based analysis since the entire output pool included single copies of each potential binder, all sequences were recovered early from qPCR amplification. Simultaneously, melt curves of the output samples from each library were measured to monitor the structural complexity in the post-selection libraries of each round based on changes in melting temperatures and loss of diversity based on the width of the curves. Increased remelting curves indicated highly folded structures, and the narrowing width of the parabolic melt curves indicated convergence of sequence families in the binder outputs. Relatively high C_T values of the NTC samples were used to confirm a lack of contamination during selections. Although qPCR analysis provided insight during selection rounds, the success of the three swap-codon aptamer evolutions was inconclusive prior to sequencing in the final step after six rounds of selections for each library were complete.

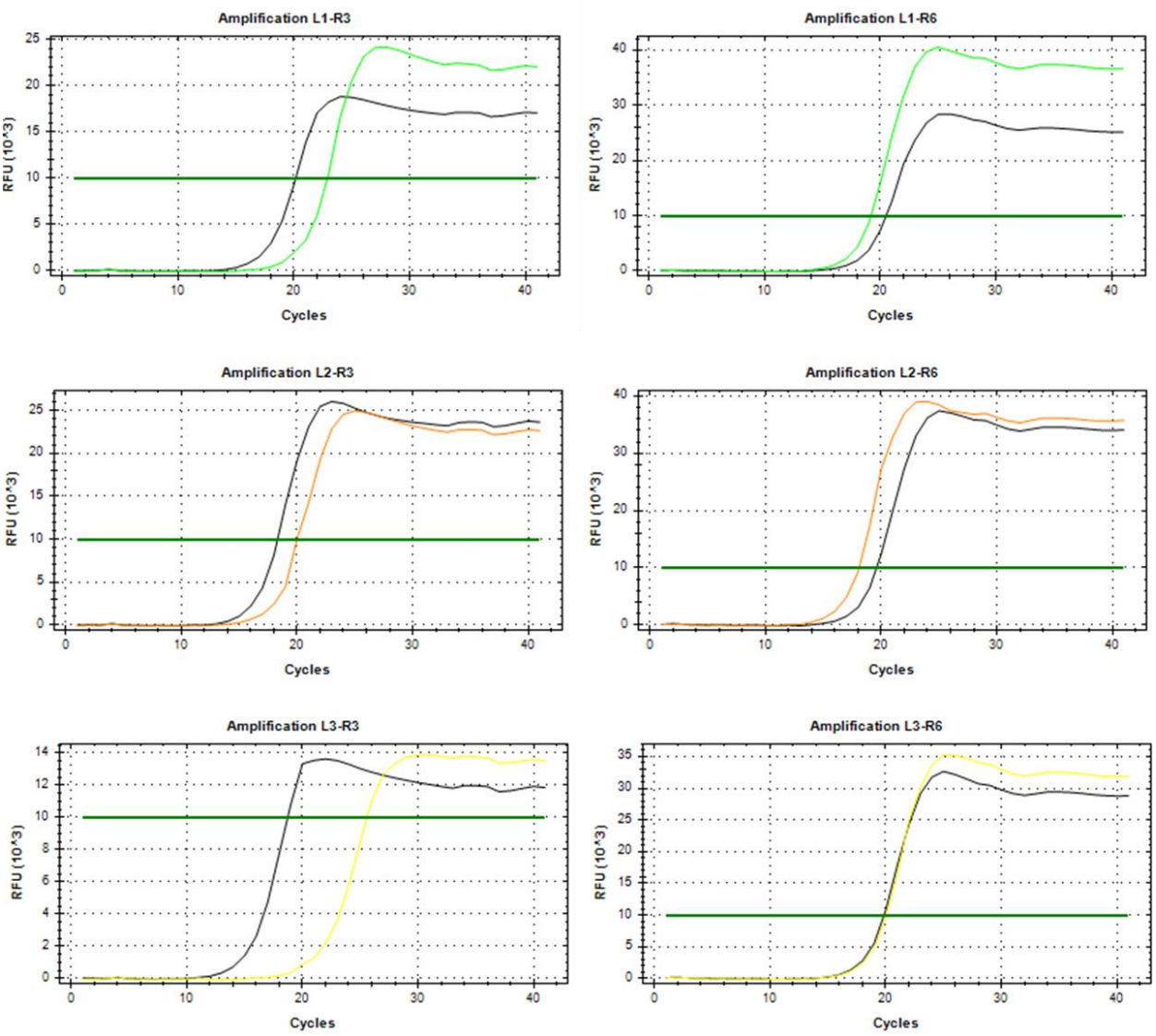


Figure 3-12: Real-time qPCR amplification curves of swap-codon library outputs (binder curves in color) relative to flow-through (non-binder curves in black) from rounds 3 and 6; library1 output is represented in orange, library2 output is represented in green, and library3 output is represented in yellow.

These selections were attempted three times, each time making corrections to the method in pursuit of a successful evolution. High throughput DNA sequencing enabled a more conclusive analysis of the post-selection libraries from rounds 3 and 6. Each post-selection library was amplified with the appropriate ion-code barcodes for Ion Gene Studio S5 Systems[125]. Sequencing data was evaluated using the FASTAptamer[126,127] developed by the Burke lab as a user guide to count the

number of reads for each sequence in the post-selection pools providing correspondent frequencies in reads per million (RPM) to evaluate enrichment.

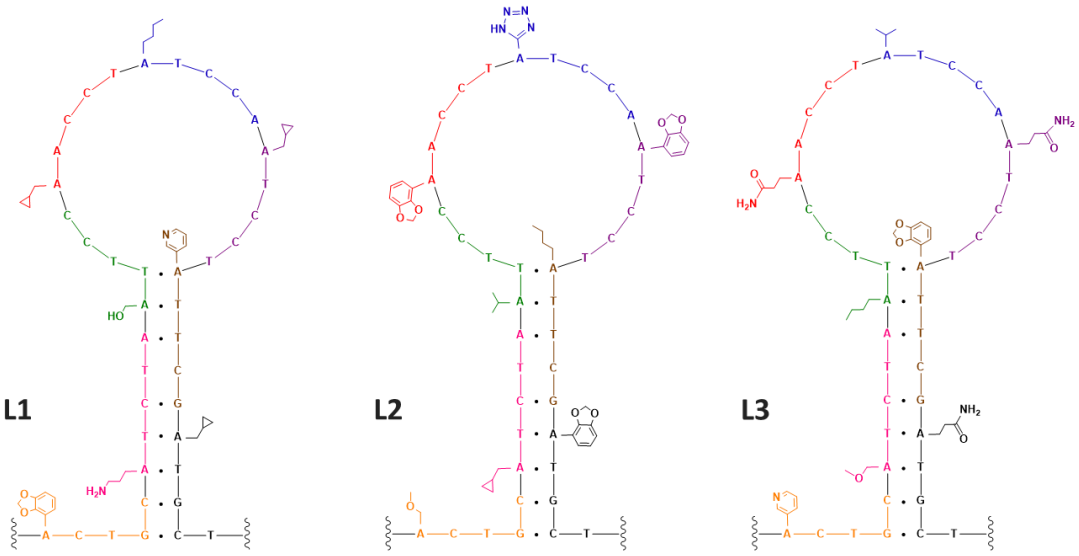


Figure 3-13: Enriched sequences of the TBL1-doped selection of swap-codon libraries against Thrombin; where L1 corresponds to library 1, L2 corresponds to library 2, and L3 corresponds to library 3.

The first attempt at the swap-codon aptamer evolutions resulted in a unanimous convergence onto the same DNA sequence for all three post-selection libraries, from both round 3 and round 6 sequencing data. This winning aptamer sequence was the same DNA sequence as TBL1. Hence, the LOOPER-SELEX evolutions of the three libraries against Thrombin resulted in highly modified functional aptamers displaying completely different modifications along the same hairpin loop structure (Figure 3-13). These results were due to overlapping synthesis of TBL1 in the lab, which corroborated a minor contamination that quickly took over all three libraries after a few rounds of SELEX and PCR amplifications. Initially, the primers used for the swap codon libraries were conserved from the TBL1 selection to enable a direct comparison. However, the second attempt for swap-codon selections removed the potential of TBL1 contaminations *via* newly designed primers of similar T_m to that of TBL1. Version two (V2) selections although promising through qPCR monitoring during selections, demonstrated no enrichment of LOOPER-

aptamers for all three swap-codon libraries in the sequencing data. The output libraries of all three swap-codon libraries demonstrated low frequencies of reads; Table 3-8 represents a sample of top ten sequences from library-3 from round 3 and round 6 of the selections. This consistent lack of success drew attention to the selection method as written in the protocol.[53]

Table 3-8: Sequencing data of V2 selections representing Library 3 outputs from round 3 and round 6 of LOOPER-SELEX; the data was analyzed by FASTAptamer count.

	Sequence (description line -> rank. reads. reads per million)
Library 3 (Round 3)	>1-14-3.59 TCGACGCGAGCTCTGACGATATGACCTTATTTAATGACAAGTGAATATAATACATGTATGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGATTCAATCAGACTCAAGCATATGTCAAATTAATACAATGATGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGACAACAAACTAGTGAACCAGATTAGACTATACCACACCGATGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGACTTAACTGGACTTTAAGTCACATAAGATGACGTGATATGTGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGACTGCACTATATATTATCTAATGTAAGGCGACTTCAAGGCTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGAATTCAGTATACTGCACTACAAACAATTTCACACTACGAGTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGAATTCAACTAACTAGAATGCATTGTAGCCGATGTCAAAGCTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGACCCAACTTAATCAAATTTGACAACAAAGTATACGAAACTTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGACCTTACGGAATCGCATAAAAGCTAAATCTACAGTATAATTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGACGTTAACGCAAATCAGTTCATCGAACTGCAAATAAGTCCTGGCACGAGCAGGACCTG
Library 3 (Round 6)	>1-14-5.19 TCGACGCGAGCTCTGACGATCTAATTCTATGAGATGAGATTTAATTTAATTTGAGAAGTGGCACGAGCAGGACCTG
	>2-13-4.82 TCGACGCGAGCTCTGACGATTTGATGCGAGAGGATAAAACAGTATCATATTTTAGTAGTGGCACGAGCAGGACCTG
	>2-13-4.82 TCGACGCGAGCTCTGACGAAACTATCCGACCTGAGACTAAGAGATGCCAAGAAATAAGTGGCACGAGCAGGACCTG
	>2-13-4.82 TCGACGCGAGCTCTGACGAAATGAACTGATACCAACCGATTTGATCCTATCCGACTCATGGCACGAGCAGGACCTG
	>2-13-4.82 TCGACGCGAGCTCTGACGATACCAACCGAGATCAATTTATTAGAATCAAAGATATATTTGGCACGAGCAGGACCTG
	>3-12-4.45 TCGACGCGAGCTCTGACGATTGGATCTTATGTTATGGCATTGGATATAAGTAGATTTCTGGCACGAGCAGGACCTG
	>3-12-4.45 TCGACGCGAGCTCTGACGACGTAATGGCAACTAAAGCAAGGTAACGCCAGCAGATGCTTGGCACGAGCAGGACCTG
	>3-12-4.45 TCGACGCGAGCTCTGACGACGTCACTCTAACTCACATCATTAGATTACAACTCACAAGTGGCACGAGCAGGACCTG
	>3-12-4.45 TCGACGCGAGCTCTGACGAGTGGAAAGGAAGATAGCATACGAAAGGTTATGTAATCAATGGCACGAGCAGGACCTG
	>4-11-4.08 TCGACGCGAGCTCTGACGAGATAATAGCATACTATTAAAAAGCAGCCAAAGTCATTACTGGCACGAGCAGGACCTG

Finally, the third attempt (V3) of the swap-codon library selections incorporated changes to the protocol. This selection used 8 times more Thrombin in the first selection round to avoid loss of potential binders. Further to this, the amount of template amplified and strand-separated for follow-up rounds was held at a minimum 15 pmol scale, which required additional PCR amplifications. To avoid potential contaminations between libraries, the swap-codon libraries each had unique primer sets.

Table 3-9: Sequencing data of V3 selections representing Library 1 outputs from round 1 and round 6 of LOOPER-SELEX; the data was analyzed by FASTAptamer count.

	Sequence (description line -> rank. reads. reads per million)
Library 1 (Round 1)	>1-14-3.59 TCGACGCGAGCTCTGACGATATGACCTTATTTAATGACAAGTGAATATAATACATGTATGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGATTCAATCAGACTCAAGCATATGTCAAATTAATACAATGATGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGACAACAAACTAGTGAACCAGATTAGACTATACCACACCGATGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGACTTAACTGGACTTTAAGTCACATAAGATGACGTGATATGTGGCACGAGCAGGACCTG
	>2-13-3.34 TCGACGCGAGCTCTGACGACTGCACTATATATTATCTAATGTAAGGCGACTTCAAGGCTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGAATTCAGTATACTGCACTACAAACAATTTCACACTACGAGTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGAATTCAACTAACTAGAATGCATTGTAGCCGATGTCAAAGCTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGACCCAACTTAATCAAATTTGACAACAAAGTATACGAAACTTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGACCTTACGGAATCGCATAAAAGCTAAATCTACAGTATAATTGGCACGAGCAGGACCTG
	>3-12-3.08 TCGACGCGAGCTCTGACGACGTTAACGCAAATCAGTTCATCGAACTGCAAATAAGTCCTGGCACGAGCAGGACCTG
Library 1 (Round 6)	>1-41-20.99 CGCATCCTGAGCAGTGCGATTTGACTCTATTGTATGGTAAGATAGTGGATTGTACCGTGCGTATGAGGACGGCACC
	>2-26-13.31 CGCATCCTGAGCAGTGCGATTGGATTTTAATGCATCTTATAGGATATAACTCAATACAGCGTATGAGGACGGCACC
	>3-23-11.77 CGCATCCTGAGCAGTGCGATCTTACTGTAGAGAATTCCAGTGTAGGGCAATGCAGCACGCGTATGAGGACGGCACC
	>4-22-11.26 CGCATCCTGAGCAGTGCGACTCAATCTTAGGGTAGCGCAGTCTATTTTATGCTATCAAGCGTATGAGGACGGCACC
	>4-22-11.26 CGCATCCTGAGCAGTGCGATGGAAGTGGAGCATATGATAGCCCAGTTTAGTCTAGTTTGCGTATGAGGACGGCACC
	>6-18-9.22 CGCATCCTGAGCAGTGCGATGCTATTTCACTGGATTCAATGTCACTAAACGCCACCATGCGTATGAGGACGGCACC
	>7-17-8.70 CGCATCCTGAGCAGTGCGATGTGACGTGAGTTTATGAAACGCTAGTCTATTTTACACCGCGCGTATGAGGACGGCACC
	>8-15-7.68 CGCATCCTGAGCAGTGCGATCTAATCTTACTTAATGTAATCTGATTTTACGAAAACGTGCGTATGAGGACGGCACC
	>9-14-7.17 CGCATCCTGAGCAGTGCGATTGGAGTTGATTACACCTAAATAGAGTCTAAGTAATTCTGCGTATGAGGACGGCACC
	>9-14-7.17 CGCATCCTGAGCAGTGCTATTGCACCTTAACATATCTAATTGCATTTTAGCGGACTAAGCGTATGAGGACGGCACC

The results demonstrated a lack of enrichment of LOOPER-aptamers for all swap-codon libraries; the sequencing data demonstrated low frequencies of reads. Table 3-9 represents a sample of top ten sequences from library-1 from round 1 and round 6 of the selections. This consistent lack of success drew attention to the extensive selection cycles and amplifications required for LOOOPER-SELEX. Given the complexity of LOOPER library synthesis and the randomized design of the swap-codon libraries, multi-round selections of aptamers *via* SELEX were not achievable. Every round, due to varying fidelities of the swap-codon libraries, the re-synthesis of binders was likely compromised reducing the chances of enrichment from one round to the next. Considering the time and resources required for LOOPER-SELEX evolutions, the swap-codon selections were ultimately reported inconclusive.

