EXPANDING THE SCOPE OF T3 DNA LIGASE-CATALYZED OLIGONUCLEOTIDE POLYMERIZATIONS (T3-LOOPER)

NATALIE KHAMISSI

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

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO

April 2025

Abstract

Antibodies are currently the gold standard in the field of therapeutics but are quite costly and chemically unstable. Nucleic acid polymers can mimic antibody therapeutics but are more cost effective to produce, can reversibly denature, and have a high synthesis reproducibility. These polymers are called 'aptamers'. However, they lack in chemical diversity when compared to antibody-based therapeutics and are susceptible to renal filtration, making them inferior for specific binding. Chemical modifications can be attached to these DNA/RNA polymers to increase their diversity, though only four unique modifications can be incorporated into one polymer. Ligase-catalyzed oligonucleotide polymerization (LOOPER) is a method that synthesizes chemically modified nucleic acid libraries containing up to 16 unique chemical modifications. The method entails template directed ligations of chemically modified oligonucleotides. The purpose of this thesis is to push the boundaries of LOOPER to tolerate nucleobase modified oligonucleotides, sugar modified oligonucleotides, modification positioning on the oligonucleotide sequence, RNA oligonucleotides, and reverse transcription via LOOPER of a modified nucleic acid polymer back to its DNA form. A nucleobase modified library was synthesized to include predominantly hydrophobic modifications, allowing for an increased chance of developing an aptamer during in vitro selections against a protein target. The library generated a LOOPER yield of 19% with a fidelity of 94.8%, allowing this library to be a viable option for selections. When attempting LOOPER with sugar modified oligonucleotides, LOOPER was able to tolerate 2'-F at every position in the trinucleotide sequence and locked nucleic acids (LNA) at the middle position of the oligonucleotide. Fidelity analysis revealed that all 2'-F fidelities were over 90%. LNA oligonucleotide fidelities were unable to be characterized since the modification cannot be transcribed by commercially available polymerases. Luckily, we discovered that LOOPER can "reverse transcribe" the LNA-modified nucleic acid polymer back to DNA with a 99.0% fidelity. Lastly, the LOOPER system was not able to accurately ligate RNA oligonucleotides, as the fidelities were below 20%. LOOPER is a method that can "polymerize" and "reverse transcribe" modified nucleic acid polymers, making these libraries more accessible to the in vitro selection process.

Dedication

To my loving husband, Sebastian Romanutti, and my wonderful parents, Elias and Clara Khamissi.

Acknowledgements

I would like to thank Professor Ryan Hili for giving me such a huge opportunity. Your persistence and patience have allowed me to grow into a research scientist with unwavering curiosity, resilience, and a newfound respect for mentors like you. Thank you for taking a chance on my crazy ideas. I owe all my success to you.

To my amazing parents and husband, the most important people in my life, I cannot think of a time that you have not supported me. Thank you for your extreme patience, compassion, and understanding throughout my life. It is because of you that I am successful in all aspects of my life. I dedicate my thesis to you.

I would also like to wholeheartedly thank all of my lab mates: Dr. Matina Movahedi, Dr. Yasaman Mahdavi-Amiri, Molly Hu, Dr. Kimberley Chung Kim Chung, Morgan Cauley-Le Fevre, Ian Frankel, Emily Anacleto, Luca Di Fabio, Nicole Frias, Areeba Chaudhry, Kristina Gremi, and Luca Rustico for all of your insight, support, and wonderful memories. I have truly made life-long friends during this journey.

Thank you to Maxime Rossato and Nikita Ivanov for their help with mass spectrometry throughout my research journey and a special thank you to Professor Sergey N. Krylov and Professor Philip E. Johnson for their support and research advice throughout my research journey.

Table of Contents

List of Tables

List of Figures

Chapter 1 – LOOPER in aptamer selections[§]

[§]This chapter has been published as a review article in *Aptamers* (2021), Vol 5, 22-30. Dr. Matina Movahedi wrote the sections regarding T4 DNA ligase mediated LOOPER, I wrote the sections regarding T3 DNA ligase mediated LOOPER, and Dr. Ryan Hili wrote the introductory and concluding sections.

1.1 Aptamers (chemical antibodies)

Over the last few decades, there has been an increase in FDA-approved monoclonal antibodies (mAbs) and a decrease in small molecule therapeutics. In fact, mAbs have become the predominant treatment modality over the past 25 years[1]. This is due to the vast chemical diversity in proteins, they can bind to molecular targets with a high-level of specificity, exceptionally tight binding [2] and diffusion-controlled catalysis[3,4]. Even though their unique folding and amino acid residues give this drug class an edge, they are extremely costly and labor-intensive. In order to produce biologically functional mAbs, mammalian expression systems must be used which result in low cell growth and low product yield. Once the antibody has been expressed, biological contaminants are removed using harsh conditions, resulting in product degradation and yield loss[5]. These hurdles are what make mAbs an expensive therapeutic to manufacture.

This notwithstanding, indications of a potential functional past[6] remain extant in the form of RNAs with binding and catalytic properties[7]. These properties prompted groundbreaking work on the laboratory evolution of nucleic acid polymers with binding[8,9] and catalytic[10] functions. 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 sequence 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. Indeed, expanding the

chemical diversity of nucleic acids has been essential to improving molecular recognition and catalysis of nucleic acids[11]. While generating modified nucleic acid polymers using polymerases has enabled the incorporation of up to four different modifications[12,13],to attain the level of diversity seen in proteins requires alternative approaches (**Figure 1**).

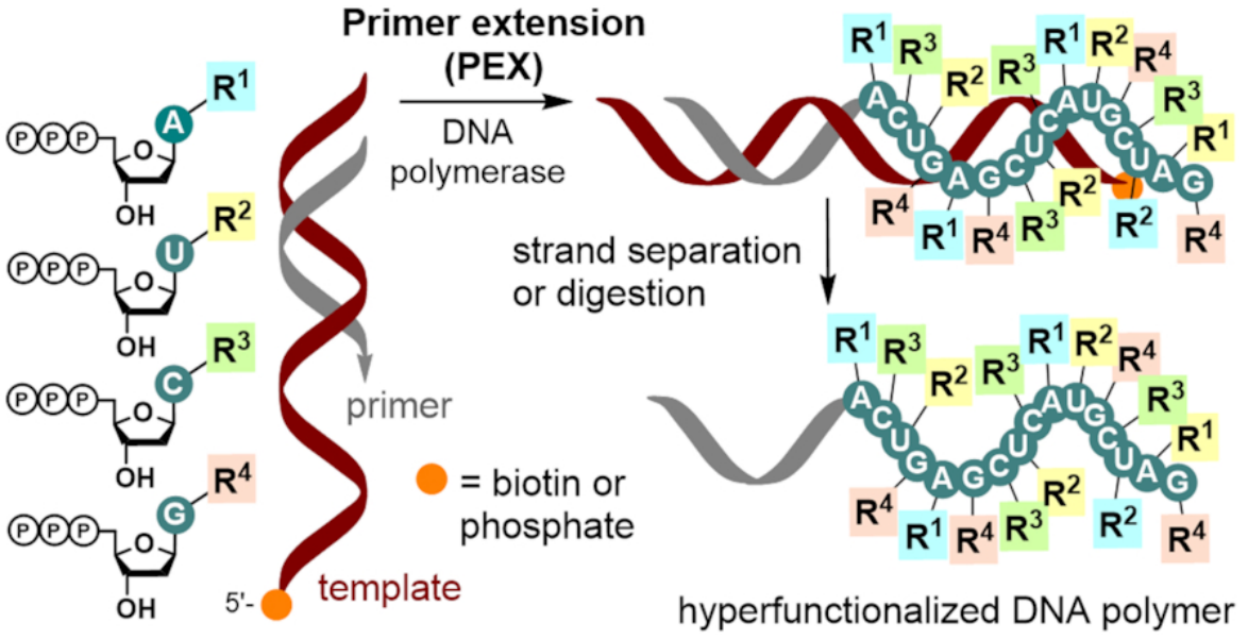


Figure 1: Polymerase mediated synthesis of chemically modified oligonucleotides. Image source: Ondruš et al., 2020[13].

1.2 Ligase-catalyzed Oligonucleotide Polymerization (LOOPER)

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[12,13], a maximum of three have been successfully implemented during *in vitro* selections[14,15]. To expand beyond this limitation requires an alternative approach. Table 1 provides a guideline of current methods researchers use to generate modified nucleic acids.

	ENGINEERED POLYMERASE-BASED METHODS	SOLID PHASE SYNTHESIS	T4-LOOPER	T3-LOOPER	CLICKMERS
SEQUENCE DIVERSITY	maximum diversity	maximum diversity	adenine must be at every fifth base	maximum diversity	maximum diversity
SELEX CAPABILITY	if the modifications can be reverse transcribed	none	if the modifications can be reverse transcribed	if the modifications can be reverse transcribed	capable of SELEX
MODIFICATION DIVERSITY	up to 4 unique modifications	High level of modification diversity	up to 16 unique modifications	up to 8 unique modifications	one unique modification per strand
MODIFICATION DENSITY	maximum density	maximum density	every fifth base	maximum density	sporadic
MODIFIED BACKBONE TOLERANCE	yes	yes	has not been tested	yes	no

Table 1: The comparison of current methods used to modify oligonucleotides for aptamer development.

By using DNA ligases to ligate short, modified oligonucleotides, 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. Thus, a trinucleotide can theoretically encode 64 modifications, whereas a pentanucleotide can accommodate up to 1024; this of course occurs at the expense of modification density. To achieve the sequence-defined ligation of short, modified oligonucleotide fragments, we developed Ligase-catalyzed Oligonucleotide Polymerization (LOOPER). LOOPER begins with a library of templates which 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 2**). 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. Thus, LOOPER enables access to highly functionalized aptamer libraries through

sequence-defined heteromultivalent display of diverse modifications. Due to the critical role played by the DNA ligase in LOOPER, different ligases have been used, including E. coli DNA ligase[16] and T3, T4, and T7 DNA ligases[17,18] each demonstrating their own unique strengths and weaknesses for the process.

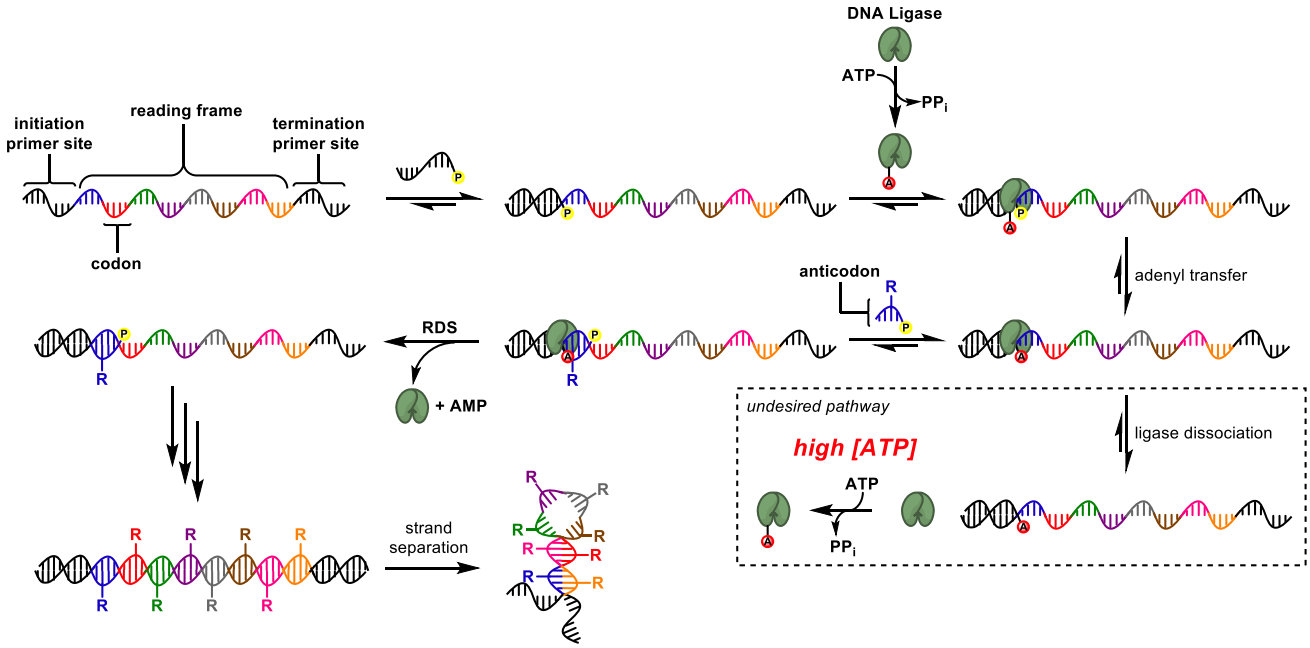


Figure 2: General process for LOOPER. A DNA template library contains two primer binding sites, and a reading frame comprising multiple instances of codons are a specific length. Upon annealing of a primer, a DNA ligase catalyzes the phosphodiester bind formation of cognate anticodons with encoded modifications 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.

1.3 T4 DNA ligase in LOOPER

T4 DNA ligase has been the standard ligase used for the LOOPER method due to its ready 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[19], so initial efforts were devoted to systematically understanding the anticodon preferences of T4 DNA ligase during LOOPER. 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[17,20], pentanucleotides were by-and-large the optimal length, with longer anticodons being prone to annealing and ligating out of frame, resulting in low yield of desired full-length products. Second, in a pentanucleotide anticodon, modifications on the Hoogsteen face of adenine, typically *via* position 8 (**Figure 3**), 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[21]. Third, high ATP concentrations resulted in a precipitous drop in polymerization efficiency, particularly with larger modifications relative to 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 (**Figure 2**). To offset this issue, the ATP concentration had to be carefully controlled, with the optimal concentration found to be 25 μM[21]. 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[18]. 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[22]. Sixth, PEG 6000 was essential as a molecular crowding reagent to increase the effective molarity and thus the rate of LOOPER polymerization reactions[17]. These early findings were essential toward the rational design of codon sets to be used in LOOPER and for porting the method into *in vitro* selections.

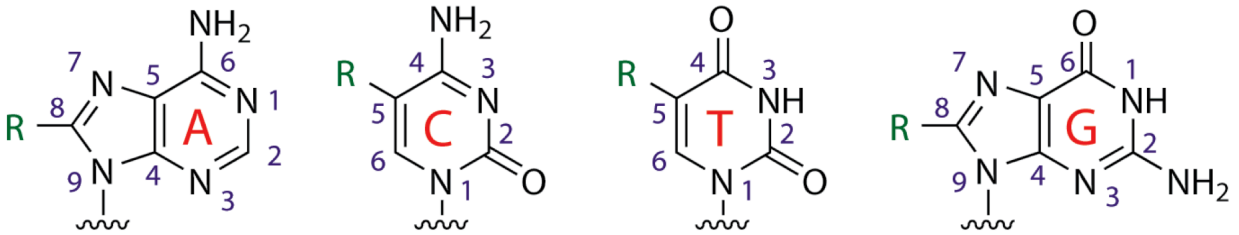


Figure 3: Common modification sites on nucleobases tested in LOOPER. "R" denotes modification site.

1.4 Designing a high-fidelity codon set

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 consistently demonstrated that modified pentanucleotides resulted in the highest yields of full-length products. With the development of a duplex DNA sequencing method that allowed for template and polymer strands to be directly compared[18], 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. Duplex DNA sequencing is compatible with standard high-throughput DNA sequencing instrumentation with full codon fidelity analysis achieved using Analooper (GitHub repository: https://github.com/HiliLab/analooper).

Using a codon set defined as NNNNN, and the corresponding 1024-membered unmodified anticodon library, LOOPER was able to incorporate the correct anticodon across from its cognate codon 81% of the time – a staggering feat considering the combinatorial complexity of the system. However, this fidelity remains too low for molecular evolution, as less than 20% of a 40-nt genotype would be correctly translated into the corresponding phenotype, resulting in an error catastrophe[23]. As anticipated, using smaller codon sets such as NNNNT, which requires the 256-

membered anticodon library ANNNN, resulted in a higher LOOPER fidelity of 87%. Fortuitously, when modifications were incorporated into the anticodon at position 8 of the adenine base, fidelities increased sharply, ranging from 95-98% depending on the location of the modification along the anticodon. This striking observation has been suggested to result from an *anti* to *syn* conformational switch about the glycosidic bond to alleviate steric clash between the modification at position 8 of adenine and the adenylated phosphate backbone[18]; such conformational switches have been observed in other *C*8-substitued adenosine derivatives[24,25]. Since the syn conformation precludes Watson-Crick-Franklin interaction of the modified adenine of the anticodon with the thymine base in the codon, the anticodon hybridization is thus more sensitive to errors in the remaining four hybridizing nucleobases towards the 3'-end.

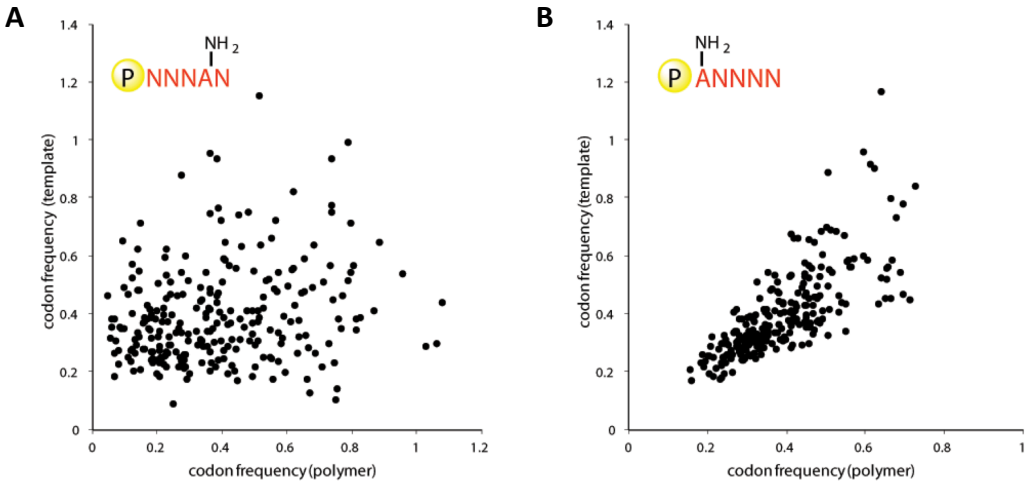


Figure 4: Codon bias observed with T4 DNA ligase mediated LOOPER using different modification sites on anticodons. Low codon bias is represented with data points lying along the diagonal. A)anticodon/codon set with high codon bias; B) anticodon/codon set with low codon bias. The amino modification was displayed as hexamethylenediamine at position 8 of adenine.

The location of a modification on an anticodon influences more than fidelity, it also had a large impact on codon bias during LOOPER. Codon bias is concerned with the frequency of a codon in the template as compared to its frequency in the product strand. Low codon bias is desirable and requires near equal frequencies of codon 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 high yield of full-length product have low codon bias. As seen in (**Figure 4**), the codon set NNNNT with the corresponding anticodon ANNNN with modification at the 5'-end has acceptable codon bias, yield, and fidelity, making it an ideal codon/anticodon system.

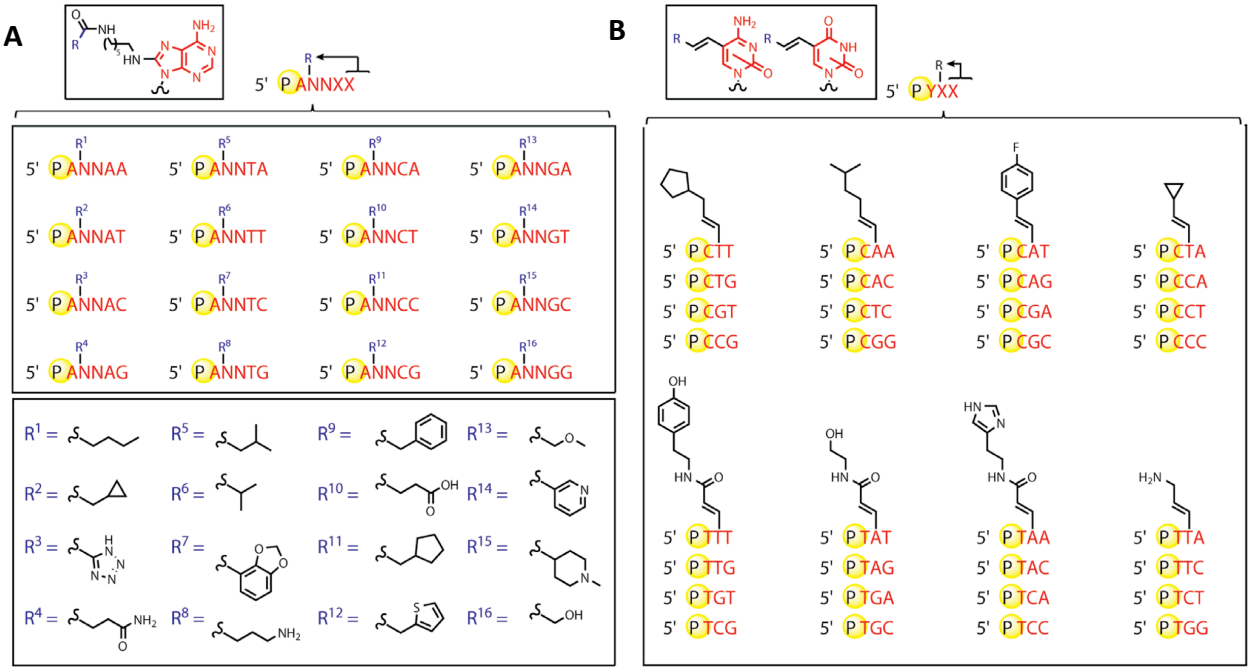


Figure 5: Modified anticodon sets used during LOOPER. A) 256-membered pentanucleotide anticodon set encoding 16 unique chemical modifications. B) 32-membered trinucleotide anticodon set encoding eight unique chemical modifications.

With high fidelity and low codon bias observed with the NNNNT codon set, the choice of how to encode the modification becomes important. Sequencing analysis of LOOPER with the NNNNT codon set revealed that most errors occurred within the two nucleotides at the 3'-end of the anticodon[26]. This can be rationalized by the preference of T4 DNA ligase to extend from the 3'-end, thus errors occur distal to the site of ligation. To take advantage of these errors during *in vitro* evolutions, encoding of anticodon modifications were designated by the two 3'-end nucleotides of

the anticodon, such that when an error occurred, not only was sequence space altered, but also the identity of the modification – akin to a missense mutation (**Figure 5a**). With a dinucleotide encoding system, up to 16 modifications can be incorporated throughout a nucleic acid polymer. Furthermore, each modification has 16 degenerate codons due to the two non-encoding nucleotides in each anticodon. This affords the opportunity for each modification to be displayed within greater sequence space. This codon/anticodon set was used to generate oligonucleotide libraries decorated with 16 different functional groups, the most to date, and has been successfully employed during *in vitro* selection of aptamers.

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 LOOPER during in vitro selections.

Fidelities ranging from 87-98% (average 94%) have been used during successful selections[26,27]. Anticodons modified with 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[21]. Since the peptide modified aptamer libraries could not be directly amplified with polymerase, a variant of duplex DNA sequencing was developed to analyze the fidelity of incorporation of these peptide-modified anticodons[22]. Using oxidative cleavage to remove the peptide fragments ahead of PCR, fidelities could readily be determined. Due to the size of the modification, some codon sets suffered from low full-length yields and concomitant codon bias; however, 16-membered codon sets resulted in good yields, low bias, and fidelities of up to 99%[22]. Generation of DNA-scaffolded peptide libraries using

LOOPER has been validated in mock selections[21], adapting previously developed[28,29] display approaches.

1.5 Recent advances in LOOPER with T4 DNA ligase

An increase in fidelity can benefit *in vitro* applications by increasing the enrichment factors and allowing access to larger reading frames. Considering the essential role that ATP plays in LOOPER, cofactor modifications can directly influence ligation kinetics, codon bias, and fidelity. 17 ATP derivatives were investigated as co-factors for T4 DNA ligase mediated LOOPER, including modifications on the nucleobase, ribose ring, and triphosphate[30] (**Figure 6**). Modifications on the ribose ring were not tolerated, resulting in no observable LOOPER. Nucleobase modifications were tolerated, including 2-amino ATP and 2-chloro ATP which provided 87% and 64% yield of full-length product at 25 μM, respectively (versus 90% for ATP). *N*6-methyl ATP required high concentrations to effect LOOPER, yielding 70% full-length product at 1 mM (versus 60% for ATP). Importantly, both 2-amino ATP and *N*6-methyl ATP showed fidelities in excess of 96%, without compromising codon bias and yield, suggesting that they provide advantages over ATP in LOOPER and could serve as promising ATP derivatives in future iterations of LOOPER.

LOOPER has traditionally been conducted using hexamethylenediamine (HMDA) linkers between the 8-position of adenine and the modification. This has largely been due to the commercial availability of amine-modified phosphoramidites used for automated solid-phase DNA synthesis. While these linkers have been successfully used in raising aptamers derived from LOOPER generated libraries, the length and flexibility of HMDA could negatively impact the thermodynamics of binding. To address this issue, the compatibility of shorter linkers in LOOPER

has been explored[31]. Using an ANNNN anticodon library as a model, various linkers were examined. Very short linkers, such as methyleneamine (RCH_2NH-) polymerized efficiently and with good fidelity (92%), albeit suffered from very high codon bias, making this linker unsuitable for selections. Ethylene diamine (EDA) linkers provided a good middle ground, allowing LOOPER with a range of modifications (including acid, base, polar, and hydrophobic) in good yield (67-77%), good codon bias, and good fidelities (92- 97%). Importantly, it was found that the EDA linker could be directly incorporated during the solid-phase DNA synthesis deprotection step using commercially available 8-bromo 2-deoxyadenosine phosphoramidites[31]. While no selections have been performed on EDA-linker LOOPER libraries, such selections would be instructive in assessing the impact of linker length on modified aptamer binding.

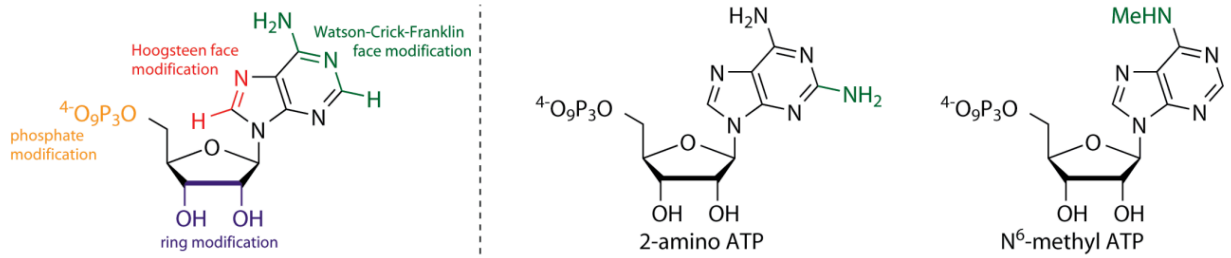


Figure 6: ATP modifications sites tested in T4 DNA ligase mediated LOOPER. Of the 17 derivatives screened, 2-amino ATP and N^6-methyl ATP provided the most promising results.

1.6 T3 DNA ligase in LOOPER

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[32], with full-length products observed for up to 45-nt reading frames (15 codons). 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[20]. Similar to T4 DNA ligase, modifications were optimal on the 5'-end of the anticodon, and accommodation of various small polar and hydrophobic groups have been shown[32]; larger groups have yet to be explored. It is important to note that since T3 DNA ligase can only tolerate small modifications in trinucleotide systems, custom modified phosphoramidites are required[32], albeit this may not be required for longer codon sets. 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 (Figure 4b). 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[32]; using duplex DNA sequencing, T3 DNA ligase was later shown to have 97% fidelity with either unmodified or modified YNN anticodons[20]. T3 DNA ligase can handle larger anticodons 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.

1.7 *In vitro* evolution of LOOPER-derived aptamers

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 Systematic Evolution of Ligands by EXponential enrichment (SELEX) (**Figure 7a**) provides access to high quality functionalized aptamers with potential improvements of affinity and selectivity against molecular targets. Provided that fidelity, codon bias, and polymerization yield of LOOPER are well optimized, 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), 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.

1.8 Successful LOOPER-SELEX evolutions

The first successful LOOPER-SELEX aptamer evolution was demonstrated using a 256-membered ANNNN anticodon library encoding 16 unique modifications (**Figure 5a**), generated using T4 DNA ligase[27]. 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; the highest frequency sequence, TBL1, represented 1.1% of the total output library (**Figure 7b**). 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. Even more surprising was that TBL1 also bound to thrombin, and with much tighter affinity (K_D = 1.6 nM). This represented the first known instance where both the genotype and phenotype achieved the same function during molecular evolution.

The modifications displayed on TBL1 were shown to be essential for binding[33]. 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 the 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 modelling studies[33]. TBL1 shares functional group similarities with known peptide-based inhibitors, such as hirugen and the FDA-approved drug bivalirudin – both are rich in acidic and hydrophobic residues similar to TBL1. These peptide inhibitors are known to bind exosite I on thrombin suggesting a potential similar mode of action for TBL1. Indeed, competitive binding assays demonstrated that TBL1 binds exosite I without interacting with nearby exosite II[33]. TBL1 exhibits considerably tighter binding compared to unmodified DNA aptamers that target exosite I of thrombin, such as the original thrombin-binding aptamer HD1[34], which exhibits a K_D of 20 nM[35]. TBL1 was also shown to inhibit thrombin activity in serum assays with an IC_{50} of 23 nM. 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.

More recently, T3 DNA ligase has been used in LOOPER-SELEX with a 32-membered YNN anticodon library encoding eight modifications (**Figure 5b**). This library was used to generate aptamers against Proprotein convertase subtilisin/ kexin type 9 serine protease (PCSK9) and interleukin-6 (IL6)[32]. After 9 rounds of selection, aptamer PCSK9-A5 was isolated, demonstrating a K_D of 98 nM (**Figure 7b**). This shows markedly higher affinity against PCSK9 than a recently evolved unmodified DNA aptamer, known as AP-1, which had a K_D of 294 nM[36]. 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 3nM 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 LOOPERSELEX 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 = 12nM (**Figure 7b**). 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[37]. 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[38]. 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.

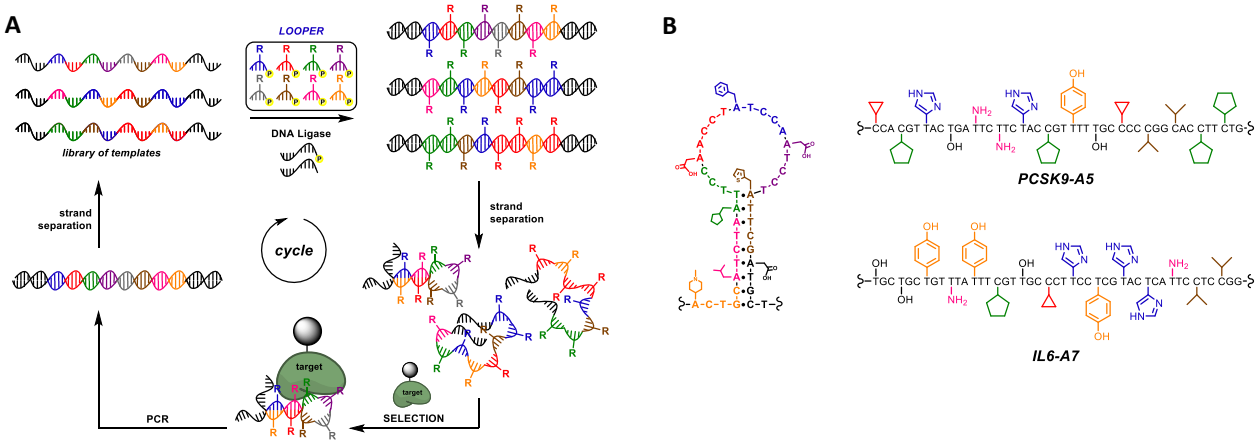


Figure 7: Evolution of sequence-defined diversely functionalized DNA aptamers. A) LOOPER-SELEX cycle used to evolve modified DNA aptamers against molecular targets. B) Aptamers evolved using LOOPER-SELEX including their binding and inhibition.

1.9 Outlook

While still in its early stages of development, LOOPER has demonstrated its ability to synthesize diversely modified oligonucleotide libraries that can be readily selected to generate aptamers against molecular targets. The diverse modifications have enabled the discovery of aptamers that break with traditional binding motifs and engage in non-canonical folding, allowing nucleic acids to bridge the functional divide toward protein-like properties. Notwithstanding, several challenges remain with LOOPER to facilitate its widespread use in the development of aptamers. The first challenge is that LOOPER currently does not explore full sequence coverage with either T3 or T4 DNA ligase-based systems. This is due to the ligase preferences for modifications on the anticodon. Expanding T4 DNA ligase modified anticodon sets beyond ANNNN and T3 DNA ligase modified anticodon sets beyond YNN is expected to improve modified aptamer evolution outcomes. Second, the synthesis of LOOPER derived aptamers involving modifications on the sugar phosphate backbone has not yet been explored, which further expand the capabilities of LOOPER to generate

aptamers with increased nuclease resistance. Solid-phase approaches that can generate LOOPER-derived aptamers without recourse to extensive custom phosphoramidite synthesis would significantly improve the utility of LOOPER-derived aptamers, while enabling deletion and minimizations, as well as structural studies by crystallography. Lastly, due to the longer process involved in LOOPER-SELEX compared with the traditional SELEX process, advances in high-efficiency selections[39] that are compatible with LOOPER-derived libraries are needed. Indeed, new selection technologies, such as ideal-filter capillary electrophoresis[40], demonstrates immense promise in achieving single-round aptamer selections.

Beyond these three major challenges, controlled evaluations are still needed that compare unmodified aptamer evolutions against LOOPER-derived aptamer evolutions. Such evaluations will provide a critical look at the scale of impact LOOPER-derived modifications provide to aptamer fitness during evolution, such as affinity and selectivity of binding. Furthermore, since LOOPER has the potential to generate nucleic acid polymers comprising other unnatural nucleoside derivatives, continued exploration into the modification preferences for ligases will instruct the design of new anticodon sets, as will results from selections against a diverse set of molecular targets, including proteins and small molecules. With the ability to incorporate a large repertoire of chemical functionality, selections of more complex functional nucleic acids, including those with enhanced selectivity or catalytic activity, are expected to yield interesting insights into the impact of expanded chemical diversity on the evolutionary fitness landscape of nucleic acid polymers.

Chapter 2 – Xeno-nucleic acids in LOOPER[§]

[§]Reproduced from ***Chem. Sci.***, 2025, Advance Article with permission from the Royal Society of Chemistry. Areeba Chaudhry and I conducted the experiments for yield analysis and duplex sequencing analysis. I conducted the experiments for hairpin sequencing, analysis, and reverse transcription LOOPER. Christopher Korfmann wrote the Python scripts for hairpin sequencing analysis and reverse transcription LOOPER sequencing analysis. Dr. Ryan Hili and I innovated the scientific concepts and experiments.

2.1 Xeno-nucleic acids

Nature has evolved to use highly specific backbones for its genetically encoded polymers. In proteins, amide-linked α-amino acid are employed, while RNA and DNA use phosphodiester linked ribose and deoxyribose sugars, respectively. These naturally occurring backbones enable efficient and specific root processing of the biopolymers of life, be it synthesis, replication, modification, or degradation.

Chemists have long explored the use of unnatural backbones to study and perturb the biology of living systems. For instance, β-peptides can form complex and predictable folding patterns, engage with protein targets, and exhibit profound resistance to proteolysis[41,42]. Inspired by their potential, efforts to engineer the protein translation machinery has enabled the translation, evolution, and ultimately the study and application of β-peptides[41,42]. Within the domain of genetic polymers, the exploration of alternative phospho-sugar backbones have been investigated for decades[41] with the ultimate goal to understand the etiology of DNA and RNA in living systems, possibly from a structurally related progenitor[42]. However, more recently, nucleic acid polymers with unnatural backbones have come to the fore as a critical advance in synthetic biology with considerable potential in various therapeutic applications. Such polymers, termed *xeno-nucleic acids* (XNA), are not recognized by endogenous nucleases, thus slowing, or precluding their degradation in

biological samples. Indeed, this property has prompted their emergence in oligonucleotides used in CRISPR guide RNA,[42] and various antisense oligonucleotide therapeutics[43].

Functional nucleic acids are conventionally discovered through a process called Systematic Evolution of Ligands by EXponential enrichment (SELEX), which involves iterative cycles of a selection pressure against a library of nucleic acids for a desired function (*e.g.*, binding a protein target), and amplifying the survivors. Recent advances in engineered polymerases have enabled the transcription and reverse transcription of several XNA forms with high fidelity – a critical step to enable the evolution of functional XNAs using traditional SELEX methods. High-affinity aptamers against protein targets have been isolated from selections involving α-L-threofuranosyl nucleic acid (TNA)[42–44], 2'-fluoro-arabinonucleic acid (FANA)[42–44], and 1'5'-anhydrohexitol nucleic acid (HNA)[42]. Despite the considerable advances in polymerase engineering, new conceptual approaches to enzymatically generate XNA are needed to fully explore their potential. In particular, general methods to transcribe and reverse transcribe broad classes of XNA, chimeric XNA, or XNAs bearing nucleobase modifications, are needed to expand the utility of this promising class of biopolymers for therapeutic, diagnostic, and other biotechnology applications. The ligase-catalyzed oligonucleotide polymerization (LOOPER) provides a platform for analogous transcription of DNA into synthetic biopolymers (**Figure 8**)[42–49].

LOOPER involves the nucleic acid-templated ligation of short oligonucleotide fragments, termed anticodons, in a sequence-defined manner and works with high fidelity when using large combinatorial libraries of templates. LOOPER has been shown to accommodate nucleobase-modified anticodons. Indeed, complex libraries can be generated with up to 16 different chemical modifications[18] or with peptide fragments up 8-residues in length[17]. LOOPER has also been implemented in SELEX to identify high-affinity aptamers against protein targets, including human

α-thrombin[32], PCSK9 and IL-6[22]. Due to the potential flexibility of LOOPER, we reasoned that DNA-templated XNA library synthesis and XNA-templated DNA library synthesis may be within the scope of the process, thus enabling a potentially generalized approach to analogous transcription and reverse transcription of XNAs that can be ported into SELEX or other applications including memory storage (**Figure 8**)[42].

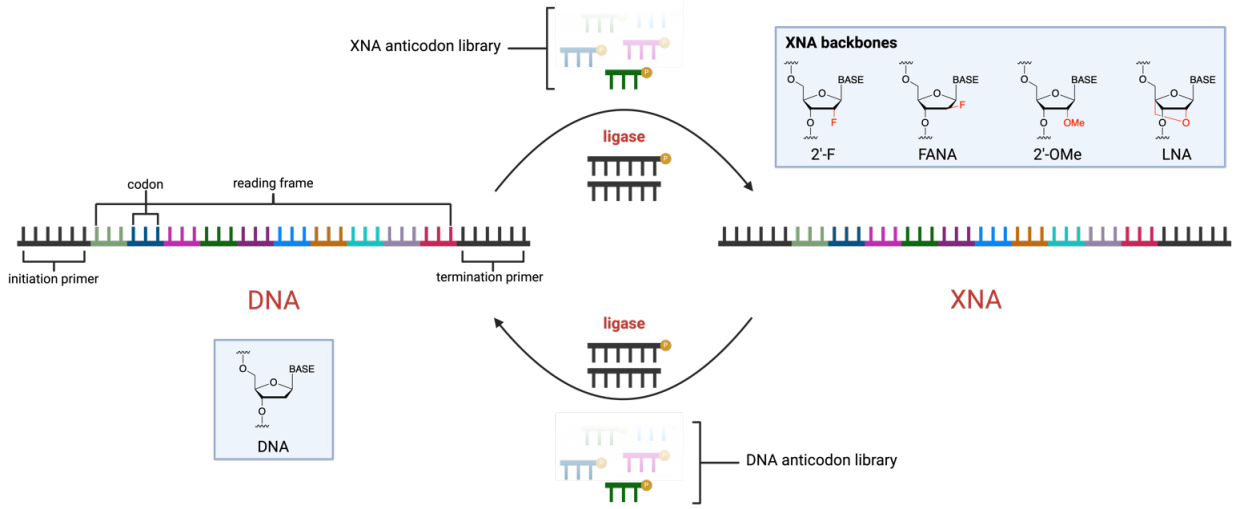


Figure 8: General schematic for the ligase-mediated "transcription" of DNA into XNA using a DNA ligase and 5'-phosphorylated XNA trinucleotides and the ligase-mediated "reverse transcription" of XNA into DNA using a DNA ligase and 5'-phosphorylated DNA trinucleotides.

The ligase used during LOOPER can have profound influence over the outcome of the process. T4 DNA ligase, which was the original enzyme used in LOOPER, can most efficiently polymerize anticodons that are at least pentanucleotides in length[42]. Importantly, the anticodon must be chemically modified to enable high fidelity. When polymerizing anticodons comprising the ANNNN sequence, whereby N = A, C, G, T, the incorporation of a *C*8 modification at the 5' adenine increases fidelity from 87% to 95%, presumably through an *anti* to *syn* conformational switch along the adenine glycosidic bond, resulting in a more stringent annealing process[20]. With fully degenerate pentanucleotide anticodons, NNNNN, which would be required for the proposed XNA system, fidelities are decreased to 81% which are too low for most applications. T3 DNA

ligase on the other hand has been shown to accommodate anticodons as short as three nucleotides in length[20,32]. Importantly, while fully degenerate unmodified anticodons such as NNNNN result in very low fidelity (67%), shorter anticodons, such as NNNN and NNN are incorporated with much high fidelities of 89% and 97%, respectively. For this reason, we pursued T3 as a potential ligase for XNA-based libraries in LOOPER.

2.2 Polymerization efficiency of XNA anticodons in LOOPER

Using degenerate trinucleotide libraries, we tested commonly used and commercially available XNA modifications including 2'-F, 2'-OMe, FANA, and LNA (Fig. 2). For each XNA type, four NNN anticodon libraries were synthesized: one with the XNA modification at every position within the anticodon, and then three corresponding libraries where the XNA modification was either in the 5', middle, or 3' position of the anticodon. This allowed us to evaluate the efficiency and positional preferences of polymerization, which has been an important parameter in earlier developments[42]. Each library was subjected to LOOPER along a templated comprising a 39-nt (13 codon) reading frame using T3 DNA ligase and the yield was evaluated by polyacrylamide electrophoresis (PAGE) and densitometry (**Figure 9**). While both 2'-OMe and FANA at any position failed to yield full-length products, T3 DNA ligase was able to generate full-length product using any of the four 2'-F libraries. Surprisingly, LNA was also well-tolerated by T3 DNA ligase, specifically when the modification was on the middle ribose sugar of the trinucleotide. For both LNA and 2'-F, the greatest yields were found with modifications at the middle position, suggesting that the modifications may be interfering with the enzymatic synthesis of the phosphodiester linkages between the nucleotides at each terminus.

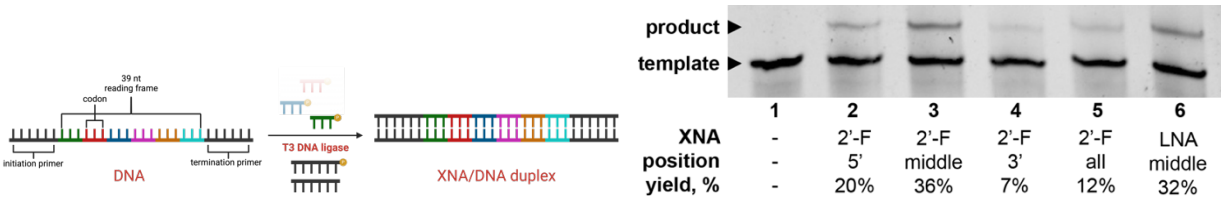


Figure 9: LOOPER of XNAs 2'F and LNA. Denaturing PAGE analysis of LOOPER process containing a DNA template comprising a 39-nucleotide random region and degenerate trinucleotides (NNN) bearing XNA modifications at noted trinucleotide positions. The product strand is PEGylated, making the product migrate slower than the template.

Although there was no full-length product observed when using 2'-OMe or FANA-modified trinucleotides in LOOPER, partially polymerized products were detected depending on the position of the modification on the anticodon (**Figure S1&S2)**; further optimization may enable these classes of XNA. 2'-OMe modifications at the middle position of the trinucleotide were tolerated the best, incorporating up to four trinucleotides. It is important to note the LOOPER with 2'-OMe-modified LOOPER on discrete trinucleotides, perform much better than in library contexts. For instance, TTT trinucleotides with 2'-OMe modifications at the middle position were well tolerated, resulting in full-length product (**Figure S2a**). While the reason for this have not been fully explored, our lab has previously observed that yield typically decreases as the LOOPER anticodon library size increases[20]. This may be due to the decrease in relative concentration of the cognate anticodon at each ligation site. Furthermore, certain anticodon sequences may have superior annealing and ligation kinetics with T3 DNA ligase, which creates challenges in library-based polymerizations. For FANA-modified trinucleotide libraries, the 3'-modified library was the most tolerated by T3 DNA ligase resulting in three trinucleotide incorporations (**Figure S1b)**. The ligation efficiency marginally increased as the modification approached the 3' position of the template sequence. This is consistent with what was observed with FANA-modified TTT sequences, as the modification at the 3' position of the TTT sequence had the most efficient polymerization (**Figure S1).**

2.3 Fidelity analysis *via* duplex sequencing

The duplex sequencing method has been previously developed by our lab to identify errors in trinucleotide/pentanucleotide incorporations in a high-throughput manner (**Figure 10**). The preparation begins with a LOOPER template that contains an extra T nucleotide at the 3' end. Primers anneal to the primer binding sites LOOPER is initiated[18]. The LOOPER products are then ligated to A-tailing adapter duplex with a 14N coding region[20]. This region enables the identification of each duplex product during sequencing. The adapter-ligated LOOPER products are then subjected to PCR followed by Illumina sequencing. The products were then submitted to Sickkids TCAG Sequencing Facility for their Illumina MiSeq.

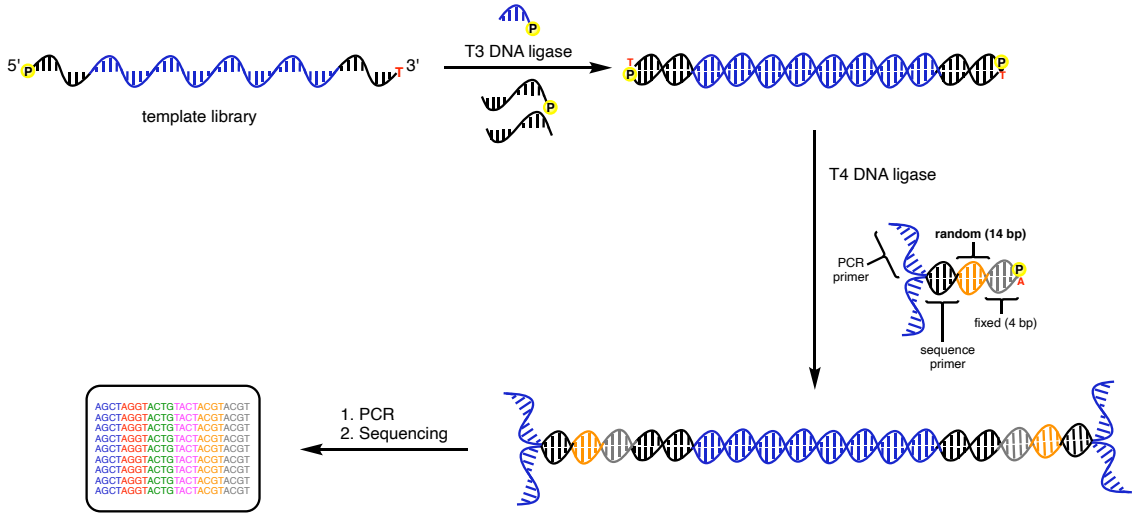


Figure 10: General schematic of the duplex sequencing preparation.

After performing LOOPER using each library, two adapters are ligated onto either end of the duplex, where each barcode pairing would be unique to each duplex. The ligation was carried out using T4 DNA ligase and 200 equivalents of the adapter duplex. The product was then imaged and extracted using 5% native PAGE. Unfortunately, the gel purified did not appear to be the same as

what was seen from previously in our lab (Figure 10). It was not clear as to why there was a reproducibility issue. Therefore, it was decided to run a control and see which band was the product band in our scenario. To do this, the two highest bands on the gel were collected together since they were in close proximity, faint, and not clearly distinguishable from one another. We initially hypothesized that the secondary duplex product generated two bands that were not well resolved due to secondary structures (**Figure 11a**, Band A). To confirm which band contains the product, the two bands were extracted together, as well as other prominent bands, and visualized the products using denaturing PAGE. Due to this, it was determined that the correct product resided in the top two bands (Band A). The product was then PCR amplified and submitted for Illumina sequencing.

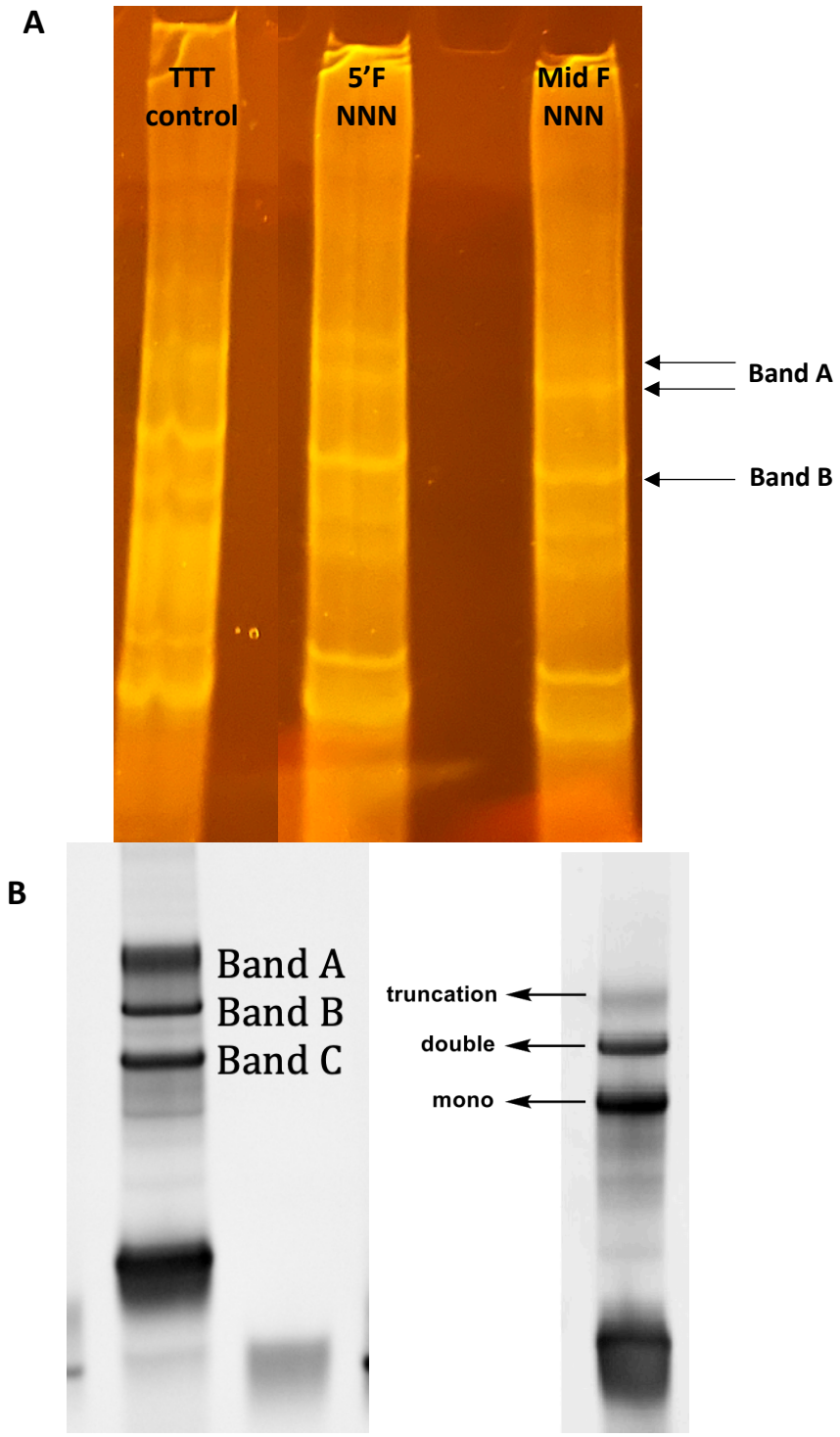


Figure 11: Native PAGE (5%, 150 V, 45 min) gel after adapter ligation. A) Three out of the five samples submitted for sequencing. The control was done to determine the product band. B) Samples done in the past by previous graduate students in the lab. These previous gels were used for comparison.

Unfortunately, the results are not what was expected. Many reads of the correct length and expected sequence were observed, however, very limited number of pairs were observed (**Table 2**). After the run, there was an attempt to look into our data analysis to determine if there were any bugs within the script. There was also an attempt to manually find pairs within the reads and not rely on the script to process the data first. We were unable to determine the issue from this angle. Then, I was able to rule out that there was not an issue in the preparation of the adapters. I visualized the sequences added to the reaction before and after Klenow extension, where it is confirmed that the extension is occurring. It was then concluded that the reason for the abundance of appropriate reads but modest pairings suggested that it was the number of diverse duplexes in each sample that was used before PCR. In the first run, 50 attomoles of each sample was PCR amplified. The abundance of sequences results in the reduced probability to identify matched pairs.

	Round 1			Round 2		
	Matched Reads	Codon Occurrence	Fidelity	Matched Reads	Codon Occurrence	Fidelity
R8-L1	15	195	97.9%	4	52	92.3%
5'F-NNN	11	143	94.4%	5	65	98.5%
middle F-NNN	40	520	78.1%	25	325	79.7%
3'F-NNN	1	13	92.3%	3	39	97.4%
all F-NNN	1	13	92.3%	1	13	100%

Table 2: Matched pairs and fidelity of each codon within each read from both sequencing attempts.

It was decided to re-attempt sequencing, keeping everything constant except for the amount of duplex product used for PCR. Instead of using 50 attomoles of product, we used 100-fold less (0.5 attomoles). The results were still unsuccessful, signifying that the issue lies elsewhere (**Table 2**). The inconsistency in our gel purification was further investigated. I first started by attempting to extract the highest two suspected product bands (Band A) separately to able to visualize them

using denaturing PAGE. There was not a visible difference between the two bands, indicating that there is not sufficient separation (**Figure 12**). In order to distinguish the products within both bands, a new method of purification needed to be optimized. Performing two purifications instead of one allowed for better desalting and higher band resolution (**Figure 13**, left gel).

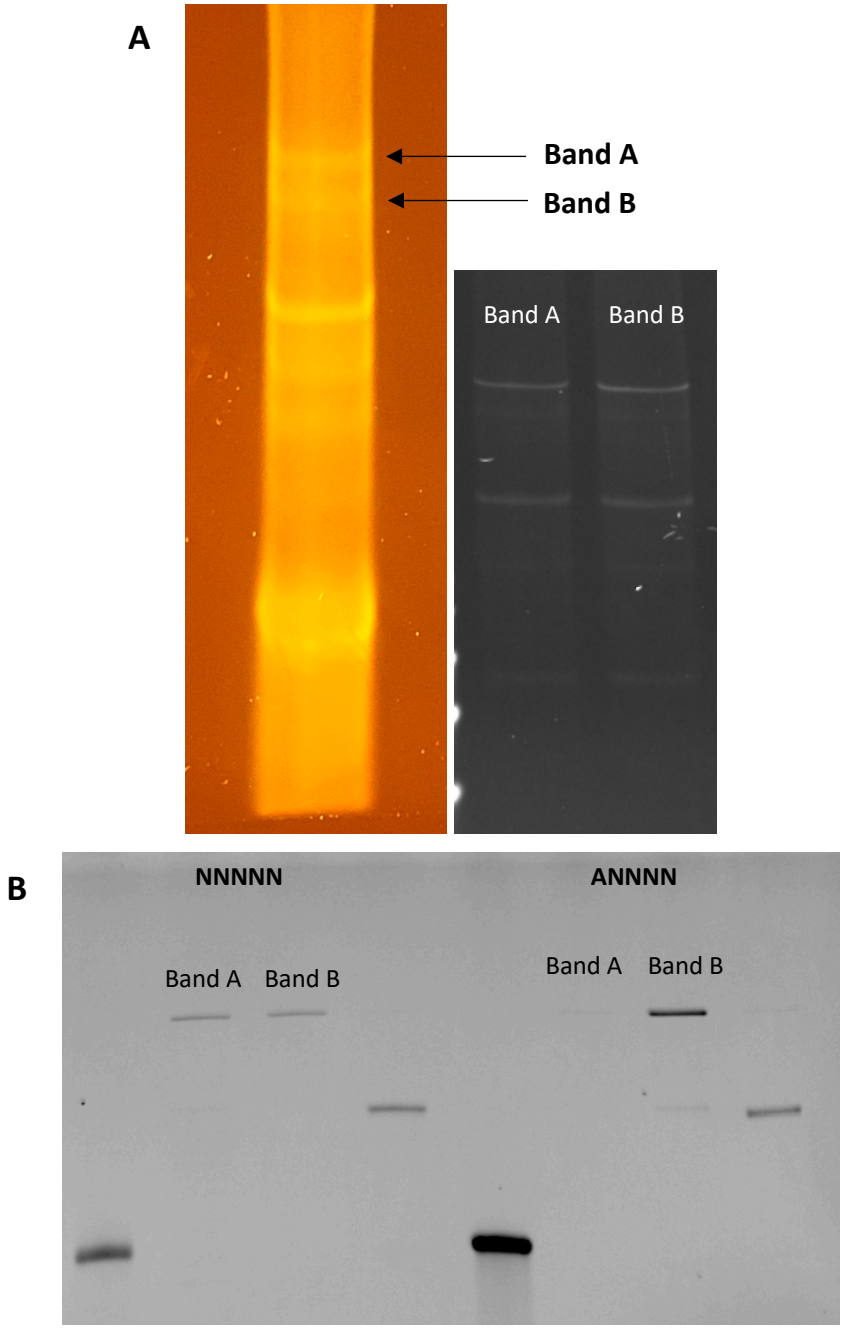


Figure 12: A) Gel extraction of Band A and Band B separately after adapter ligation. The bands were purified and subjected to 10% denaturing PAGE. B) Denaturing PAGE confirmation done by a previous graduate student (Dr. Yi Lei).

We also wanted to attempt using purchased acrylamide gels instead of casting them manually (**Figure 13**). The results are now very similar to what our lab has seen previously. I proceeded to

extract the prominent bands and confirm the product by denaturing PAGE. This confirmed that the bands are distinguishable from one another. Moreover, this result is also consistent with results seen previously by our lab (**Figure 12b**).

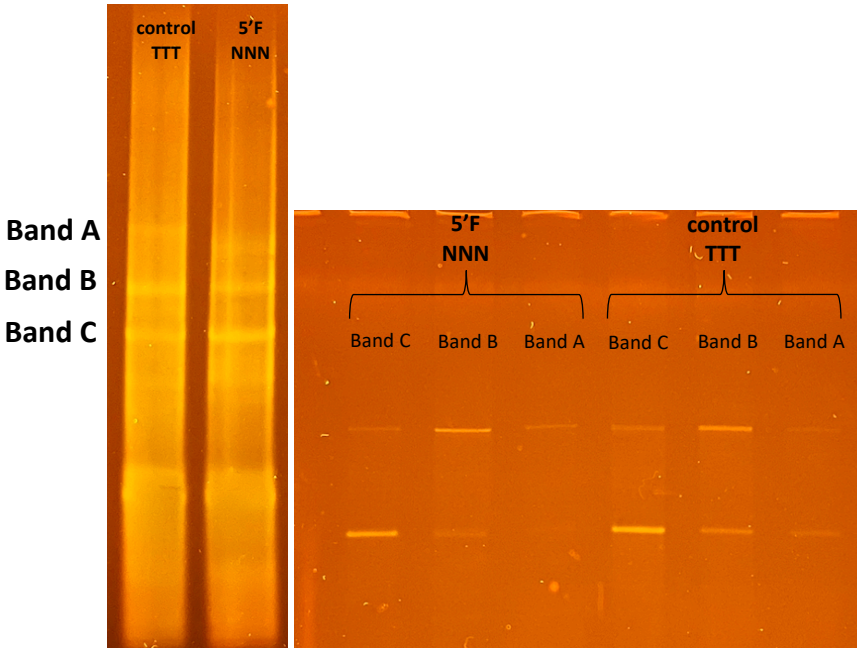


Figure 13: Native PAGE of post adapter ligation using the new purification method and gels purchased. Denaturing PAGE was then performed to confirm the products.

With the purification issue resolved, we wanted to investigate the PCR amplification process afterwards. In the past, Q5 DNA polymerase had been used due to its ability to polymerize with a high fidelity. It was thought that previous runs may have had an excess of unmodified DNA template, where Q5 DNA polymerase had a preference in amplifying the template.

Duplex sequencing has proven to be successful in previous years, however, the method has never been used to prepare libraries with densely modified nucleic acid polymers with bulky groups. Even though the purification issue was resolved, moving forward with another sequencing run was not economically feasible, since there were still suspicions about the PCR amplification efficiency of modified nucleotides.

2.4 Fidelity analysis *via* hairpin sequencing

With a degenerate trinucleotide anticodon set, there are $4^3 = 64$ trinucleotide sequences used for polymerization during LOOPER. This gives rise to a combinatorially challenging system, where errors occur due to the ligation of a misannealed anticodon across from a codon within the reading frame. Examining the fidelity of polymerases is more straightforward as they typically use just four monomers during transcription and reverse transcription, enabling the use of sequencing methods in a non-library format. For LOOPER, in order to obtain sequencing coverage of each anticodon in different sequence contexts, and to establish codon incorporation bias for each anticodon, we previously developed a duplex sequencing method to allow for the direct comparison of template and polymerized product in a library format using high-throughput sequencing[21]. Since LNA was not amplifiable using commercial polymerases, we focused on 2'-F XNA as a model system. Unfortunately, while duplex sequencing could accurately assess the fidelities of nucleobase-modified DNA anticodons[18], the method failed to yield the sequencing coverage needed for accurate fidelity characterizations with 2-'F XNA. When analyzing duplex sequencing data, 28 nt barcodes on the sense and antisense strands are paired together during the fidelity analysis. We found that very few pairs could be matched from the sequencing data. We attempted reducing the amount of input into the sequencing runs to maximize the probability of finding pairs, but this did not result in significant improvements. Q5U polymerase was then used instead of Q5 since this polymerase is reported to better tolerate uridine bases present on the 2'--F and 2'OMe modified nucleotides[42]. Unfortunately, Q5U polymerase did not improve matching rates. Encouraged by the high fidelities from the limited number of matched sequences (**Table 2**), we sought to redevelop the duplex sequencing into a more robust method for XNA.

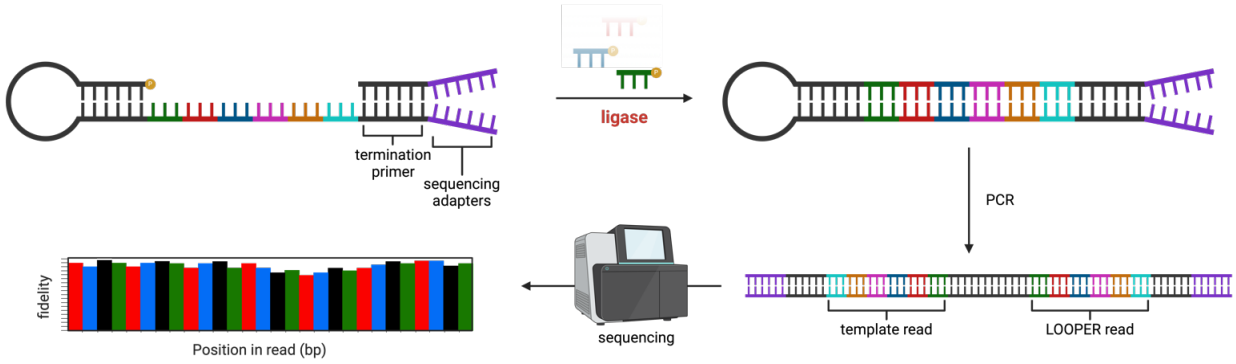


Figure 14: General pipeline for hairpin sequencing (HP-seq) to assess LOOPER fidelity with poorly amplifying XNA templates. Hairpin library template (reading frame = N39) containing primer-binding site and sequencing adapters is subjected to LOOPER. PCR amplification of the product results in an amplicon comprising the original template read and the LOOPER strand read. High-throughput sequencing and post sequencing analysis is used to compare the template and LOOPER reads to identify errors in LOOPER within a library context.

We hypothesized that the lack of pairing was resulting from the disparity of PCR efficiency between the template (DNA) and products (XNA) strands. Even with more permissive polymerases, it was difficult to control this bias. To overcome this issue, we chose to evaluate the fidelity using a hairpin architecture (**Figure 14**). Using this approach, full-length amplicons that go into sequencing would necessarily have read through both the template and product strand, allowing for straightforward pairing for fidelity analysis. An additional benefit to this method is that it does not require barcoding of the duplex product which allows for maximum diversity during sequencing.