3.3. Introduction to Capillary Electrophoresis (CE)

The selection of high-quality aptamers from a combinatorial library of oligonucleotides *via* SELEX relies upon multiple rounds of enrichment. Although an increase in diversity of the starting library conceptually increases the probable chance of evolving potential binders, the main challenge lies within the selection method enabling a robust separation of binders from nonbinders in the output library. Alternative separation methods such as Capillary Electrophoresis (CE) have been explored to allow for the one-step selection of unmodified aptamers.[128] The two predominant modes of CE-based partitioning have been developed by the Krylov lab: NECEEM[129] (Non-Equilibrium CE of Equilibrium Mixtures) and IFCE[130] (Ideal-Filter CE). Both methods enable homogenous partitioning of aptamers based on the differences in electrophoretic mobilities of target-bound complexes versus unbound sequences. Considering the high volume of non-binder sequences in the starting library, the success of single-round selection using CE is limited by the non-binder background at the output. Due to the low abundance of binders (commonly 1 binder

per million or billion non-binders), the efficiency of partitioning for a successful one-step selection requires a maximum of 1 in 10^9 non-binders in the output.[130] The method relies on the accurate and efficient filtration of binders based on engineered electro-osmotic flow (EOF).[131,132] Similar to optical filtrations, the term "transmittance" can be used as a quantitative measure for partitioning defined as the ratio between quantities of binders and non-binders at the output and input of the capillary (Figure 3-14).[130] The binder and non-binder transmittance ratios can be adjusted through the following CE experimental conditions: the value of negative charge on the oligonucleotide, the hydrodynamic size of the protein target, the ionic strength and pH of the separation buffer, the length of the sample plug, the length of the capillary, *etc*.[129,133]

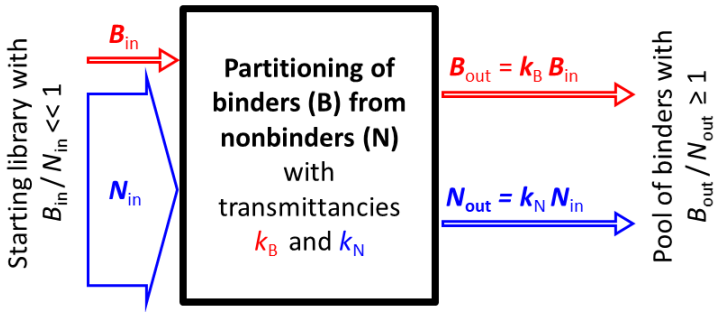


Figure 3-14: Model demonstrating single-round partitioning of binders from nonbinders in an aptamer library. Figure from: Analytical chemistry.[133]

The optimization of single-round selection using NECEEM and IFCE depends on both library design (high abundance of presumed binder sequences) and quantitative characterization of aptamer library separation using a mock library as a comparative measure.[133] The mock-library assembly includes varying concentrations of a known binder and non-binder (two different primer sets) in solution. Single-step partitioning followed by quantitative PCR of the output library enables direct characterization of partitioning efficiency.

3.3.1. LOOPER library evolution *via* CE-SELEX

Selection of high-affinity LOOPER aptamers *via* SELEX remains a challenging process. The multistep iterative rounds of LOOPER, selection against target, on-bead separation of binders from non-binders, amplification of winning aptamer pool, followed by gel extraction, create more chances for mistakes and failure. The combination of LOOPER, KOD polymerization, and Q5 polymerization can introduce mutations beyond the evolutionary requirement for aptamers after multiple rounds selection.

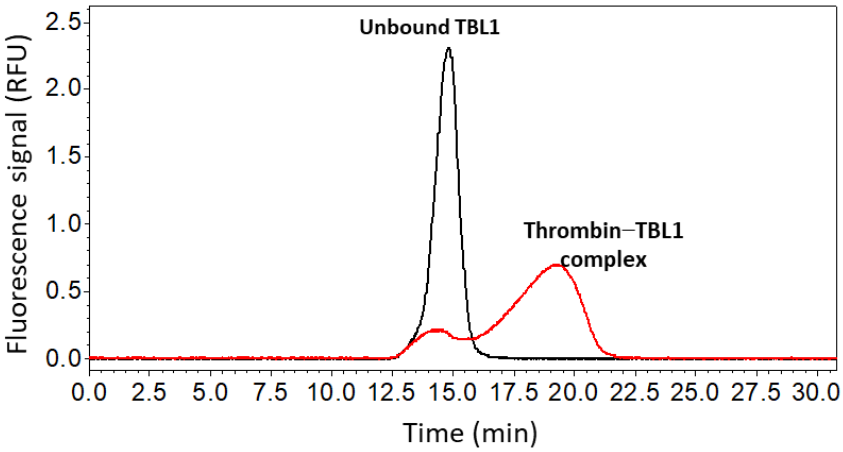


Figure 3-15: Positive control CE partitioning of LOOPER aptamer; TBL1-Thrombin complex formation for PVA-coated CE optimization. Figure from: An T. H. Le.

Herein, LOOPER libraries are the first highly modified functional aptamers to be selected using a CE partitioning method. LOOPER-CE selections were carried out as a comparative selection to the previously successful LOOPER-SELEX evolution against Thrombin for a proof-of-concept. In collaboration with the Krylov group, An T. H. Le designed a mock selection CE setup allowing the quantitative characterization of LOOPER aptamers partitioning in a selection against Thrombin. The previously mentioned TBL1 aptamer was synthesized and used as a reference for complex formation and partitioning in CE (Figure 3-15). Considering the positively charged

surface of Thrombin, the negatively charged silica walls of the capillary resulted in undesirable interactions of the protein target with the capillary. The loss of target (binder-complex) and high inevitable background required a PVA-coated capillary (polyvinyl alcohol).[134,135]

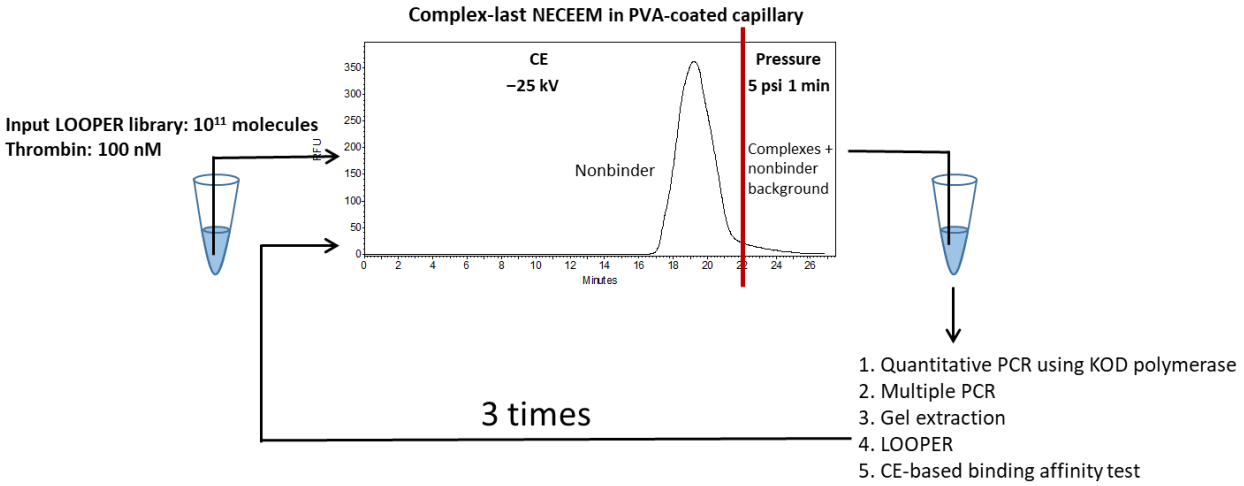


Figure 3-16: Schematic of LOOPER-CE aptamer evolution workflow. Figure from: An T. H. Le.

The PVA coated capillary avoids charge interactions, simultaneously removing the EOF and desirable partitioning efficiencies of IFCE. Hence, single-round selections were not possible due to the reduced enrichment; from as high as 10^9-fold to 10^3-fold per round. The LOOPER-CE selections were carried out in 3 rounds similar to SELEX, however the selection step *via* CE enabled homogenous immobilization-free target-binding (Figure 3-16). In this collaboration all LOOPER synthesis of the starting libraries and output regenerations were carried out in the Hili lab by me, while the three rounds of CE partitioning for aptamer selections were carried out in the Krylov lab by An T. H. Le.

3.3.2. Materials and methods

3.3.2.1. General information

Unless otherwise noted, water was purified with the Direct-Q® 3 UV Water Purification System. All primers and templates were purchased from Integrated DNA Technologies. All LOOPER

polymerizations and sequencing preparations were carried out as mentioned in section 3.2.4. for selections and Ion torrent NGS sequencing. Buffer recipes are included below.

Ligation reaction buffer 4X pH 7.6

(40 mM $MgCl_2$, 24% w/v PEG6000, 40 mM DTT, 264 mM Tris)

Strand separation buffer pH 7.5

(40 mM HEPES, 125 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.05% TWEEN- 20)

Selection buffer pH 7 (SELEX buffer)

(20 mM Tris-HCl, 140 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$)

Elution buffer pH 8

(40 mM Tris-HCl, 10 mM EDTA, 3.5 M urea and 0.02% TWEEN20)

3.3.2.2. Primers and Templates

LOOPERtemp /5BioTinTEG/GAT TCG CCT GCC GTC GCA NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TCA CGT GGA GCT CGG ATC C

poly_pr1 5′- /5Phos/TGC GAC GGC AGG CGA ATC -3′

poly_pr2 5′-/5Alex488N/GGA TCC GAG CTC CAC GTG -3′

PrBt_PCR /5BioTinTEG/GATTCGCCTGCCGTCGCA

PEG-prB AACAACAACAACAA/iSp18/GGATCCGAGCTCCACGTG

LOOPERtemp-V2 /5BioTinTEG/CAG GTC CTG CTC GTG CCA NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TCG TCA GAG CTC GCG TCG A

poly_pr1-V2 5′- /5Phos/TGG CAC GAG CAG GAC CTG -3′

poly_pr2-V2 5′-/5Alex488N/TCG ACG CGA GCT CTG ACG -3′

PrBt_PCR_V2 5′-/5BiotinTEG/CAG GTC CTG CTC GTG CCA

PEG-prB_V2 AACAACAACAACAA/iSp18/TCG ACG CGA GCT CTG ACG

3.3.2.3. CE selection

All functionalized pentamers and LOOPER polymerizations were carried out as previously mentioned in section 3.2.4. All CE experiments were performed with a P/ACE MDQ apparatus (SCIEX, Concord, ON, Canada) equipped with a laser-induced fluorescence (LIF) detection system. Fluorescence was excited with a solid-state laser blue light at 488 nm and detected at 520 nm using a spectrally optimized emission filter system. The capillaries had an inner diameter of 75 μm and an outer diameter of 360 μm. The poly(vinyl alcohol) (PVA)-coated capillary was prepared as described elsewhere. The total length of the capillary was 80 cm. The detector was located 10 cm away from the elution end. Prior to every fraction-collection experiment, a new capillary was installed and rinsed with the running buffer (50 mM Tris-Acetate at pH 8.2) at 20 psi for 10 min prior to the fraction-collection experiment.

In round 1, the sample contained 1.3 μM annealed LOOPER library (melted at 90 °C for 4 min, snap-cooled on ice for 10 min and then maintained at room temperature for 30 min) and 100 nM Thrombin. In round 2 and 3, the concentration of DNA in the sample was 0.33 μM. The sample mixture was injected with a pressure pulse of 1 psi × 28 s to yield a 3.7-mm long sample plug. The injected sample plug was propagated through the uncooled part of the capillary at the inlet by injecting a 5.4 cm long plug of the buffer with a pressure pulse of 0.9 psi × 45 s. CE was carried out at an electric field of −312.5 V/cm (−25 kV over 80 cm) in reverse polarity mode (negative

electrode at the injection end). Once the unbound DNA peak eluted out from the capillary, a pressure pulse of 5 psi × 60 s was applied to push the bound DNA (protein−DNA complexes) out from the capillary into a collection vial containing 10 µL of the running buffer.

3.3.3. Results and discussion

LOOPER-CE evolutions of the original LOOPER library against Thrombin were carried out in three rounds. The success of the selections was monitored prior to sequencing analysis *via* CE-based Thrombin binding affinity assay of the output libraries parallel to each other alongside a negative control and the starting library. The CE-based affinity assay from this first selection demonstrated the volume of target-bound complexes (Figure 3-17). The original library showed no observable affinity towards Thrombin, while the complex peak was consistently enhanced from round 1 to 3 outputs.

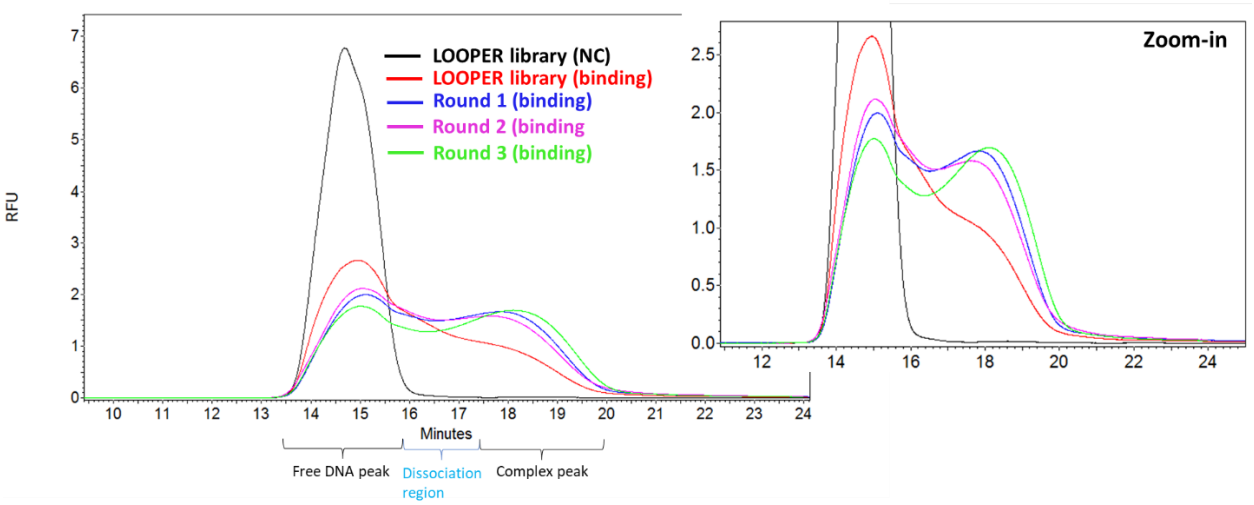


Figure 3-17: Affinity tests of 3-round selection for constant 100 nM Thrombin using LOOPER library. Sample: 10 nM DNA + 50 nM Thrombin. Sample buffer = Run buffer = 50 mM Tris Acetate pH 8.2, 0.1% Tween20. Capillary: length = 80 cm (length from the detection window to the capillary outlet = 10.2 cm), I.D. = 75 µm. The sample was injected to the capillary inlet at 1 psi for 28 s. The sample plug was propagated away from the uncooled region of the capillary at 0.9 psi for 45 s. Voltage = –25 kV. Coolant temperature = 15 °C. Figure from: An T. H. Le.

The CE affinity assay examined the quality of target-bound complexes of enriched aptamers; however, high throughput DNA sequencing enabled a more conclusive analysis of the post-selection libraries from each round compared to the starting library. The LOOPER-CE evolutions against Thrombin were carried out twice, in pursuit of successful selection results. The first attempt was treated as a mock selection due to CE contaminations with TBL1; remnant TBL1 in the capillary from the control experiments was included in the selection and survived amplification due to primer overlap (the selection library contained the same primer binding sites and TBL1). The post-selection library outputs were amplified with the appropriate ion-code barcodes for Ion Gene Studio S5 Systems[125]. Sequencing data was evaluated using the count and cluster library scripts within the FASTAptamer[126,127] program.

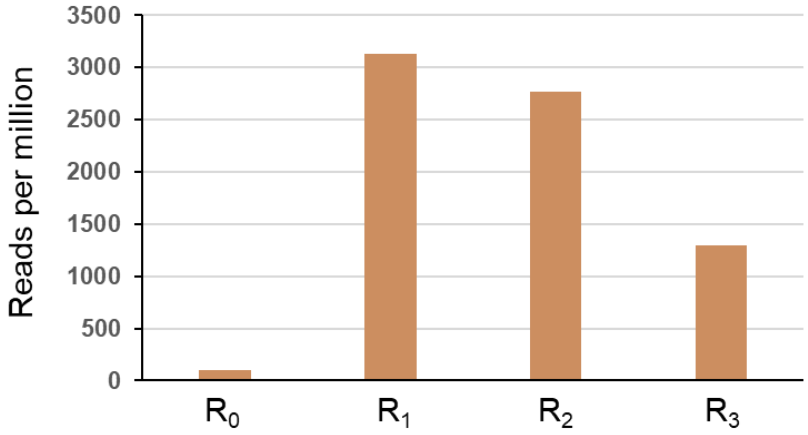


Figure 3-18: Sequencing data of mock selection for TBL1 and its SNPs (Single Nucleotide Polymorphism); The frequencies are represented in RPM (reads per million), where R_0 (starting library) represents an RPM of 97, R_1 (round1 output) represents an RPM of 3134, R_2 (round2 output) represents an RPM of 2765, and R_3 (round3 output) represents an RPM of 1300.