LOOPER was performed along a hairpin template library and PCR amplified before high-throughput sequencing. An advantage to purifying the hairpin product is that the fully extended product is well resolved during PAGE from the partial products, allowing for efficient purification. The product was validated by attaching a fluorophore to the 5'-end of the primer. This allowed for confirmation of the full-length product band for extraction. The correct amplicon length prior to sequencing further validated the fully extended product. All modified hairpin products

demonstrated identical PAGE gels before purification, except for the completely modified 2'-F product. Although the modification was successful during LOOPER, PCR amplification using Q5 DNA polymerase did not result in any product. This could be due to the highly structured hairpin in conjunction with the densely modified strand that makes the polymerization difficult. BST 3.0 polymerase, a polymerase with a larger, more accommodating active site, did not improve PCR efficiency. KOD DNA polymerase had a faint product band that was not well resolved; however, sequencing confirmed that this band was not the product. Therefore, we were unable to determine the fidelity of the completely modified 2'F product using this current approach. In addition, the LNA product could not be PCR amplified by Q5 DNA polymerase, so we were not able to perform sequencing on these products.

Anticodon	*Fidelity(N)*[b]	*Reads*[c]	*Fidelity*[d]	*Norm. Fidelity*[e]
NNN	94.5%	3790	87.9%	95.9%
N*NN*	95.5%	8500	86.3%	94.0%
*N***N***N*	92.6%	226298	77.3%	84.3%
*NN***N**	95.5%	31357	84.3%	92.0%
NNN	N/A	N/A	N/A	N/A
Q5 control[f]	97%	5117	91.7%	100%

Table 3: Fidelities of LOOPER using trinucleotide 2'F-XNA. [a]Fidelities were determined by Ion Torrent sequencing on an Ion GeneStudio S5 Plus. All LOOPER experiments were conducted on a template comprising a 13 codon (39 mer) reading frame. [b]Fidelity was calculated at the single nucleotide level. [c]Number of observed codon reads. [d]Fidelity at the trinucleotide codon level. [e]Fidelity at the trinucleotide level normalized to Q5 standard, which was benchmarked at 100%. [f]Q5 DNA polymerase control.

After fidelity sequencing of the hairpin products, we noticed that the sequencer had difficulties sequencing through the hairpin structure. We determined this using a control hairpin that was extended using Q5 DNA polymerase and sequenced alongside the modified products. This resulted in a single nucleotide fidelity of 97.0% and a trinucleotide fidelity of 91.7%; we normalized the

modified fidelities to this Q5 DNA control (**Table 3**). Analysis of the LOOPER control (unmodified NNN) resulted in a 95.9% normalized fidelity, which is comparable to the fidelity obtained from duplex sequencing (97.0%)[20]. The data shows that the middle-modified 2-'F trinucleotide system, while having the highest polymerization yield, resulted in the lowest fidelity, which is consistent with our preliminary data using duplex sequencing (**Table 2**). This may indicate that this anticodon set has a stronger polymerization bias for subsets of anticodons. While the single nucleotide fidelities using anticodons with the 2'-F at either the 3'-end or the 5'-end were equivalent, at the trinucleotide level, the trinucleotide library with the 2'-F at the 5'-end resulted in the highest fidelities. Overall, these levels of fidelities are consistent with those used for LOOPER-based aptamer selections[27], which highlights the utility of this method.

We sought to explore trends in the sequencing data to better understand the strengths and shortcomings of the XNA library synthesis method (**Figure 15**). We first evaluated the effect of GC content in each trinucleotide library. We observed in the unmodified DNA control that increasing GC content resulted in decreasing fidelity; this can be rationalized by higher GC-content facilitating the ligation of stable misannealed anticodons (**Figure 15a**). However, with the 2'-F modified libraries the effect was surprisingly reversed. Upon further analysis, we noticed that there are a select few trinucleotides with 0% GC-content in each of the 2'-F libraries that have unusually poor fidelity. For example, the average fidelity of all 0% GC-containing anticodons in the 5'-modified 2'-F library is just 77.3%. However, when the ATT anticodon is removed from the average, the 0%-GC fidelity increases to 86.0%. Indeed, analysis of sub-libraries in each trinucleotide library (**Figure 15b**) show that the A and T sub-libraries have significantly lower fidelities at all positions within the trinucleotide anticodon. It is known that the incorporation of 2'-F modifications have considerable effects on DNA thermodynamics, including the increase in

thermal melting[42], and changes in conformational equilibrium[42], both of which may be at play here.

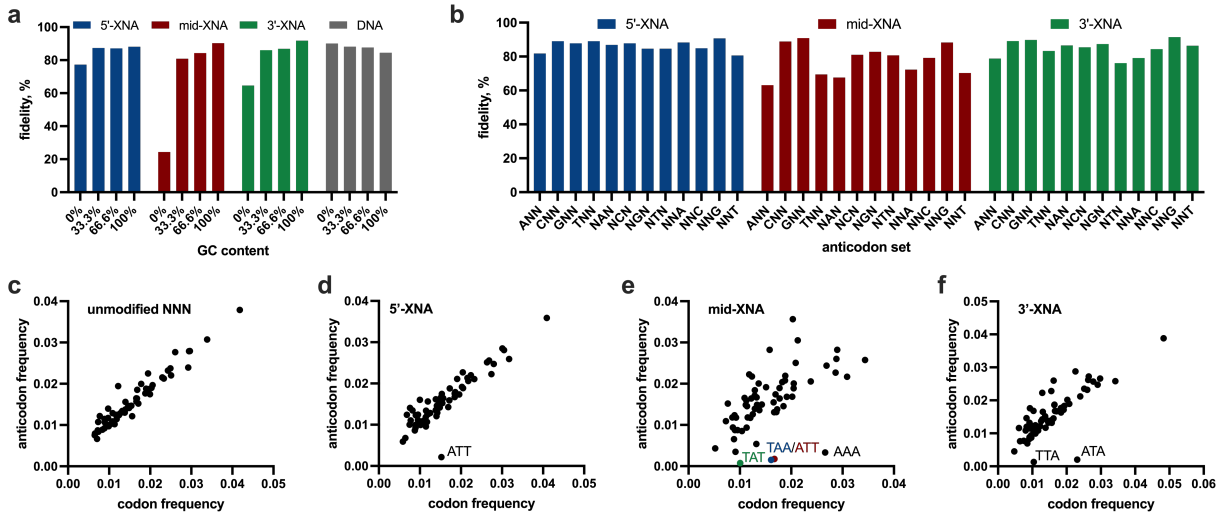


Figure 15: Analysis of codon sets used during LOOPER of 2'-F XNA trinucleotide anticodons. a) GC-content dependence on LOOPER fidelity for 2'-F XNA at various positions within the NNN anticodon; b) LOOPER fidelity of anticodon sub-libraries for 2'-F XNA at various positions within the NNN anticodon; c) Codon bias observed for LOOPER involving the control unmodified NNN anticodon; d) Codon bias observed for LOOPER involving the NNN anticodon with the 2'-F XNA at the 5'-position; e) Codon bias observed for LOOPER involving the NNN anticodon with the 2'-F XNA at the middle position; f) Codon bias observed for LOOPER involving the NNN anticodon with the 2'-F XNA at the 3'-position. The anticodon sequence of significant outliers are denoted in each graph.

While fidelity plays a large role in oligonucleotide polymerizations, so does polymerization bias. Biases can occur when specific anticodons are incorporated into the product strand at higher frequencies than the starting template. This can happen in low fidelity systems but can also happen in high fidelity systems when polymerization yields are lower. High bias can convolute analysis in downstream applications, such as *in vitro* selections of functional nucleic acids. Here, bias analysis has revealed several trends (**Figure 15c-f**). While the unmodified DNA control library displayed little bias (as indicated by data distributed along the diagonal), the XNA-based systems exhibit some outliers (**Figure 15d-f**, noted in plots). Amongst the XNA libraries, having the 2'F-XNA at the 5'-end of the anticodon resulted in the lowest level of library bias, with just one

significant outlier, namely ATT, which had a considerable negative bias. Across all libraries we observed that trinucleotide sequences with a 5'-G base tend to have positive biases, where T3 DNA ligase prefers incorporating these trinucleotide sequences over others. This positive bias has in part resulted in several 0% GC sequences having low fidelities. For example, the ATT trinucleotide sequence found in the 5'-modified 2'-F library resulted in a fidelity of only 11.0%. Upon analysis, we observed that GTT has 4-fold more incorporations across from the TAA codon in the template compared to the cognate ATT anticodon.

2.5 LOOPER as a method for reverse transcription

The faithful conversion of XNA into DNA is an essential process in the development of XNA-based technologies. This is largely due to traditional amplification and sequencing methods requiring DNA as the genetic medium. While LOOPER has been extensively explored to generate nucleobase-modified oligonucleotides, the ability to use LOOPER to effectively reverse transcribe XNA would be of considerable value to the field.

During our fidelity analysis of XNA synthesis using LOOPER, we experienced difficulty with the reverse transcription of the completely modified 2'-F products and LNA products, and thus we explored the reverse transcription ability of LOOPER within this context. In this method, polymerization occurs using an unmodified trinucleotide DNA library along a defined XNA-modified template (**Figure 16**). We attempted an LNA-modified template modified at every third base, preserving the modification structure of the successful middle-modified LNA transcription product; in our hands this template could not be PCR amplified by Q5 DNA polymerase (**Figure S6**). LOOPER was able to successfully reverse transcribe the LNA-modified template into DNA (**Figure 16a**, lane 6). A considerable amount of duplex remains intact during the denaturing PAGE

conditions, likely due to the known increase in thermal melting stability of LNA-containing duplexes[42]. We also tested completely modified XNA templates including 2'-F, 2'-OMe, and FANA, the latter two of which could not be directly amplified by Q5 (**Figure S6**), were all found to be within the scope of the LOOPER method. Similar to LNA, they also exhibited varying levels of duplex:single strand ratios on denaturing PAGE, which can be attributed to their known increases in thermal stability relative to canonical DNA[42,43,50]. Unfortunately, the completely-modified LNA synthesis could not be included in this study due to difficulties in its synthesis and lack of available commercial availability.

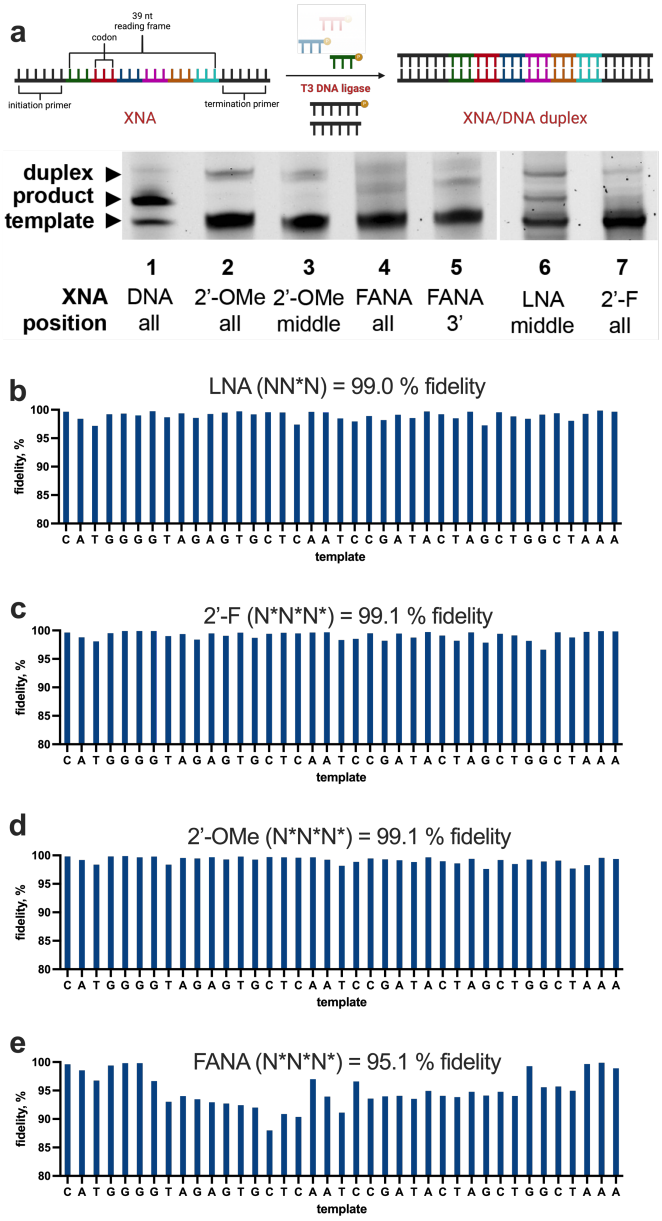


Figure 16: Evaluation of LOOPER of unmodified DNA anticodons along XNA-containing templates. a) Denaturing PAGE analysis of LOOPER process of degenerate DNA trinucleotides (NNN) along a corresponding XNA template comprising a discrete 39-nucleotide region. LOOPER fidelities at each position for templates containing b) LNA at middle position of codon, c) 2'-F at every position in codon, d) 2'-OMe at every position in codon, and e) FANA at every position in codon.

The fidelities of the LOOPER-mediated reverse transcriptions were purified and assessed by Illumina sequencing and were found to be between 95-99% average at the single-nucleotide level (**Figure 16b-e**). Interesting, FANA was exhibited considerably lower fidelity during the reverse

transcription, particularly within the middle of the reading frame. Further investigation into this is ongoing as it is unclear if this effect is dependent on sequence or length of the hybrid duplex. Nonetheless, these results provide a platform for the evolution of XNA-containing nucleic acid polymers.

T3 DNA ligase has been shown to enable LOOPER-mediated synthesis of XNA from DNA templates and DNA from XNA templates by using trinucleotide anticodon libraries. Both LNA and 2'-F libraries were successfully synthesised from corresponding DNA template libraries in moderate yields. A hairpin-based sequencing method was developed to calculate the fidelity and bias of trinucleotide incorporation, which were consistent with those used in LOOPER-mediated *in vitro* selections, highlighting the strength of this approach. While LOOPER with 2'-OMe and FANA trinucleotides were only partially polymerised, further investigation and optimisation may yield improved conditions to make these XNAs products tractable. Lastly, the reverse transcription of XNAs, including LNA, FANA, 2'-OMe, and 2'-F were achieved using LOOPER in moderate yields. A discrete template was used as a model to assess the fidelity of the method, which showed average fidelities over 95% for all XNA types, which is within the range of LOOPER methods that have been used to evolve aptamers. Taken together, we have shown that LOOPER is capable of the transliteration of various genetic polymers and can potentially serve as a mediator for their use in various biotechnology applications.

2.6 Sequences

<u>LOOPER Sequences</u>

Poly_pr1-PEG
/phos/TGCGACGGCAGGCGAATC/iSp18/AACAACAACAACAA

3P-F
/56-FAM/AGGATCCGAGCTCCACGTG

3P5P(AAA)13
GATTCGCCTGCCGTCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AACACGTGGAGCTCGGATTCT

3P5P(NNN)13
GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNCACGTGGAGCTCGGATTCT

<u>Duplex Sequencing</u>

PR5
/phos/GGATCCGAGCTCCACGTG

PR6
/phos/TGCGACGGCAGGCGAATCT

PRIMER B
AATGATACGGCGACCACCGAG

iTruS_i7_D701
CAAGCAGAAGACGGCATACGAGATATTACTCGGTGACTGGAGTTCAG

iTruS_i7_D702
CAAGCAGAAGACGGCATACGAGATTCCGGAGAGTGACTGGAGTTCAG

iTruS_i7_D703
CAAGCAGAAGACGGCATACGAGATCGCTCATTGTGACTGGAGTTCAG

iTruS_i7_D704
CAAGCAGAAGACGGCATACGAGATGAGATTCCGTGACTGGAGTTCAG

iTruS_i7_D705
CAAGCAGAAGACGGCATACGAGATATTCAGAAGTGACTGGAGTTCAG

AdapterA
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
T

AdapterB
/phos/ACTGNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

TP(NNN)13P
/phos/GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCCT

Hairpin Sequencing
5HP_39mer_IT1
/phos/TGCGACGGCAGGCGAATCCAATGATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCCGCGTGCGGACTCCAGCTA

3P_IT1
AGGCACGGCGAGCGATCTGGATCCGAGCTCCACGTG

3P-F_IT1 **used to confirm hairpin product – not for sequencing*
/56-FAM/AGGCACGGCGAGCGATCTGGATCCGAGCTCCACGTG

3HP_39polymerase
/phos/NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCC **CAAT**GGATCCGAGCTCCACGTG

HPprimerA
/phos/TGCGACGGCAGGCGAATCGCGTGCGGACTCCAGCTA

HPprimerArev
AGGCACGGCGAGCGATCTGATTCGCCTGCCGTCGCA

Fl-HPprimerArev **used to confirm hairpin product – not for sequencing*
/56-FAM/AGGCACGGCGAGCGATCTGATTCGCCTGCCGTCGCA

Hairpin Sequencing PCR Primers
IT1_3PRev_Seq
CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
TAGCTGGAGTCCGCACGC

IT1_IC_0130
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCCAAGCTGC GGTGAT
AGGCACGGCGAGCGATCT

IT1_IC_0131
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCTTACACAC GGTGAT
AGGCACGGCGAGCGATCT

IT1_IC_0132
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TTCTCATTGAAC GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0133
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCGCATCGTTC GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0134
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TAAGCCATTGTC GGTGAT AGGCACGGCGAGCGATCT

Reverse Transcription – XNA templates

RTLtemp_unmod
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

F-RTLtemp Green=2'F modified
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

LNA-RTLtemp Red=LNA modified
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

allOMe-RTLtemp Blue=2'OMe modified
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

midOMe-RTLtemp Blue=2'OMe modified
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

allFANA-RTLtemp Orange=FANA modified
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

3'FANA-RTLtemp Orange=FANA modified
TCCAAGCGTGCACGAGCAG CAT GGG GTA GAG TGC TCA ATC CGA TAC TAG CTG GCT AAA AGGCTCGCGACGGCATAG

RTL_3P
CTATGCCGTCGCGAGCCT

RTL_5P_peg
/phos/CTGCTCGTGCACGCTTGGA/iSp18/AACAACAACAACAA

RTL_AmpEZp1
ACACTCTTTCCCTACACGACGCTCTTCCGATCT
GGAGTCTGATCGATCGTCGAACGGTCGCCTTGACGTGGGCTAGAGCG
TCCAAGCGTGCACGAGCAG

RTL_AmpEZp3
GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
GGAGTCTGATCGATCGTCGAACGGTCG CTATGCCGTCGCGAGCCT

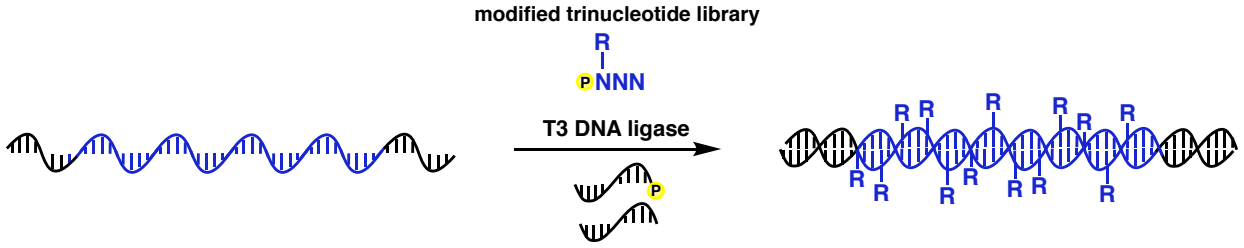
2.7 Methods

Synthesis of trinucleotide libraries and templates

Trinucleotides for LOOPER were synthesized on a ABI 394 DNA/RNA synthesizer using a DMT-ON protocol on a 1μmol scale (1000 Å CPG column). 2'F-modified phosphoramidites (Glen Research, 10-3430, 10-3420, 10-3415, 10-3400), LNA-modified phosphoramidites (Glen Research, 10-2030, 10-2000, 10-2029, 10-2011), 2'OMe-modified phosphoramidites (Glen Research, 10-3121, 10-3130, 10-3115, 10-3100), FANA-modified phosphoramidites (Glen Research, 10-3800, 10-3820, 10-3830, 10-3815), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (Glen Research 10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin according to the manufacturer's protocols. The cleaved resin was filtered off, and 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 [10 % acetonitrile in 0.1 M TEAA, pH 7] to [80 % acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45 °C. 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 and lyophilized. The oligonucleotide was incubated in 500 μl 30 % ammonium hydroxide at room temperature for 15 min 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 a [10 % acetonitrile in 0.1M TEAA, pH 7] to [80 % acetonitrile in 0.1M TEAA, pH 7] solvent gradient with a column temperature of 45 °C. The purified oligonucleotide was resuspended in H_2O for subsequent use in LOOPER.

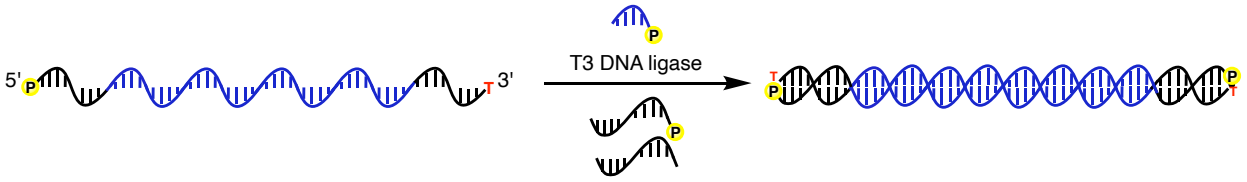
Ligase-catalyzed oligonucleotide polymerizations (LOOPER) and efficiency analysis



The ligation protocol begins with mixing 1 µL of 3P5P(NNN)13 (10 µM), 0.5 µL 3P-F (30 µM), 0.5 µL Polu_pr1-PEG (30 µM), 1.3 µL of the modified trinucleotide library (1 mM), and 5 µL of 2X T3 DNA ligase buffer (NEB M0317L) in a PCR tube. The mixture is then incubated at 95°C 10s, 65°C for 4 min, and cooled using a ramp from 65°C to 4°C at 0.1°C/s. After incubation, 1 µL T3 DNA ligase (3000 U/µL, NEB M0317L) and 0.7 µL of water were added. The reaction was then incubated in a thermal cycler at 4°C for 24 h. The product was then purified using the E.Z.N.A. Omega Biotek® purification kit (Omega Bio- tek, D6492-02) and eluted in 30 µL of water. The samples were denatured at 98°C for 2 min and loaded while still hot onto a 10% denaturing PAGE gel (55 °C, 150V, 40 min). The gels were imaged before and after ethidium bromide staining. The full-length product yield was quantified by gel densitometry against the template band after ethidium bromide staining.

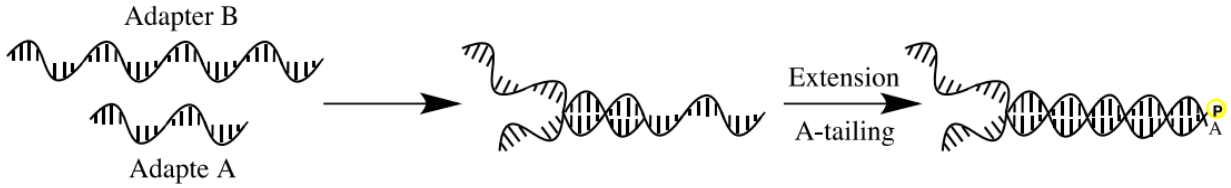
Duplex Sequencing

LOOPER for Duplex Sequencing



The LOOPER protocol begins with mixing 1 µL of TP(NNN)13P (10 µM), 0.5 µL PR5 (30 µM), 0.5 µL PR6 (30 µM), 1.3 µL of the modified trinucleotide library (1 mM), and 5 µL of 2X T3 DNA ligase buffer (NEB M0317L) in a PCR tube. The mixture is then incubated at 95°C 10s, 65°C for 4 min, and cooled using a ramp from 65°C to 4°C at 0.1°C/s. After incubation, 1 µL T3 DNA ligase (3000 U/µL, NEB M0317L) and 0.7 µL of water were added. The reaction was then incubated in a thermal cycler at 4°C for 24 h. The product was then purified using the E.Z.N.A. Omega Biotek® purification kit (Omega Bio- tek, D6492-02) and eluted in 30 µL of water.

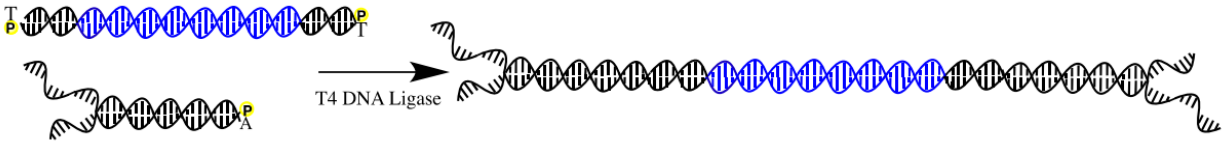
Adapter duplex synthesis and A-tailing



In a PCR tube, 15 µL of Adapter A (100 µM in H_2O) and 15 µL of Adapter B (100 µM in H_2O) were mixed. The mixture was incubated for 5 minutes at 95 °C, then cooled to room temperature over 1 hour. In the same PCR tube, 4 µL NEBuffer2 (10X, New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo-, New England Biolabs, M0212L), and 1 uL dNTP mix were added. The mixture was incubated at 37 °C for 1 hour and the adapter duplex was purified using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs, T1030S).

In a PCR tube, 30 µL of the purified adapter duplex, 5 µL of NEBuffer2 (10X, New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo-, New England Biolabs, M0212L), 5 µL dATP (10 mM, Thermo Scientific), and 5 µL H_2O were added. The mixture was incubated at 37 °C for 1 hour, and the A-tailed adapter duplex was purified using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs, T1030S) and eluted in 30 µL of water.

Adapter ligation



In a PCR tube, 10 pmol of LOOPER product, 200 pmol of the A-tailing adapter duplex, 10 µL NEBNext® Quick Ligation Reaction Buffer, 2.5 µL BSA (2 mg/mL in H_2O), 1000U T4 DNA ligase, and 21 µL of water were mixed and incubated at 16 °C for 16 hours. The ligated products were purified with the E.Z.N.A. Omega Biotek® purification kit (Omega Bio- tek, D6492-02), and native PAGE (5% native, 150 V, 45 minutes). The band containing the correct product was extracted and incubated overnight at 37 °C in 100 µL 0.3 M NaCl solution. The product was lastly purified using Centri-Sep columns (Princeton Separations, CS-901) and quantified by Qubit.

PCR and sequencing

In a PCR tube, 50 attomoles of the purified ligation product in 20 µL water, 2.5 µL Primer B (10 µM), 2.5 µL of the corresponding iTruS_i7_D7XX primer (10 µM), 25 µL Q5® High-Fidelity Master Mix (2X, New England Biolabs, M0492S) were transferred to a thermocycler where it was incubated for 10s 98 °C, 30s 55 °C, and then 30s 72 °C for the first two cycles, and then 10s 98

°C, 30s 71 °C, and then 72 °C for the rest of the cycles (20-22 cycles). The PCR products were purified with the E.Z.N.A. Omega Biotek® purification kit (Omega Bio- tek, D6492-02), and native PAGE (5% native, 150 V, 40 minutes). The band containing the correct product was extracted and incubated overnight at 37 °C in 100 µL 0.3 M NaCl solution. The products were lastly purified using Centri-Sep columns (Princeton Separations, CS-901) and quantified by Qubit.

Paired-end Illumina® sequencing was performed on an Illumina® MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing) at the TCAG Seuquencing Facility at the Hospital for Sick Children.

Hairpin Sequencing

Formation of the hairpin product

In a PCR tube, 1 µL 5HP_39mer_IT1 (10 µM), 0.5 µL nuclease-free water, 0.5 µL 3P_IT1 (30 µM), 1.3 µL trinucleotide (1 mM), and 5 µL StickTogether DNA Ligase Buffer (2X, NEB M0317L) were added. The mixture was then incubated at 98 °C for 10 s, 65 °C for 4 min, then a ramp from 65 °C to 4 °C at 0.1 °C/s. Lastly, 0.7 µL nuclease-free water and 1 µL T3 DNA ligase (3000 U/µL, NEB M0317L) were added to the mixture. The PCR tube was then incubated at 4 °C for 24 hours. The resulting mixture was then purified with Omega E.Z.N.A.® Cycle Pure Kit (Omega Bio- tek, D6492-02) and eluted in 30 µL nuclease-free water. A 15% denaturing PAGE gel (55 °C, 150 V, 80 min, denatured at 98 °C for 2 min and loaded hot) was then run to visualize and extract the product. The product was then purified using a Centri-Sep column (Princeton Separations, CS-901) and then quantified by Qubit and diluted to 10 pM before PCR.

Hairpin Sequencing – polymerase control

On ice and in a PCR tube, 1 µL 3HP_39polymerase (10 µM), 24 uL nuclease-free water, and 25 µL Q5 High-Fidelity Master Mix (2X, NEB M0492S) were added and transferred to a preheated thermocycler at 98 °C. The mixture was incubated at 98 °C for 10s, 71 °C for 30s, then 71°C for 5 min. The product was then purified with Omega E.Z.N.A.® Cycle Pure Kit (Omega Bio-tek, D6492-02) and eluted in 30 µL nuclease-free water. The sample was then dried and reconstituted in 11.5 µL of nuclease-free water.

In a PCR tube, 2 µL HPprimerA (100 uM in nuclease-free water), 2 µL HPprimerArev (100 µM in nuclease-free water). The primers were denatured at 95 °C for 5 min, then annealed at room temperature for 1 h. Then, 11.5 µL of polymerized sample, 2 µL T4 DNA ligase buffer (10X, NEB B0202S), and 2.5 µL T4 DNA ligase (400U/µL, NEB M0202S) were added to the primer mixture. The mixture is then incubated at 16 °C for 16h, then 65 °C for 10 min for enzyme deactivation). Lastly the product was purified with Omega E.Z.N.A.® Cycle Pure Kit (Omega Bio-tek, D6492-02) and eluted in 30 µL nuclease-free water.

Hairpin Sequencing – PCR and Ion Torrent sequencing

In a PCR tube and on ice, 10 µL corresponding hairpin product (10 pM), 2.5 µL IT1_3PRev_Seq (10 µM), 2.5 µL corresponding IT1_IC primer (10 µM), 10 µL nuclease-free water, and 25 µL Q5 High-Fidelity Master Mix (2X, NEB M0492S) or NEBNext® Q5U MasterMix (2X, NEB M0597S) (2'F modified products required Q5U polymerase) were added and transferred to a preheated thermocycler at 98 °C. The mixture was then incubated at 98 °C for 2 min, then remain incubating for PCR [98 °C for 10s, 55 °C for 30s, 72 °C for 30s] for the first 2 cycles, then [98 °C for 10s, 70 °C for 30s, 72 °C for 30s] for the rest of the cycles. The product was then purified with Omega E.Z.N.A.® Cycle Pure Kit (Omega Bio-tek, D6492-02) and eluted in 30 µL nuclease-free water

before gel extraction (15% native PAGE, 150 V, 100 min). All purified samples were subjected to Ion Torrent sequencing using the 530 chip and kit.

Hairpin Sequencing – script development

Sequences are read from a FASTQ file. Reads with lengths outside the specified tolerance (read length ±(N-1), where N is the modified oligonucleotide length) are rejected. For example, if the specified line length is 190 bases and N=3, reads <188 bases and >192 bases are rejected. This is to ensure that the reads represent the accurate full-length product that may or may not have indels. Reads are further filtered to ensure the Should we primer sequences are present verbatim.

For each accepted read, the LOOPER coding region and its corresponding template are extracted and global alignment is performed. Alignment is done using the 'PairwiseAligner' object in Biopython's 'Align' module ('mode="global"', 'scoring="blastn"'). The read is rejected if 'PairwiseAligner' is unable to align the two sequences. Insertions are removed from the LOOPER sequence.

Each sequence is then split into Nmer codons. Comparison is performed between each codon pair. Codon pairs are skipped if the template contains an indel. Unique global LOOPER codon and unique global template codon instance counts are tracked across all accepted and aligned reads. For each template codon, the Nmer error count (i.e. in a codon pair, the template codon does not equal the LOOPER codon) is tracked, as well as the mononucleotide error count (i.e. in a codon pair, the positional mismatch for each base between template codon and LOOPER codon).
From these metrics, the LOOPER codon bias is calculated for each codon by dividing the global LOOPER codon instance count by the respective template codon instance count. Template codon

fidelity is calculated by dividing the template codon error count by the global template codon instance count and subtracting the result from 1.

Reverse transcription LOOPER, PCR, and sequencing

The ligation protocol begins with mixing 1 µL of the modified DNA template (10 µM), 0.5 uL RTL_3P, 0.5 uL RTL_5P_peg, 1.3 µL of the unmodified trinucleotide library (1 mM), and 5 µL 2X T3 DNA ligase buffer (New England Biolabs, M0317L) in a PCR tube. The mixture is then incubated at 95°C 10s, 65 °C for 4 min, and cooled using a ramp from 65°C to 4°C at 0.1°C/s. After incubation, 0.7 µL nuclease-free water and 1 µL T3 DNA ligase (NewEngland Biolabs, M0317L) were added. The reaction was then incubated at 4°C for 24 h. The product was then purified using the E.Z.N.A. Omega Biotek® purification kit. The samples were denatured at 98°C for 2 min and loaded while still hot onto a 10% denaturing PAGE gel. The gels were imaged before and after ethidium bromide staining. The full-length product yield was gel extracted (10% denaturing PAGE, 55 °C, 150 V, 40 min).