Upon sequencing the LOOPER-CE output libraries, the mock selection was partially successful in selecting TBL1. Sequencing results confirmed that TBL1 was enriched about 30-fold in the first round and further de-enriched (relative decrease in frequency of reads) gradually in the second and

third rounds of selection as the sequence is replaced by non-binder background (Figure 3-18). The TBL1 sequence and its SNPs were de-enriched by about 10 % from round one to two, and about 50 % from round two to three. On the contrary, sequencing results revealed enrichment of two unique sequences of potential aptamers from round two to three. The new emerging sequences, depicted in Figure 3-19, demonstrated higher single round enrichment compared to TBL1. Although these two sequences show promising secondary structures and functional groups for potential Thrombin-binding interactions, both were represented at random without any overlapping families of sequences. The TBL1 de-enrichment results along with the abrupt enrichment of the two unique sequences after round one raise a few concerns; notably for high non-binder background resulting in poor partitioning of binder-complexes, non-specific binding to Thrombin due to the target's DNA binding ability, and PCR by-products during library regeneration.

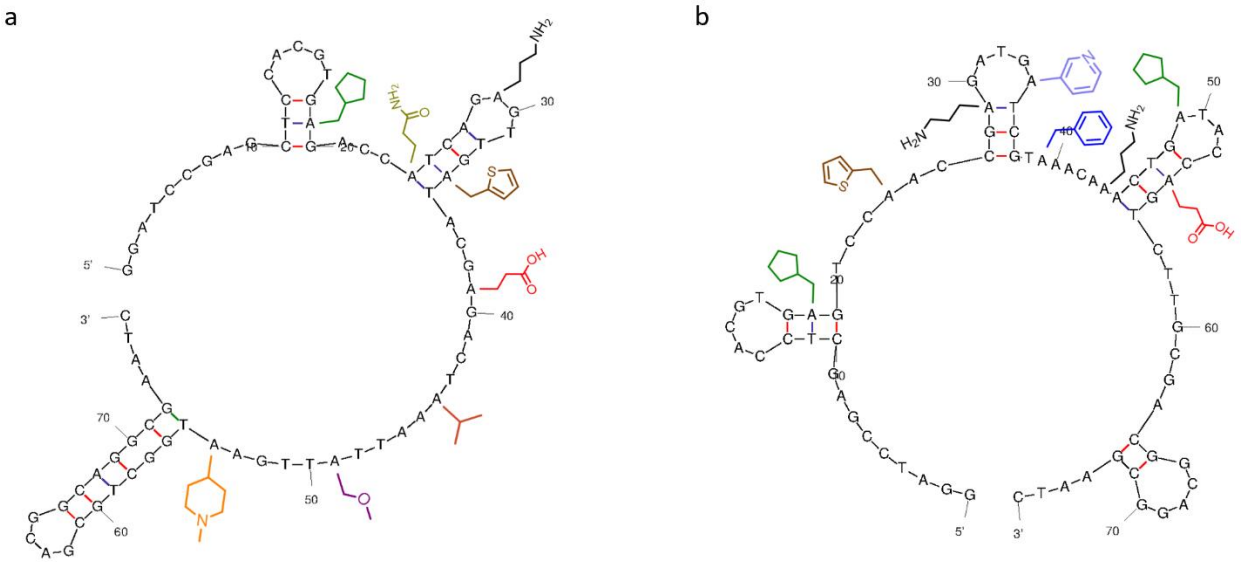


Figure 3-19: Emerging aptamer sequences from round 3 of the LOOPER-CE mock selection, a) demonstrated a ~150-fold enrichment from R2 to R3, and b) demonstrated a ~52-fold enrichment from R2 to R3.

In the subsequent LOOPER-CE selection against Thrombin, new primers were designed to avoid any contaminations from TBL1. The output libraries were amplified for the CE-based affinity assay to determine the quality of enrichment (Figure 3-20). The peaks representative of binder-complex was much smaller relative to the non-binder background peaks in all three post-selection outputs. The CE-data post-selection was not promising, drawing attention to the compromised partitioning quality when using a coated capillary for CE selections. Further to this, the library outputs were sequenced and analyzed as previously.

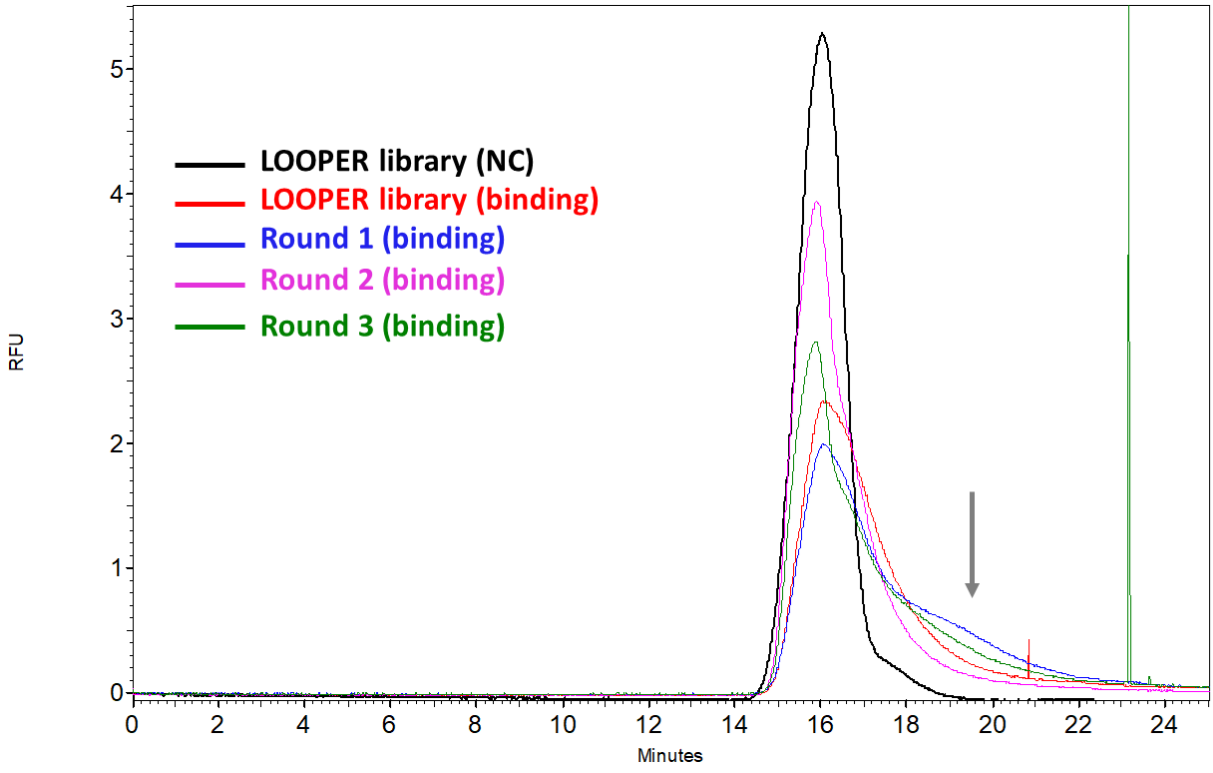


Figure 3-20: Affinity tests of 3-round selection for constant 100 nM Thrombin using LOOPER library. Sample: 10 nM DNA + 50 nM Thrombin. Sample buffer = Run buffer = 50 mM Tris Acetate pH 8.2, 0.1% Tween20. Capillary: length = 80 cm (length from the detection window to the capillary outlet = 10.2 cm), I.D. = 75 µm. The sample was injected to the capillary inlet at 1 psi for 28 s. The sample plug was propagated away from the uncooled region of the capillary at 0.9 psi for 45 s. Voltage = –25 kV. Coolant temperature = 15 °C. The complex peak is represented by the arrow. Figure from: An T. H. Le.

The sequencing data demonstrated a lack of enrichment of binders in the output of every round, as anticipated based on the CE-affinity assay. The output libraries represented a diverse array of sequences and no overlap between rounds. The method was rendered susceptible to a much higher non-binder background due to lower efficiencies in partitioning with the coated capillary hindering the presence of an EOF, and the small size of the target protein providing a small difference in non-binder versus binder-complex mobility.

3.4. Conclusions

The evolution of highly modified LOOPER aptamers requires a robust selection method enabling single-round selections. LOOPER-SELEX is a time-consuming and challenging process, vulnerable to PCR bias during selections and codon bias during LOOPER. Enrichment of high-quality aptamers relies on stringent separation of non-binders as well as accurate regeneration of every output candidate for follow-up rounds. Considering the potential loss of binders during LOOPER-SELEX, future selections should explore single-round selection methods. Capillary electrophoresis is a capable separation method for single-round aptamer selection; provided the presence of an optimized EOF for efficient partitioning. The EOF enables IFCE and optimal partitioning efficiencies to eliminate highly abundant non-binders in the starting LOOPER library. LOOPER-CE selections against Thrombin were herein designed using the original library to be compared with the LOOPER-SELEX results from the TBL1 selection. Although Thrombin is a great preliminary target for method development in other cases, however, in CE selections the positive charge and small molecular weight of the protein (37 kDa) make it challenging to achieve homogenous partitioning of binders from non-binders. Holding considerable promise, LOOPER-CE requires further optimization against other DNA-binding protein targets. Future LOOPER-CE selections will explore protein targets of larger size and more appropriate physiochemical

properties enabling non-coated capillary electrophoresis to reduce background signal and significantly increase enrichment. The qualified target must enable IFCE or NECEEM, where LOOPER binders are selected in a single step and avoid de-enrichment upon regeneration.

CHAPTER 4. SINGLE ROUND APTAMER SELECTIONS

4.1. Introduction

Aptamers are less common among FDA-approved oligonucleotide therapeutics compared to ASOs (antisense oligonucleotides), siRNAa (short interfering RNAs), and mRNA vaccines due to the challenges associated with aptamer selection processes for selective and high affinity target binding.[136] Selection of high-affinity aptamers *via* SELEX has no limit to the number of rounds; however, more selection steps facilitate inherent sequence-specific polymerization and amplification bias and failure. Not to mention the additional time and reagent consumption with increasing selection rounds. Ideally, the output library after multiple rounds of SELEX is analyzed for enrichment, with the dominant sequence families identified as successful candidates for further binding analysis. Numerous factors beyond high-quality target binding can influence enrichment during the selection rounds. Using highly diverse libraries requires PCR amplification between selection rounds; the first round of selection is the deciding factor for the success of the overall selection due to low copy numbers in the starting library.[137] Notably, PCR bias significantly impacts the evolution of winning sequences in each round and enrichment does not always correlate with the quality of target binding.[120] The consequence is competitive enrichment of low-quality binders that emerge over multiple rounds of SELEX; whereby high-quality binders are identified in the initial rounds of SELEX but become less prevalent in subsequent rounds due to de-enrichment. This PCR bias can be avoided by eliminating the amplification step in SELEX. Alternative non-SELEX methods minimize the rounds of *in vitro* selection without compromising affinity and avidity towards the target.[130,138]

The ideal solution is a method for single-round selection with stringent parameters to accommodate robust separation of binders out of the initial library pool, which contains mostly non-binders. The success of such selections relies heavily on the quality of the starting library to

promote binding interactions. Single-round selections in combination with next-generation sequencing and bioinformatic analysis is a powerful tool for aptamer selections compared to traditional SELEX. For instance, DNA-encoded libraries (DELs) of small molecules are selected against pharmaceutical targets in a single batch manner, where the starting library undergoes three rounds of target incubation and non-binder partitioning for optimal elution of high-quality binders.[139] The method relies upon higher copy numbers of each library member in the starting pool since the DEL library cannot be amplified between rounds. The resulting output molecules are often analyzed based on common structures and functional capabilities for binding interactions. In principle, the same method can be utilized for structured aptamer libraries, where the starting library focuses mainly on quality toward potential binding rather than an extremely diverse sequence selection as explored in the following chapter.

4.2. SHAPE aptamer libraries and single-round selections

Although commonly present as a B-form helix, cellular DNA is found in non-canonical forms and unique secondary structures. Regions of single-stranded DNA, during replication and repair or infection by viruses, prompt DNA folding into secondary structures recognized by proteins. DNA sequences with palindromic inverted repeats are predisposed to form hairpin structures; cruciforms are observed in double-stranded DNA.[140] Indeed, synthetic single-stranded DNA with inverted repeats behaves similarly; structural studies highlight the significance of loop regions for DNA-ligand binding. Conversely to DNA, naturally occurring RNA is single stranded enabling secondary structures (Figure 4-1). The cellular interactions between the transcriptome and the proteome are studied based on the enabling structures of RNA. Varying RNA types are classified based on the abundance of double-stranded regions corresponding to folded structures, creating bulges for protein contacts.[141]

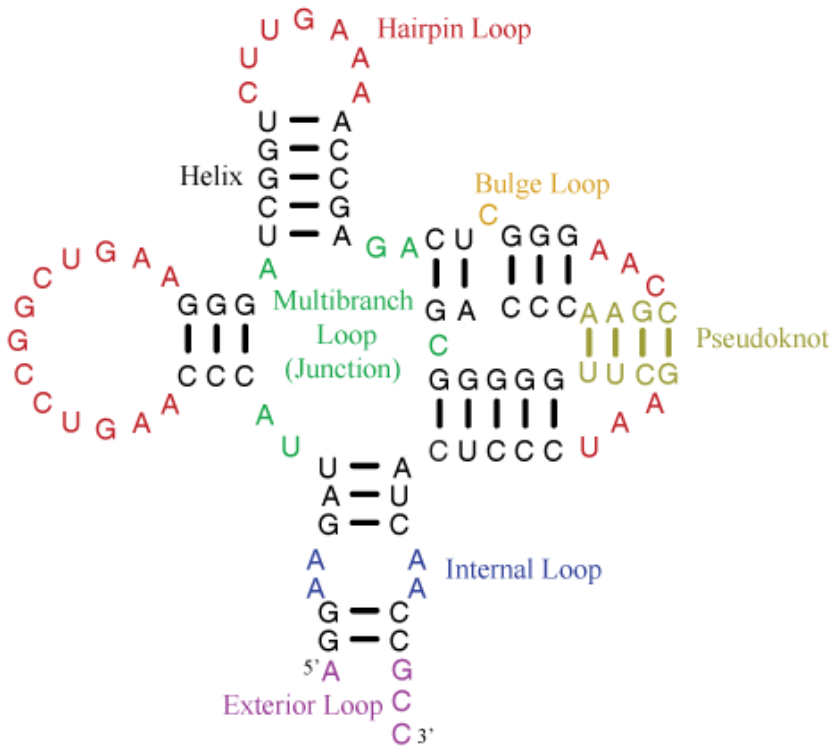


Figure 4-1: Common examples of RNA secondary structures. Figure adapted from: Nearest neighbor database.

Such prevalent secondary structures display nucleobase residues for direct interactions with epitope binning sites along protein targets.[107,142] The predominant idea behind SHApE (Structured Hairpin Aptamer Elution) includes a pre-designed hairpin structure for selective ligand binding. The preliminary SHAPE aptamer libraries contain a pre-defined 8-mer stem sequence and a randomized 12-mer loop for diversity, 5′-GGCTTCGT NNNNNNNNNNN ACGAAGCC AAACTT-3′ (Figure 4-2). The smaller library size, relative to frequently used 40-mer aptamer libraries in SELEX, provides multiple copies of the 17 million unique sequences in the starting pool. The combination of enforced secondary structure and abundance of sequences in the starting aptamer library yield an opportunity to undertake single-round selections.

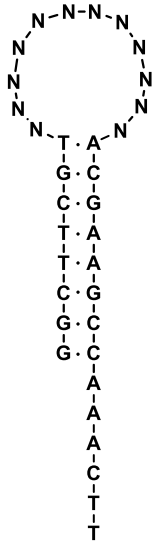


Figure 4-2: SHApE (Structured Hairpin Aptamer Elution) library design.

4.2.1. Materials and methods

4.2.1.1. General information

Unless otherwise noted, water was purified with the Direct-Q® 3 UV Water Purification System. All primers and templates were purchased from Integrated DNA Technologies. The RNA library was purchased from Horizon Discovery. All sequencing preparations were carried out as mentioned in section 3.2.4. for Ion torrent NGS sequencing. Buffer recipes are included below.

Annealing buffer Ph 8

(50 mM Tris-HCl, 5 mM EDTA, 100 mM $MgCl_2$)

Selection buffer pH 7 (SELEX buffer)

(20 mM Tris-HCl, 140 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$)

Elution buffer pH 8

(40 mM Tris-HCl, 10 mM EDTA, 3.5 M urea and 0.02% TWEEN20)

4.2.1.2. Sequences and primers

STEMLOOP-DNA 5′P-GGC TTC GT NNN NNN NNN NNN AC GAA GCC AAA CTT-3′

SL-DNA-PR1 5′-GCC AGT CAA GCT GCT GGC TGC AAA GTC TCC AAG TTT-3′

SL-DNA-PR2 5′P-GGA GAC TTT GCA CGT CGC AGG CGC TGA AGA-3′

STEMLOOP-RNA

5′- /5Phos/GGC UUC GU NNN NNN NNN NNN AC GAA GCC AAA CUU-3′

SL-DNApr1-RNAlig

5′- /5Phos/GCC AGT CAA GCT GCT GGC TGC AAA GTC TCC AAG TrUrU-3′

SL-DNApr2-RNAlig

5′- /5Phos/rGrGA GAC TTT GCA CGT CGC AGG CGC TGA AGA-3′

CNT-STEMLOOP-DNA

5′- /5Phos/GGC TTC GT AAA AAA AAA AAA AC GAA GCC AAA CTT-3′

CNT-STEMLOOP-RNA

5′- /5Phos/GGC UUC GU AAA AAA AAA AAA AC GAA GCC AAA CUU-3′

SHORT-SL

5′- /5Phos/CTCTGAATCT AAG TTT GGC TTC GT AC GAA GCC AGGTTAATGC-3′

SHORT-SL-1A

5′- /5Phos/CTCTGAATCT AAG TTT GGC TTC GT A AC GAA GCC AGGTTAATGC-3′

SHORT-SL-2A

5′- /5Phos/CTCTGAATCT AAG TTT GGC TTC GT AA AC GAA GCC AGGTTAATGC-3′

SHORT-SL-3A

5′- /5Phos/CTCTGAATCT AAG TTT GGC TTC GT AAA AC GAA GCC AGGTTAATGC-3′

SHORT-SL-4A

5′- /5Phos/CTCTGAATCTAAGTTTGGCTTCGTAAAAACGAAGCCAGGTTAATGC-3′

PR1-SPLINT 5′- /5Phos/GCG TTG CCA GAC CGT AGC GCATTAACCT-3′

PR2-SPLINT 5′- /5Phos/AGATTCAGAG TCC GAC TAG CGA GCC GTG-3′

PR2-SPLINT-ddC 5′phos/AGATTCAGAG TCC GAC TAG CGA GCC GTG ddC-3′

PR1-RNA-SPLINT 5′-GCG TTG CCA GAC CGT AGC GCATTAArCrCrU-3′

4.2.1.3. Protein immobilization

Transfer 25 µL beads in a new 1.5 mL DNA low binding tube. Wash beads with 200 µL 1× SELEX buffer, then pull the beads down on a magnetic rack and carefully remove the supernatant without touching the beads. Repeat this step twice. Resuspend beads with 125 µL SELEX Buffer (SB). Transfer 25 µL of suspended beads to a new 1.5 mL DNA low binding tube labeled as **Blank Beads**. Pull the ~100 µL remaining beads down on a magnetic rack and carefully remove the supernatant without touching the beads. Make the solution by diluting 6 µg of protein in 120 µL SB. Transfer 20 µL of the protein solution to a new 1.5 mL DNA low binding tube labeled as **Input**.

Note: *The quantity of protein used in this protocol is based on a protein weighing 30,000 Daltons. You will need to adjust the amount of protein used based on its mass; thus in this step you would use 12* µg *of a 60'000 Da protein.*

Resuspend the beads from step 4 with the remaining protein solution (~100 µL) from step 5, and incubate at 25°C for 30 minutes, on a tube rotator. Centrifuge the protein-beads complex for 5 seconds with E-Centrifuge. Pull the beads down on a magnetic rack and carefully transfer the supernatant to a new 1.5 mL DNA low binding tube labeled as **Flow-through**. Wash beads with 100 µL SB, pull the beads down on a magnetic rack and carefully transfer the supernatant to a new

1.5 mL DNA low binding tube labeled as **Wash**. Resuspend the beads from step 8 with 100 µL SB buffer. Transfer 50 µL of the suspension to a new 1.5 mL DNA low binding tube labeled as Beads. Transfer the remaining 50 µL suspension to a dry bath. Incubate at 95 ℃ for 10 minutes. Centrifuge the heated sample for 5 seconds on a mini centrifuge. Pull the beads down on a magnetic rack and transfer the supernatant to a new 1.5 mL tube labeled as Heated Elution. Resuspend beads with 50 µL SB and label as Heated Beads.

Table 4-1: Samples were prepared for denaturing gel electrophoresis to confirm protein immobilization.

Marker	Input	Flowthrough	Wash	Beads	Heated Elution	Heated Beads	Blank Beads
5 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

SDS PAGE gel. Separating 15% (8 mL total): Combine 3 mL 40% acrylamide, 2 mL 1.5M Tris pH 8.8, 80 µL 10% SDS, 80 µL10% APS and bring the volume to 8 mL with water. Then add 8 µL TEMED to polymerize. Add isopropanol, let it polymerize for 35 min.

Stacking 6% (5 mL total): Combine 0.75 mL 40% acrylamide, 1.25 mL 0.5M Tris pH 6.8, 50 µL 10% SDS, 50 µL 10% APS and bring the total volume to 5 mL with water. Then add 5 µL TEMED to polymerize. Let it cast for 35 min.

Gel imaging: Run the gel for 80 min (use 5× SDS loading dye) and use a protein ladder. Stain with Coomassie blue on a shaker (gently) 15 min at room temperature. Wash with de-stain solution (wait 15 min and pour out), final rise with just water (purple color). Image with white background on gel imager (protein gel setting).

4.2.1.4. SHApE selection (DNA)

Dilute 10 µL SL-DNA-library (10 µM) with 30 µL water (total volume 40 µL) and incubate at 95 ℃ for 5 min. Add 10 µL of 5× SELEX buffer to 40 µL SL-DNA-library and incubate at 37 ℃ for 15 min. Prewash 6 µL (in round1)/ 5 µL (in round2)/ 4 µL (in round3) of Thrombin-bound agarose

beads (1 μg/μL) with 50 μL 1× SELEX buffer. If using larger radius beads, such as Sepharose or agarose beads, ensure that a wide bore pipette tip is used. Normal pipette tips can be made wide-bore by cutting off the tip of the pipette using a clean razor. Add 50 μL aptamer library to prewashed Thrombin-bound agarose beads (dry) and incubate for 30 min on rotator at room temperature. Spin down and remove supernatant (keep for flow through). Wash with 50 μL 1× SELEX buffer three times (if doing 3 rounds), or 6 times if doing 1 round (keep for wash).

CRITICAL STEP: Centrifugation during wash should be kept at 376 RCF for 1 min. Low speeds results in poor separation. High speeds results in aggregation of agarose beads.

Add 50 μL elution buffer (40 mM Tris-HCl pH 8, 10 mM EDTA, 3.5 M urea and 0.02 % TWEEN20) and incubate at 90 °C for 4 minutes. Remove entire supernatant, which contains eluted binders. If some beads are pipetted up with the supernatant, this is not an issue since they will be removed in by the Princeton Separation column. Purify elution with Princeton Separation columns. These are Selection winners. Bring sample to 10 μL and save ~30 pmol each round (3 μL), then repeat steps 1-14 with new beads for the eluted library (left over 7 μL) for up to 3 rounds of DEL-style (no pcr) selection.

CRITICAL STEP: Ensure complete purification. Urea is a strong denaturant and can prevent amplification during PCR.

4.2.1.5. SHApE selection (RNA)

Prewash 6 μL (in round1)/ 5 μL (in round2)/ 4 μL (in round3) of bead-bound protein (1 μg /μL) with 50 μL 1× SELEX buffer, three times. Dilute 10 μL SL-RNA-library (10 μM) with 30 μL water (total volume 40 μL) and incubate at 90 ℃ for 2 min to denature sequences. Then add 10 μL of 5× SELEX buffer to 40 μL SL-RNA-library and incubate at 37 ℃ for 15 min to anneal stems. Add 50 μL aptamer library (in buffer) to prewashed bead-bound protein (dry) and incubate

for 30 min on rotator at room temperature. Spin down and remove supernatant (keep for flow through). Wash with 50 µL 1× SELEX buffer three times (if doing 3 rounds), or 6 times if doing 1 round (keep for wash). Add 50 µL elution buffer (40 mM Tris-HCl pH 8, 10 mM EDTA, 3.5 M urea and 0.02% TWEEN20) and incubate at 90 °C for 2 minutes. Remove entire supernatant, which contains eluted binders. If some beads are pipetted up with the supernatant, this is not an issue since they will be removed in by the Princeton Separation column.

CRITICAL STEP: Do not vortex or centrifuge. Pipette to mix and use a magnetic rack to discard elution from washes.

Purify elution with Princeton Separation columns. These are Selection winners. Save 30% each round, then repeat steps 1-14 with new beads for the eluted library (LEFT OVER 70%) for up to 3 rounds of DEL-style (no PCR) selection.

CRITICAL STEP: Ensure complete purification. Urea is a strong denaturant and can prevent amplification during PCR.

4.2.1.6. Splint Ligation of primer binding sites (DNA)

Add the following in a microcentrifuge tube: do this for the starting library and the outputs from R1 (round1), R2 (round2), R3 (round3), and SR (single round). For DNA outputs Combine 1 µL SL-DNA-lib (10 µM), 3 µL 4A-splint (10 µM), 2.5 µL PR1-SPLINT (10 µM), 2.5 µL PR2-SPLINT-ddC (10 µM) and incubate at 95 °C for 5 min. Add 1 µL of 10× annealing buffer [50 mM Tris-HCl (pH 8), 5 mM EDTA, 100 mM MgCl2] and incubate at 37°C for 15 min. Then add 4 µL ligation buffer 5X, 2.5 µL BSA (2mg/mL), 2.5 µL nuclease free water, 1 µL T4 DNA ligase and incubate at 16°C for 16 hr, followed by cleanup with OMEGA kit.

4.2.1.7. Splint Ligation of primer binding sites (RNA)

Add the following in a microcentrifuge tube: do this for the starting library and the outputs from R1 (round1), R2 (round2), R3 (round3), and SR (single round). For RNA outputs Combine 1 µL SL-RNA-lib (10 µM), 3 µL 4A-splint (10 µM), 2.5 µL PR1-RNA-SPLINT (10 µM), 2.5 µL PR2-SPLINT-ddC (10 µM) and incubate at 90 ℃ for 2 min. Add 1 µL of 10× annealing buffer [50 mM Tris-HCl (pH 8), 5 mM EDTA, 100 m M $MgCl_2$] and incubate at 37 °C for 15 min. Then add 1 µL ligation buffer 10X, 8 µL nuclease free water, 1 µL T4 RNA ligase II and incubate at 37 °C for 1 h and 4 °C for 15 h, followed by cleanup with Monarch kit.

4.2.1.8. Sequencing preparation (RNA): RT-PCR

Combine 10 µL of sample (0.01 µM) using 1 µL of each primer and 25 µL 2× reaction mix, and 2 µL SuperScript Taqmix to amplify sequences for 10 cycles. Purify with OMEGA Kit.

4.2.1.9. BioLayer Interferometry (BLI) assay

The streptavidin biosensor tips are hydrated in 1× selection buffer for at least 10 min. Samples are diluted in 1× selection buffer accordingly, and biotinylated aptamers are heated to 95 ℃ for 5 min prior to setup. Sample steps include a baseline for 120 s, followed by loading for 300 s (200-250 nM bt-DNA), a second baseline for 60 s, then association for 600 s (low nM to low µM Thrombin, 2-fold DS) and dissociation for 600 s. The shake speed and temperature were set to 1,000 RPM (on the Sartorius 4-Channel System: Octet® R4 BLI) and 30 °C.

4.2.2. Results and discussion

The SHApE selections were carried out using a DNA hairpin library against Thrombin, and an RNA hairpin library against YTHDF1, YTHDF2, and YTHDF3 proteins. The method involved the elution of binders from a pool of non-binders without the requirement for amplification, where many copies of each unique sequence in the starting library provide sufficient binding

opportunities. The starting pools of SHApE sequences were individually incubated against the appropriate bead-immobilized target in pairs, where single and triple incubations with the target were attempted as shown in Figure 4-3.

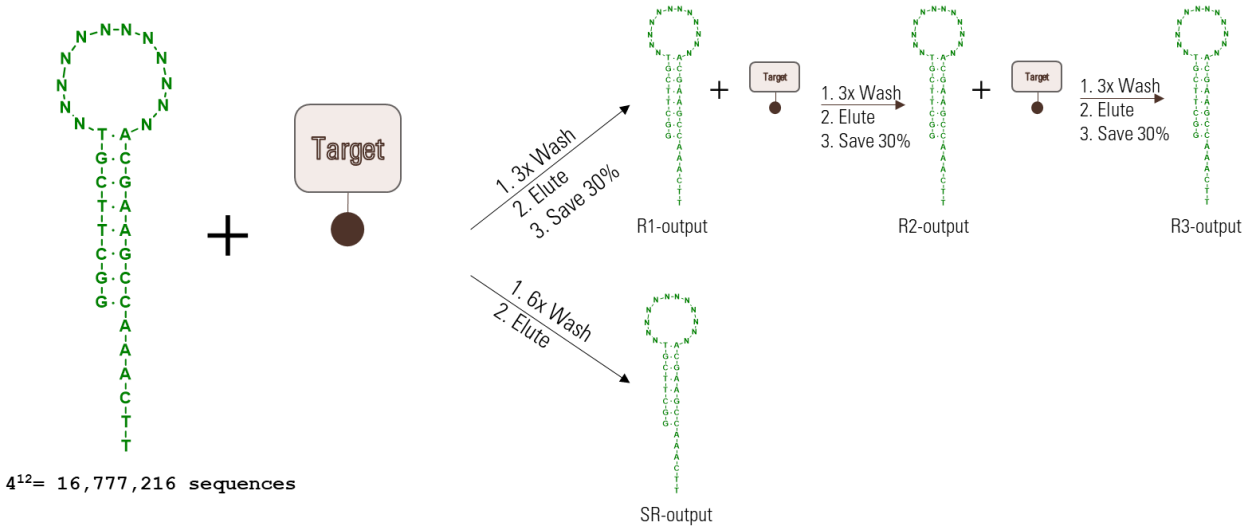


Figure 4-3: Illustration of single round selections *via* SHApE libraries against Thrombin.

The library sequence design incorporated a 6-mer overhang, initially designed to enable the ligation of primer-binding sequences for NGS preparations post-selection (Figure 4-4). However, sequencing results demonstrated a challenge with long hairpin stems in Ion Torrent sequencing. An overwhelming number of truncated reads corresponded to the Y-adaptor primer overhangs for every library. To eliminate sequencing failure, a new method for primer ligation without lengthening the 8-mer SHApE library stem was critical. This method utilized a hybridizing internal hairpin acting as a splint, providing overhangs to facilitate the ligation of primer binding sites (Figure 4-5).