In a PCR tube and on ice, 10 µL corresponding reverse transcription product (10 pM), 2.5 µL RTL_AmpEZp1 (10 µM), 2.5 µL RTL_AmpEZp3 (10 µM), 10 µL nuclease-free water, and 25 µL Q5 High-Fidelity Master Mix (2X, NEB M0492S) were added. The mixture was then incubated for PCR [98 °C for 10s, 55 °C for 30s, 72 °C for 30s] for the first 2 cycles, then [98 °C for 10s, 70 °C for 30s, 72 °C for 30s] for the rest of the cycles. The product was then purified with Omega E.Z.N.A.® Cycle Pure Kit (Omega Bio-tek, D6492-02) and eluted in 30 µL nuclease-free water. The sample was then concentrated to 25 µL and submitted for paired-end Illumina sequencing.

Chapter 3 – LOOPER involving base modifications

Aptamers are single-stranded nucleic acid polymers that exhibit many uses in therapeutics and diagnostics due to their binding and catalytic functions. In biological systems, proteins are responsible for these types of cellular functions due to their high functionality; whereas DNA and RNA polymers are used for information storage[51]. Proteins are superior to nucleic acids with respect to chemical functionality but lack in flexibility, thermal stability, and reproducibility[52]. Aptamers also have the advantage of having a much smaller size than antibodies, allowing for recognition of smaller binding domains that might be inaccessible to larger antibodies[52]. However, due to its smaller size it is more susceptible to renal filtration and nuclease digestion, chemical modifications would improve their pharmacokinetic properties[52]. Current methods for generating modified nucleic acid sequences involve the use of a polymerase and modified dNTPs[20,51]. However, only uniquely modified dNTPs have been able to be successfully incorporated[53].

The utilization of a ligase instead of a polymerase and oligonucleotides instead of dNTPs can allow for the generation of chemically-modified aptamers with a higher modification density and greater chemical diversity. Ligase-catalyzed oligonucleotide polymerization (LOOPER) is a method that allows for chemically-modified oligonucleotides to be sequence specifically added onto a template strand. The oligonucleotides are then ligated together to generate a modified DNA sequence library that can be used to select an aptamer against a target molecule. The template consists of a 40-nucleotide coding region that is flanked by 18-nucleotide primer binding sites. The template is mixed with a chemically-modified ANNNN pentanucleotide library where the appropriate sequences anneal to the coding region of the template strand. T4 DNA ligase then ligates the oligonucleotides together to generate the 76-mer product sequence. LOOPER generates DNA sequences that contain a modification at every fifth base, adding up to 8 unique modifications. Our

lab has previously been able to develop the first aptamer against human alpha-thrombin that does not contain a G-quadruplex, suggesting a novel method of binding. Unfortunately, chemical modifications are only tolerated by T4 DNA ligase when they are incorporated on an adenine base at the most 5'-position of the pentanucleotide sequence[26]. Liu and coworkers have introduced a LOOPER method by using T3 DNA ligase and modified trinucleotides[32]. T3 DNA ligase enables trinucleotide polymerization with a high yield and a high fidelity, making the combination the most ideal for LOOPER. However, the method is quite costly due to purchasing custom phosphoramidites[20]. Our method involves generating trinucleotide libraries that are made entirely from commercially available reagents.

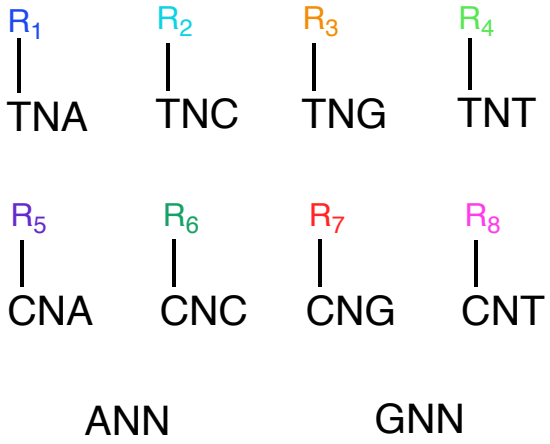


Figure 17: Trinucleotide codon library to be made with 8 possible chemical modifications (N = A, C, G, or T). Note that ANN and GNN will remain unmodified.

The T3 LOOPER method begins with a DNA template, consisting of a 39-base coding region flanked by 18-base and 19-base primer binding sites. The 39-base coding region allows for 13 trinucleotide incorporations to generate the modified LOOPER strand. The 5'-end of the modified sequence is attached to a fluorescein in order to confirm the presence of the LOOPER strand once the ligation is completed. After the full-length product is generated, the yield will be determined

by denaturing PAGE analysis. The band intensity of the product will be compared to that of the template band within the same lane. Typically, the LOOPER strand will appear as a band that is higher on the gel than the template.

All trinucleotide sequences were synthesized using an ABI 394 DNA synthesizer with commercially available phosphoramidites that were purchased. Amino-modified C and T bases were purchased for the bases that would eventually be modified. This allows chemical groups to be later attached to the base via amide coupling. The amino-modified linker can vary in length; however, we aimed to have the shortest possible linker lengths that were commercially available. The amino-modified phosphoramidites (bases) are shown below in Figure 18.

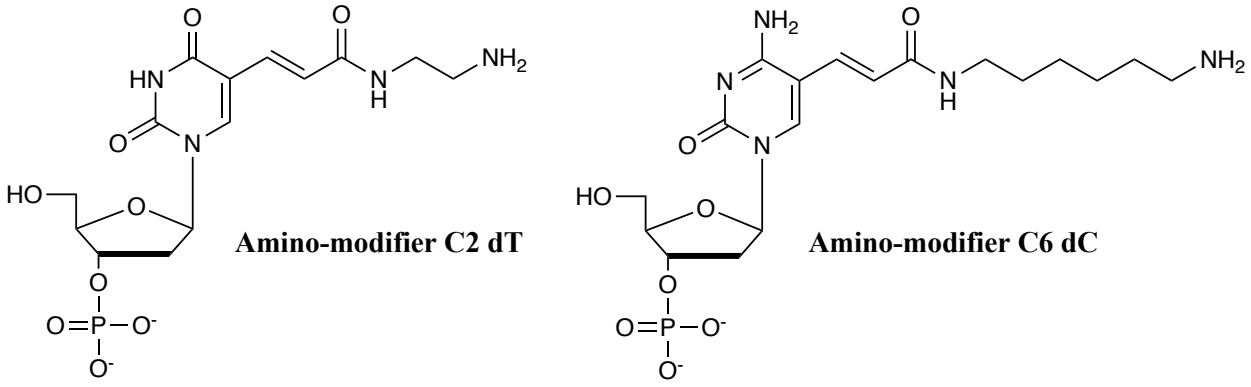


Figure 18: Amino-modified phosphoramidites purchased for the T and C sub-libraries.

An increase in the length of the modifier can result in a decrease in polymerization efficiency. Dr. Chun Guo has already attempted polymerization using amino-modifier C6 dT (the same modifier length as C, shown in **Figure 18**), which had already shown promising yields for the majority of the trimers. Since a shorter modifier was commercially available, this should help further improve the efficiency. The library was synthesized by evaluating various modifications on each individual sub-library then pooling together the best modified sub-libraries.

3.1 The T sub-library

3.1.1 TNT

The TNT sub-library containing the amino-modifier (amino-TNT) was shown to ligate with 100% efficiency. The efficiency of T3 DNA ligase to generate the fully polymerized LOOPER strand was confirmed using 10% denaturing PAGE (**Figure 19**).

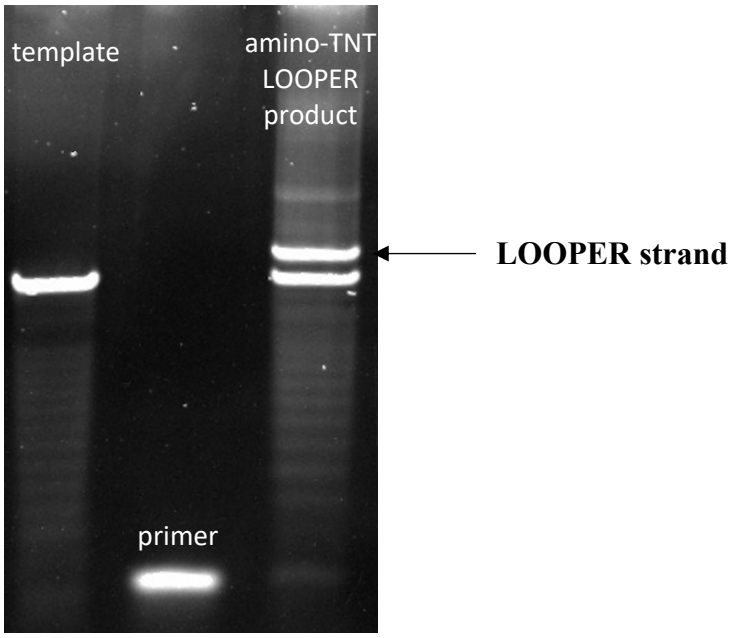


Figure 19: 10% denaturing PAGE gel (150V, 55 °C, 40 min) of the LOOPER product after 13 incorporations of amino-modified TNT trimers ligated to generate the full-length product.

All LOOPER yields were quantified by comparing the band intensities of the LOOPER strand to the template strand on the EtBr-stained gel. After proving that our optimized LOOPER method can enable T3 DNA ligase to function with high efficiency, a cyclopentyl modification was then attached to the amino-modifier to test the tolerability of the ligase. This was synthesized by amide coupling using 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide and N-hydroxysuccinamide (EDC/NHS). The product was purified by reverse-phase HPLC and confirmed by mass spectrometry. After subjecting cyclopentyl-TNT to LOOPER, the yield was found to remain at 100%. However, when a phenyl modification was used instead, the yield decreased to 58%. In the

final library, the cyclopentyl modification or the phenyl modification can be used, as they differ in binding properties. The modification that will be chosen will be determined before sequencing analysis.

3.1.2 TNG

The TNG sub-library was first tested only with the amino modifier attached, and the yield of that was found to be 100%. Therefore, a larger modification was tested. Phenyl-TNG was then synthesized and subjected to LOOPER. The yield was evaluated to also be 100% so an even larger modification was used. An indole group was used to modify TNG and this was also 100%. In the final library, indole-TNG was chosen.

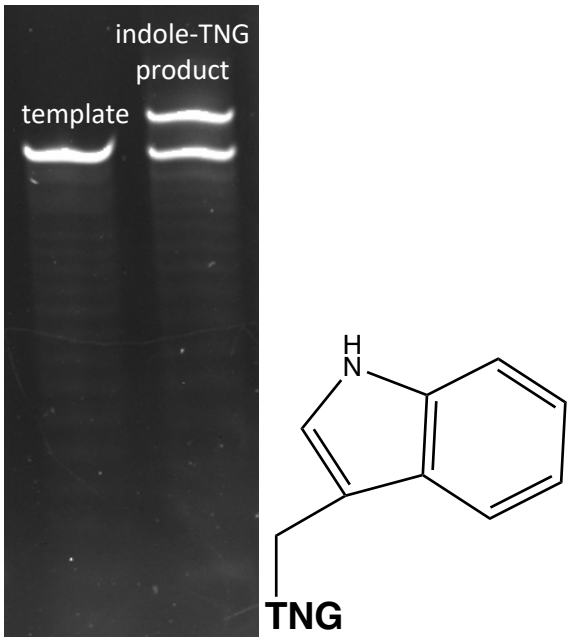


Figure 20: 10% denaturing PAGE gel (150V, 55 °C, 40 min) of the LOOPER product after 13 incorporations of indole-TNG trimers ligated to generate the full-length product. The image on the right is the general structure of indole-TNG.

3.1.3 TNC

First, TNC was tested with only the amino-modifier, which had a yield of 85%. Due to the high efficiency, a cyclopropyl modification was attached to the amino-modifier and attempted. The yield was found to be 61%, and so the cyclopropyl modification was chosen for the final library.

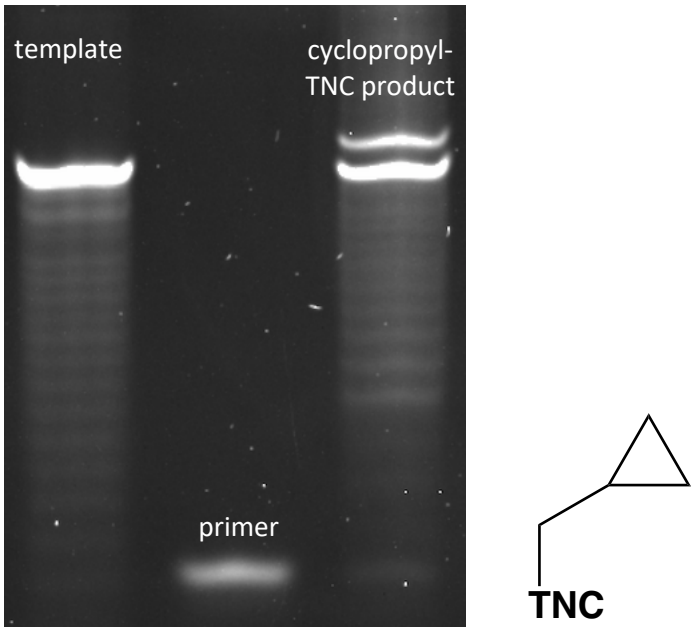


Figure 21: 10% denaturing PAGE gel (150V, 55 °C, 40 min) of the LOOPER product after 13 incorporations of cyclopropyl-TNC trimers ligated to generate the full-length product. The image on the right is the general structure of cyclopropyl-TNC.

3.1.4 TNA

Amino-TNA was functionalized using pentanoic acid to create a 'butyl' modification and evaluated by LOOPER and denaturing PAGE. The yield of butyl-TNA was found to be 10%. However, the yield of amino-TNA was found to be 21%, signifying that T3 DNA ligase is unable to tolerate the amino-modifier on the T base for this sub-library. Therefore, TNA would need to have an even shorter modifier or be left as the natural base. The natural T base was first attempted, and surprisingly, the yield appears to be below 50%. The exact value of the yield could not be quantified since when using natural bases, the bands of the LOOPER strand and the template strand overlap. The LOOPER strand was confirmed to be present by imaging the gel before ethidium bromide staining. It is unclear as to why T3 DNA ligase has trouble tolerating the natural base, as further research would need to be done. In the future, a carboxy-modifier will be attempted. A carboxylate group will introduce different electrostatic properties that could make TNA more tolerable to T3 DNA ligase.

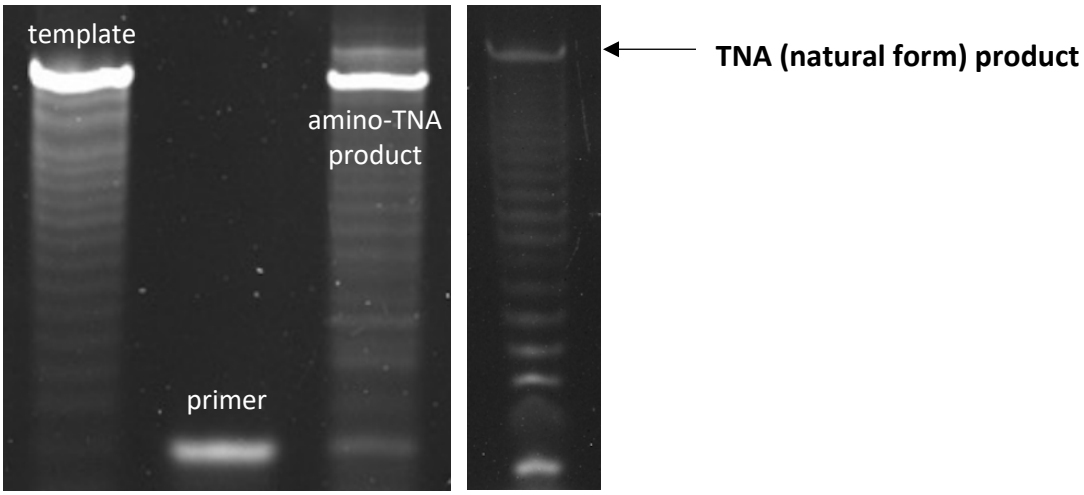


Figure 22: 10% denaturing PAGE gel (150V, 55 °C, 40 min) of the LOOPER product after 13 incorporations of amino-TNA trimers ligated to generate the full-length product. The image to the right is the gel showing the product using TNA with natural bases (before EtBr staining).

3.1.5 LOOPER optimizations on unmodified TNA

It was unexpected that T3 DNA ligase is not able to ligate the unmodified TNA sub-library efficiently. In an attempt to determine the interaction between TNA and T3 DNA ligase, TNA (using natural, unmodified bases) was subjected to LOOPER using T3 DNA ligase in combination with T4 DNA ligase. This would be able to help us understand if T3 DNA ligase was struggling with either the adenylation of the trimer, or catalyzing the formation of the phosphodiester bond. Our hopes were that T4 DNA ligase will be able to assist T3 ligase in the step that was less efficient. After performing this experiment, a denaturing PAGE gel was done. The results show no observable change in comparison to LOOPER using only T3. We can therefore hypothesize that T3 DNA ligase is specific to the base located at the 3'-position of the trinucleotide sequence since TNT, TNG, and TNC are able to be tolerated by T3 DNA ligase very efficiently.

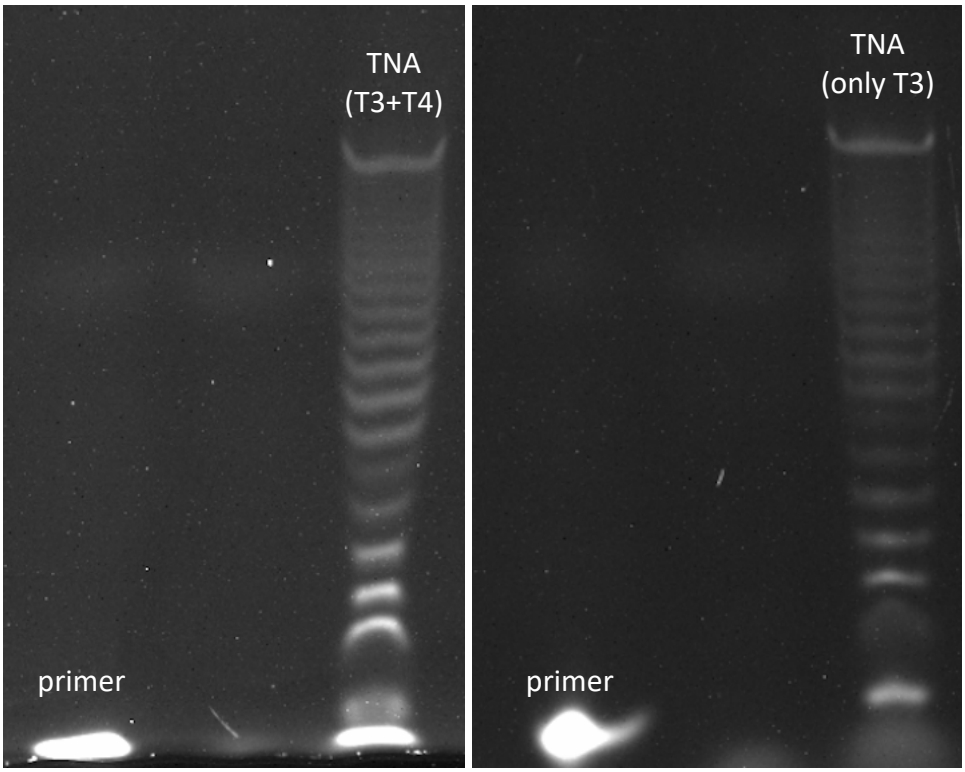


Figure 23: 10% Denaturing PAGE (150 V, 55 °C, 40 min) before ethidium bromide staining. In the T3+T4 method, 1 μL of each enzyme was added (300,000 U of T3 DNA ligase, 400 U of T4 DNA ligase) during LOOPER.

3.2 ATP optimizations

We hypothesized that decreasing the ATP concentration would help improve the performance of T3 DNA ligase, since T4 DNA ligase has been shown to stall the polymerization with excess ATP. In order to perform this experiment, a buffer was made that was identical to the buffer purchased from New England Biolabs, except without ATP. I then added ATP separately to create final concentrations of 500 μM, 100 μM, 50 μM, and 25 μM in the reaction. The 1 mM ATP control was done by using the buffer from the manufacturer (New England Biolabs, NEB). This was tested on amino-modified TNA, as the yield of this library is in the most need of improvement. The results show that the yield decreased when the ATP concentration fell below 1 mM. As a control, the buffer that was made was compared to the NEB buffer (both having an ATP concentration of 1 mM). The buffer made by us in the lab had a slightly lower yield. The reason for this is unknown.

Concentrations greater than 1 mM were attempted next. For this experiment, I used the NEB buffer containing 1 mM ATP, and supplemented with additional ATP.

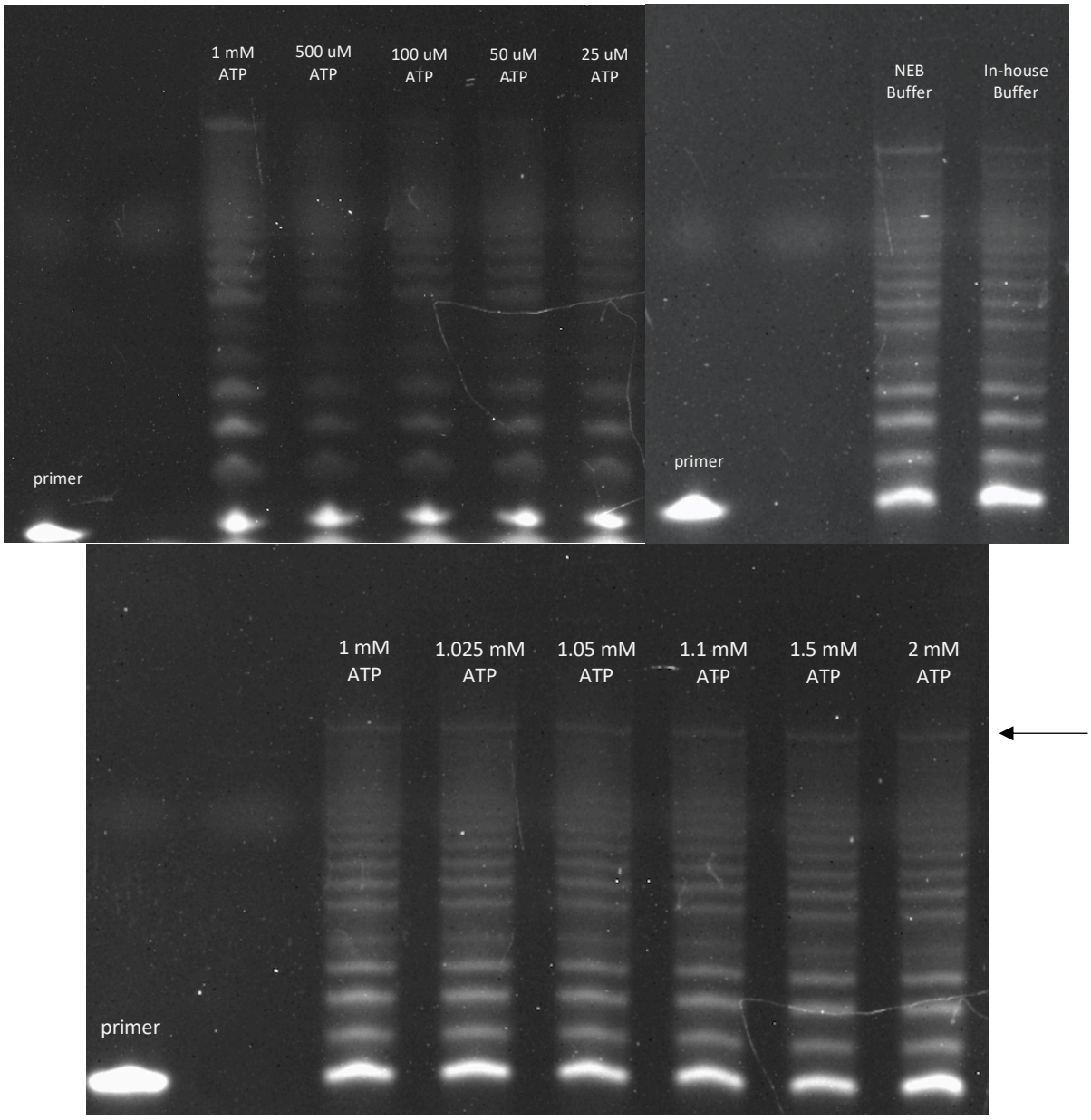


Figure 24: 10% Denaturing PAGE (150 V, 55 °C, 40 min) before ethidium bromide staining. The amino-modified TNA LOOPER products after using various concentrations of ATP. The buffer that was made in the lab was also compared against the buffer purchased from New England Biolabs (also compared using LOOPER with amino-TNA).

As the concentration of ATP increased, the yield remained the same. We can therefore conclude that changing the ATP concentration does not increase the performance of T3 DNA ligase when ligating the TNA sub-library.

3.3 The C sub-library

3.3.1 CNT

The C sub-libraries were functionalized by using amino-modifier C6 dC (purchased from Glen Research). Amino CNT was first tested and resulted in a yield of 86%. It was thought that functionalizing CNT with a relatively small chemical modification will slightly decrease the yield, but still result in sufficient ligation efficiency (>50%). Butanoic acid was synthesized to amino-CNT using EDC/NHS amide coupling to create a "propyl" modification. After LOOPER, the efficiency to generate the ligated product was found to be 19%. Therefore, an addition of 4 carbons significantly dropped the yield, which means that the modification would need to be even smaller. Smaller modifications may be accommodated by T3 DNA ligase.

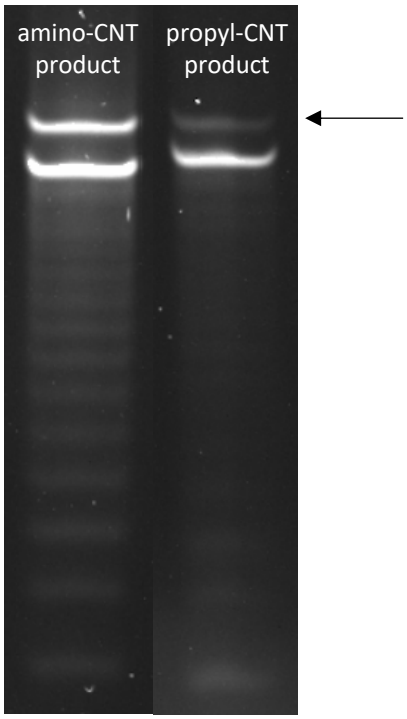


Figure 25: 10% denaturing PAGE gel (150V, 55 °C, 40 min) of the LOOPER product after 13 incorporations of amino-CNT trimers ligated to generate the full-length product. The image on the right is the LOOPER product after 13 incorporations of propyl-CNT trimers ligated to generate the full-length product.

3.3.2 CNG

The CNG sub-library was the most difficult to evaluate, due to difficulties denaturing the CG rich template strand. The template strand was smeared on the gel, making the yield unquantifiable (**Figure 26**). Future experiments involving the CNG sub-library were evaluated with a template consisting of a shorter reading frame to facilitate complete denaturation.

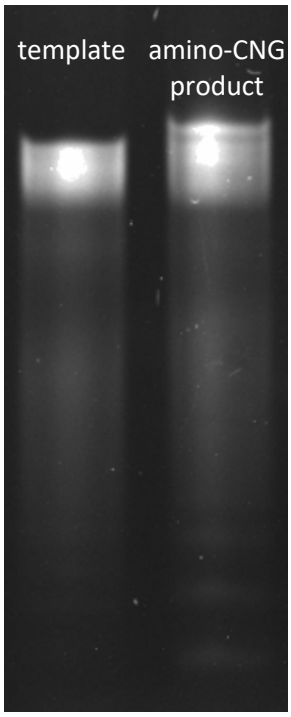


Figure 26: 10% denaturing PAGE gel (150V, 55 °C, 40 min) of the LOOPER product after 13 incorporations of amino-CNG trimers ligated to generate the full-length product. The template band smearing makes this sub-library difficult to quantify.

We attempted cyclopropylacetic acid and methoxyacetic acid modifications were each coupled on to amino-modified CNG. Once the products were collected and isolated, they were each subjected to T3 DNA ligase using a shorter DNA template facilitating 6 or 8 CNG iterations instead of 13 (**Figure 27&28**). Sadly, the shorter template did not help with denaturation of the LOOPER duplex, as the product appears as multiple bands on the gel. The product is suspected to be the band just above the template band, since the modified product migrates slower. Since some of the product remained in duplex form, it makes quantification impossible using our standard method.

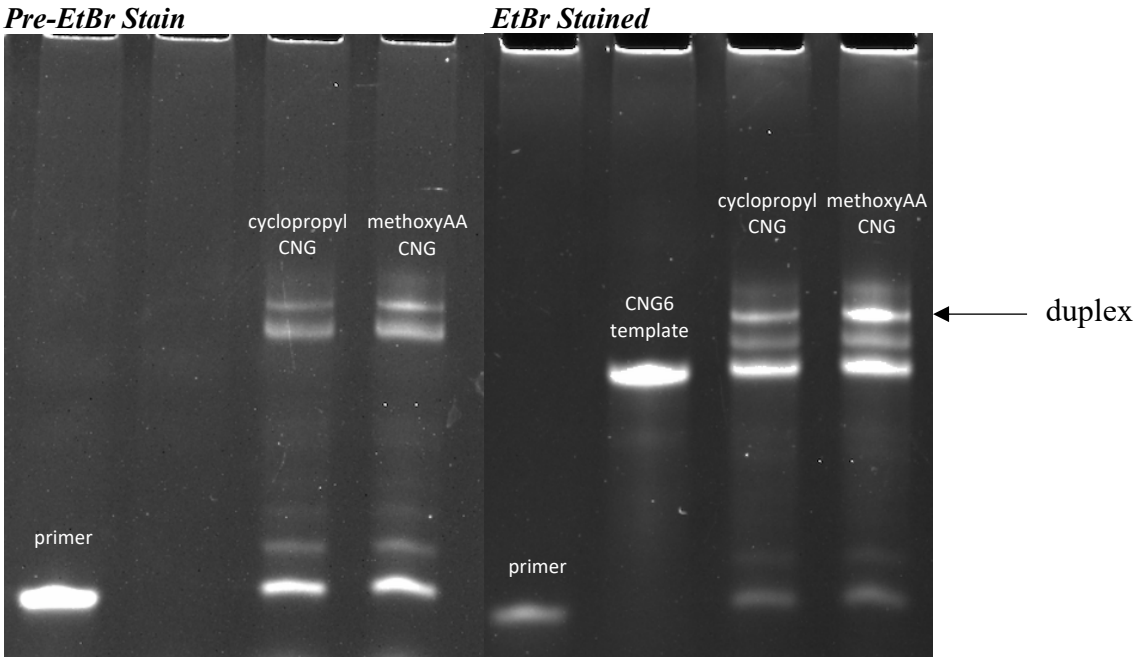


Figure 27: 10% Denaturing PAGE before and after ethidium bromide staining. Cyclopropyl-CNG and methoxyacetic acid-CNG (methoxyAA-CNG) LOOPER products after 6 incorporations. The denaturation issues persisted even after shortening the template.

Both templates were initially run on a denaturing PAGE gel to determine its ability to denature. Both templates appeared to denature properly due to the presence of one well-defined band per template. Therefore, we attempted LOOPER with amino-modified CNG and carboxy-modified CNG on the CNG8 template. The result is shown on the gel below.

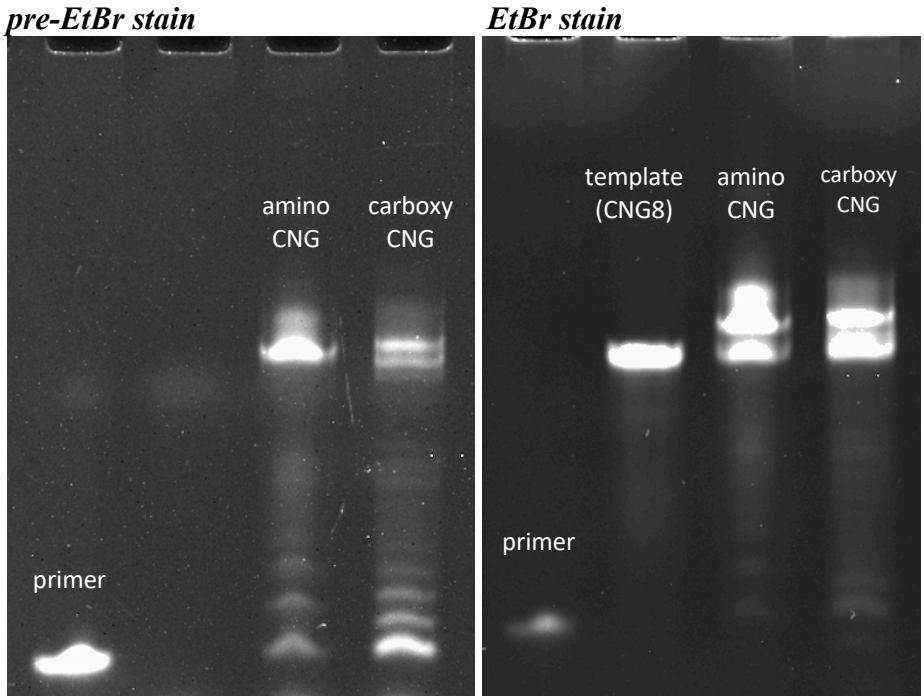


Figure 28: 10% Denaturing PAGE before ethidium bromide staining. Analysis of denaturation difficulties between CNG containing the amino modifier and CNG with a carboxylic acid modification after 8 iterations of ligation.