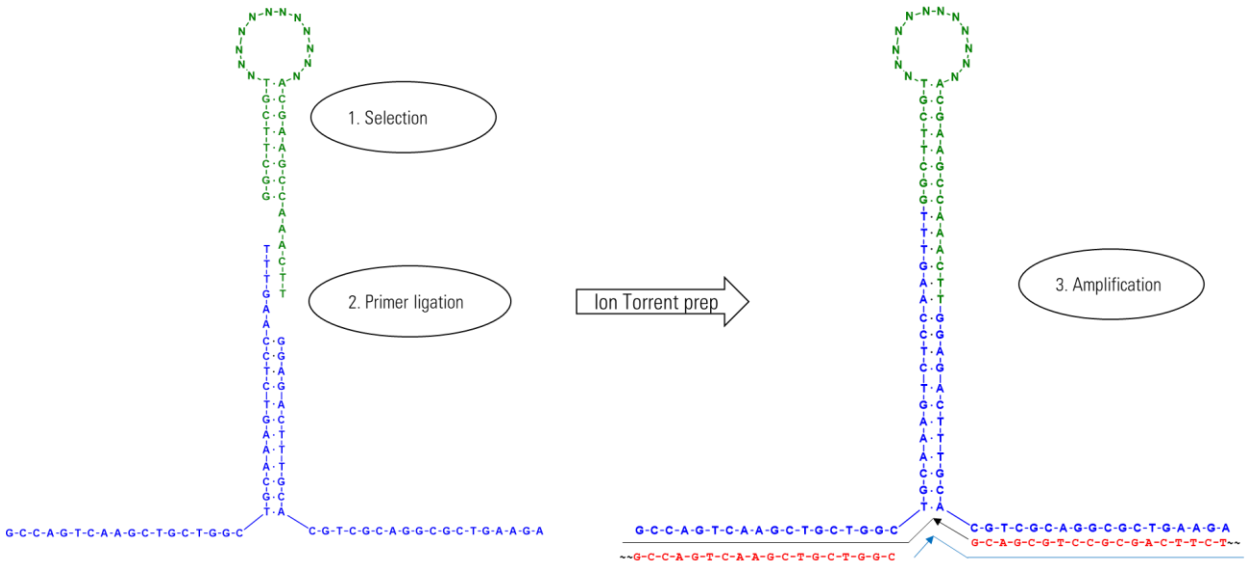


Figure 4-4: Illustration of the initial SHApE primer ligation design for sequencing preparations.

Splint ligation was optimized using sequences with varying hairpin bulge lengths of Adenosine bases flanked by the stem sequence. Splint sequences containing no Adenosine, 2A, and 4A long loop regions were tested and the best ligation results were observed with the 4A-containing splint sequence; where the splint ligation resulted 61 % yield without a loop region, 69 % yield with the 2A-containing splint, and a 76 % yield with the 4-A containing splint. To eliminate unwanted ligations, the splint sequence and primer-1 were not phosphorylated and primer-2 contained a 3′-dideoxy cytosine base. Finally, the sequences were amplified using Q5 polymerase[143] (low bias and high fidelity) and the appropriate Ion Torrent barcoded primers for sequencing.

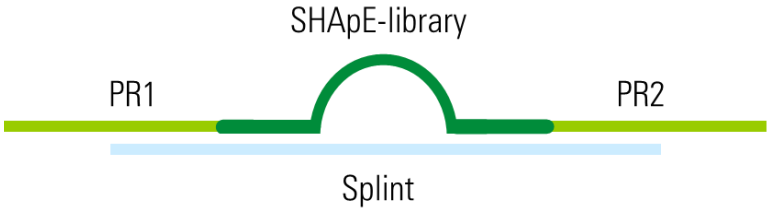


Figure 4-5: SHApE library sequencing preparation *via* splint-ligation; the splint and primer-1 are unphosphorylated, and primer-2 contains a ddC at the 3′-end.

Upon sequencing the SHApE output libraries, the data was analyzed extensively to distinguish potential binders. Considering the SHApE library selections were not an evolution but rather an elution of potential aptamers, the outlook for sequencing data analysis was to evaluate similarities in sequences *via* Levenshtein distances to group families of aptamers.[127] The Thrombin selection showed the most promising results and was the focus of preliminary analysis. The output sequences were generally T-rich, and raised attention for potential PCR bias; however, sequencing data from the starting library carried out in parallel eliminated this concern and demonstrated a variety of sequences without bias.

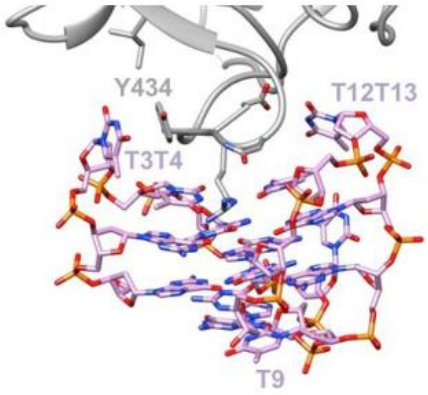


Figure 4-6: Crystal structure of the α-Thrombin (grey) complex with TBA; only a portion of the protein is shown. Figure adapted from: Nucleic acids research.[144]

Interestingly, the Thrombin binding aptamer, TBA, is a G-quadruplex with thymine bases displayed around the corners of the structure, and these thymine bases are known to be responsible for the binding interaction with the target (Figure 4-6).[145] This trend of T-rich sequences was common in both selections against Thrombin. To choose sequences for further analysis, due to a lack of enrichment of individual sequences, the overlapping sequences between the two selections against Thrombin were considered. Ultimately the ten sequences in Table 4-2 were selected and ordered with a biotin tag for binding kinetic analysis on BLI[111] (BioLayer Interferometry). These

sequences represented common motifs and were members of clusters commonly present in the output library. However, preliminary BLI analysis against Thrombin at low micromolar concentrations showed no binding to the target. The aptamers may be able to bind Thrombin with much higher micromolar affinities. This potentially low-affinity binding was not justifiable for the use of additional resources, and the selections were ultimately unsuccessful.

Table 4-2: SHApE output sequences from selection against Thrombin.

Name	Sequence
SHApE-C4r195-R1Thr	/5Biosg/-GGC TTC GT **TTTTTACTTTTC** AC GAA GCC AAA CTT-3′
SHApE-C15r185-R1Thr	/5Biosg/-GGC TTC GT **TTTGTTTGTTTT** AC GAA GCC AAA CTT-3′
SHApE-C4r10-R1Thr	/5Biosg/-GGC TTC GT **TTTTTATTTTTT** AC GAA GCC AAA CTT-3′
SHApE-C15r178-R1Thr	/5Biosg/-GGC TTC GT **TTTATTTCTTTT** AC GAA GCC AAA CTT-3′
SHApE-C4r249-R1Thr	/5Biosg/-GGC TTC GT **TCTTTCTTTCTT** AC GAA GCC AAA CTT-3′
SHApE-C59r124-R1Thr	/5Biosg/-GGC TTC GT **TCTTCTTTCTTC** AC GAA GCC AAA CTT-3′
SHApE-C15r1-R1Thr	/5Biosg/-GGC TTC GT **TTTGTTTCTTTT** AC GAA GCC AAA CTT-3′
SHApE-C218r1-R1Thr	/5Biosg/-GGC TTC GT **CTTTCATTGTTT** AC GAA GCC AAA CTT-3′
SHApE-C19r3-R1Thr	/5Biosg/-GGC TTC GT **TTTTCGTTTTTC** AC GAA GCC AAA CTT-3′
SHApE-C218r78-R1Thr	/5Biosg/-GGC TTC GT **CTTTGTTTGTTG** AC GAA GCC AAA CTT-3′

4.3. G-quadruplex aptamer libraries and single-round selections

Naturally occurring structures such as the G-quadruplex have demonstrated biological significance upon genome mapping.[146,147] DNA sequences abundant with guanine bases augment stability *via* the four-stranded inherent G-quadruplex structure (G4s) in single-stranded regions. Upon stacking guanine tetrads, facilitated by dispersion forces of π-π interactions in the presence of monovalent cations, form non-canonical Hoogsteen hydrogen bonds amongst four guanine residues enabling folding (Figure 4-7). Early sequencing studies reveal the consensus sequence $G_{3–5}\,N_{1–7}\,G_{3–5}\,N_{1–7}\,G_{3–5}\,N_{1–7}\,G_{3–5}$ enabling the formation of parallel, antiparallel, and hybrid G4s.[123] A variety of G4 topologies emerge from different loop sequences and strand polarities. This thermodynamically

stable folded structure is prevalent in regions of the genome that directly affect regulatory interactions with proteins. Remarkable protein interactions of high avidity, with K_D as low as picomolar, are recognized involving active promoters, enhancers, and telomeres. Selective protein engagement of G4 structured oligonucleotide sequences plays critical roles in cellular regulatory functions; the stability of G4s can have implications for diseases such as cancer and neurodegeneration. G4 landscapes are acquired epigenetic features in cells, selectively formed in unmethylated active alleles to enhance transcription and maintain cell identity.[146]

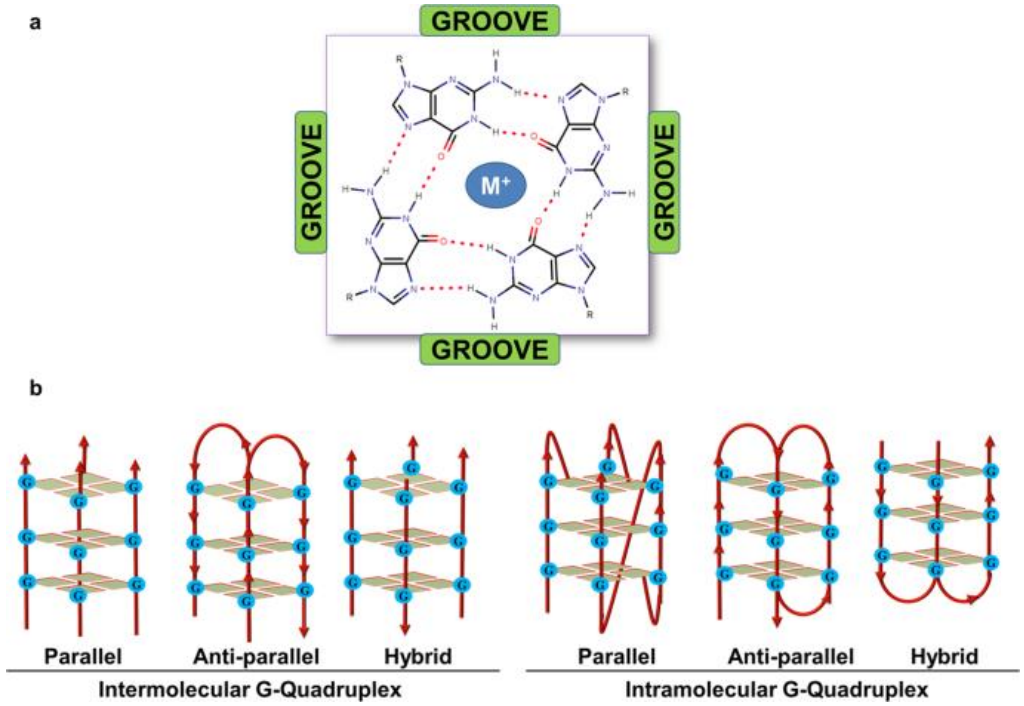


Figure 4-7: G-quadruplex structures; a) an overtop view of the G4 H-bonding, b) various topologies of G4 structures. Figure from: Nature scientific reports.[123]

The diverse array of biological interactions facilitated by the variety of G4 oligonucleotide structures is herein utilized to select high-quality aptamers against human α-Thrombin. The G-quadruplex structure as previously mentioned enhances aptamer stability, but also avoids immunogenicity and shows resistance to numerous serum nucleases in some cases.[148–150] The first and most thoroughly studied Thrombin-binding aptamer, TBA (Figure 4-8a), was discovered in

1992 using SELEX.[151] The 15-mer sequence, 5′-GGTTGGTGTGGTTGG-3′, folds *via* two G-tetrads stacking into an antiparallel G4 chair-like conformation. TBA demonstrates anticoagulation activity upon exosite-I binding; epitope binning is facilitated by one of the TT loops enabling a pincer-complex with the ligand.[108] Another well-studied Thrombin binding aptamer, HD22 (Figure 4-8b), also folds into a G4 structure mixed with duplex DNA. The 27-mer aptamer impedes clotting efficiency by obstructing access to factor V and factor VIII activation and non-catalytic fibrin polymerization *via* exosite-II binding.[152] Provided the known binding ability of Thrombin to G4 DNA sequences, such as TBA and HD22, an innovative G4 aptamer library design is instrumental in developing single-round selections against Thrombin. The aptamer library shape is procured for the target of choice.

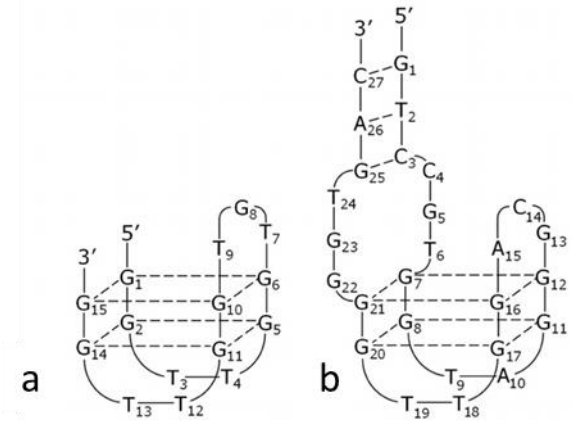


Figure 4-8: Schematic representation of the G-quadruplex forming Thrombin binding aptamers, (a) structure of TBA-15mer, and (b) structure of HD22–27mer. Figure adapted form: Nucleic acids research.[145]

Although previously unsuccessful with SHApE libraries, the same method for single-round selection against Thrombin was attempted with a modified structured anti-parallel G-quadruplex DNA library. Single strand aptamers forming unimolecular G-quadruplexes often contain at least four G-tracts, where the number of Guanines in each tract dictates stability.[147,153] Although both TBA and HD22 represent G4 structures forming from GG repeats, the stability of the structure can

be improved with new G4 aptamer libraries. The library design for this selection contains four G-tracts (triple Guanine repeats) with 3mer loop regions in between, along a 24-mer G-quadruplex sequence flanked by 18mer primer binding sites, 5′-GCGTTGCCAGACCGTAGC **NN** GGG **NNN** GGG **NNN** GGG **NNN** GGG **N** CCGACTAGCGAGCCGTG-3′. The increased structural stability due to increased stacking provides a more rigid structure, which may favour binding. Another influential factor for G4 stability is the loop region of ssDNA in between the tandem G repeats. Every loop generally contains 1-7 nucleotides, where in DNA G4 structures the longer loop region design provides additional stability. Meanwhile, G4 aptamers such as TBA and HD22 contain short loop regions of 1-2 nucleotides.[147] The trinucleotide loop region in this new G4 library also provides greater opportunity for binding interactions with the target protein. The aptamer library also incorporates modified bases along the diversity regions to promote intermolecular interactions along target binding sites.

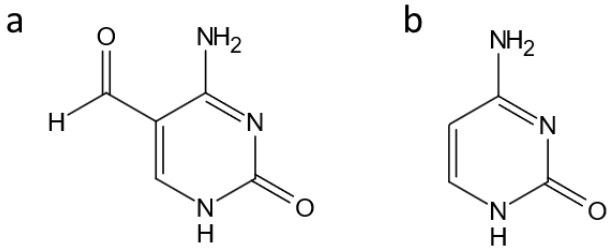


Figure 4-9: Schematic of 5-Formylcytosine (a) compared to Cytosine (b).

Thrombin is a DNA-binding protein with positively charged amino acid residues in its active sites, namely Lysine and Arginine.[106] The modification incorporated in this library was 5-Formylcytosine (5fC) shown in Figure 4-9, where the aldehyde modification enables reversible cross-conjugation to Lysine and Arginine residues on protein targets *via* a primary amine to form a Schiff base intermediate (imine), Figure 4-10.[154] The covalent bond between the amino acid residue and 5fC *via* imine formation is reversible by hydrolysis. The regions of sequence diversity

within the G-quadruplex loops are selected for binding, where NN represents any of A, T, G, or 5fC (5-Formylcytosine) at random.

Figure 4-10: Schematic of 5-Formylcytosine reversible cross-linkage with Lysine *via* aldehyde reactivity with amine sidechain forming a Schiff base (imine).

4.3.1. Materials and methods

4.3.1.1. General information

Unless otherwise noted, water was purified with the Direct-Q® 3 UV Water Purification System. 5fC modified G-quadruplex library was synthesized on the Applied Biosystems DNA/RNA synthesizer model 394. All materials and reagents used for oligo-synthesis were purchased from Glen Research™. All unmodified libraries and primers were purchased from Integrated DNA Technologies IDT. Functional G4 library products were purified by reverse-phase high-performance liquid chromatography (HPLC, 1260 Infinity II LC System) using a C18 stationary phase (Eclipse-XDB C18, 5 µm, 4.6 × 150 mm) and an acetonitrile/ 0.1 M triethylammonium acetate (TEAA buffer) gradient. Purified products were lyophilized using a FreeZone 2.5 L -84 °C Benchtop Freeze Dryer. Coupled product concentrations were quantified from A_{260} values using a Nanodrop One UV-Vis Spectrophotometer. Mass spectrometry of coupled products was carried out using the Velos Pro Mass spectrometer (Orbitrap Elite MS). The mass accuracy of Velos Pro Orbitrap ranges from 0-6.62 ppm. The 5fC-modified G4 library was synthesized by a previous undergraduate student, Kristina Gremi, alongside Natalie Khamissi, a PhD. Fellow. All sequencing preparations were carried out as mentioned in section 3.2.4. for Ion torrent NGS sequencing. Buffer recipes are included below.

SELEX-K buffer pH 7

(20 mM Tris-HCl, 150 mM NaCl, 4 mM KCl, 1 mM $MgCl_2$, and 1 mM $CaCl_2$)

Elution buffer pH 8

(40 mM Tris-HCl, 10 mM EDTA, 3.5 M urea and 0.02% TWEEN20)

4.3.1.2. Sequences and primers

5fC G-quad library (where N* = A/G/T/5fC)

GCGTTGCCAGACCGTAGC**N*N*GGGN*N*N*GGGN*N*N*GGGN*N*N*GGGN*** CCGACTAGCGAGCCGTG

G-quad library

GCGTTGCCAGACCGTAGC**NN**GGG**NNN**GGG**NNN**GGG**NNN**GGG**N**CCGACTAGCGAGC CGTG

4.3.1.3. Single-round selection against Thrombin using 5fC G-quadruplex library

Two single-round selections are carried out in parallel, *via* a single-incubation and a triple-incubation for selection against the target. During the triple-incubation selection, save 30 % of every output post-target-incubation for sequencing. For the triple-incubation selection repeat the following protocol using the eluted output and fresh protein target.

Dilute 10 µL 5fC G-quad library (10 µM) with 30 µL water (total volume 40 µL) and 10 µL of 5× SELEX-K buffer (1× contains 20 mM Tris-HCl at pH 7, 150 mM NaCl, 4 mM KCl, 1 mM $MgCl_2$, and 1 mM $CaCl_2$) to 40 µL SL-DNA-library and incubate at 95 ℃ for 5 min then room temperature for 30 minutes. Prewash 6 µL of Thrombin-bound agarose beads (1 µg/µL) with 50 µL 1× SELEX-K buffer three times. If using larger radius beads, such as Sepharose or agarose beads, ensure that a wide bore pipette tip is used to measure out the 6 µL. Normal pipette tips can

be made wide-bore by cutting off the tip of the pipette using a clean razor. Add the 50 µL aptamer library to prewashed Thrombin-bound agarose beads (dry) and incubate for 30 min on rotator at room temperature. Spin down and remove supernatant (keep the flow through in a labeled PCR tube). Wash with 50 µL 1× SELEX-K buffer three times (if doing 3 rounds), or 6 times if doing 1 round (keep every wash).

CRITICAL STEP: Centrifugation during wash should be kept at 376 RCF for 1 min. Low speeds results in poor separation. High speeds results in aggregation of agarose beads.

Add 50 µL elution buffer (40 mM Tris-HCl pH 8, 10 mM EDTA, 3.5 M urea, and 0.02% TWEEN20) and incubate at 90 °C for 4 minutes. Remove the entire supernatant, which contains eluted binders. If some beads are pipetted up with the supernatant, this is not an issue since they will be removed in the desalting step. Purify elution with Princeton Separation columns. These are selection winners.

CRITICAL STEP: Ensure complete purification. Urea is a strong denaturant and can prevent amplification during PCR.

4.3.1.4. BioLayer Interferometry (BLI) assay

The streptavidin biosensor tips are hydrated in 1× selection buffer for at least 10 min. Samples are diluted in 1× SELEX-K buffer accordingly, and biotinylated aptamers are heated to 95 ℃ for 5 min prior to setup. Sample steps include a baseline for 60 s, followed by loading for 300 s (200-250 nM bt-DNA), a second baseline for 120 s, then association for 600 s (low nM to low µM Thrombin, 4-fold DS) and dissociation for 600 s. The shake speed and temperature were set to 1,000 RPM (on the Sartorius 4-Channel System: Octet® R4 BLI) and 30 °C.

4.3.2. Results and discussion

The G-quadruplex library contained 5fC modified members, along with unmodified members, to enhance target binding. Two single-round (no amplification between rounds) selections were carried out in parallel, one with a single target-incubation step where the corresponding output library was designated SR and another with a three target-incubation steps where the output libraries were designated R1, R2, and R3 accordingly. The differentiating factor between these two selections were the number of target incubations and the number of washes post selection (prior to output elution), where the single incubation method followed six washes post selection while the triple incubation method was followed by three washes after each target incubation as illustrated in Figure 4-11.

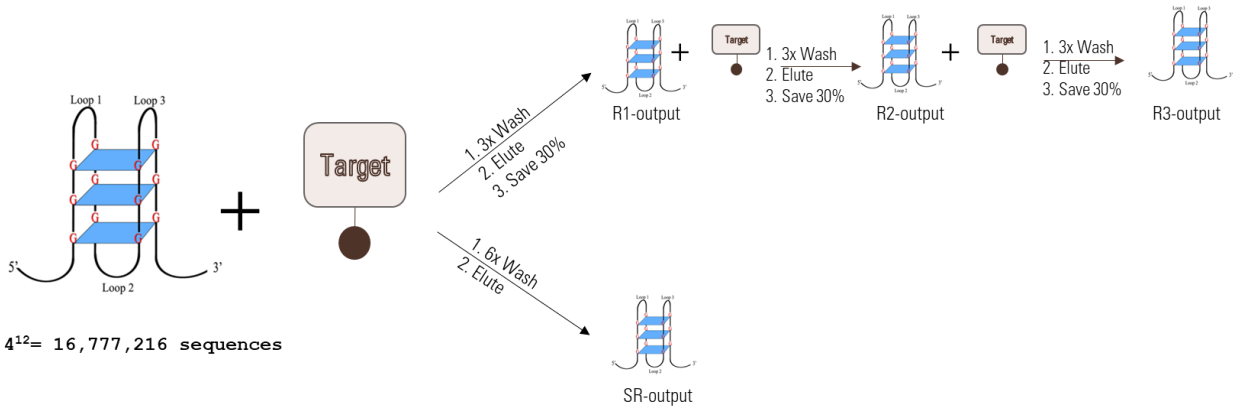


Figure 4-11: Illustration of single round selections *via* 5fC-modified antiparallel G-quad DNA library against Thrombin.

After selections against Thrombin, the sequencing preparation required PCR amplification of the output libraries. Although previously tested, 5fC-containing sequences had shown amplification success *via* Q5 polymerase, in the case of a library containing both modified and unmodified sequences the polymerase revealed preference against 5fC-containing sequences. The G-quadruplex library sequencing data of both starting and post-selection libraries demonstrated an

absence of C-containing sequences, where every C in the randomized region corresponded to 5fC. Regardless of the PCR-eliminated modified sequences, the sequencing data showed promising results of enriched sequences in the post-selection libraries. Sequencing data was analyzed first by removing short reads using the Java Trimmomatic[155] script, followed by removing the primer binding sites using the Python Cutadapt[156] script, and finally counted and clustered through the FASTAptamer[126,127] scripts in Perl. The winning sequences were consistently enriched over three rounds of target incubation. Five unique sequences (Table 4-3) were selected for further kinetic analysis based on the sequencing data, the evaluated enrichments and clustered family representatives.

Table 4-3: Selected winning G-quadruplex library output sequences and enrichments represented relative to the starting library.

Name	Sequence	Enrichment (SR/LIB)	Enrichment (R1/LIB)	Enrichment (R2/LIB)
MSR-1-1-0	TT GGG TGT GGG GTT GGG TTG GGG T	46.0	20.1	471
MSR-2-1-0	TA GGG AAA GGG TGG GGG GTT GGG T	35.4	7.03	445
MSR-4-1-0	AT GGG TAT GGG GTT GGG TTG GGG T	30.1	10.1	74.2
MSR-2-3-1	TA GGG AAT GGG TGG GGG GTT GGG T	28.3	9.04	469
MSR-9-1-0	TT GGG TTT GGG GTT GGG TGG GGG T	19.5	1.01	81.8

Enrichment was evaluated as a ratio based on frequency of reads per million (RPM) in the output libraries relative to the starting library. Comparing the two parallel selections against Thrombin, the output libraries after one round of target incubation demonstrated a correlation between enrichment and the number of washes post-selection, where SR enrichments of the same sequences were double or more than the R1 resulting enrichments. The most enriched sequence was consistent in both SR and R2 outputs; MSR-1-1-0 was enriched about 46-fold in the SR output and 471-fold in the R2 output relative to the starting library after selection against Thrombin. The only highly enriched sequence present in R3 sequencing data was MSR-2-3-1, enriched about 7-fold from the starting library to R1 and about 63-fold from R1 to R2; the sequence was enriched

469-fold by R2 relative to the starting library. Following enrichment analysis, the post-selection output libraries of SR and R2 were clustered based on a Levenshtein distance of 1 (maximum number of sing-character edits) to group closely related families of sequences. The family 1 and 2 clustered sequences overlapped between SR and R2 (Table 4-4). The G-quadruplex post-selection library sequences contain conserved loop regions in each cluster, these loop sequences can provide structural stability or selective target-binding avidity.

Table 4-4: Top 10 highly enriched sequences from SR and R2 and the corresponding cluster families. Family 1 sequences are highlighted in blue, and family 2 sequences are highlighted in green, the overlapping sequences are the highlighted in the same shade. Note, the sequencing data are clustered for each output library individually and cluster numbers of families do not overlap perfectly.

	Sequence	RPM	Enrichment	Cluster
SR	TTGGGTGTGGGGTTGGGTTGGGGT	68.1	46.0	1
	ATGGGTGTGGGGTTGGGTTGGGGT	56.3	38.0	1
	TAGGGAAAGGGTGGGGGGTTGGGT	52.4	35.4	2
	ATGGGTATGGGGTTGGGTTGGGGT	44.5	30.1	4
	TAGGGATAGGGTGGGGGGTTGGGT	44.5	30.1	2
	ATGGGTTTGGGGTTGGGTTGGGGT	41.9	28.3	4
	TAGGGAATGGGTGGGGGGTTGGGT	41.9	28.3	2
	TGGGGTTTGGGGTTGGGTTGGGGT	39.3	26.5	5
	GTGGGTGTGGGGTTGGGTTGGGGT	39.3	26.5	4
	TTGGGTTAGGGGTTGGGTTGGGGT	30.1	20.3	6
R2	TTGGGTGTGGGGTTGGGTTGGGGT	696	471	1
	TAGGGAATGGGTGGGGGGTTGGGT	695	469	2
	TAGGGAAAGGGTGGGGGGTTGGGT	659	445	2
	ATGGGTGTGGGGTTGGGTTGGGGT	530	358	1
	GTGGGTGTGGGGTTGGGTTGGGGT	360	243	1
	TTGGGTGTGGGGTTGGGTTTGGGG	320	216	4
	TAGGGATAGGGTGGGGGGTTGGGT	303	205	5
	TGGGGTTTGGGGTTGGGTTGGGGT	282	191	6
	TTGGGTTTGGGGTTGGGTTGGGGT	259	175	1
	TAGGGATAGGGTGGGGGGTTGGGA	244	165	5

Family 1 sequences from both SR and R2 outputs contain the same frequently conserved regions along their sequence, corresponding to MSR-1-1-0 in Table 4-3. These sequences are potentially stabilizing the antiparallel G4 structure *via* wobble base-pairs at the TGT (loop1) and TTG (loop3). Family 2 clustered sequences commonly contained the following conserved regions: 5′-TA sequence, followed by AAN loop1 sequence (where N was often found to be T in R2 outputs and A in SR outputs), TGG loop2 sequence, GTT sequence at loop3. and a 3′-T. The antiparallel loop1 and loop3 in family 2 sequences likely provide stability to the G4 structure *via* hydrogen-bonding between the canonical AA and TT bases. Sequences corresponding to cluster 1 and 2 are abundant in both SR and R2 outputs and are listed among the high-frequency enriched hits.

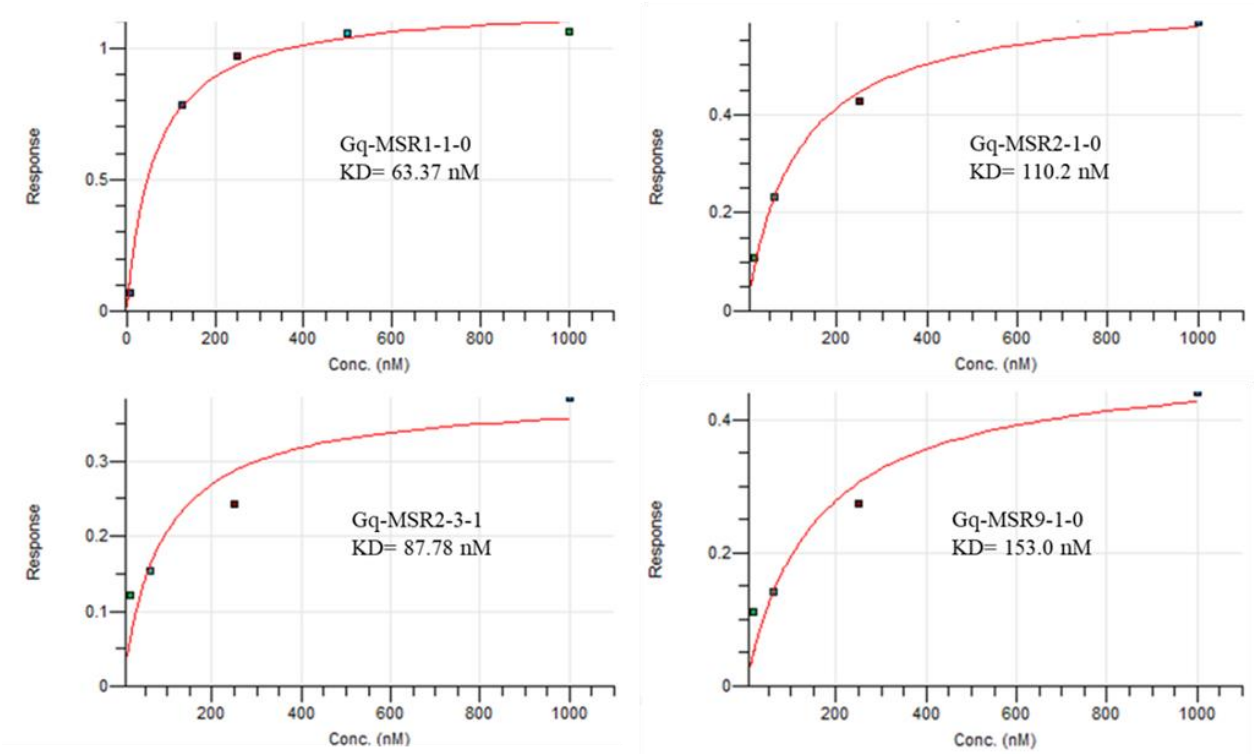


Figure 4-12: Binding kinetics of G-quadruplex library winners based on BLI data; x-axis represents protein concentration, y-axis represents BLI response upon target binding. The processed data demonstrates Gq-MSR1-1-0 with a K_D= 63.37 nM and R^2= 0.99, Gq-MSR2-3-1 with a K_D= 87.78 nM and R^2= 0.84, Gq-MSR2-1-0 with a K_D= 110.2 nM and R^2= 0.99, Gq-MSR9-1-0 with a K_D= 153.0 nM and R^2= 0.92.

The five sequences in Table 4-3 were selected from SR hits as cluster representatives from family 1, 2, 4, and 9. The biotinylated aptamers were purchased for kinetic analysis *via* BioLayer Interferometry analysis against Thrombin to compare families with varying enrichments. The BLI assays were assembled with streptavidin biosensor tips for aptamer immobilization, and kinetic binding analysis at varying protein target concentrations. Initially, single-concentration BLI experiments were set up at 1 μM Thrombin to determine the presence of high-affinity target binding. This binding threshold was set to identify aptamers with acceptable nM to μM dissociation constants. All five sequences demonstrated binding at the 1 μM threshold and qualified for further kinetic analysis. Follow-up BLI experiments were conducted *via* 4-fold dilution series of Thrombin ranging from 15.6 nM to 1 μM.