Although the template no longer has a denaturation issue, the modified strands appear to be contributing to the denaturation problem. Although, the yield of amino-CNG appears to be approximately 100%. We purchased a primer that is PEGylated for future LOOPER experiments where the product is difficult to denature or the LOOPER strand overlaps the template strand on the gel. We attempted a methoxyacetic acid-CNG using the CNG8 template and the PEG-ylated primer. The denaturation appears to improve with the addition of the PEG-ylated primer. The yield of this polymerization is 92%. Methoxyacetic acid will be included in the final library.

3.3.3 CNA

Amino-CNA showed a ligation efficiency of 100%, potentially allowing for the addition of a large chemical modification. Therefore, 1,4-methylenedioxyphenylacetic acid (MDPA) was attached to amino-CNA by EDC/NHS amide coupling. A 52% efficiency was seen. A phenyl group was also tested and showed a yield of 54%. Considering a 2% difference, the larger MDPA modification will be used in the final LOOPER library.

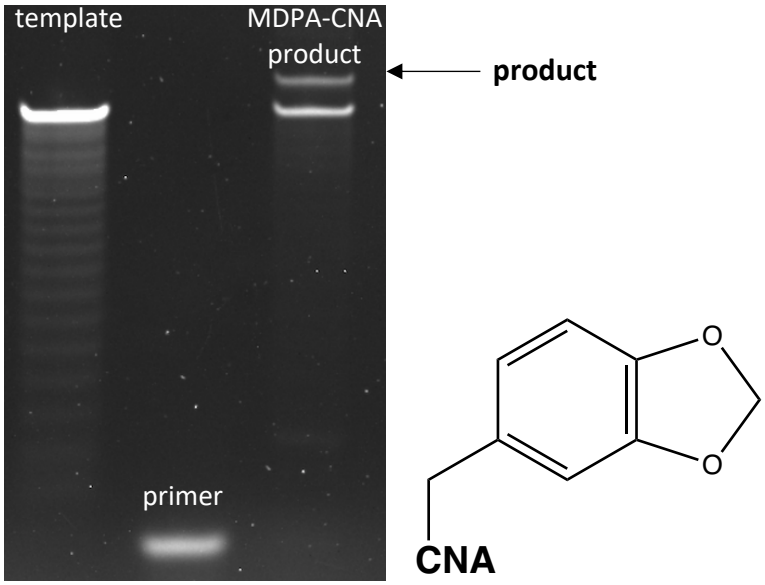


Figure 29: 10% denaturing PAGE gel of the LOOPER product after 13 incorporations of MDPA-CNA trimers ligated to generate the full-length product. The image on the right is the general structure of MDPA-CNA.

3.3.4 CNC

T3 DNA ligase cannot tolerate the CNC sub-library as well as others. Using only the amino-linker, CNC has an efficiency of 55%. Since this yield is already quite low, we attempted a smaller modification so that the yield is not impacted severely. Since a smaller modification was also needed for the CNT sub-library, we decided to couple a hydroxymethyl modification (that does not contain the C6-amino linker) to both sub-libraries to see which modification performed the best.

Figure 30: 10% Denaturing PAGE before ethidium bromide staining. LOOPER product after 13 incorporations of hydroxymethyl-CNT and hydroxymethyl-CNC, resulting in yields of 95% and 54%, respectively.

Due to the small modification and the absence of a positive charge, the LOOPER strand and template strand overlapped on the gel, and can therefore can only be quantified by approximation (relative quantity of the LOOPER strand to the primer before staining). The yields resulted in approximately 95% for hydroxymethyl-CNT and 54% for hydroxymethyl-CNC. Either of these sub-libraries could be used, considering that the yield is above 50%. After this, I realized that the yield for CNC only decreases by 1% when a hydroxymethyl modification was used instead of the C6-amino linker. This could potentially represent that the lower yields seen with the CNC sub-library could be sequence dependant, not size dependant.

Considering that hydroxymethyl-CNC has a polymerization yield of 54%, and a much larger amino-linker produces a very similar yield (55%), it was decided that the addition of a small

modification to the amino linker should not greatly affect the yield. I attempted to add a propyl group and monitor the yield by denaturing PAGE.

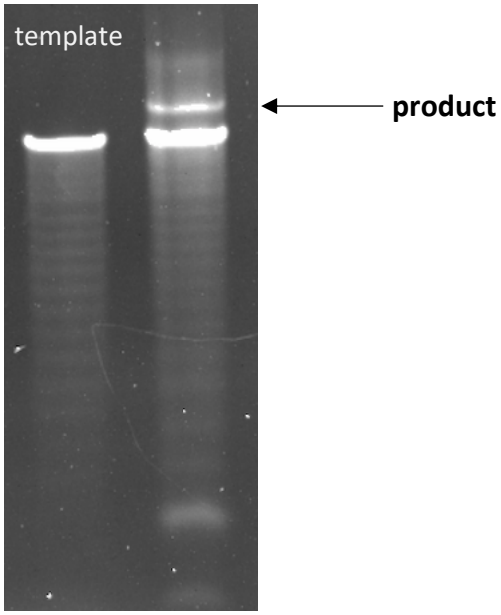


Figure 31: 10% Denaturing PAGE after ethidium bromide staining. LOOPER product after 13 incorporations of propyl-CNC.

Considering the yield is above 50%, propyl-CNC will be used in the final library. In order to complete the final library, we attempted incorporating a modification onto CNG. As we began to see the finalized library form, we agreed that a cyclopentyl modification may be critical for high affinity binding to a protein target, since the modification has been present in TBL1, and in the IL-6 and PCSK9 binding aptamers from Liu et al.'s 2018 publication. A few modifications were attempted on various sub-libraries to maximize the number of candidates for the library.

3.4 ANN & GNN sub-libraries

The unmodified A and G sub-libraries comprises 16 sequences. These two sub-libraries were tested. However, the LOOPER strand and the template strand bands overlap on a denaturing PAGE gel, and therefore could not be quantified. Since the LOOPER strand is fluorescently labeled, the gel was able to be imaged before staining in order to confirm the presence of the product. The pre-stained image of the gel is shown below. The yield is estimated to be 100%.

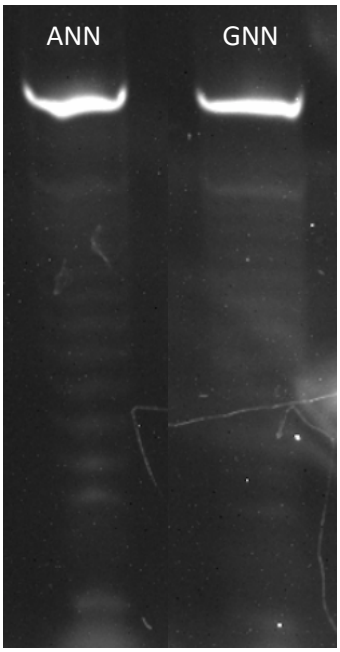


Figure 32: 10% denaturing PAGE gel of the LOOPER products after 13 incorporations of ANN and GNN trimers each ligated to generate the full-length product.

3.5 The R8-L1 library – yield and fidelity

The final library was pooled together at last. The table below indicates all the modified sub-libraries. The final library has been named R8-L1. R8-L1 resulted in a LOOPER yield of 19%. Considering the complexity of the system, a yield all together is impressive. T3 DNA ligase would need to ligate the appropriate trinucleotide anticodon 13 times to 10^{23} possible templates.

TRIMER	MODIFIED PHOSPHORAMIDITE	AMIDE-COUPLED MODIFICATION	YIELD
TNT	amino-modifier C2-dT	cyclopentyl	100%
TNG	amino-modifier C2-dT	indole	100%
TNC	amino-modifier C2-dT	cyclopropyl	61%
TNA	amino-modifier C2-dT	N/A	21%
CNT	hydroxymethyl-dC	N/A	95%
CNG	amino-modifier C6-dC	methoxyacetic acid	92%
CNC	amino-modifier C6-dC	butyric acid	53%
CNA	amino-modifier C6-dC	methylenedioxy phenylacetic acid	52%
ANN	unmodified	N/A	41%
GNN	unmodified	N/A	100%

Table 4: Finalized R8-L1 library that has been combined in equimolar amounts.

The R8-L1 library resulted in an overall fidelity of 86.9% prior to normalization to the Q5 DNA polymerase control. After normalization, the fidelity is 94.8%, making this library sufficient to use for selections. Like the unmodified LOOPER control, increasing GC content causes the fidelity to decrease by approximately 9%. This could be in relation to the bias, where anticodons with G at the 5' position generally had a higher positive bias. Further analysis indicated that fidelity decreases if cytosine is positioned at the 3'-position of the trinucleotide. The molecular basis for this will be investigated in the future.

We were then curious to evaluate the fidelity of each modified sub-library. In both T and C modified sub-libraries, the sub-library with the lowest fidelity was TNC and CNC. Even within the unmodified sub-libraries, the ANC sub-library had the lowest fidelity. Since this trend has been noticed in the unmodified sub-library as well as both modified sub-libraries, it is unlikely that the fidelity is lower due to the unique modification attached. In addition to this, the raw data output suggests that most errors occur due to A→G errors as well as C→T errors. Since the bases are modified, it is possible that T3 DNA ligase has some issues with distinguishing between specific pyrimidine or purine bases. Each sub-library has a fidelity above 78% prior to

normalization, meaning that T3 DNA ligase can be used during *in vitro* aptamer selection using this library system. If the fidelity is low (<80%), *in vitro* evolutions would not be able to occur, since we would not be able to accurately enrich sequence hits between rounds.

3.6 Future work on the R8-L1 library

Luca Rustico, an undergraduate, is currently working on performing *in vitro* aptamer selections using this system. We decided to move forward with a modified G-quadruplex library to select for an aptamer against human α-thrombin, a process developed by Dr. Matina Movahedi. The method allows for 1-3 rounds of selection without the need for PCR amplification in between rounds for enrichment. The process uses a 17 million membered library and relies on the process of removing non-binders and retaining binders. After selection and sequencing of the output, hits are analyzed by the number of copies of each sequence that survived each round.

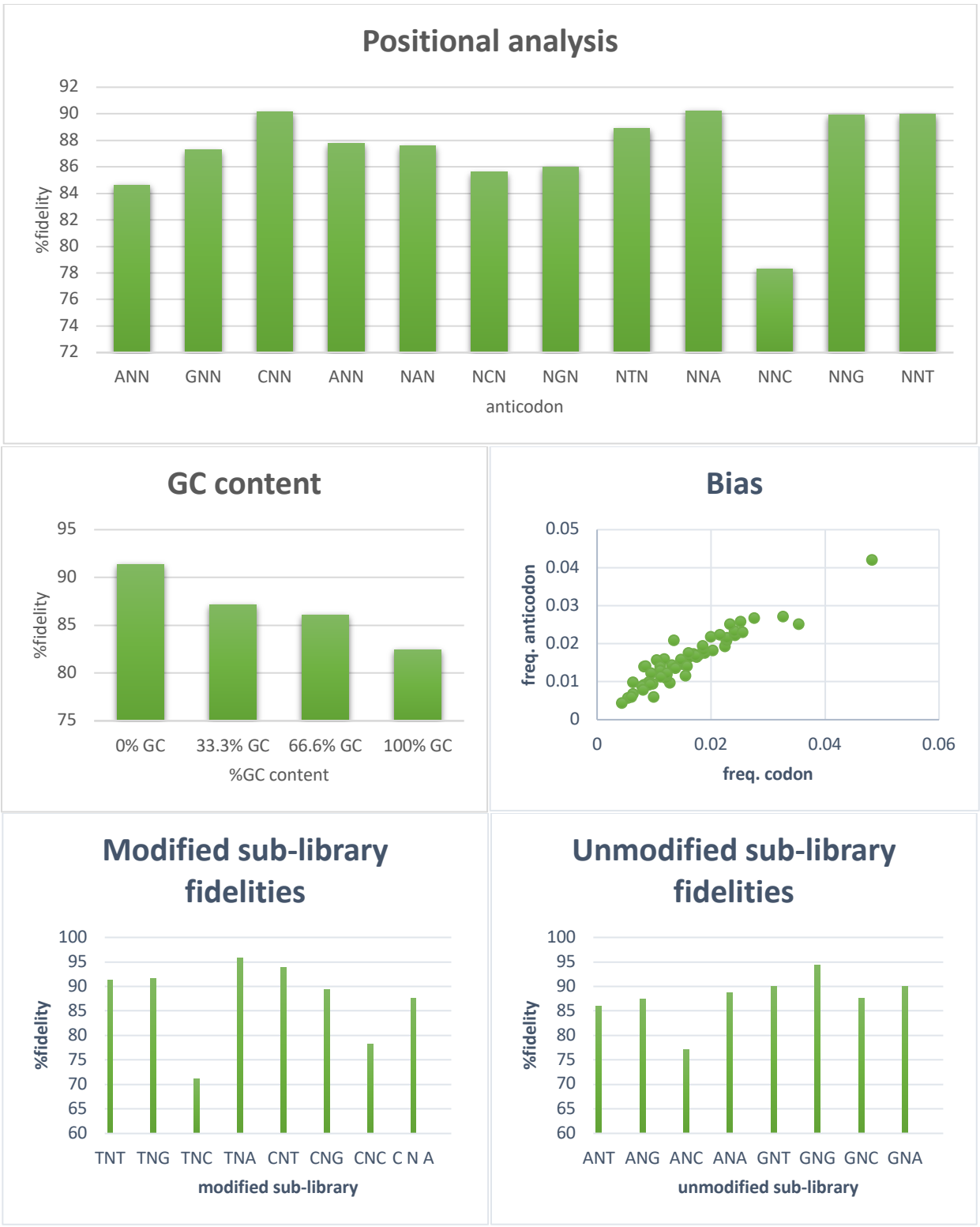


Figure 33: Hairpin sequencing fidelity and bias analysis of the R8-L1 library.

Previous iterations of LOOPER have suffered from limited sequence diversity and modification density. By completely removing the limitations on sequence diversity, the library diversity increases by four orders of magnitude. A library with maximum diversity allows for every sequence to be considered for binding to the target, potentially increasing the chances of finding an aptamer against the chosen target. Increasing modification density will allow for a more dynamic selection, where new folding properties of the nucleic acid polymer as well as new binding mechanisms to targets can be considered. The R8-L1 library is a high fidelity and mostly hydrophobic library to hopefully allow for an efficient aptamer selection against a protein target.

3.7 Sequences

Poly_pr1
/phos/TGCGACGGCAGGCGAATC

Poly_pr1_PEG
TGCGACGGCAGGCGAATC/iSp18/AACAACAACAACAA

3P-F
/56-FAM/AGGATCCGAGCTCCACGTG

TNT template
GATTCGCCTGCCGTCGCAANAANAANAANAANAANAANAANAANAANAANAANAANAA
NACACGTGGAGCTCGGATCCT

TNG template
GATTCGCCTGCCGTCGCACNACNACNACNACNACNACNACNACNACNACNACNACN
ACACGTGGAGCTCGGATCCT

TNC template
GATTCGCCTGCCGTCGCAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAG
NACACGTGGAGCTCGGATCCT

TNA template
GATTCGCCTGCCGTCGCATNATNATNATNATNATNATNATNATNATNATNATNATNA
CACGTGGAGCTCGGATCCT

CNT template
GATTCGCCTGCCGTCGCAANGANGANGANGANGANGANGANGANGANGANGA
NGCACGTGGAGCTCGGATCCT

CNG6 template
GATTCGCCTGCCGTCGCACNGCNGCNGCNGCNGCNGCACGTGGAGCTCGGATCCT

CNG8 template
GATTCGCCTGCCGTCGCACNGCNGCNGCNGCNGCNGCNGCNGCACGTGGAGCTCGG
ATCCT

CNG13 template
GATTCGCCTGCCGTCGCACNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCN
GCACGTGGAGCTCGGATCCT

CNC template
GATTCGCCTGCCGTCGCAGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG
NGCACGTGGAGCTCGGATCCT

CNA template
GATTCGCCTGCCGTCGCATNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNG
CACGTGGAGCTCGGATCCT

ANN template
GATTCGCCTGCCGTCGCANNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNT
CACGTGGAGCTCGGATCCT

GNN template
GATTCGCCTGCCGTCGCANNCNNCNNCNNCNNCNNCNNCNNCNNCNNCNNCNN
CCACGTGGAGCTCGGATCCT

TTT template
GATTCGCCTGCCGTCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AACACGTGGAGCTCGGATTCT

NNN template
GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNCACGTGGAGCTCGGATTCT

3.8 Methods

Synthesis of the R8-L1 sub-libraries

The library was synthesized using an ABI 394 DNA/RNA synthesizer, which facilitates solid-phase synthesis from the 3' to 5' using glass beads. All phosphoramidite bases used for synthesis were purchased from Glen Research. Once completed, the beads attached to the oligonucleotides were incubated at 65°C for 15 min in 400 μL of AMA solution, a 1:1 mixture of methylamine and ammonium hydroxide. After incubation, the supernatant was extracted and dried using a SpeedVac. The dried product was then purified using reverse-phase HPLC. The purified sample was frozen using dry ice and lyophilized. The dried, purified sample was then incubated in 1 mL of 40% aqueous acetic acid for 1 hour at room temperature to cleave off the DMT protecting groups. The sample was frozen and lyophilized once again, and then incubated in 500 μL concentrated ammonium hydroxide for 15 minutes at room temperature. The sample was then dried and purified once again using reverse-phase HPLC with a solvent gradient 10% acetonitrile in 0.1M TEAA (pH7) to 80% acetonitrile in 0.1M TEAA (pH7).

Deprotection of carboxy-dT and carboxy-dC sub-libraries begins with incubating the solid support in 400 μL 0.4 M NaOH in 4:1 methanol/water for 17 hours at room temperature. The supernatant was extracted and dried using a SpeedVac. The dried product was then reconstituted in 100 μL water and purified by HPLC. The HPLC product was lyophilized and then reconstituted in 400 μL of AMA solution, where it was incubated for 15 min at 65 °C. The sample was dried using a SpeedVac. Then, the dried sample was incubated in 1 mL of 40% aqueous acetic acid for 1 hour at room temperature to cleave off the DMT protecting groups. The sample was lyophilized and then incubated in 500 μL concentrated ammonium hydroxide for 15 minutes at room temperature. The sample was then dried and purified once again using reverse-phase HPLC with a solvent gradient 10% acetonitrile in 0.1M TEAA (pH7) to 80% acetonitrile in 0.1M TEAA (pH7).

Amide-coupled modifications for each amino-modified sub-library

Synthesis of the carboxylic acid modification onto the amino-modified trimer was done using 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC) and N-hydroxysuccinimide (NHS) to facilitate amide-coupling. The reaction began with mixing 25 μL of 100 mM carboxylic acid modification (100 mM in DMSO), 25 μL of NHS (100 mM in 1:1 DMSO/water), 5 μL of EDC (100 mM in DMSO) and 7.5 μL of DMSO. The mixture was incubated and mixed for 30 minutes at room temperature. Then, 30 μL of Na_2CO_3/NaOH buffer (pH 9.0) and 20 μL of 1 mM of the amino-modified DNA trimer was added to the mixture and incubated overnight (approx. 16 h) while mixing at room temperature. After incubation, the reaction was quenched with 15 μL Tris-HCl buffer (pH 8.0) and incubated on vortex at room temperature for 1 hour. The mixture was then lyophilized and purified by reverse-phase HPLC with a solvent gradient of 10% acetonitrile in 0.1M TEAA (pH7) to 80% acetonitrile in 0.1M TEAA (pH7). Finally, the product was confirmed by mass spectrometry.

LOOPER via T3 DNA ligase

The LOOPER protocol requires T3 DNA ligase and 2X T3 DNA ligase buffer (New England Biolabs, M0317S). A DNA template was also used, having a sequence of 13 repetitions of the complementary trimer sequence and two primer binding regions. All DNA templates and primers were purchased from Integrated DNA Technologies (IDT). The ligation protocol begins with mixing 1 μL of the DNA template (10 μM), 0.5 μL Poly_pr1 (30 μM), 0.5 μL 3P-F (30 μM), 1.3 μL of the trimer sub-library (1 mM), and 5 μL 2X T3 DNA ligase buffer in a PCR tube. The mixture is then incubated at 95°C 10s, 65°C for 4 min, and cooled using a ramp from 65°C to 4°C

at 0.1°C/s. After incubation, 0.7 µL MilliQ water and 1 µL T3 DNA ligase (3000 U/µL) were added. The reaction was then incubated at 4°C for 24 h. The product was then purified using the E.Z.N.A. Omega Biotek® purification kit. The samples were denatured at 95°C for 5 min and loaded while still hot onto a 10% denaturing PAGE gel. The gels were imaged before and after ethidium bromide staining.

Hairpin Sequencing – Formation of the hairpin product

In a PCR tube, 1 µL 5HP_39mer_IT1 (10 µM), 0.5 µL nuclease-free water, 0.5 µL 3P_IT1 (30 µM), 1.3 µL trinucleotide (1 mM), and 5 µL StickTogether DNA Ligase Buffer (2X, NEB M0317L) were added. The mixture was then incubated at 98 °C for 10 s, 65 °C for 4 min, then a ramp from 65 °C to 4 °C at 0.1 °C/s. Lastly, 0.7 µL nuclease-free water and 1 µL T3 DNA ligase (3000 U/µL, NEB M0317L) were added to the mixture. The PCR tube was then incubated at 4 °C for 24 hours. The resulting mixture was then purified with Omega E.Z.N.A.® Cycle Pure Kit (Omega Bio- tek, D6492-02) and eluted in 30 µL nuclease-free water. A 15% denaturing PAGE gel (55 °C, 150 V, 80 min, denatured at 98 °C for 2 min and loaded hot) was then run to visualize and extract the product. The product was then purified using a Centri-Sep column (Princeton Separations, CS-901) and then quantified by Qubit and diluted to 10 pM before PCR.

Chapter 4 – RNA LOOPER

Dr. Matina Movahedi and I contributed equally to all aspects of this project.

4.1 Purpose

Modified DNA has been vastly explored with the LOOPER system, considering various nucleobase modifications, nucleobase linker lengths, and 2' sugar modifications. An important nucleic acid polymer that has never been explored by LOOPER is RNA synthesis and modified RNA synthesis. RNA polymers play a crucial role in biological function and RNA-based therapeutics can mimic natural RNA functions such as catalytic processes and regulating gene expression[54]. Moreover, RNA modifications are vital in regulating cell biology since they influence all aspects of RNA biological functions, such as generation, transportation, function, and metabolization[55]. In recent years, it has been discovered that RNA modifiations are not only dynamic and reversible, being catalyzed by reader, writer, and eraser proteins, but they are also regulated by environmental factors[56]. Unfortunately, RNA modifications and their implications in diseases are not well known[55,56]. The development of a chemically-modified RNA library allows for the selection of RNAzymes as well as targets for RNA binding proteins which can be used to further study RNA modifications and its implications in diseases.

Dr. Matina Movahedi and I sought to develop a LOOPER system that can accommodate RNA oligonucleotides to later develop an m^6A modified oligonucleotide library as m^6A is the most prevalent and abundant RNA modification in eukaryotes[55]. This modified library can then be used to select for sequence hits against m^6A reader, writer, or eraser proteins. This developed system can potentially allow for further research on naturally-occurring RNA modifications and as well as their cellular functions.

4.2 Development of the RNA LOOPER system

A suitable ligase and primers as well as a specific oligonucleotide length are the most necessary parameters for RNA LOOPER to be successful. However, the process has been proven to be difficult to validate the product, so a method of validation was necessary before optimizing. Our validation method consisted of using a defined DNA template, DNA primers, and a defined 40-mer RNA strand to anneal to the coding region of the template (**Figure 34**). T4 RNA ligase 2 has been reported by New England Biolabs to ligate RNA on a DNA template. This control was done to validate the location of the full-length product strand on the acrylamide gel.

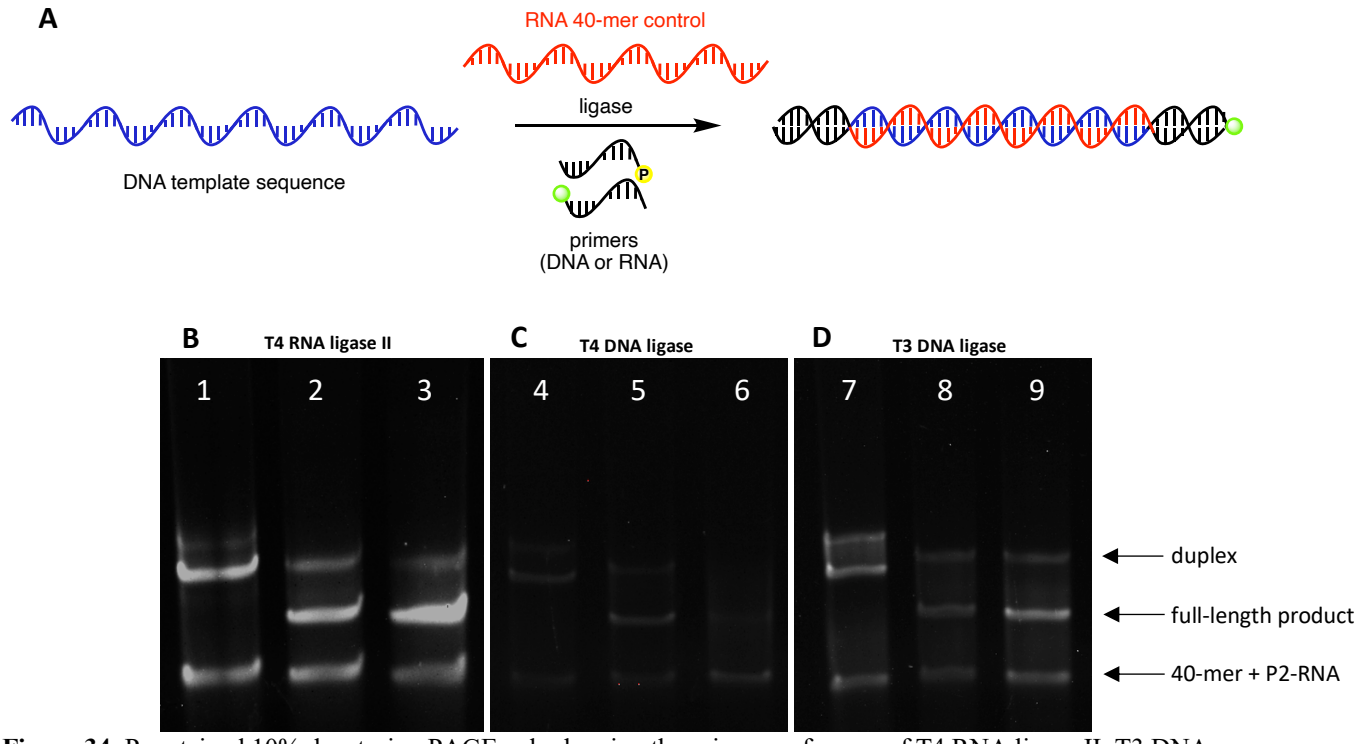


Figure 34: Pre-stained 10% denaturing PAGE gels showing the primer preference of T4 RNA ligase II, T3 DNA ligase, and T4 DNA ligase. A) RNA LOOPER control ligation scheme using a defined DNA template, a 40 nucleotide RNA sequence, and a variety of DNA/RNA primer combinations. B) T4 RNA ligase II control ligation using DNA and RNA primers. Lane 1: PEG-primer1-DNA + fluorescent-primer2-RNA, Lane 2: primer1-DNA + fluorescent-primer2-RNA, Lane 3: primer1-RNA + fluorescent-primer2-DNA. C) T4 DNA ligase control ligation using DNA and RNA primers. Lane 4: PEG-primer1-DNA + fluorescent-primer2-RNA, Lane 5: primer1-DNA + fluorescent-primer2-RNA, Lane 6: primer1-RNA + fluorescent-primer2-RNA. D) T3 DNA ligase control ligation using DNA and RNA primers. Lane 7: PEG-primer1-DNA + fluorescent-primer2-DNA, Lane 8: primer1-DNA + fluorescent-primer2-DNA, Lane 9: primer1-RNA + fluorescent-primer2-DNA.

We then attempted using this LOOPER system with a combination of DNA and RNA primers to see which one gave the best result. All three combinations attempted were able to produce the full-length product (**Figure 34**). These combinations were then tested with T4 RNA ligase II and rN3, rN4, and rN5 RNA anticodon libraries (**Figure 35**). Unfortunately, T4 DNA ligase and T3 DNA ligase were not able to achieve full-length product, which is expected since they are enzymes designed to ligate DNA polymers. Regarding T4 RNA ligase II, the pentamer library performs the best regardless of the primer combination. The DNA-RNA primer mix seems to have the most full-length product when compared to having both RNA primers (**Figure 35**, lanes 1&4). We have chosen to optimize the system containing both RNA primers (**Figure 35**, lane 4) and the anticodon is the largest size (rN5). This would result in a product strand that is completely RNA where we can then DNAse degrade the DNA template and visualize the product on its own. This was the most optimal visualization strategy since previous experiments suggested that the product duplex had difficulty denaturing prior to PAGE analysis. If there is a DNA primer (*e.g.* **Figure 35**, Lane 1), DNAse degradation would not be feasible since the final product would be contaminated with partially-ligated product, resulting in an overestimated yield.

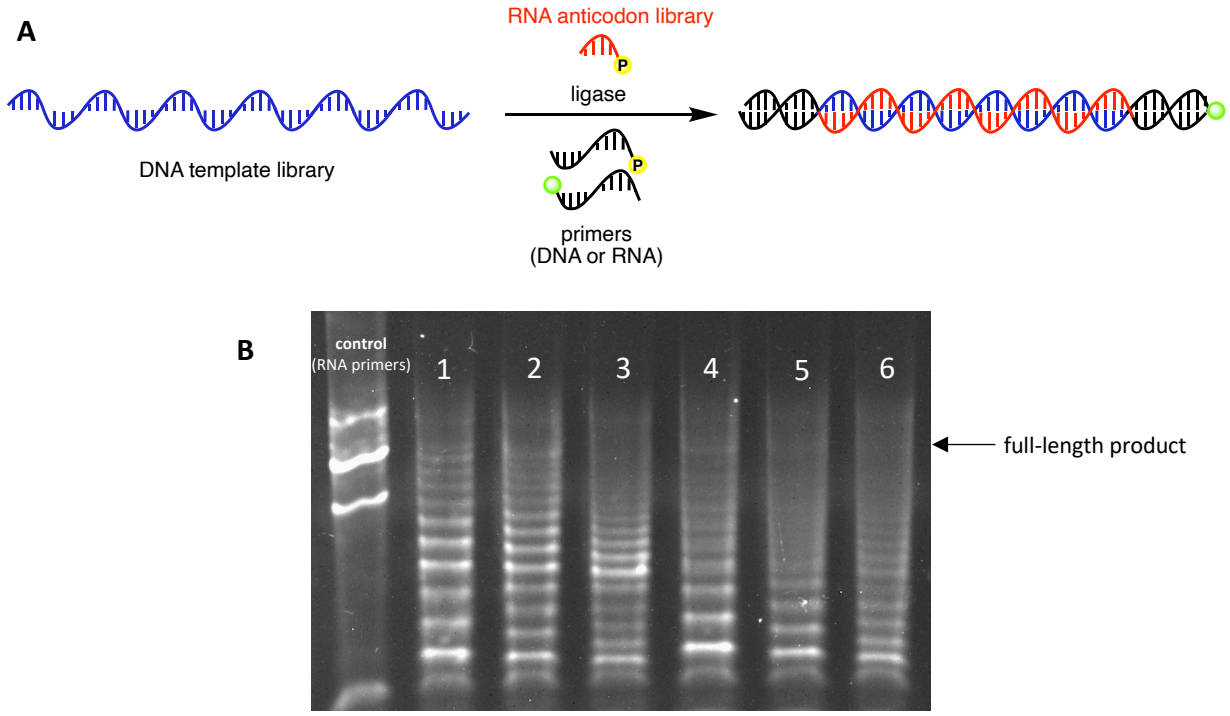


Figure 35: A) Schematic of the RNA LOOPER process. B) Pre-stained 10% denaturing PAGE gels showing the primer and codon preference of T4 RNA ligase II. Lane 1: rN5 anticodon library, primer1-DNA + fluorescent-primer2-RNA, Lane 2: rN4 anticodon library, primer1-DNA + fluorescent-primer2-RNA, Lane 3: rN3 anticodon library, P1-DNA primer + fluorescent-primer2-RNA, Lane 4: rN5 anticodon library, primer1-RNA primer + fluorescent-primer2-RNA, Lane 5: rN4 anticodon library, primer1-RNA + fluorescent-primer2-RNA, Lane 6: DNA template library, rN3 anticodon library, primer1-RNA + fluorescent-primer2-RNA.

Once these tests were completed, we wanted to confirm that the highest band seen in the controls was the non-denatured duplex. To do this, we utilized a control that used both RNA primers. The best method to obtain this product was to use T4 RNA ligase II. We then DNAse degraded the template after the ligation. The result confirmed the presence of the non-denatured duplex, further suggesting that using only RNA primers is the best option to move forward with.

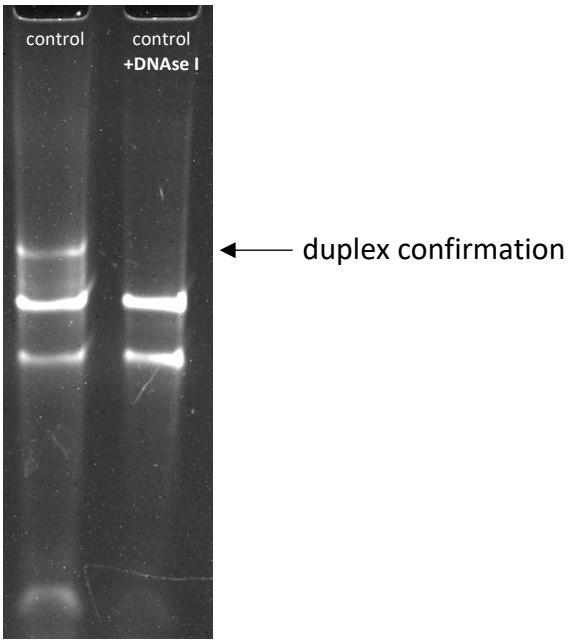


Figure 36: Control experiment using T3 DNA ligase and RNA primers with and without DNAse I degradation.

We then began with the optimization of incubation temperature. The incubation that was originally used was 25 °C for 18 h, which was used as our standard. The incubation temperatures attempted ranged from 4-37 °C are shown in the gel below. We found that that although the product band was faint, 30 °C was the most promising temperature.

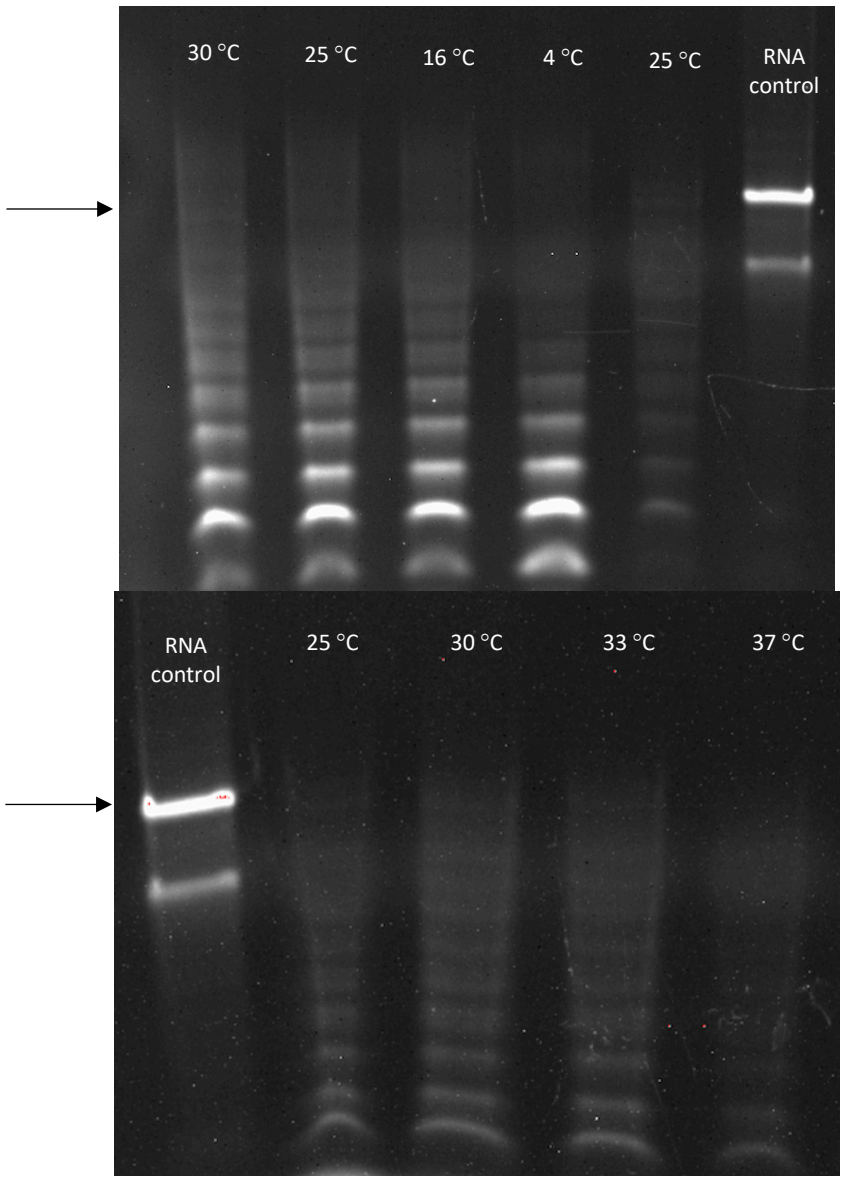


Figure 37: Optimization of the RNA LOOPER incubation temperature.

We then proceeded to optimize incubation time at 30 °C. The 6 h incubation seemed to have the most product with the least amount of smearing. We hypothesized that an increased incubation time would allow for an increase in product yield, though that was not the case. An interesting discovery is that increased incubations caused degradation of the product. In hindsight,

degradation of the RNA product over longer periods of time reasonable due to the unstable nature of RNA.

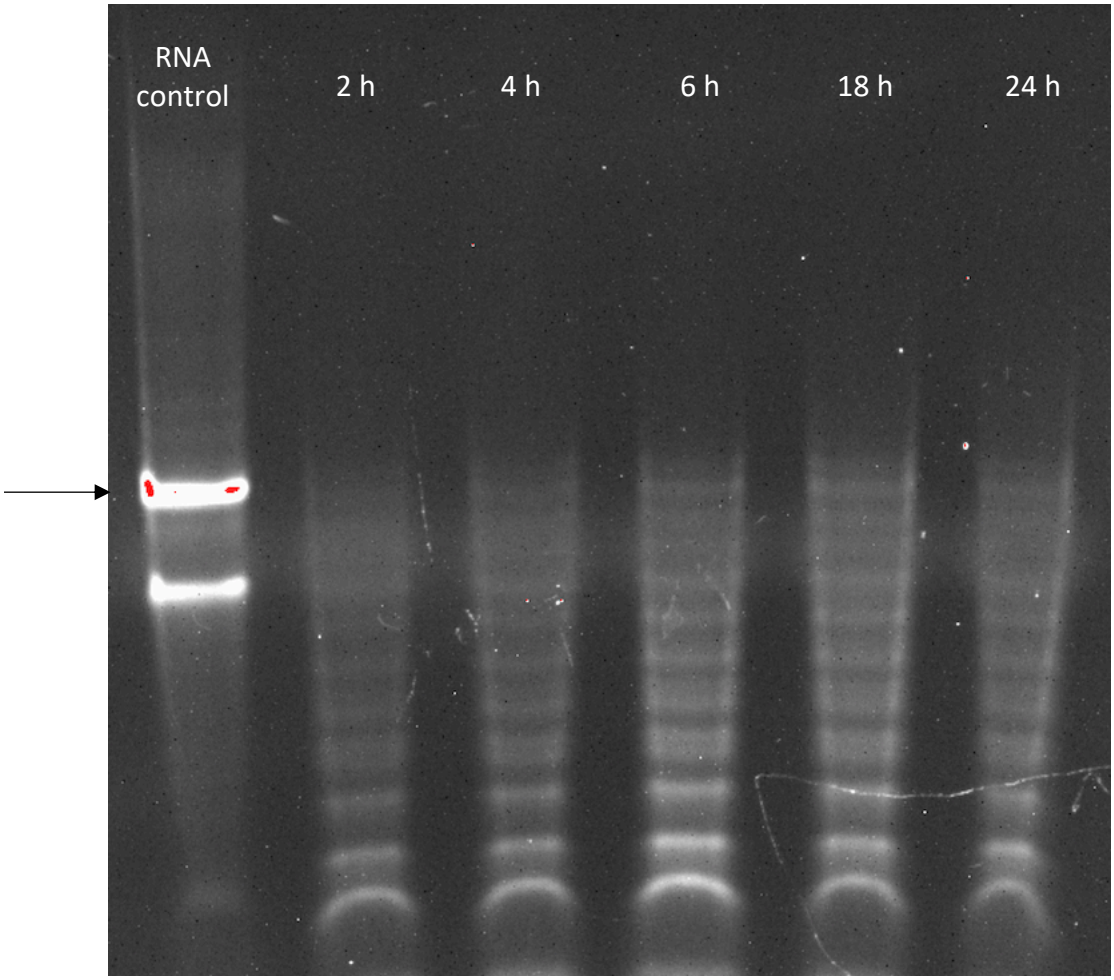


Figure 38: Optimization of the RNA LOOPER incubation time at 33 °C.

During this incubation experiment, we noticed that there is a faint band right above the product band. We hypothesized this could be due to folding of the fully ligated RNA product since we can also see some faint bands above the control product as well (**Figure 38**). Further analysis would need to be done to determine how to mitigate this issue. This denaturation issue has also resulted in the inability to accurately quantify the LOOPER product.

We then attempted this time incubation experiment over the course of 6 h with temperature of 30 °C, 33 °C, and 37 °C; the gel is shown below. The highest polymerization efficiency was seen with a 4 h incubation at 33 °C.

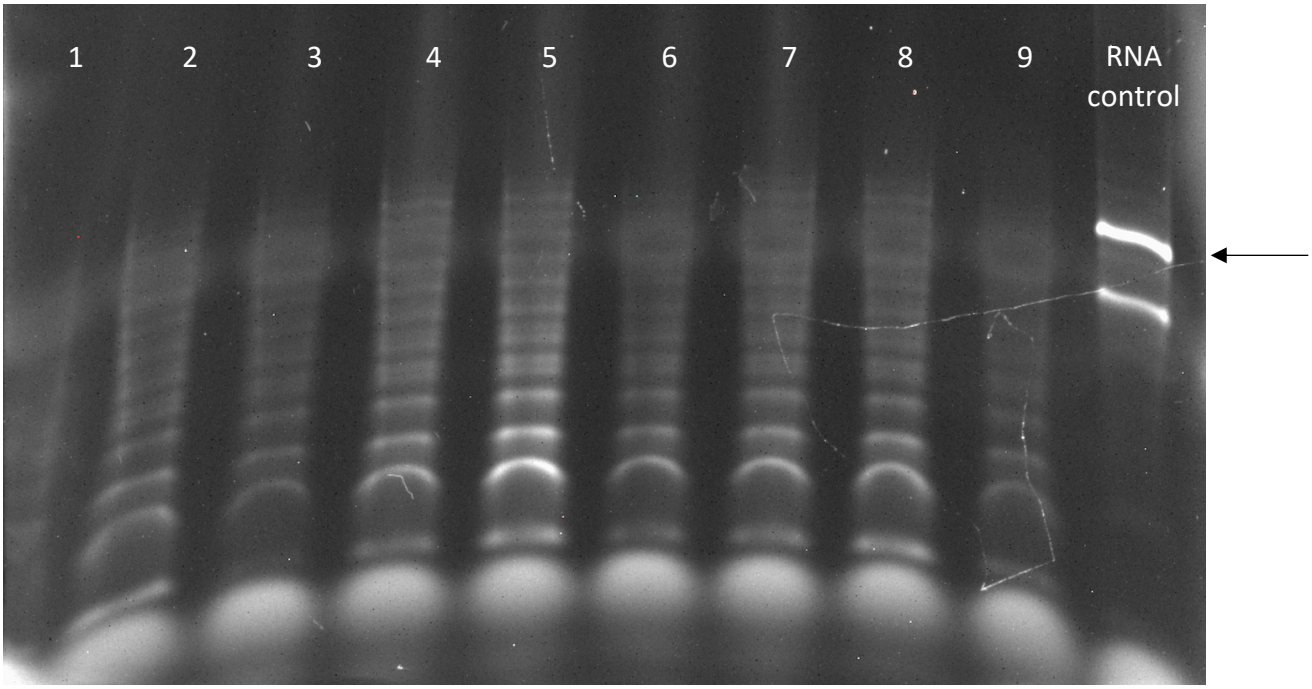


Figure 39: Optimization of temperature and incubation time. Lane 1: 6 h incubation at 37 °C; Lane 2: 4 h incubation at 37 °C; Lane 3: 2 h incubation at 37 °C; Lane 4: 6 h incubation at 33 °C; Lane 5: 4 h incubation at 33 °C; Lane 6: 2 h incubation at 33 °C; Lane 7: 6 h incubation at 30 °C; Lane 8: 4 h incubation at 30 °C; Lane 9: 2 h incubation at 30 °C.

After optimizing the time and temperature, we wanted to try incorporating a second ligase addition to boost the polymerization efficiency. We attempted doing this with variations in time and temperature, but keeping the overall incubation time at 6 h. Even though all conditions that we attempted generated full length product, we opted to move forward with another ligase addition after a 2 h incubation at 33 °C and then a 4 h incubation at 30 °C after that (**Figure 40**, lane 2), since it had the lowest incubation temperature to avoid RNA degradation.

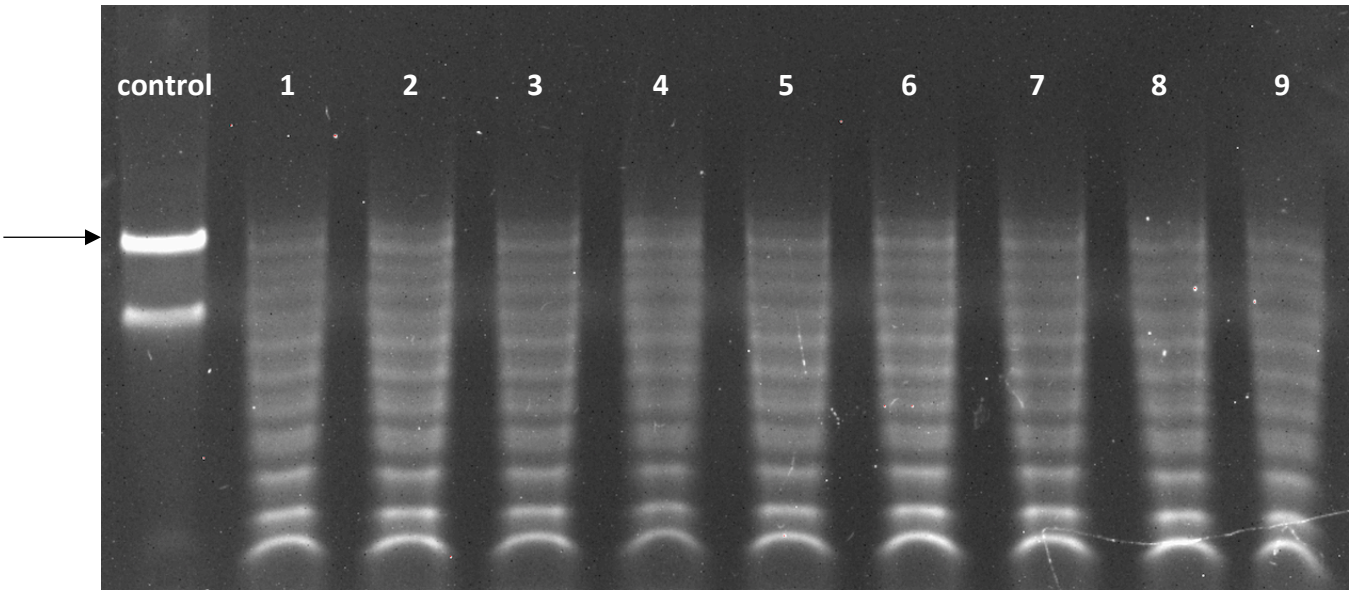


Figure 40: Optimization for incubation time and extra T4 RNA ligase II additions. Lane 1: 33 °C for 2 h, 10 U ligase addition, 30 °C for 2 h; Lane 2: 33 °C for 2 h, 10 U ligase addition, 30 °C for 4 h; Lane 3: 33 °C for 2 h, 10 U ligase addition, 33 °C for 2 h; Lane 4: 33 °C for 2 h, 10 U ligase addition, 33 °C for 4 h; Lane 5: 33 °C for 2 h, 10 U ligase addition, 37 °C for 2 h; Lane 6: 33 °C for 2 h, 10 U ligase addition, 37 °C for 4 h; Lane 7: 33 °C for 4 h, 10 U ligase addition, 30 °C for 2 h; Lane 8: 33 °C for 4 h, 10 U ligase addition, 33 °C for 2 h; Lane 9: 33 °C for 4 h, 10 U ligase addition, 37 °C for 2 h.

Lastly, we attempted three experiments, a denaturing experiment, adding 20 U of T4 RNA ligase II at the start of LOOPER rather than 10 U, and overnight incubations at 25 °C and 30 °C. We noticed that adding 20 U of T4 RNA ligase II prior to incubating resulted in a higher full-length product yield as opposed to adding 10 U (**Figure 41**, lane 2). The yield of the 10 U ligase addition and longer incubation system was similar to the 20 U ligase addition to the start (**Figure 41**, lanes 2&3). We therefore proceeded with using the 20 U addition prior to incubation since it was easier to perform a shorter incubation and to add the ligase from the start of the incubation.

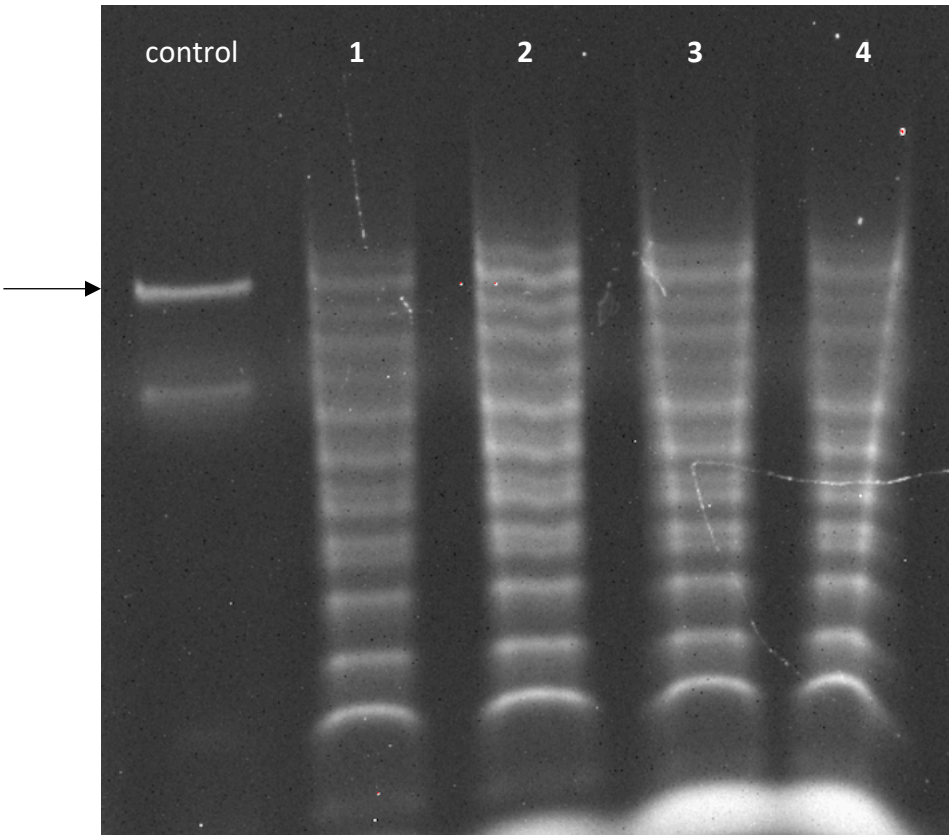


Figure 41: Optimization of the amount of ligase to add prior to and during incubation. Lane 1: 33 °C for 2 h, 10 U ligase addition, 30 °C for 4 h; Lane 2: 20 U ligase addition (from the start), then incubated at 33 °C for 2 h, 30 °C for 4 h; Lane 3: 33 °C for 2 h, 10 U ligase addition, 30 °C for 4 h, 25 °C for 12 h; Lane 4: 33 °C for 2 h, 10 U ligase addition, 30 °C for 16 h.

After optimizing the incubation conditions, we wanted to begin optimizing the components added to the system. The next two parameters that we tested were the anticodon stoichiometry with respect to the template and the importance of adding PEG8000 and extra $MgCl_2$. We attempted anticodon equivalents of 10, 15, and 20. In addition to this, we also removed either PEG8000, $MgCl_2$, or both (at 10 equivalents). 15 equivalents seemed to produce the highest yield so this stoichiometry was used in future experiments. Since we started this LOOPER system by using the protocol derived from the ligase manufacturer, the protocol required an additional 8 mM of $MgCl_2$ to the $MgCl_2$ already found in the reaction buffer. However, the removal of supplementary $MgCl_2$ dramatically increased the yield. Therefore, supplementary $MgCl_2$ was not added to the reaction buffer in future experiments. We then wanted to combine

the parameters that worked best. Not surprisingly, we discovered that 15 equivalents of rN5 in addition to not adding extra $MgCl_2$ visually produced the highest yield (**Figure 42**, lane 8).

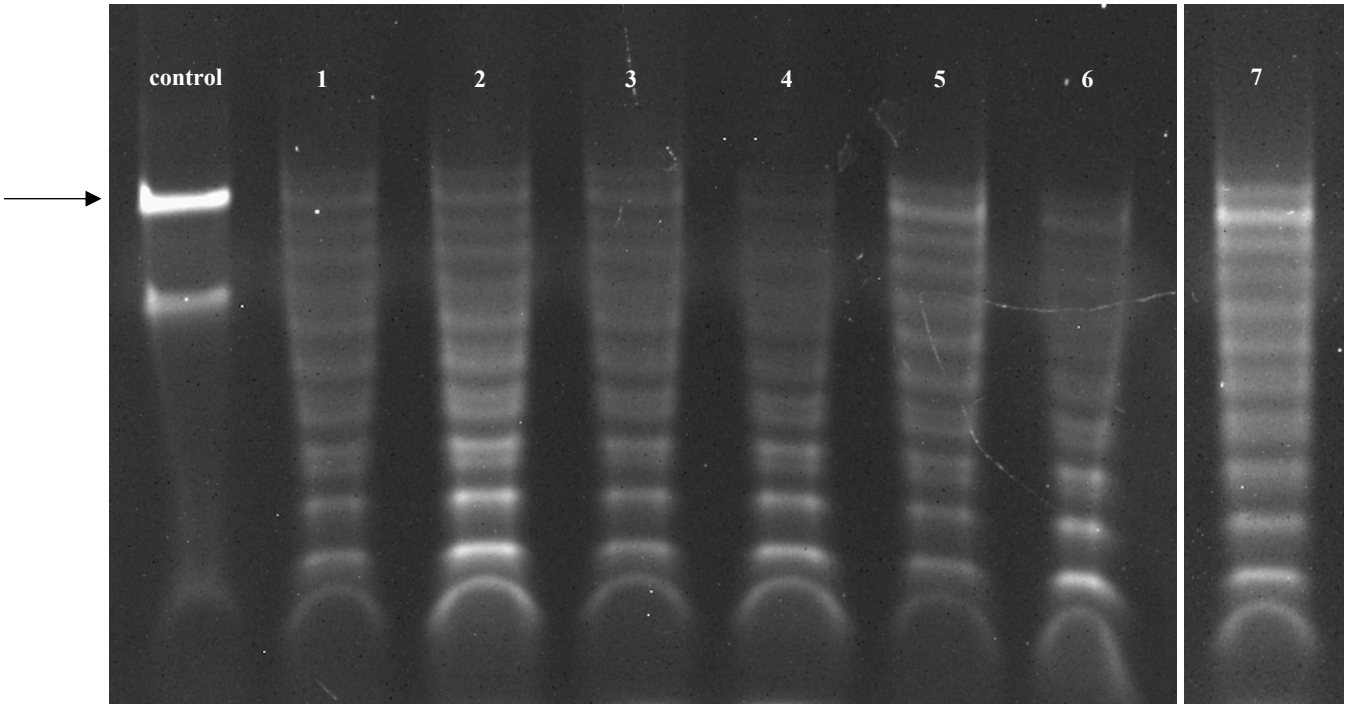


Figure 42: Optimizations of the anticodon stoichiometry as well as understanding the significance of $MgCl_2$ and PEG8000. Lane 1: 10 equivalents of rN5; Lane 2: 15 equivalents of rN5; Lane 3: 20 equivalents of rN5; Lane 4: removal of PEG8000; Lane 5: removal of supplementary $MgCl_2$; Lane 6: removal of both PEG8000 and supplementary $MgCl_2$; Lane 7: 15 equivalents of rN5 and removal of supplementary $MgCl_2$.

We then began optimizing the concentration of PEG8000 in the reaction. We usually add 4 μL of a 50% solution (10% in the total reaction volume). We tried reactions with 0%, 2.5%, 5%, 7.5%, 10% (control), and 12.5% of PEG8000 in the total reaction volume. As the concentration increases, the yield also increases. Attempting concentrations that are significantly higher are not feasible due to solubility issues.

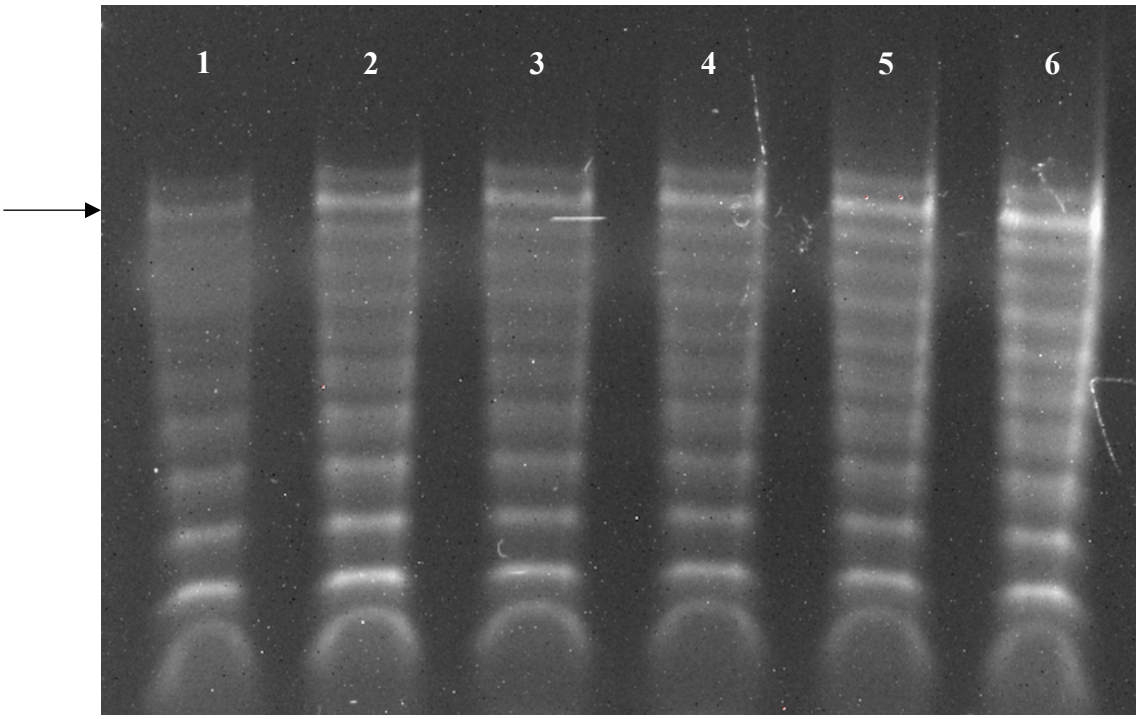


Figure 43: PEG8000 optimization for the RNA LOOPER system. Lane 1: 0% PEG8000, Lane 2: 2.5% PEG8000, Lane 3: 5% PEG8000, Lane 4: 7.5% PEG8000, Lane 5: 10% PEG8000, Lane 6: 12.5% PEG8000.

Previously, we have attempted not incorporating an additional 8 mM of $MgCl_2$ and the yield increased dramatically. Therefore, we wanted to attempt various amounts of $MgCl_2$ to determine the optimal concentration of $MgCl_2$ within the reaction. We discovered that 3 mM $MgCl_2$ is the most optimal concentration for the system.

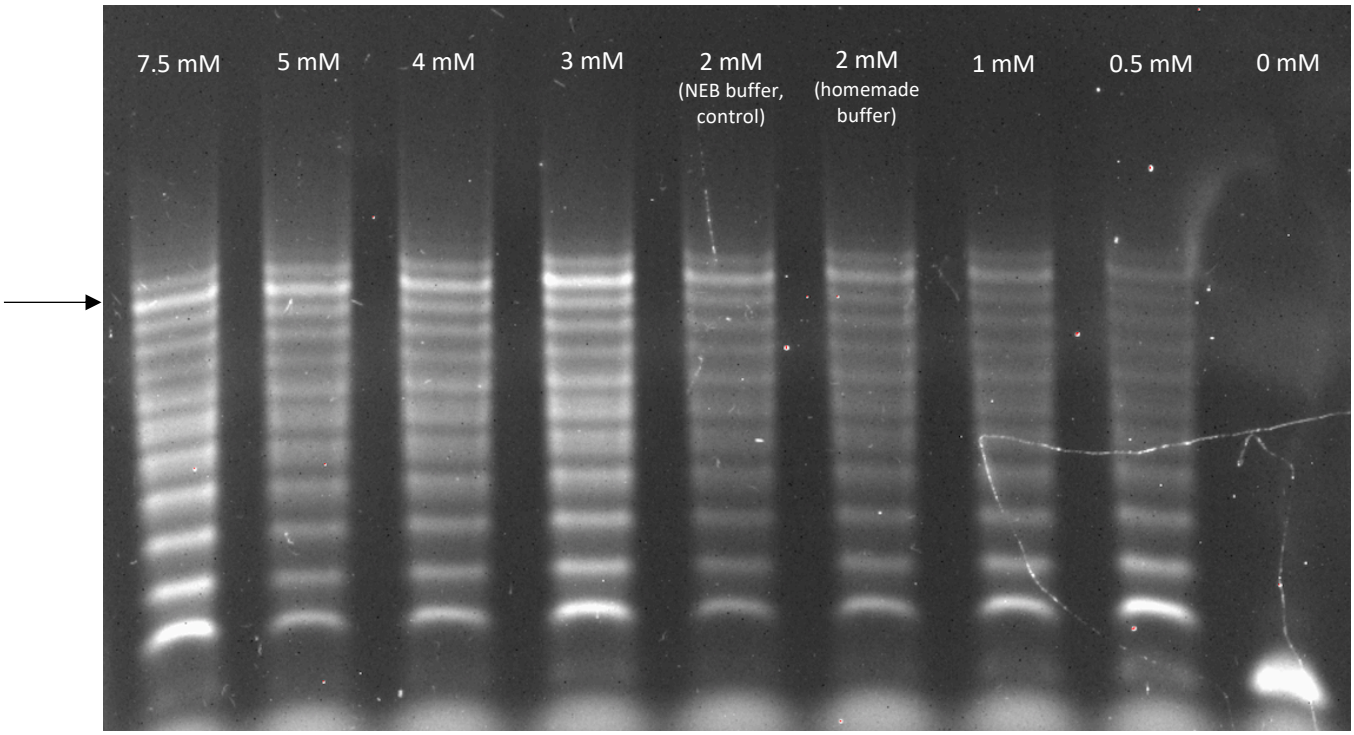


Figure 44: $MgCl_2$ concentration optimization for the RNA LOOPER system.

The NEB buffer has a DTT concentration of 1 mM. Matina attempted a lower concentration of 0.5 mM and a higher concentration of 1.5 mM DTT to evaluate its influence on ligation efficiency. As a result, the 1 mM concentration appeared to be the appropriate concentration for LOOPER as it resulted in the highest full-length product yield.