The top four sequences with the higher RPM frequencies demonstrated tight binding; BLI results showed binding curves at the lowest nM concentrations. The processed BLI data revealed dissociation constants (K_D) ranging from 63.73 nM to 153.0 nM for the winning sequences (Figure 4-12). The dissociation constants correlate with the enrichment of selected sequences in R2 sequencing data. The BLI binding curves for MSR-4-1-0 (lower relative enrichment in R2 compared to other four sequences) demonstrated irregular binding curves, likely due to low-quality epitope binning, and the dissociation constant for this aptamer was not extrapolated. The BLI data conclusively supports the enrichments patterns observed from the multiple incubation selection method, however SR outputs favourably enriched family 1 and 2 sequences similar to the R2 outputs. Although stringent selection parameters such as increased washes enhance enrichment, however multiple rounds of target-incubation promote substantially selective enrichment of relatively tighter binders in the output library.

4.4. Conclusions

Aptamer selection is often an evolution process through multiple rounds of *in vitro* selection, leading to excessive time and materials consumption. To overcome this issue single-round selections are introduced using structured oligonucleotide libraries, where the quality of the starting library increases the ratio of binders to non-binders prior to selection. Beginning with SHApE, the structured hairpin aptamer libraries design with a single loop were unsuccessful for selections against Thrombin, YTHDF1, YTHF2, and YTHDF3. Although the results demonstrate almost no enrichment in the outputs, the commonly T-rich sequences representative of the Thrombin selection outputs revealed the aptitude for separation by this method. Considering the simplicity of the SHApE libraries, the follow-up single-round selection library design incorporates complexity *via* the G-quadruplex structure and added opportunity for protein-binding *via* the 5fC modifications. Although the 5fC modifications were eliminated in sequencing data due to PCR bias against the modification, the success of the single-round G-quadruplex library selection compared to the SHApE libraries relies heavily on the added complexity of structure in the starting library. Provided the consistent overlap of the highest frequency hits in R1 and SR outputs, future selections can benefit from triple-incubation selections with increased number of washes prior to the R1 output elution for increased overall enrichment. Further to this, the sequencing preparation can be modified to only sequencing the starting library and R3 output to avoid losing potential binders; previously 30 % of the outputs from R1 and R2 were saved for sequencing and may have negatively impacted the enrichment results.

CONCLUSION AND OUTLOOK

The present frontrunners in custom oligonucleotide production are disrupting traditional solid-phase synthesis with enzymatic methods, aiming to reduce both the expense and environmental impact of waste generation. These enzymatic methods are incapable of library synthesis due to the challenges with the design and evolution of new mutated versions of the required enzymes to accommodate various modifications. The LOOPER method provides a robust synthetic route combining solid-phase and enzymatic synthesis of modified DNA oligonucleotide libraries for aptamer selections. While the LOOPER method has shown success in synthesizing modified RNA libraries, its effectiveness is hindered by low fidelities, possibly stemming from RNA anticodon preferences during annealing on a DNA template. Future advances to the method will require more-expensive RNA templates and mutated ligases.

The multiple attempts with LOOPER libraries for SELEX and CE evolutions highlight challenges based on the PCR-amplified regeneration of templates for follow-up rounds of selection. LOOPER libraries provide access to a diverse array of modifications in the starting library, however, the abundance of non-binders and the required amount of template sequences for output regeneration throughout iterative rounds of *in vitro* selection can outcompete enrichment from round to round. With respect to the challenges of aptamer selections resulting from intercycle amplifications, the ideal method requires single-round selections. Both CE and bead-based selections show promising results. Although unsuccessful, CE-based selections facilitate the stringent separation of binders from non-binders. The choice of target is critical in the quality of CE separation, enabling an EOF without tight interactions with the silica coating of the capillary and providing a large size difference relative to the aptamer sequences.

Ultimately, the single-round selection success observed by the G-quadruple library against Thrombin provides insight into a well-balanced selection strategy. Sequence diversity is relatively less important in selecting high-quality aptamers than functionality based on structure and modifications in the starting selection library. Considering the presence of commonly one binder per million or billion non-binders, increased sequence diversity promotes an increase in the number of non-binders significantly more than binders. On the contrary, the target-based design of the aptamer library structure and modifications improve the ratio of binders to non-binders. Thereby in a selection using a smaller library size, the presence of a target-based blueprint and higher copy numbers per molar concentration largely promote single-round selection success.

LIST OF REFERENCES

1. Crick, F. & Watson, J. 1953. *Chinese Educ.* **11**, 76–106 (1978).

2. Markham, R. & Smith, J. D. Structure of ribonucleic acid. *Nature* **168**, 406–408 (1951).

3. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122**, 565–581 (2008).

4. Minchin, S. & Lodge, J. Understanding biochemistry: Structure and function of nucleic acids. *Essays Biochem.* **63**, 433–456 (2019).

5. Zhu, H. *et al.* PCR past, present and future. *Biotechniques* **69**, 317–325 (2020).

6. Kumar, K. R., Cowley, M. J. & Davis, R. L. Next-Generation Sequencing and Emerging Technologies. *Semin. Thromb. Hemost.* (2024) doi:10.1055/s-0044-1786397.

7. Villiger, L. *et al.* CRISPR technologies for genome, epigenome and transcriptome editing. *Nat. Rev. Mol. Cell Biol.* **25**, 464–487 (2024).

8. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. *Cell* **187**, 1076–1100 (2024).

9. Stayton, P. S. *et al.* Streptavidin-biotin binding energetics. *Biomol. Eng.* **16**, 39–44 (1999).

10. Koenig, S. H. & Brown, R. D. H 2 CO 3 as substrate for carbonic anhydrase in the dehydration of HCO 3 . *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2422–2425 (1972).

11. Chou, K. C. & Zhou, G. P. Role of the Protein Outside Active Site on the Diffusion-Controlled Reaction of Enzyme. *J. Am. Chem. Soc.* **104**, 1409–1413 (1982).

12. Xu, M. & Lewis, R. V. Structure of a protein superfiber: Spider dragline silk. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7120–7124 (1990).

13. Cech, T. R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **4**, 1–5 (2012).

14. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776–790 (2007).

15. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific

ligands. *Nature* **346**, 818–822 (1990).

16. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–10 (1990).

17. Robertson, D. L. & Joyce, G. F. cleaves single-stranded DNA. *Lett. to Nat.* **344**, 467–468 (1990).

18. Veedu, R. N. & Wengel, J. Locked nucleic acids: Promising nucleic acid analogs for therapeutic applications. *Chemistry and Biodiversity* (2010) doi:10.1002/cbdv.200900343.

19. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science (80-.).* **336**, 341–344 (2012).

20. Ni, S. *et al.* Chemical modifications of nucleic acid aptamers for therapeutic purposes. *International Journal of Molecular Sciences* (2017) doi:10.3390/ijms18081683.

21. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **5**, e15004 (2010).

22. Eid, C., Palko, J. W., Katilius, E. & Santiago, J. G. Rapid Slow Off-Rate Modified Aptamer (SOMAmer)-Based Detection of C-Reactive Protein Using Isotachophoresis and an Ionic Spacer. *Anal. Chem.* **87**, 6736–6743 (2015).

23. Volk, D. E. & Lokesh, G. L. R. Development of Phosphorothioate DNA and DNA Thioaptamers. *Biomedicines* (2017) doi:10.3390/biomedicines5030041.

24. Lokesh, G. L. *et al.* X-Aptamer selection and validation. in *Methods in Molecular Biology* (2017). doi:10.1007/978-1-4939-7138-1_10.

25. Shen, X. & Corey, D. R. Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gkx1239.

26. Veedu, R. N. & Wengel, J. Locked nucleic acid as a novel class of therapeutic agents. *RNA Biol.* (2010) doi:10.4161/rna.6.3.8807.

27. Kang, H. *et al.* Small interfering RNA (siRNA)-based therapeutic applications against viruses: principles, potential, and challenges. *J. Biomed. Sci.* **30**, 1–18 (2023).

28. Hu, B. *et al.* Therapeutic siRNA: state of the art. *Signal Transduct. Target. Ther.* **5**, (2020).

29. Klussmann, S., Nolte, A., Bald, R., Erdmann, V. A. & Fürste, J. P. Mirror-image RNA that binds D-Adenosine. *Nat. Biotechnol.* (1996) doi:10.1038/nbt0996-1112.

30. Eulberg, D. & Klussmann, S. Spiegelmers: Biostable aptamers. *ChemBioChem* (2003) doi:10.1002/cbic.200300663.

31. Vater, A. & Klussmann, S. Turning mirror-image oligonucleotides into drugs: The evolution of Spiegelmer® therapeutics. *Drug Discovery Today* (2015) doi:10.1016/j.drudis.2014.09.004.

32. Gao, S., Zheng, X., Jiao, B. & Wang, L. Post-SELEX optimization of aptamers. *Analytical and Bioanalytical Chemistry* (2016) doi:10.1007/s00216-016-9556-2.

33. McKenzie, L. K., El-Khoury, R., Thorpe, J. D., Damha, M. J. & Hollenstein, M. Recent progress in non-native nucleic acid modifications. *Chem. Soc. Rev.* **50**, 5126–5164 (2021).

34. Jäger, S. *et al.* A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.* **127**, 15071–15082 (2005).

35. Ondruš, M., Sýkorová, V., Bednárová, L., Pohl, R. & Hocek, M. Enzymatic synthesis of hypermodified DNA polymers for sequence-specific display of four different hydrophobic groups. *Nucleic Acids Res.* **48**, 11982–11993 (2020).

36. Obexer, R., Nassir, M., Moody, E. R., Baran, P. S. & Lovelock, S. L. Modern approaches to therapeutic oligonucleotide manufacturing. *Science (80-.).* **384**, (2024).

37. Lauffer, M. C., van Roon-Mom, W. & Aartsma-Rus, A. Possibilities and limitations of antisense oligonucleotide therapies for the treatment of monogenic disorders. *Commun. Med.* **4**, 1–11 (2024).

38. Catani, M. *et al.* Oligonucleotides: Current Trends and Innovative Applications in the Synthesis, Characterization, and Purification. *Biotechnol. J.* **15**, (2020).

39. Andrews, B. I. *et al.* Sustainability Challenges and Opportunities in Oligonucleotide Manufacturing. *J. Org. Chem.* **86**, 49–61 (2021).

40. Matteucci MD, Caruthers, M. Synthesis of Deoxyoligonucleotides on a Polymer Support. *J. Am. Chem. Soc.. Am. Chem. Soc.* **103**, 3185–3191 (1980).

41. Pichon, M. & Hollenstein, M. Controlled enzymatic synthesis of oligonucleotides. *Commun. Chem.* **7**, 1–11 (2024).

42. Wiegand, D. J. *et al.* Template-independent enzymatic synthesis of RNA oligonucleotides. *Nat. Biotechnol.* (2024) doi:10.1038/s41587-024-02244-w.

43. Jul, G. T. & Content, C. D. T. C. Oligonucleotide Synthesis Market worth $ 19 . 7 billion by 2029. *barchart.com* 13–15 (2024).

44. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).

45. Robert L. Letsinger and V. Mahadevan. Stepwise Synthesis of Oligodeoxyribonucleotides on an Insoluble Polymer Support. *J. Am. Chem. Soc.* 5319–5324 (1966).

46. Pon, R. T. Solid-Phase Supports for Oligonucleotide Synthesis. *Curr. Protoc. Nucleic Acid Chem.* **00**, 1–28 (2000).

47. Rublack, N. *et al.* Synthesis of specifically modified oligonucleotides for application in structural and functional analysis of RNA. *J. Nucleic Acids* **2011**, (2011).

48. Wegner, K., Barnes, D., Manzor, K., Jardine, A. & Moran, D. Evaluation of greener solvents for solid-phase peptide synthesis. *Green Chem. Lett. Rev.* **14**, 152–163 (2021).

49. Palluk, S. *et al.* De novo DNA synthesis using polymerasenucleotide conjugates. *Nat. Biotechnol.* **36**, 645–650 (2018).

50. Hollenstein, M., Hipolito, C. J., Lam, C. H. & Perrin, D. M. A DNAzyme with three protein-like functional groups: Enhancing catalytic efficiency of M2+-independent RNA cleavage. *ChemBioChem* **10**, 1988–1992 (2009).

51. Wang, Y., Liu, E., Lam, C. H. & Perrin, D. M. A densely modified M2+-independent DNAzyme that cleaves RNA efficiently with multiple catalytic turnover. *Chem. Sci.* **9**, 1813–1821 (2018).

52. Hili, R., Niu, J. & Liu, D. R. DNA ligase-mediated translation of DNA into densely

functionalized nucleic acid polymers. *J. Am. Chem. Soc.* **135**, 98–101 (2013).

53. Kong, D., Yeung, W. & Hili, R. In Vitro Selection of Diversely Functionalized Aptamers. *J. Am. Chem. Soc.* **139**, 13977–13980 (2017).

54. Guo, C., Kong, D., Lei, Y. & Hili, R. Expanding the Chemical Diversity of DNA. *Synlett* (2018) doi:10.1055/s-0036-1591959.

55. Kong, D. *et al.* Evolutionary Outcomes of Diversely Functionalized Aptamers Isolated from in Vitro Evolution. *ACS Synth. Biol.* **9**, 43–52 (2020).

56. Movahedi, M., Khamissi, N. & Hili, R. Ligase-Catalyzed Oligonucleotide Polymerization (LOOPER): Evolution of chemically diverse aptamer libraries. *Aptamers* **5**, 22–30 (2021).

57. James, K. D., Boles, A. R., Henckel, D. & Ellington, A. D. The fidelity of template-directed oligonucleotide ligation and its relevance to DNA computation. *Nucleic Acids Res.* **26**, 5203–5211 (1998).

58. Lei, Y., Kong, D. & Hili, R. A High-Fidelity Codon Set for the T4 DNA Ligase-Catalyzed Polymerization of Modified Oligonucleotides. *ACS Comb. Sci.* **17**, 716–721 (2015).

59. Chen, Z., Lichtor, P. A., Berliner, A. P., Chen, J. C. & Liu, D. R. Evolution of sequence-defined highly functionalized nucleic acid polymers. *Nat. Chem.* **10**, 420–427 (2018).

60. Shi, K. *et al.* T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. *Nucleic Acids Res.* **46**, 10474–10488 (2018).

61. Lei, Y., Washington, J. & Hili, R. Efficiency and fidelity of T3 DNA ligase in ligase-catalysed oligonucleotide polymerisations. *Org. Biomol. Chem.* **17**, 1962–1965 (2019).

62. Guo, C., Watkins, C. P. & Hili, R. Sequence-Defined Scaffolding of Peptides on Nucleic Acid Polymers. *J. Am. Chem. Soc.* **137**, 11191–11196 (2015).

63. Guo, C. & Hili, R. Fidelity of the DNA Ligase-Catalyzed Scaffolding of Peptide Fragments on Nucleic Acid Polymers. *Bioconjug. Chem.* **28**, 314–318 (2017).

64. Kong, D., Yeung, W. & Hili, R. Generation of Synthetic Copolymer Libraries by Combinatorial Assembly on Nucleic Acid Templates. *ACS Combinatorial Science* vol. 18 355–370 (2016).

65. Kim, Y. RNA Therapy : Current Status and Future Potential. 87–93 (2020).

66. Kim, H. *et al.* Oligonucleotide therapeutics and their chemical modification strategies for clinical applications. *J. Pharm. Investig.* **54**, 415–433 (2024).

67. Jeong, M., Lee, Y., Park, J., Jung, H. & Lee, H. Lipid nanoparticles (LNPs) for in vivo RNA delivery and their breakthrough technology for future applications. *Adv. Drug Deliv. Rev.* **200**, 114990 (2023).

68. Paunovska, K., Loughrey, D. & Dahlman, J. E. Drug delivery systems for RNA therapeutics. (2022) doi:10.1038/s41576-021-00439-4.

69. Hooper, A. J. & Burnett, J. R. Anti-PCSK9 therapies for the treatment of hypercholesterolemia. *Expert Opin. Biol. Ther.* **13**, 429–435 (2013).