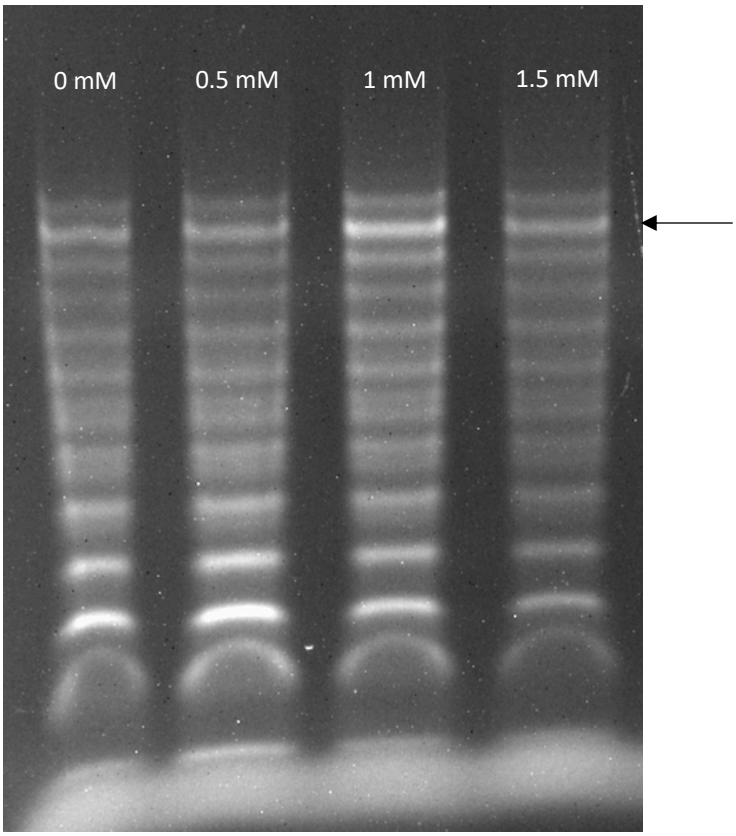


Figure 45: DTT concentration optimization for the RNA LOOPER system.

Since PEG6000 is also considered an optimal crowding agent for ligation, I decided to revisit the PEG optimizations and compare various concentrations of PEG 8000 with PEG 6000. Surprisingly, 15% PEG 6000 resulted in the highest full-length product yield compared to all other PEG6000 and PEG8000 concentrations.

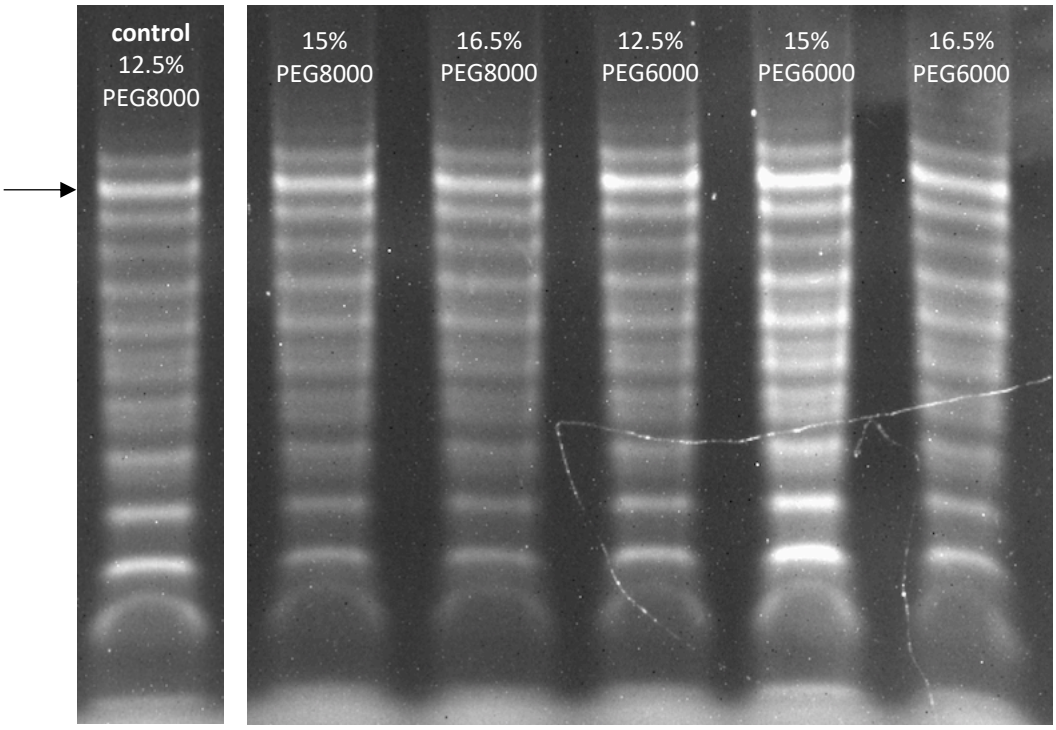


Figure 46: Further PEG optimizations to the RNA LOOPER system. PEG6000 was evaluated and compared to the effectiveness of PEG8000.

The last parameter to optimize was the ATP concentration within the reaction. The NEB buffer contains 0.4 mM ATP, so I tested concentrations higher and lower than 0.4 mM. I tested 0.2 mM, 0.4 mM (control), 0.6 mM (NEB buffer + 0.2 mM), 0.8 mM (NEB buffer + 0.4 mM), and 1 mM (NEB buffer + 0.6 mM). The results suggest that 0.8 mM ATP in the total reaction resulted in the highest full-length product yield (qualitatively).

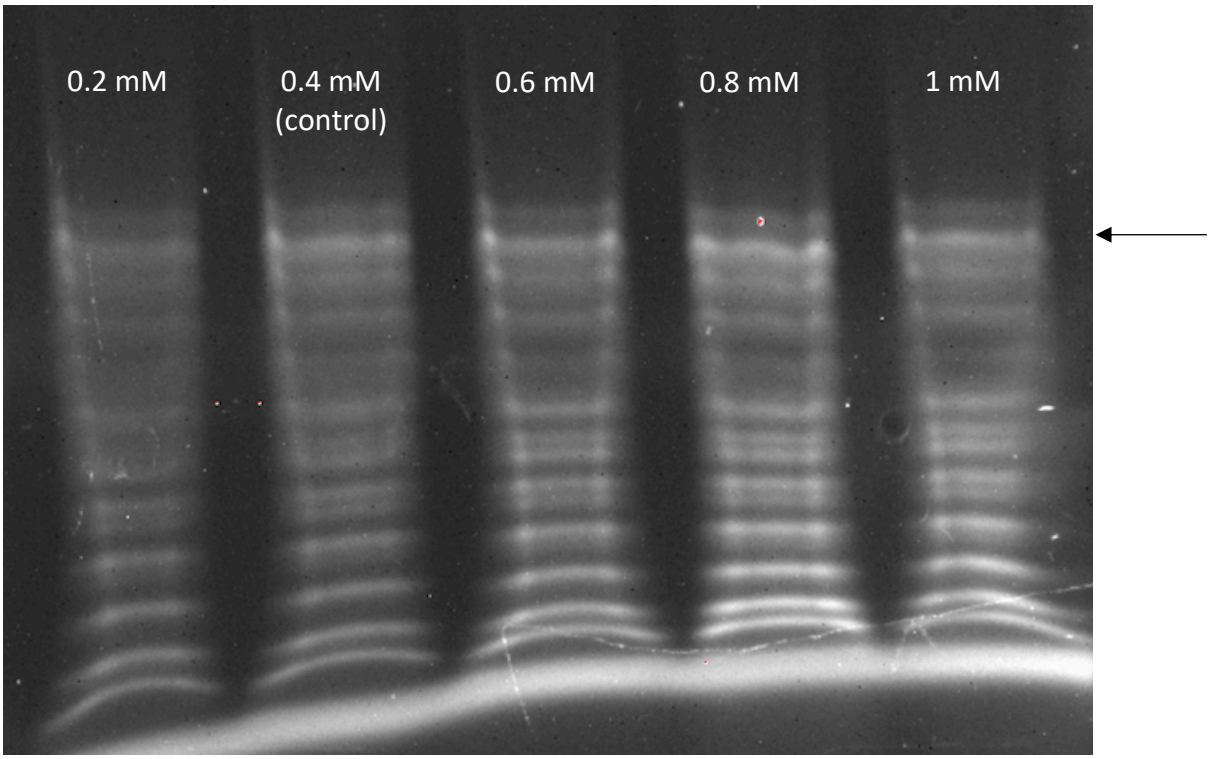


Figure 47: ATP concentration optimization for the RNA LOOPER system.

4.3 RNA LOOPER – m^{6}A-modified anticodon experiment with all optimized conditions

Since we now have an optimized RNA LOOPER to achieve the highest yield possible, we decided to move forward with using an *N*^{6}-methyl adenosine modification on the pentanucleotide sequence. Since this system has never been used before, we were curious to determine the modification positional preference of T4 RNA ligase II. We therefore conducted experiments to determine if T4 RNA ligase preferred where the modification was located. The pentamer RNA sequence libraries used were (m^{6}A)NNNN, N(m^{6}A)NNN, NN(m^{6}A)NN, NNN(m^{6}A)N, and NNNN(m^{6}A). Surprisingly, we noticed full length product for every position tested. We then decided to proceed with hairpin sequencing to analyze the fidelity of each RNA library in the developed system.

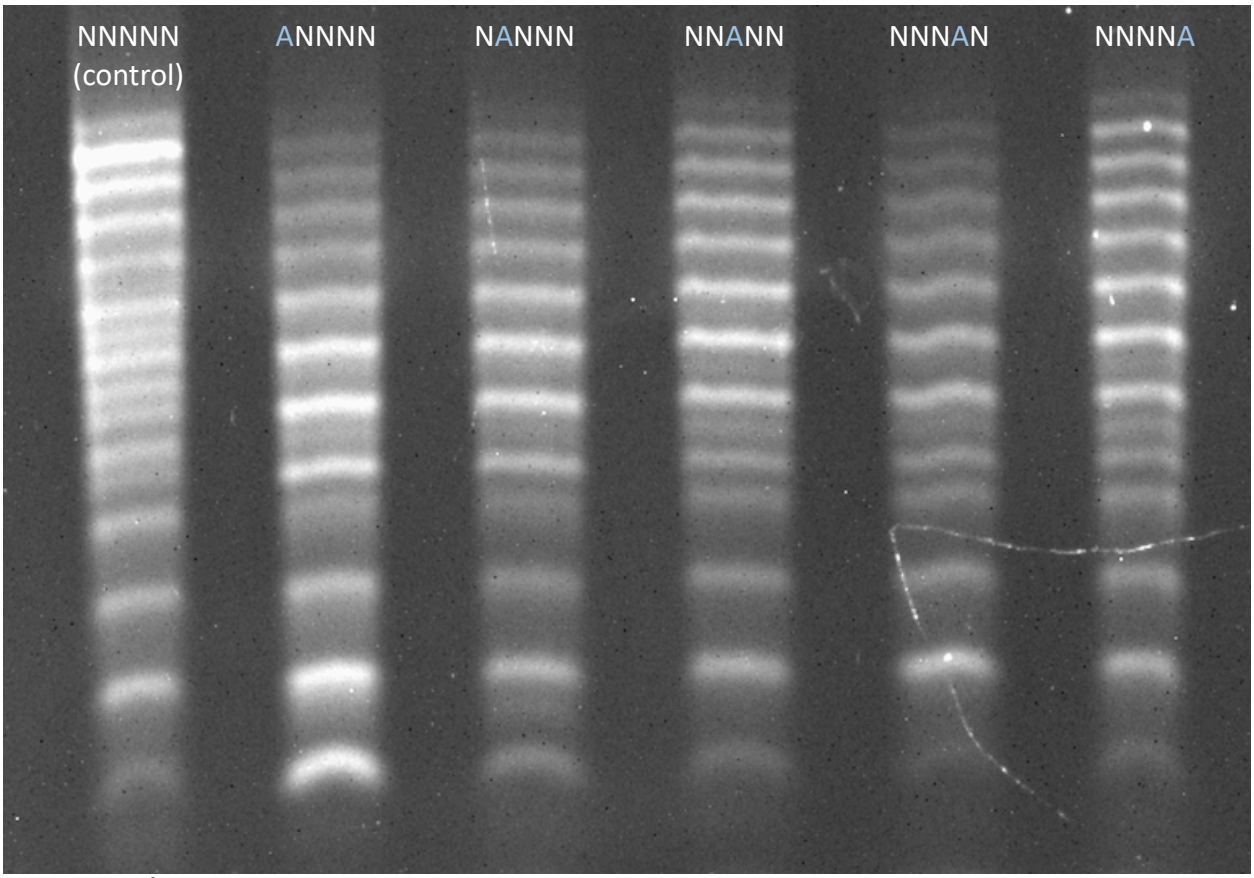


Figure 48: m^{6}A-modified pentamer libraries during RNA LOOPER along the DNA N40 template.

4.4 Fidelity and bias analysis (from HPSeq)

Compiling our optimizations from earlier, we decided to evaluate the fidelities of various RNA LOOPER products; the unmodified pentanucleotide anticodon (N5), m^{6}A-modified pentanucleotide anticodons modified at each position, and trinucleotide and tetranucleotide products. Unfortunately, the fidelities were much lower than expected. As seen previously with the 2'-F products, a higher yield does not necessarily mean that there will be a higher fidelity. Since the fidelities are low, it is instructive to extrapolate any trends from the data. The reason for the low fidelity has not been fully explored; however, a few parameters could explain this issue. We could reason that an incorrect pentanucleotide is annealing instead due to proximity in space. Before the correct pentanucleotide can anneal instead, T4 RNA ligase II ligates the incorrect one.

Further analysis would need to be done on RNA LOOPER fidelity involving T3 and T4 DNA ligases, since these have already proven to be successful in DNA oligonucleotide polymerizations.

library	fidelity
rN5	11.00%
NNN (DNA)	88.50%
(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%
rN3	16.60%
rN4	8.60%

Table 5: Fidelities of the RNA LOOPER system using various RNA anticodon libraries.

RNA LOOPER was shown to be unsuccessful using T4 RNA ligase II and the optimized method. However, further analysis will need to be done using other enzymes to conclude that the LOOPER system does not work with RNA polymerization. T3 and T4 DNA ligases have proven not only to be successful enzymes involved in DNA-based LOOPER systems, but they have also shown to successfully ligate RNA strands in our control experiments. Finding an enzyme that allows for accurate incorporations of RNA oligonucleotides will be crucial before exploring modified anticodons.

4.5 Sequences

Anticodons
cont-40mer-RNA
5'/phos/ACCUAACCCAUUGCACAAAAACUGCUAAAACGACAGUACC

RNA trimer
5'/phos/rNrNrN

RNA tetramer
5'/phos/rNrNrNrN

RNA pentamer
5'/phos/rNrNrNrNrN

Templates
PEG-CONT-40mer-temp
5'GATTCGCCTGCCGTCGCAGGTACTGTCGTTTTAGCAGTTTTTGTGCAATGGGTT
AGGTCACGTGGAGCTCGGATCCT/iSp18/AACAACAACAACAA

N39temp
GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNCACGTGGAGCTCGGATCCT

N40temp
GATTCGCCTGCCGTCGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNCACGTGGAGCTCGGATCCT

5HP_RNAlooper_N40
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCCGCGTGCGG
ACTCCAGCTA

5HP _IT1_N32
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNCACGTGGAGCTCGGATCCGCGTGCGGACTCCAGC
TA

5HP_IT1_N24
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANNNNNNNNNNN
NNNNNNNNNNNNNCACGTGGAGCTCGGATCCGCGTGCGGACTCCAGCTA

5HP_IT1_TNNNN8
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCATNNNNTNNNNTN
NNNTNNNNTNNNNTNNNNTNNNNTNNNNCACGTGGAGCTCGGATCCGCGTGCGGAC
TCCAGCTA

5HP_IT1_NTNNN8
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANTNNNNTNNNNT
NNNNTNNNNTNNNNTNNNNTNNNNTNNNCACGTGGAGCTCGGATCCGCGTGCGGAC
TCCAGCTA

5HP_IT1_NNTNN8
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANNTNNNNTNNN
NTNNNNTNNNNTNNNNTNNNNTNNNNTNNCACGTGGAGCTCGGATCCGCGTGCGGA
CTCCAGCTA

5HP_IT1_NNNTN8
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANNNTNNNNTNN
NNTNNNNTNNNNTNNNNTNNNNTNNNNTNCACGTGGAGCTCGGATCCGCGTGCGGA
CTCCAGCTA

5HP_IT1_NNNNT8
5'/phos/TGCGACGGCAGGCGAATC**CAAT**GATTCGCCTGCCGTCGCANNNNTNNNNTN
NNNTNNNNTNNNNTNNNNTNNNNTNNNNTCACGTGGAGCTCGGATCCGCGTGCGGA
CTCCAGCTA

<u>Primers</u>
primer1-DNA
5'/phos/TGCGACGGCAGGCGAATC

PEG-primer1-DNA
5'/phos/TGCGACGGCAGGCGAATC/iSp18/AACAACAACAACAA

primer1-RNA
5'/phos/rUrGrCrGrArCrGrGrCrArGrGrCrGrArArUrC

Fluorescent-primer2-DNA
5'/56FAM/AGGATCCGAGCTCCACGTG

Fluorescent-primer2-RNA
5'Fl-rGrGrArUrCrCrGrArGrCrUrCrCrArCrGrUrG

3P_IT1
5'AGGCACGGCGAGCGATCT<u>rGrGrArUrCrCrGrArGrCrUrCrCrArCrGrUrG</u>

Ion Torrent PCR primers
IT1_3PRev_Seq
CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT
TAGCTGGAGTCCGCACGC

IT1_IC_0130
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCCAAGCTGC
GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0131
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCTTACACAC
GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0132
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG
TTCTCATTGAAC GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0133
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TCGCATCGTTC
GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0134
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG
TAAGCCATTGTC GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0135
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG AAGGAATCGTC
GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0136
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG
CTTGAGAATGTC GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0137
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG
TGGAGGACGGAC GGTGAT AGGCACGGCGAGCGATCT

IT1_IC_0138
CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TAACAATCGGC
GGTGAT AGGCACGGCGAGCGATCT

4.6 Methods

4.6.1 T4 RNA ligase II protocols

Control reaction

In a PCR tube, a mixture of 2μL T4 Rnl2 Reaction Buffer (10X), 3 μL primer1 (primer1-RNA or primer1-DNA or PEG-primer1-DNA, 10 μM in H_2O), 3 μL primer2 (Fluorescent-primer2-RNA, 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 10U T4 RNA ligase II (New England Biolabs, M0239L). The mixture was allowed to polymerize for 18 hours at 25 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

Pre-optimized anticodon LOOPER

In a PCR tube, a mixture of 2μL T4 Rnl2 Reaction Buffer (10X), 3 μL primer1 (primer1-RNA or primer1-DNA or PEG-primer1-DNA, 10 μM in H_2O), 3 μL primer2-RNA (Fluorescent-primer2-RNA, 10 μM in H_2O), 2 μL template (N39/40temp, 10 μM in H_2O), 1.6 μL $MgCl_2$ (100 mM), 4 μL 50% PEG 8000, 0.8/1.4/1.8 μL H_2O and 2.6/2/1.6 μL of rN3/rN4/rN5 anticodons accordingly (1 mM 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 10U T4 RNA ligase II (New England Biolabs, M0239L). The mixture was allowed to polymerize for 18 hours at 25 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) before denaturing PAGE.

Optimized anticodon LOOPER

In a PCR tube, a mixture of 2µL T4 Rnl2 Reaction Buffer (10X), 3 µL primer1 (primer1-RNA or primer1-DNA or PEG-primer1-DNA, 10 µM in H_2O), 3 µL primer2-RNA (Fluorescent-primer2-RNA, 10 µM in H_2O), 2 µL template (N39/40temp_PEG, 10 µM in H_2O), 0.6 µL H_2O, 6 µL 50% PEG6000, 2.4 µL of rN5 anticodon (1 mM 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 20U T4 RNA ligase II (New England Biolabs, M0239L). The mixture was allowed to polymerize for 2 hours at 33 °C then 4 hours at 30 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) and template was DNAse degraded (New England Biolabs, M0303S) before denaturing PAGE.

4.6.2 T4 DNA ligase protocols

Control reaction

In a PCR tube, a mixture of 5µL ligation reaction buffer 4X (40 mM $MgCl_2$, 24% w/v PEG6000, 40 mM DTT, 264 mM Tris pH 7.6), 3 µL primer1 (RNA or DNA or PEG-DNA, 10 µM in H_2O), 3 µL primer2 (Fluorescent-RNA or Fluorescent-DNA, 10 µM in H_2O), 2 µL template (PEG-CONT-39/40MER-TEMP, 10 µM in H_2O), 3 µL of RNA (CONT-39/40MER-RNA, 10 µM in H_2O) and 2 µL ATP (0.25 mM 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 BSA (2 mg/mL in H_2O) and 1 µL 400U T4 DNA ligase (New England Biolabs, M0202L). The mixture was allowed to polymerize for 18 hours at 25 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

Anticodon reaction

In a PCR tube, a mixture of 5µL ligation reaction buffer 4X (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 (39/40MER-TEMP, 10 µM in H_2O), 3.05/3.5/3.8 µL H_2O and 1.95/1.5/1.2 µL of rN3/rN4/rN5 anticodons accordingly (1 mM in H_2O) and 2 µL ATP (0.25 mM 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 BSA (2 mg/mL in H_2O) and 1 µL 400U T4 DNA ligase (New England Biolabs, M0202L). The mixture was allowed to polymerize for 18 hours at 25 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

4.6.3 T3 DNA ligase protocols

Control reaction

In a PCR tube, a mixture of 5µL T3 DNA ligase buffer (2X), 0.5 µL primer1 (RNA or DNA or PEG-DNA, 30 µM in H_2O), 0.5 µL primer2 (Fluorescent-RNA or Fluorescent-DNA, 30 µM in H_2O), 1 µL template (PEG-CONT-39/40MER-TEMP, 10 µM in H_2O), 0.5 µL of RNA (CONT-39/40MER-RNA, 30 µM in H_2O) and 0.5 µL 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 H_2O and 1 µL 3000U T3 DNA ligase (New England Biolabs, M0317L). The mixture was allowed to polymerize for 18 hours at 25 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

Anticodon reaction

In a PCR tube, a mixture of 5μL T3 DNA ligase buffer (2X), 0.5 μL primer1 (RNA or DNA, 30 μM in H_2O), 0.5 μL primer2 (3PF, 30 μM in H_2O), 1 μL template (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) 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 H_2O and 1 μL 3000U T3 DNA ligase (New England Biolabs, M0317L). The mixture was allowed to polymerize for 18 hours at 25 °C. The products were then purified with a Monarch® RNA Cleanup Kit (#T2030L) before PAGE analysis.

Chapter 5 – Impact and future work

LOOPER has proven to be a diverse method for generating modified oligonucleotides. As most nucleic acid therapeutics have chemical modifications, it proves to be a powerful and cost effective, and more accessible method for the research and development of modified oligonucleotides. Since the method has tolerated hydrophobic modifications with steric bulk on C and T nucleobases as well as modifications at the 2'-position, LOOPER can be further evaluated using phosphorothioate modifications and unnatural backbones and sugars. In addition to testing more diverse XNA modifications, combining nucleobase modifications with backbone modifications in the trinucleotide libraries could allow for chemically diverse aptamers that have an increased resistance to nuclease degradation. Further analysis can then be done using this system to confirm its compatibility with SELEX.

In addition to the generation of these diversely-modified sequences, LOOPER has the ability to reverse transcribe them back to their DNA form, enabling PCR amplification and sequencing analysis. This significance of reverse transcribing modified nucleic acids is immense as it allows for the characterization and amplification of these oligonucleotides. Since LOOPER has been able to reverse transcribe these four unique XNA modifications, this provides a method that amplifies a broad range of modifications using only commercially available reagents. Reverse-transcription LOOPER enables XNA research more accessible.

The nucleobase modified sub-libraries have been carefully selected to generate a 64-membered trinucleotide library that is sufficient for in vitro selections against a protein target. While the yield is rather low (19%), the complexity of the system is immense. The resulting fidelity from hairpin sequencing is 94.8%, which is an ideal fidelity for the evolution of aptamers. The G-quadruplex structured selection that Luca Rustico is working on will hopefully be able to result in a modified aptamer against human α-thrombin.

Unfortunately, LOOPER is not suitable for RNA polymers using T4 RNA ligase II due to the low fidelities. Although T3 and T4 DNA ligases have been proven to work successfully in LOOPER, T4 RNA ligase II does not have the ability to accurately polymerize RNA oligonucleotides in a library format. Fidelity analysis of LOOPER products using various ligases will be necessary before proceeding to evaluate modified RNA. As seen with the sugar modifications, a higher LOOPER yield does not equate to a higher fidelity.

LOOPER is a method that not only generates modified nucleic acid polymers for aptamer selections. It is a system that can transform aptamer therapeutics. Its tolerance for complex systems can result in therapeutics that not only surpass antibody therapeutic abilities but also can be tailored to display modifications that are not typically found in nature. Synthesizing these polymers on a large scale will present its challenges with the current LOOPER system. Hopefully future iterations of this system will be able to overcome this challenge.

References

(1) Lu, R. M.; Hwang, Y. C.; Liu, I. J.; Lee, C. C.; Tsai, H. Z.; Li, H. J.; Wu, H. C. Development of Therapeutic Antibodies for the Treatment of Diseases. *Journal of Biomedical Science 2020 27:1* **2020**, *27* (1), 1–30. https://doi.org/10.1186/S12929-019-0592-Z.

(2) Stayton, P. S.; Freitag, S.; Klumb, L. A.; Chilkoti, A.; Chu, V.; Penzotti, J. E.; To, R.; Hyre, D.; Le Trong, I.; Lybrand, T. P.; Stenkamp, R. E. Streptavidin-Biotin Binding Energetics. *Biomol Eng* **1999**, *16* (1–4), 39–44. https://doi.org/10.1016/S1050-3862(99)00042-X.

(3) Koenig, S. H.; Brown Iii, R. D. H2CO3-as Substrate for Carbonic Anhydrase in the Dehydration of HCO3 (Metalloenzymes/Enzyme Mechanism/CO2 Hydration). *Proc Natl Acad Sci USA* **1972**, *69* (9), 2422–2425.

(4) Chou, K. C.; Zhou, G. P. Role of the Protein Outside Active Site on the Diffusion-Controlled Reaction of Enzyme. *J Am Chem Soc* **1982**, *104* (5), 1409–1413. https://doi.org/10.1021/JA00369A043/ASSET/JA00369A043.FP.PNG_V03.

(5) Chung, C.; Kudchodkar, S. B.; Chung, C. N.; Park, Y. K.; Xu, Z.; Pardi, N.; Abdel-Mohsen, M.; Muthumani, K. Expanding the Reach of Monoclonal Antibodies: A Review of Synthetic Nucleic Acid Delivery in Immunotherapy. *Antibodies 2023, Vol. 12, Page 46* **2023**, *12* (3), 46. https://doi.org/10.3390/ANTIB12030046.

(6) Gesteland RF; Cech TR; Atkins JF. *The RNA World, Third Edition*. Cold Spring Harbor Monograph Series. https://www.cshlpress.com/default.tpl?action=full&--eqskudatarq=536 (accessed 2025-01-23).

(7) Serganov, A.; Patel, D. J. Ribozymes, Riboswitches and beyond: Regulation of Gene Expression without Proteins. *Nature Reviews Genetics 2007 8:10* **2007**, *8* (10), 776–790. https://doi.org/10.1038/nrg2172.

(8) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346*, 818–822.

(9) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249* (4968), 505–510. https://doi.org/10.1126/SCIENCE.2200121.

(10) Robertson, D. L.; Joyce, G. F. Selection in Vitro of an RNA Enzyme That Specifically Cleaves Single-Stranded DNA. *Nature* **1990**, *344*, 467–468.

(11) 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* **2021**, *50* (8), 5126–5164. https://doi.org/10.1039/D0CS01430C.

(12) Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M. A Versatile Toolbox for Variable DNA Functionalization at High Density. *J Am Chem Soc* **2005**, *127* (43), 15071–15082. https://doi.org/10.1021/JA051725B.

(13) 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* **2020**, *48* (21), 11982. https://doi.org/10.1093/NAR/GKAA999.

(14) 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* **2009**, *10* (12), 1988–1992. https://doi.org/10.1002/CBIC.200900314.
(15) 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* **2018**, *9* (7), 1813–1821. https://doi.org/10.1039/C7SC04491G.
(16) 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* **1998**, *26* (22), 5203. https://doi.org/10.1093/NAR/26.22.5203.
(17) Hili, R.; Niu, J.; Liu, D. R. DNA Ligase-Mediated Translation of DNA into Densely Functionalized Nucleic Acid Polymers. *J Am Chem Soc* **2013**, *135* (1), 98–101. https://doi.org/10.1021/JA311331M/SUPPL_FILE/JA311331M_SI_001.PDF.
(18) Lei, Y.; Kong, D.; Hili, R. A High-Fidelity Codon Set for the T4 DNA Ligase-Catalyzed Polymerization of Modified Oligonucleotides. *ACS Comb Sci* **2015**, *17* (12), 716–721. https://doi.org/10.1021/ACSCOMBSCI.5B00119.
(19) Shi, K.; Bohl, T. E.; Park, J.; Zasada, A.; Malik, S.; Banerjee, S.; Tran, V.; Li, N.; Yin, Z.; Kurniawan, F.; Orellana, K.; Aihara, H. T4 DNA Ligase Structure Reveals a Prototypical ATP-Dependent Ligase with a Unique Mode of Sliding Clamp Interaction. *Nucleic Acids Res* **2018**, *46* (19), 10474–10488. https://doi.org/10.1093/NAR/GKY776.
(20) Lei, Y.; Washington, J.; Hili, R. Efficiency and Fidelity of T3 DNA Ligase in Ligase-Catalysed Oligonucleotide Polymerisations. *Org Biomol Chem* **2019**, *17* (7), 1962. https://doi.org/10.1039/C8OB01958D.
(21) Guo, C.; Watkins, C. P.; Hili, R. Sequence-Defined Scaffolding of Peptides on Nucleic Acid Polymers. *J Am Chem Soc* **2015**, *137* (34), 11191–11196. https://doi.org/10.1021/JACS.5B07675.
(22) Guo, C.; Hili, R. Fidelity of the DNA Ligase-Catalyzed Scaffolding of Peptide Fragments on Nucleic Acid Polymers. *Bioconjug Chem* **2017**, *28* (2), 314–318. https://doi.org/10.1021/ACS.BIOCONJCHEM.6B00647/SUPPL_FILE/BC6B00647_SI_001.PDF.
(23) Eigen, M. Selforganization of Matter and the Evolution of Biological Macromolecules. *Naturwissenschaften* **1971**, *58* (10), 465–523. https://doi.org/10.1007/BF00623322/METRICS.
(24) Sarma, R. H.; Lee, C. H.; Evans, F. E.; Yathindra, N.; Sundaralingam, M. Probing the Interrelation between the Glycosyl Torsion, Sugar Pucker, and the Backbone Conformation in C(8) Substituted Adenine Nucleotides by 1H and 1H-(31P) Fast Fourier Transform Nuclear Magnetic Resonance Methods and Conformational Energy Calculations. *J Am Chem Soc* **1974**, *96* (23), 7337–7348. https://doi.org/10.1021/JA00830A028.
(25) Luyten, I.; Herdewijn, P. Hybridization Properties of Base-Modified Oligonucleotides within the Double and Triple Helix Motif. *Eur J Med Chem* **1998**, *33* (7–8), 515–576. https://doi.org/10.1016/S0223-5234(98)80016-0.
(26) Kong, D.; Lei, Y.; Yeung, W.; Hili, R. Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angew Chem Int Ed Engl* **2016**, *55* (42), 13164–13168. https://doi.org/10.1002/ANIE.201607538.
(27) Kong, D.; Yeung, W.; Hili, R. In Vitro Selection of Diversely Functionalized Aptamers. *J Am Chem Soc* **2017**, *139* (40), 13977–13980. https://doi.org/10.1021/jacs.7b07241.