70. Fitzgerald, K. *et al.* A Highly Durable RNAi Therapeutic Inhibitor of PCSK9. *N. Engl. J. Med.* **376**, 41–51 (2017).

71. Cicero, A. F. G., Fogacci, F., Zambon, A., Toth, P. P. & Borghi, C. Efficacy and safety of inclisiran a newly approved FDA drug: a systematic review and pooled analysis of available clinical studies. *Am. Hear. J. Plus Cardiol. Res. Pract.* **13**, 100127 (2022).

72. Holm, A., Hansen, S. N., Klitgaard, H. & Kauppinen, S. Clinical advances of RNA therapeutics for treatment of neurological and neuromuscular diseases. *RNA Biol.* **19**, 594–608 (2022).

73. Boccaletto, P. *et al.* MODOMICS : a database of RNA modification pathways . 2021 update. **50**, 8–10 (2022).

74. Cells, H., Desrosiers, R., Friderici, K. & Rottman, F. Identification of Methylated Nucleosides in Messenger RNA from Novikoff. **71**, 3971–3975 (1974).

75. Zou, J. *et al.* Dynamic regulation and key roles of ribonucleic acid methylation. *Front. Cell. Neurosci.* **16**, 1–25 (2022).

76. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).

77. Ping, X. L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-

methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).

78. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).

79. Zheng, G. *et al.* ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49**, 18–29 (2013).

80. Zijlstra, A. & Di Vizio, D. Size matters in nanoscale communication. *Nat. Cell Biol.* **20**, 228–230 (2018).

81. Csepany, T., Lin, A., Baldick, C. J. & Beemon, K. Sequence specificity of mRNA N6-adenosine methyltransferase. *J. Biol. Chem.* **265**, 20117–20122 (1990).

82. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).

83. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).

84. Shi, R. *et al.* Linking the YTH domain to cancer: the importance of YTH family proteins in epigenetics. *Cell Death Dis.* **12**, (2021).

85. Theler, D., Dominguez, C., Blatter, M., Boudet, J. & Allain, F. H. T. Solution structure of the YTH domain in complex with N6-methyladenosine RNA: A reader of methylated RNA. *Nucleic Acids Res.* **42**, 13911–13919 (2014).

86. Luo, S. & Tong, L. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13834–13839 (2014).

87. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* **10**, 927–929 (2014).

88. Wang, N., Yue, Z., Liang, D. & Ma, F. Genome-wide identification of members in the YTH domain-containing RNA-binding protein family in apple and expression analysis of their responsiveness to senescence and abiotic stresses. *Gene* **538**, 292–305 (2014).

89. Zhang, C. *et al.* YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation. *Oncogene* **39**,

4507–4518 (2020).

90. Paris, J. *et al.* Targeting the RNA m6A Reader YTHDF2 Selectively Compromises Cancer Stem Cells in Acute Myeloid Leukemia. *Cell Stem Cell* **25**, 137-148.e6 (2019).

91. Bai, Y. *et al.* YTHDF1 regulates tumorigenicity and cancer stem cell-like activity in human colorectal carcinoma. *Front. Oncol.* **9**, (2019).

92. Shi, Y. *et al.* YTHDF1 links hypoxia adaptation and non-small cell lung cancer progression. *Nat. Commun.* **10**, (2019).

93. Zhang, M., Song, J., Yuan, W., Zhang, W. & Sun, Z. Roles of RNA Methylation on Tumor Immunity and Clinical Implications. *Front. Immunol.* **12**, 1–13 (2021).

94. E. Wang, R., Wu, H., Niu, Y. & Cai, J. Improving the Stability of Aptamers by Chemical Modification. *Curr. Med. Chem.* **18**, 4126–4138 (2011).

95. Nandakumar, J., Shuman, S. & Lima, C. D. RNA Ligase Structures Reveal the Basis for RNA Specificity and Conformational Changes that Drive Ligation Forward. 71–84 (2006) doi:10.1016/j.cell.2006.08.038.

96. Lesnik, E. A. & Freier, S. M. Relative Thermodynamic Stability of DNA , RNA , and DNA : RNA Hybrid Duplexes : Relationship with Base Composition and Structure. 10807–10815 (1995).

97. Liu, J. *et al.* Structural Flexibility of DNA-RNA Hybrid Duplex : Stretching and Twist-Stretch Coupling. *Biophysj* **117**, 74–86 (2019).

98. Quail, M. A. *et al.* A tale of three NGS sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 13 (2012).

99. Xu, D., Landon, T., Greenbaum, N. L. & Fenley, M. O. The electrostatic characteristics of G · U wobble base pairs. *Nucleic Acids Res.* **35**, 3836–3847 (2007).

100. Kong, D., Lei, Y., Yeung, W. & Hili, R. Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angew. Chemie - Int. Ed.* **55**, 13164–13168 (2016).

101. Kuznetsova, A. A. & Kuznetsov, N. A. Direct Enzyme Engineering of B Family DNA

Polymerases for Biotechnological Approaches. *Bioengineering* **10**, 1–16 (2023).

102. Skrzypczak-Jankun, E. *et al.* Structure of the hirugen and hirulog 1 complexes of α-thrombin. *J. Mol. Biol.* **221**, 1379–1393 (1991).

103. Ofosu, F. A. *et al.* Inhibition of the amplification reactions of blood coagulation by site-specific inhibitors of α-thrombin. *Biochem. J.* **283**, 893–897 (1992).

104. Warkentin, T. E. & Koster, A. Bivalirudin: A review. *Expert Opin. Pharmacother.* **6**, 1349–1371 (2005).

105. Bates, S. M. & Weitz, J. I. Direct thrombin inhibitors for treatment of arterial thrombosis: Potential differences between bivalirudin and hirudin. *Am. J. Cardiol.* **82**, 12P-18P (1998).

106. Chahal, G., Thorpe, M. & Hellman, L. The importance of exosite interactions for substrate cleavage by human thrombin. *PLoS One* **10**, (2015).

107. Troisi, R., Balasco, N., Autiero, I., Vitagliano, L. & Sica, F. Exosite binding in thrombin: A global structural/dynamic overview of complexes with aptamers and other ligands. *Int. J. Mol. Sci.* **22**, (2021).

108. Russo Krauss, I. *et al.* High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res.* (2012) doi:10.1093/nar/gks512.

109. Smirnov, I., Kolganova, N., Troisi, R., Sica, F. & Timofeev, E. Expanding the recognition interface of the thrombin-binding aptamer HD1 through modification of residues T3 and T12. *Mol. Ther. Nucleic Acids* **23**, 863–871 (2021).

110. Maréchal, E., Roy, S. & Lafanechère, L. *Chemogenomics and Chemical Genetics*. (2011).

111. Jug, A., Bratkovič, T. & Ilaš, J. Biolayer interferometry and its applications in drug discovery and development. *TrAC - Trends Anal. Chem.* **176**, (2024).

112. Hall, M. D. *et al.* Fluorescence polarization assays in high-throughput screening and drug discovery: A review. *Methods and Applications in Fluorescence* (2016) doi:10.1088/2050-6120/4/2/022001.

113. Bauke Albada, H., Golub, E. & Willner, I. Computational docking simulations of a DNA-

aptamer for argininamide and related ligands. *J. Comput. Aided. Mol. Des.* **29**, 643–654 (2015).

114. Ozvoldik, K., Stockner, T. & Krieger, E. YASARA Model-Interactive Molecular Modeling from Two Dimensions to Virtual Realities. *J. Chem. Inf. Model.* **63**, 6177–6182 (2023).

115. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA - A self-parameterizing force field. *Proteins Struct. Funct. Genet.* (2002) doi:10.1002/prot.10104.

116. Sattari, R., Palizban, A. & Khanahmad, H. Single-Strand DNA-Like Oligonucleotide Aptamer Against Proprotein Convertase Subtilisin/Kexin 9 Using CE-SELEX: PCSK9 Targeting Selection. *Cardiovasc. Drugs Ther.* **34**, 475–485 (2020).

117. Spiridonova, V. A., Novikova, T. M. & Snigirev, O. V. Obtaining DNA aptamers to human interleukin-6 for biomagnetic immunoassay nanosensors. *Moscow Univ. Phys. Bull.* **71**, 135–138 (2016).

118. Lichtor, P. A., Chen, Z., Elowe, N. H., Chen, J. C. & David, R. Side-chain determinants of biopolymer function during selection and replication. **15**, 419–426 (2019).

119. Kohlberger, M. SELEX : Critical factors and optimization strategies for successful aptamer selection. 1771–1792 (2022) doi:10.1002/bab.2244.

120. Tatjana, S. *et al.* Probing the SELEX Process with Next-Generation Sequencing. **6**, 1–10 (2011).

121. Musheev, M. U. & Krylov, S. N. Selection of aptamers by systematic evolution of ligands by exponential enrichment: Addressing the polymerase chain reaction issue. *Anal. Chim. Acta* **564**, 91–96 (2005).

122. Binet, T., Lefèvre, S. P., Octave, S., Avalle, B. & Maffucci, I. Comparative Study of Single - stranded Oligonucleotides Secondary Structure Prediction Tools. *BMC Bioinformatics* 1–20 (2023) doi:10.1186/s12859-023-05532-5.

123. Mishra, S. K. *et al.* Characterization of highly conserved G-quadruplex motifs as potential

drug targets in Streptococcus pneumoniae. *Sci. Rep.* **9**, 1–13 (2019).

124. Luo, Z., He, L., Wang, J., Fang, X. & Zhang, L. Developing a combined strategy for monitoring the progress of aptamer selection. *Analyst* **142**, 3136–3139 (2017).

125. Merriman, B., Torrent, I. & Rothberg, J. M. Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis* **33**, 3397–3417 (2012).

126. Alam, K. K., Chang, J. L. & Burke, D. H. FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids* **4**, e230 (2015).

127. Kramer, S. T., Gruenke, P. R., Alam, K. K., Xu, D. & Burke, D. H. FASTAptameR 2.0: A web tool for combinatorial sequence selections. *Mol. Ther. Nucleic Acids* **29**, 862–870 (2022).

128. Berezovski, M., Musheev, M., Drabovich, A. & Krylov, S. N. Non-SELEX selection of aptamers. *J. Am. Chem. Soc.* **128**, 1410–1411 (2006).

129. Kochmann, S., Le, A. T. H., Hili, R. & Krylov, S. N. Predicting efficiency of NECEEM-based partitioning of protein binders from nonbinders in DNA-encoded libraries. *Electrophoresis* **39**, 2991–2996 (2018).

130. Le, A. T. H., Krylova, S. M. & Krylov, S. N. Ideal-filter capillary electrophoresis: A highly efficient partitioning method for selection of protein binders from oligonucleotide libraries. *Electrophoresis* **40**, 2553–2564 (2019).

131. Tavares, M. F. M. & McQuffln, V. L. Theoretical Model of Electroosmotic Flow for Capillary Zone Electrophoresis. *Anal. Chem.* **67**, 3687–3696 (1995).

132. Ivanov, N. A., Panferov, V. G. & Krylov, S. N. Nonsteady-State Electric Circuit in Electrophoresis on Paper: Thermal Consideration of Electrophoretic Lateral-Flow Assays. *Anal. Chem.* (2023) doi:10.1021/acs.analchem.3c05832.

133. Le, A. T. H. *et al.* How to Develop and Prove High-Efficiency Selection of Ligands from Oligonucleotide Libraries: A Universal Framework for Aptamers and DNA-Encoded Small-Molecule Ligands. *Anal. Chem.* **93**, 5343–5354 (2021).

134. De Jong, S. & Krylov, S. N. Pressure-based approach for the analysis of protein adsorption in capillary electrophoresis. *Anal. Chem.* **84**, 453–458 (2012).

135. Le, A. T. H., Teclemichael, E., Krylova, S. M. & Krylov, S. N. Quantitative Characterization of Partitioning Stringency in SELEX. *ChemRxiv* **27**, 1–9 (2024).

136. Vinjamuri, B. P., Pan, J. & Peng, P. A Review on Commercial Oligonucleotide Drug Products. *J. Pharm. Sci.* **113**, 1749–1768 (2024).

137. Szeitner, Z., András, J., Gyurcsányi, R. E. & Mészáros, T. Is less more? Lessons from aptamer selection strategies. *J. Pharm. Biomed. Anal.* **101**, 58–65 (2014).

138. Lozoya-Colinas, A., Yu, Y. & Chaput, J. C. Functionally Enhanced XNA Aptamers Discovered by Parallelized Library Screening. *J. Am. Chem. Soc.* **145**, 25789–25796 (2023).

139. Song, Y. & Li, X. Evolution of the Selection Methods of DNA-Encoded Chemical Libraries. *Acc. Chem. Res.* **54**, 3491–3503 (2021).

140. Bikard, D., Baharoglu, Z. & Mazel, D. Folded DNA in Action : Hairpin Formation and Biological Functions in Prokaryotes. **74**, 570–588 (2010).

141. Sanchez de Groot, N. *et al.* RNA structure drives interaction with proteins. *Nat. Commun.* **10**, 1–13 (2019).

142. Moore, P. B. Structural Motifs in RNA. 287–300 (1999).

143. Pezza, J. A., Kucera, R. K. & Sun, L. Polymerase Fidelity : What is it , and what does it mean for your PCR ? Introduction : What is fidelity ? *Neb* 1–3 (2017).

144. Dolot, R. *et al.* Crystal structures of thrombin in complex with chemically modified thrombin DNA aptamers reveal the origins of enhanced affinity. *Nucleic Acids Res.* **46**, 4819–4830 (2018).

145. Krauss, I. R., Spiridonova, V., Pica, A., Napolitano, V. & Sica, F. Different duplex / quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. (2015) doi:10.1093/nar/gkv1384.

146. Sato, K. & Knipscheer, P. G-quadruplex resolution : From molecular mechanisms to

physiological relevance. *DNA Repair (Amst).* **130**, 103552 (2023).

147. Yuan, W. *et al.* The influencing factors and functions of DNA G-quadruplexes. 524–532 (2020) doi:10.1002/cbf.3505.

148. Cao, Z., Huang, C. C. & Tan, W. Nuclease resistance of telomere-like oligonucleotides monitored in live cells by fluorescence anisotropy imaging. *Anal. Chem.* **78**, 1478–1484 (2006).

149. Bishop, J. S. *et al.* Intramolecular G-quartet motifs confer nuclease resistance to a potent anti-HIV oligonucleotide. *J. Biol. Chem.* **271**, 5698–5703 (1996).

150. Dailey, M. M., Clarke Miller, M., Bates, P. J., Lane, A. N. & Trent, J. O. Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Res.* **38**, 4877–4888 (2010).

151. Bock, L., Griffin, L., Latham, J., Vermaas, E. & Toole JJ. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).

152. Derszniak, K. *et al.* Comparison of effects of anti-thrombin aptamers HD1 and HD22 on aggregation of human platelets, thrombin generation, fibrin formation, and thrombus formation under flow conditions. *Front. Pharmacol.* **10**, 1–13 (2019).

153. 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).

154. Ji, S., Shao, H., Han, Q., Seiler, C. L. & Natalia, Y. Reversible DNA-protein cross-linking at epigenetic DNA marks. *Angew. Chemie* **56**, 14130–14134 (2017).

155. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

156. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Eur. Mol. Biol. Netw.* **17**, 10–12 (2013).

LIST OF PUBLICATIONS

Articles published or accepted in peer-reviewed journals (*ACS Synth. Biol.* 2020, **9, **1**, 43–52)**

Kong D.; **Movahedi M.;** Mahdavi-Amiri Y.; Yeung W.; Tiburcio T.; Chen D.; Hili R. "Evolutionary Outcomes of Diversely-Functionalized Aptamers Isolated from In Vitro Evolution" *ACS Synth. Biol.* 2019, 9, 43-52

Articles published or accepted in peer-reviewed journals (JAptamers, 2021, **5**, p22**)**

Matina Movahedi, Natalie Khamissi and Ryan Hili "Ligase Catalyzed Oligonucleotide Polymerization (LOOPER): Evolution of chemically diverse aptamer libraries"

Articles published or accepted in peer-reviewed journals (*Org. Biomol. Chem.*, 2023,21**, 1463-1467)**

Yasaman Mahdavi-Amiri, Molly S. J. Hu, Nicole Frias, **Matina Movahedi**, Adam Csakai, Lisa A. Marcaurelle, and Ryan Hili **"**Photoredox-catalysed hydroaminoalkylation of on-DNA N-arylamine"