(28) Brudno, Y.; Birnbaum, M. E.; Kleiner, R. E.; Liu, D. R. An In Vitro Translation, Selection, and Amplification System for Peptide Nucleic Acids. *Nat Chem Biol* **2009**, *6* (2), 148. https://doi.org/10.1038/NCHEMBIO.280.
(29) MacPherson, I. S.; Temme, J. S.; Habeshian, S.; Felczak, K.; Pankiewicz, K.; Hedstrom, L.; Krauss, I. J. Multivalent Glycocluster Design through Directed Evolution. *Angew Chem Int Ed Engl* **2011**, *50* (47), 11238. https://doi.org/10.1002/ANIE.201105555.
(30) Lei, Y.; Hili, R. Structure–Activity Relationships of the ATP Cofactor in Ligase-Catalysed Oligonucleotide Polymerisations. *Org Biomol Chem* **2017**, *15* (11), 2349–2352. https://doi.org/10.1039/C6OB02792J.
(31) Guo, C.; Mahdavi-Amiri, Y.; Hili, R. Influence of Linker Length on Ligase-Catalyzed Oligonucleotide Polymerization. *Chembiochem* **2019**, *20* (6), 793–799. https://doi.org/10.1002/CBIC.201800616.
(32) 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* **2018**, *10* (4), 420. https://doi.org/10.1038/S41557-018-0008-9.
(33) 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* **2020**, *9* (1), 43–52. https://doi.org/10.1021/ACSSYNBIO.9B00222.
(34) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. Selection of Single-Stranded DNA Molecules That Bind and Inhibit Human Thrombin. *Nature* **1992**, *355* (6360), 564–566. https://doi.org/10.1038/355564A0.
(35) 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* **2021**, *23*, 863. https://doi.org/10.1016/J.OMTN.2021.01.004.
(36) 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* **2020**, *34* (4), 475–485. https://doi.org/10.1007/S10557-020-06986-Y.
(37) Spiridonova, V. A.; Novikova, T. M.; Snigirev, O. V. Obtaining DNA Aptamers to Human Interleukin-6 for Biomagnetic Immunoassay Nanosensors. *Moscow University Physics Bulletin* **2016**, *71* (1), 135–138. https://doi.org/10.3103/S002713491601015X/METRICS.
(38) Lichtor, P. A.; Chen, Z.; Elowe, N. H.; Chen, J. C.; Liu, D. R. Side Chain Determinants of Biopolymer Function during Selection and Replication. *Nat Chem Biol* **2019**, *15* (4), 419–426. https://doi.org/10.1038/S41589-019-0229-2.
(39) Le, A. T. H.; Krylova, S. M.; Beloborodov, S. S.; Wang, T. Y.; Hili, R.; Johnson, P. E.; Li, F.; Veedu, R. N.; Belyanskaya, S.; Krylov, S. N. 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* **2021**, *93* (13), 5343–5354. https://doi.org/10.1021/ACS.ANALCHEM.1C00601.
(40) Le, A. T. H.; Krylova, S. M.; Kanoatov, M.; Desai, S.; Krylov, S. N. Ideal-Filter Capillary Electrophoresis (IFCE) Facilitates the One-Step Selection of Aptamers. *Angew Chem Int Ed Engl* **2019**, *58* (9), 2739–2743. https://doi.org/10.1002/ANIE.201812974.

(41) Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. β-Peptides: From Structure to Function. *Chem Rev* **2001**, *101* (10), 3219–3232. https://doi.org/10.1021/CR000045I/ASSET/CR000045I.FP.PNG_V03.
(42) Seebach, D.; Gardiner, J. β-Peptidic Peptidomimetics. *Acc Chem Res* **2008**, *41* (10), 1366–1375. https://doi.org/10.1021/AR700263G/ASSET/IMAGES/MEDIUM/AR-2007-00263G_0014.GIF.
(43) Katoh, T.; Suga, H. Ribosomal Incorporation of Consecutive β-Amino Acids. *J Am Chem Soc* **2018**, *140* (38), 12159–12167. https://doi.org/10.1021/JACS.8B07247/SUPPL_FILE/JA8B07247_SI_003.XLSX.
(44) Katoh, T.; Sengoku, T.; Hirata, K.; Ogata, K.; Suga, H. Ribosomal Synthesis and de Novo Discovery of Bioactive Foldamer Peptides Containing Cyclic β-Amino Acids. *Nature Chemistry 2020 12:11* **2020**, *12* (11), 1081–1088. https://doi.org/10.1038/s41557-020-0525-1.
(45) Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure. *Science (1979)* **1999**, *284* (5423), 2118. https://doi.org/10.1126/SCIENCE.284.5423.2118.
(46) Chaput, J. C. Redesigning the Genetic Polymers of Life. *Acc Chem Res* **2021**, *54* (4), 1056–1065. https://doi.org/10.1021/ACS.ACCOUNTS.0C00886/ASSET/IMAGES/MEDIUM/AR0C00886_0009.GIF.
(47) Cromwell, C. R.; Sung, K.; Park, J.; Krysler, A. R.; Jovel, J.; Kim, S. K.; Hubbard, B. P. Incorporation of Bridged Nucleic Acids into CRISPR RNAs Improves Cas9 Endonuclease Specificity. *Nat Commun* **2018**, *9*, 1448. https://doi.org/10.1038/s41467-018-03927-0.
(48) Brad Wan, W.; Seth, P. P. The Medicinal Chemistry of Therapeutic Oligonucleotides. *J Med Chem* **2016**, *59* (21), 9645–9667. https://doi.org/10.1021/ACS.JMEDCHEM.6B00551/ASSET/IMAGES/MEDIUM/JM-2016-00551S_0023.GIF.
(49) Dunn, M. R.; Mccloskey, C. M.; Buckley, P.; Rhea, K.; Chaput, J. C. Generating Biologically Stable TNA Aptamers That Function with High Affinity and Thermal Stability. *J Am Chem Soc* **2020**, *142* (17), 7721–7724. https://doi.org/10.1021/JACS.0C00641/SUPPL_FILE/JA0C00641_SI_001.PDF.
(50) Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M. Unexpected Origins of the Enhanced Pairing Affinity of 2′-Fluoro-Modified RNA. *Nucleic Acids Res* **2011**, *39* (8), 3482–3495. https://doi.org/10.1093/NAR/GKQ1270.
(51) Movahedi, M.; Khamissi, N.; Hili, R. Ligase-Catalyzed Oligonucleotide Polymerization (LOOPER): Evolution of Chemically Diverse Aptamer Libraries. *Aptamers* **2021**, *5*, 22–30.
(52) Zhou, J.; Rossi, J. Aptamers as Targeted Therapeutics: Current Potential and Challenges. *Nature Reviews Drug Discovery 2016 16:3* **2016**, *16* (3), 181–202. https://doi.org/10.1038/nrd.2016.199.
(53) Ondruš, M.; Sýkorová, V.; Hocek, M. Traceless Enzymatic Synthesis of Monodispersed Hypermodified Oligodeoxyribonucleotide Polymers from RNA Templates. *Chemical Communications* **2022**, *58* (80), 11248–11251. https://doi.org/10.1039/D2CC03588J.
(54) Fang, Z.; Feng, X.; Tang, F.; Jiang, H.; Han, S.; Tao, R.; Lu, C. Aptamer Screening: Current Methods and Future Trend towards Non-SELEX Approach. *Biosensors 2024, Vol. 14, Page 350* **2024**, *14* (7), 350. https://doi.org/10.3390/BIOS14070350.

(55) Cui, L.; Ma, R.; Cai, J.; Guo, C.; Chen, Z.; Yao, L.; Wang, Y.; Fan, R.; Wang, X.; Shi, Y. RNA Modifications: Importance in Immune Cell Biology and Related Diseases. *Signal Transduction and Targeted Therapy 2022 7:1* **2022**, *7* (1), 1–26. https://doi.org/10.1038/s41392-022-01175-9.
(56) Jonkhout, N.; Tran, J.; Smith, M. A.; Schonrock, N.; Mattick, J. S.; Novoa, E. M. The RNA Modification Landscape in Human Disease. *RNA* **2017**, *23* (12), 1754–1769. https://doi.org/10.1261/RNA.063503.117/-/DC1.

Appendix

FIGURES

Figure S1. LOOPER yields involving FANA modifications. **A**) LOOPER products involving TTT trinucleotides with FANA modifications. **B**) LOOPER products involving NNN trinucleotide libraries with FANA modifications.

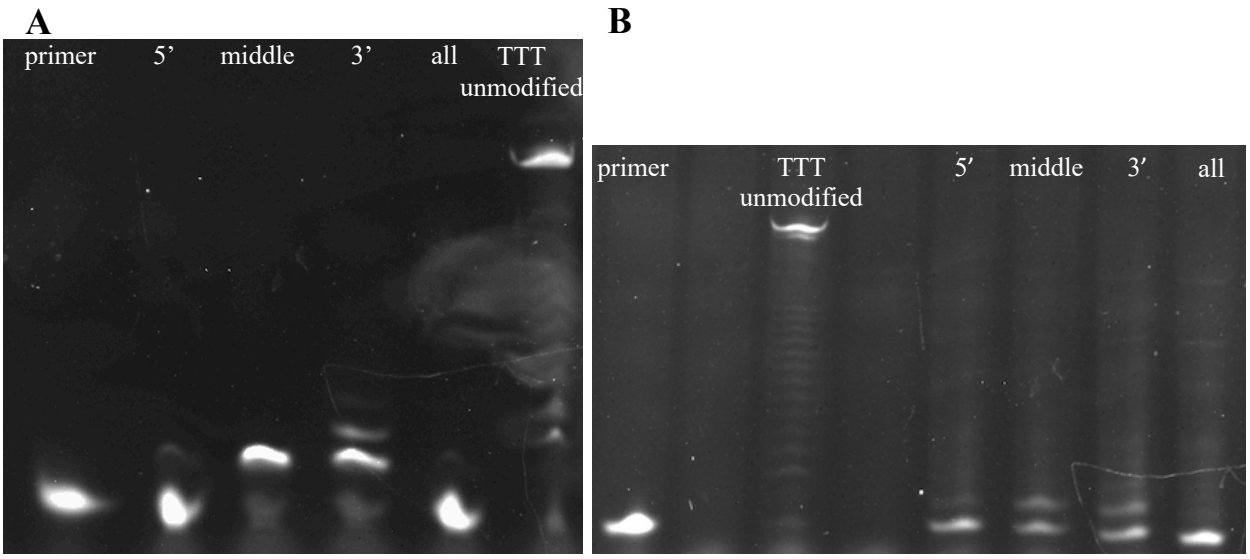


Figure S2. LOOPER yields involving 2'OMe modifications. **A**) LOOPER products involving TTT trinucleotides with 2'OMe modifications. **B**) LOOPER products involving NNN trinucleotide libraries with 2'OMe modifications.

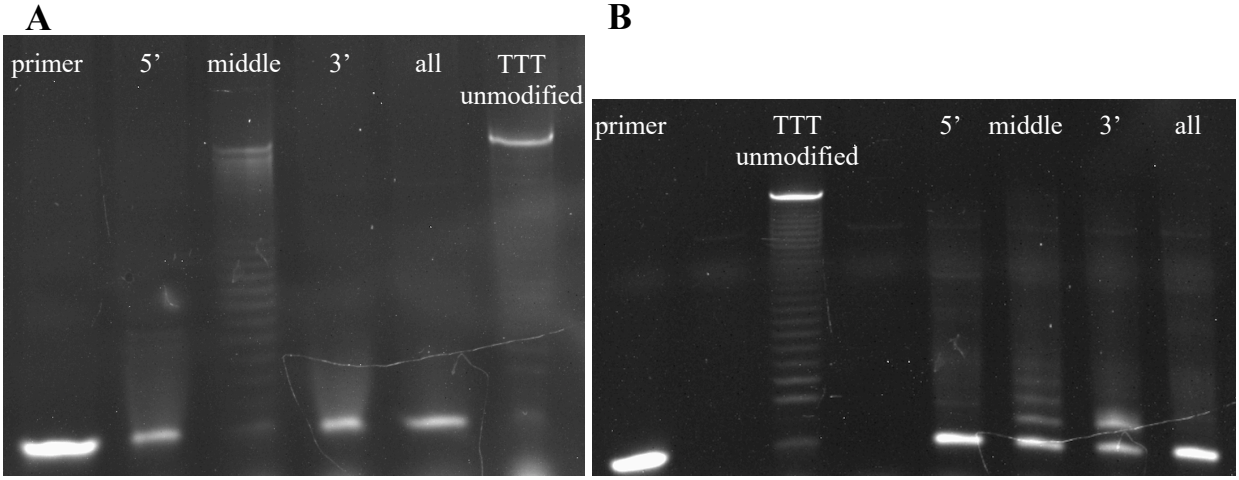


Figure S3. LOOPER yields involving LNA modifications. **A**) LOOPER products involving TTT trinucleotides with LNA modifications. **B**) LOOPER products involving NNN trinucleotide libraries with LNA modifications.

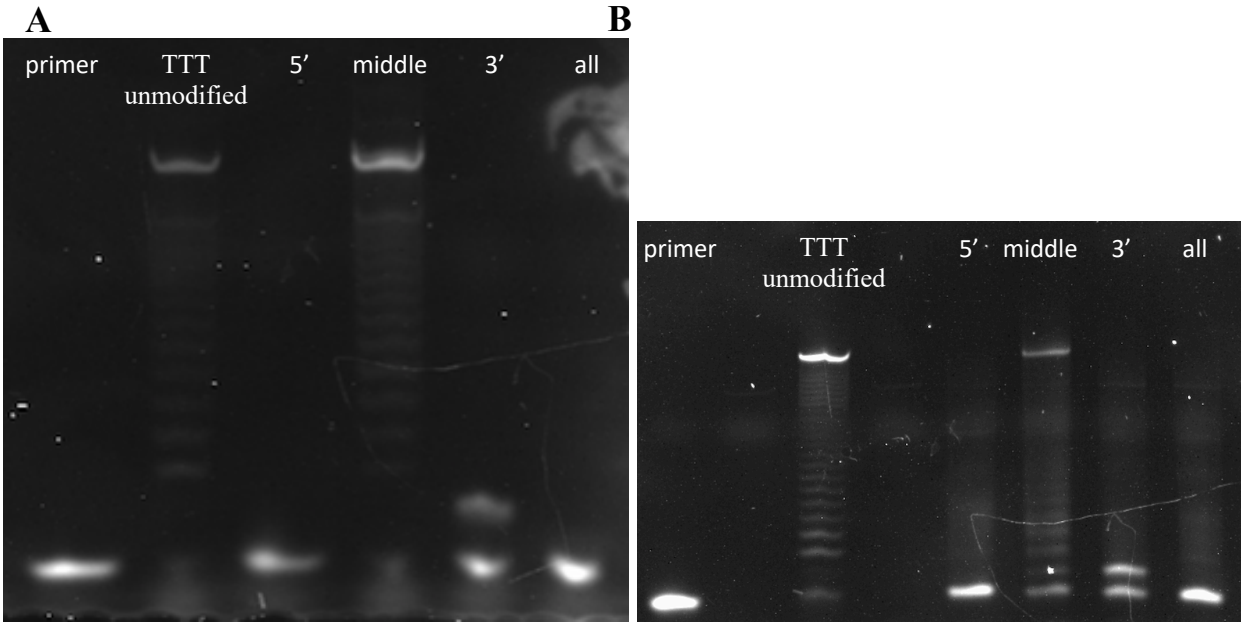


Figure S4. The hairpin product after LOOPER before and after staining. The product band is fluorescently labelled.

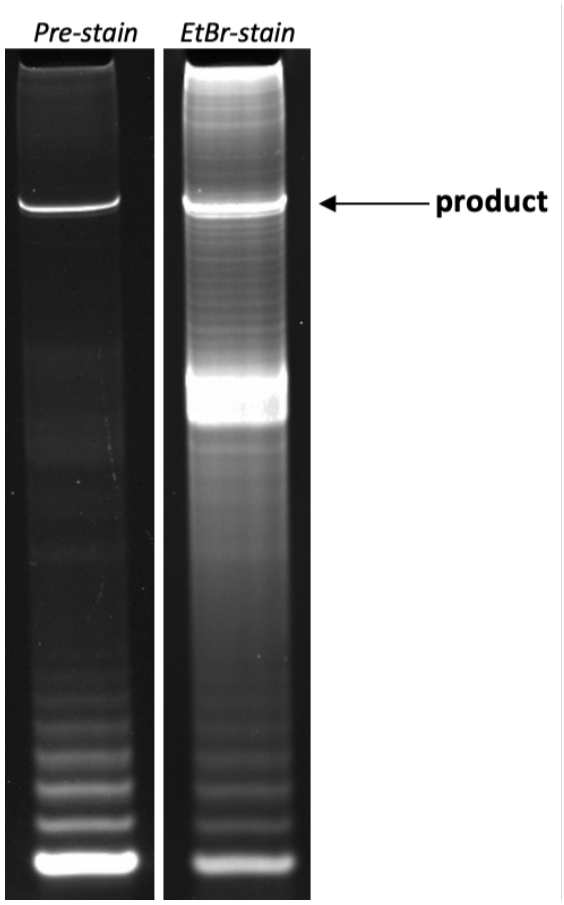
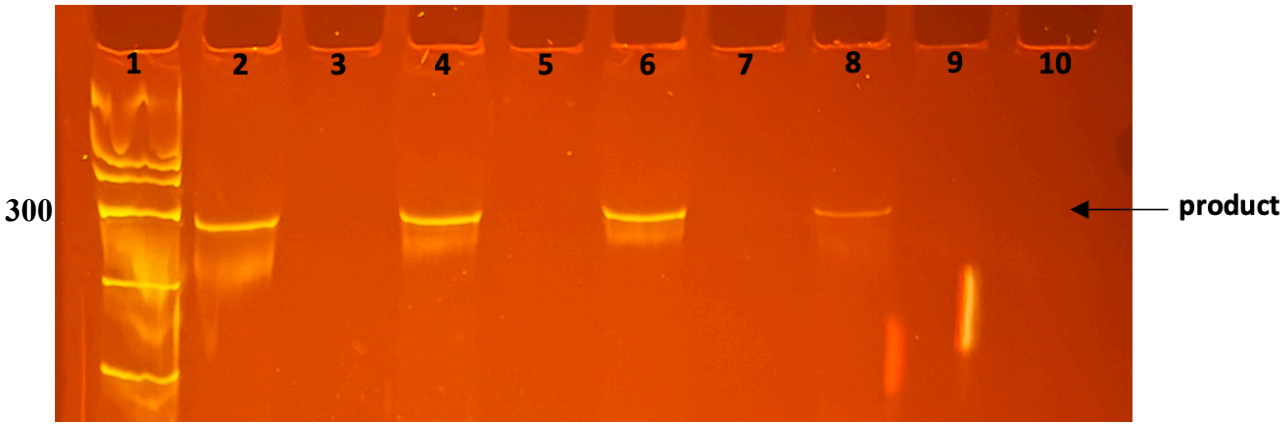
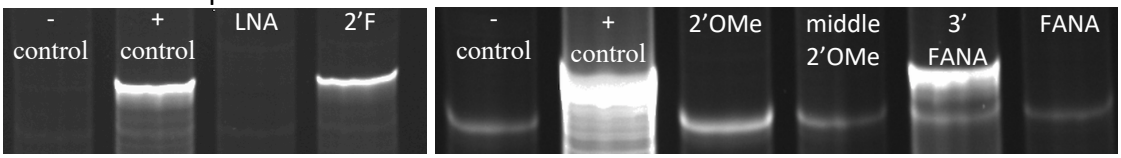


Figure S5. PCR amplified products of the LOOPER hairpins.



Lane 1: dsDNA ladder (ThermoFisher, SM1193), Lane 2: NNN unmodified LOOPER – IC_0130, Lane 4: 5'F-NNN LOOPER – IC_0131, Lane 6: midF-NNN LOOPER – IC_0132, Lane 8: 3'F-NNN LOOPER – IC_0133, Lane 10: allF-NNN LOOPER – IC_1034

Figure S6. PCR amplification of the XNA-modified templates that were subsequently involved in reverse-transcription LOOPER.



TABLES

Table S1. Duplex sequencing matched pairs and fidelity of each codon within each read from both sequencing runs attempted.

	Round 1			Round 2		
	Matched Reads	**Codon Occurrence**	**Fidelity**	**Matched Reads**	**Codon Occurrence**	**Fidelity**
R8-L1	15	195	97.9%	4	52	92.3%
5'F-NNN	11	143	94.4%	5	65	98.5%
middle F-NNN	40	520	78.1%	25	325	79.7%
3'F-NNN	1	13	92.3%	3	39	97.4%
all F-NNN	1	13	92.3%	1	13	100%

Table S2. Hairpin Sequencing output data for the unmodified control

Observed template	Template occurences	1mer errors	3mer errors	3mer bias	Expected LOOPER
TTT	925	67	64	0.947027	AAA
GTT	1063	178	169	0.907808	AAC
CTT	617	72	63	0.951378	AAG
ATT	788	122	112	0.878173	AAT
TGT	1312	162	156	0.90625	ACA
GGT	919	250	238	0.817193	ACC
CGT	720	119	111	0.938889	ACG
AGT	732	128	121	0.912568	ACT
TCT	763	92	88	0.954128	AGA
GCT	559	84	79	1.121646	AGC
CCT	366	49	44	1.071038	AGG
ACT	465	76	70	0.984946	AGT
TAT	600	53	44	0.98	ATA
GAT	591	79	73	0.939086	ATC
CAT	522	63	63	0.944444	ATG

AAT	465	47	44	0.956989	ATT
TTG	931	74	68	0.941998	CAA
GTG	819	60	56	1.059829	CAC
CTG	529	30	24	1.100189	CAG
ATG	626	100	95	0.875399	CAT
TGG	782	86	76	0.952685	CCA
GGG	490	149	145	0.777551	CCC
CGG	414	47	42	1.016908	CCG
AGG	439	53	46	0.984055	CCT
TCG	537	97	88	0.888268	CGA
GCG	346	68	64	0.956647	CGC
CCG	206	34	29	1.208738	CGG
ACG	311	61	55	0.980707	CGT
TAG	280	22	20	1.196429	CTA
GAG	439	33	29	1.116173	CTC
CAG	374	37	37	1.005348	CTG
AAG	339	39	36	0.967552	CTT
TTC	638	67	61	0.935737	GAA
GTC	521	87	82	0.992322	GAC
CTC	339	51	50	1.0059	GAG
ATC	467	61	59	0.950749	GAT
TGC	611	45	39	1.155483	GCA
GGC	383	46	42	1.592689	GCC
CGC	278	26	25	1.294964	GCG
AGC	307	15	13	1.429967	GCT
TCC	389	53	51	1	GGA
GCC	230	29	28	1.465217	GGC
CCC	227	29	26	1.202643	GGG
ACC	204	40	35	1.161765	GGT
TAC	460	40	39	0.986957	GTA
GAC	305	24	22	1.206557	GTC
CAC	243	7	7	1.572016	GTG
AAC	295	22	18	1.071186	GTT
TTA	597	46	42	0.988275	TAA
GTA	527	93	85	0.979127	TAC
CTA	371	48	42	0.967655	TAG
ATA	356	63	63	0.898876	TAT
TGA	646	64	59	0.955108	TCA
GGA	433	78	72	0.942263	TCC

CGA	278	34	30	1.039568	TCG
AGA	397	35	33	1.010076	TCT
TCA	484	63	62	0.952479	TGA
GCA	341	47	43	1.184751	TGC
CCA	232	35	31	1.133621	TGG
ACA	344	53	52	0.976744	TGT
TAA	410	40	37	1.012195	TTA
GAA	328	41	38	0.984756	TTC
CAA	263	28	25	1.060837	TTG
AAA	222	34	30	0.941441	TTT

Table S3. Hairpin Sequencing output data for 5'-F NNN library

Observed template	Template occurences	1mer errors	3mer errors	3mer bias	Expected LOOPER
TTT	1731	249	233	0.883882	AAA
GTT	1963	429	404	0.817117	AAC
CTT	1251	165	153	0.947242	AAG
ATT	1437	208	200	0.90675	AAT
TGT	2530	372	364	0.877075	ACA
GGT	1695	393	373	0.812389	ACC
CGT	1334	154	143	0.988756	ACG
AGT	1269	164	156	0.919622	ACT
TCT	1630	242	234	0.953374	AGA
GCT	1180	242	233	1.105932	AGC
CCT	858	106	105	1.165501	AGG
ACT	949	167	156	1.139094	AGT
TAT	1062	146	136	0.925612	ATA
GAT	969	121	113	1.048504	ATC
CAT	874	107	106	1.033181	ATG
AAT	942	864	830	0.141189	ATT
TTG	1860	134	130	0.947849	CAA
GTG	1661	174	163	0.951836	CAC
CTG	1069	82	77	1.086997	CAG
ATG	1211	173	168	0.885219	CAT
TGG	1885	216	204	0.922016	CCA
GGG	868	189	186	0.860599	CCC
CGG	922	76	69	1.093275	CCG
AGG	917	102	93	0.955289	CCT
TCG	1142	191	182	0.904553	CGA

GCG	773	110	108	1.028461	CGC
CCG	507	45	43	1.64497	CGG
ACG	709	207	203	0.834979	CGT
TAG	708	49	47	1.035311	CTA
GAG	732	34	32	1.061475	CTC
CAG	747	51	45	1.042838	CTG
AAG	628	47	45	0.977707	CTT
TTC	1387	147	141	0.981255	GAA
GTC	1151	186	178	0.961772	GAC
CTC	712	93	91	1.16573	GAG
ATC	881	163	157	0.905789	GAT
TGC	1264	137	127	1.110759	GCA
GGC	737	103	100	1.313433	GCC
CGC	556	32	31	1.370504	GCG
AGC	600	76	72	1.093333	GCT
TCC	939	173	166	1.096912	GGA
GCC	618	77	74	1.605178	GGC
CCC	423	33	31	1.815603	GGG
ACC	552	156	148	1.03442	GGT
TAC	903	76	75	1.024363	GTA
GAC	590	42	38	1.169492	GTC
CAC	480	27	26	1.420833	GTG
AAC	469	50	46	1.857143	GTT
TTA	1142	126	118	0.950963	TAA
GTA	1046	239	230	0.844168	TAC
CTA	670	70	69	1.010448	TAG
ATA	725	108	105	0.910345	TAT
TGA	1346	147	136	0.946508	TCA
GGA	857	72	67	0.97783	TCC
CGA	552	43	40	1.11413	TCG
AGA	665	47	40	1.007519	TCT
TCA	957	135	129	0.984326	TGA
GCA	589	72	68	1.307301	TGC
CCA	463	41	37	1.336933	TGG
ACA	580	63	61	1.086207	TGT
TAA	677	61	60	0.986706	TTA
GAA	542	55	52	0.983395	TTC
CAA	399	34	31	1.050125	TTG
AAA	363	25	22	1.008264	TTT

Table S4. Hairpin Sequencing output data for mid-F NNN library

Observed template	Template occurences	1mer errors	3mer errors	3mer bias	Expected LOOPER
TTT	26504	24468	23402	0.124321	AAA
GTT	28544	8073	7549	0.791935	AAC
CTT	18708	3611	3465	1.106372	AAG
ATT	16672	15842	14994	0.105086	AAT
TGT	34250	10050	9775	0.748847	ACA
GGT	30758	10717	10495	0.701931	ACC
CGT	23724	5887	5794	0.863261	ACG
AGT	18418	4873	4707	0.786622	ACT
TCT	20784	5049	4864	1.200491	AGA
GCT	20391	6619	6417	0.921975	AGC
CCT	13285	2833	2772	1.117125	AGG
ACT	13210	3515	3290	1.258668	AGT
TAT	13144	8359	7902	0.407638	ATA
GAT	17208	4462	4235	0.798291	ATC
CAT	14219	2193	2115	1.047683	ATG
AAT	9104	6147	5738	0.3817	ATT
TTG	28856	2600	2452	0.974459	CAA
GTG	28695	3895	3664	0.903224	CAC
CTG	18182	1151	1078	1.116269	CAG
ATG	19065	3288	3139	0.878259	CAT
TGG	26776	3553	3418	0.906894	CCA
GGG	17684	2922	2879	0.92513	CCC
CGG	17651	1606	1525	1.070534	CCG
AGG	16514	1859	1744	0.934601	CCT
TCG	20414	2360	2279	0.978299	CGA
GCG	16871	2667	2581	1.028155	CGC
CCG	11094	895	856	1.334325	CGG
ACG	11991	2191	2137	0.973063	CGT
TAG	9003	1054	977	0.9679	CTA
GAG	13748	1050	937	0.983343	CTC
CAG	12852	411	359	1.079365	CTG
AAG	9516	1179	1098	0.916982	CTT
TTC	20232	2058	1952	1.754893	GAA
GTC	18772	3174	3044	1.163115	GAC
CTC	12183	972	941	1.343347	GAG

ATC	12855	2102	1992	1.42746	GAT
TGC	21191	2048	1968	1.434288	GCA
GGC	15742	1342	1269	1.785097	GCC
CGC	12298	388	362	1.760774	GCG
AGC	10858	748	694	1.518143	GCT
TCC	13614	1222	1179	1.466579	GGA
GCC	11761	1115	1057	1.885724	GGC
CCC	7614	201	199	1.988442	GGG
ACC	8512	1297	1247	1.382989	GGT
TAC	12839	1385	1318	1.327829	GTA
GAC	11222	791	707	1.414543	GTC
CAC	8921	279	251	1.379442	GTG
AAC	7248	676	621	1.501932	GTT
TTA	16016	15337	14575	0.09528	TAA
GTA	16625	5204	4847	0.783459	TAC
CTA	11337	1834	1769	1.277939	TAG
ATA	10019	9915	9365	0.075057	TAT
TGA	20131	4450	4315	0.833838	TCA
GGA	17036	4881	4793	0.767434	TCC
CGA	11473	1663	1616	1.025713	TCG
AGA	11267	2546	2502	0.825686	TCT
TCA	15046	2320	2263	1.268643	TGA
GCA	12455	3119	3024	1.005379	TGC
CCA	9049	1049	1003	1.307437	TGG
ACA	9206	1925	1871	1.275038	TGT
TAA	8819	2718	2564	0.741354	TTA
GAA	10359	2589	2466	0.818515	TTC
CAA	8615	731	686	1.085781	TTG
AAA	5176	1270	1201	0.827859	TTT

Table S5. Hairpin Sequencing output data for 3'-F NNN library

Observed template	Template occurences	1mer errors	3mer errors	3mer bias	Expected LOOPER
TTT	5215	750	696	0.8907	AAA
GTT	5823	1094	1019	0.851623	AAC
CTT	3799	509	467	0.947355	AAG
ATT	3681	528	488	0.888346	AAT
TGT	9679	2130	2063	0.803905	ACA
GGT	6855	1894	1835	0.754778	ACC

CGT	5307	754	729	1.028453	ACG
AGT	4161	509	491	0.9118	ACT
TCT	5275	770	734	0.995071	AGA
GCT	4049	958	940	0.996542	AGC
CCT	2754	399	385	1.102033	AGG
ACT	2902	437	408	0.954859	AGT
TAT	4633	4437	4259	0.087632	ATA
GAT	3838	619	567	0.895518	ATC
CAT	3250	315	292	1.600923	ATG
AAT	2347	274	250	0.927567	ATT
TTG	5568	536	497	0.925108	CAA
GTG	5036	538	500	0.936259	CAC
CTG	3302	232	211	1.13507	CAG
ATG	3137	593	563	0.842843	CAT
TGG	5966	813	767	0.893899	CCA
GGG	3032	537	514	0.91095	CCC
CGG	3289	255	233	1.134387	CCG
AGG	2950	429	393	0.898644	CCT
TCG	3672	426	402	0.962963	CGA
GCG	2586	264	248	1.104022	CGC
CCG	1671	80	73	1.43447	CGG
ACG	1947	400	389	0.894196	CGT
TAG	2081	267	244	0.983662	CTA
GAG	2460	193	177	0.993902	CTC
CAG	2523	138	124	1.086405	CTG
AAG	1820	229	214	0.911538	CTT
TTC	3878	516	476	0.946622	GAA
GTC	3421	581	560	0.978077	GAC
CTC	2069	241	234	1.147414	GAG
ATC	2186	371	353	0.910796	GAT
TGC	4529	495	452	1.273129	GCA
GGC	2585	246	214	1.728046	GCC
CGC	2072	75	69	1.626448	GCG
AGC	1881	174	159	1.222222	GCT
TCC	2620	296	278	1.264504	GGA
GCC	1861	134	118	1.895755	GGC
CCC	1222	39	38	1.900164	GGG
ACC	1285	225	213	1.189105	GGT
TAC	3111	343	330	1.469945	GTA

GAC	2100	202	180	1.167619	GTC
CAC	1840	69	62	1.434239	GTG
AAC	1503	162	155	1.02994	GTT
TTA	3180	455	417	0.892453	TAA
GTA	2776	610	566	0.842579	TAC
CTA	2120	302	277	0.981604	TAG
ATA	1713	385	356	0.813193	TAT
TGA	4806	731	693	0.882855	TCA
GGA	3246	508	484	0.895564	TCC
CGA	1988	104	93	1.226358	TCG
AGA	2219	223	209	0.936458	TCT
TCA	3328	367	351	1.009916	TGA
GCA	2184	274	252	1.139652	TGC
CCA	1614	95	90	1.346344	TGG
ACA	1837	257	248	0.961894	TGT
TAA	2092	1888	1859	0.123805	TTA
GAA	1934	228	211	0.975181	TTC
CAA	1667	96	88	1.754649	TTG
AAA	946	101	100	0.959831	TTT

Table S6. Hairpin Sequencing output data for the Q5 DNA polymerase control

Expected LOOPER	Observed template	Template occurrences	3mer errors
AAA	TTT	86	1
AAC	GTT	64	1
AAG	CTT	96	6
AAT	ATT	155	7
ACA	TGT	73	0
ACC	GGT	17	2
ACG	CGT	46	7
ACT	AGT	84	2
AGA	TCT	107	5
AGC	GCT	31	0
AGG	CCT	76	12
AGT	ACT	114	3
ATA	TAT	118	7
ATC	GAT	62	2
ATG	CAT	138	14
ATT	AAT	187	5

CAA	TTG	64	3
CAC	GTG	20	2
CAG	CTG	61	8
CAT	ATG	71	3
CCA	TGG	29	3
CCC	GGG	7	1
CCG	CGG	25	2
CCT	AGG	25	2
CGA	TCG	41	13
CGC	GCG	24	4
CGG	CCG	28	6
CGT	ACG	47	2
CTA	TAG	56	10
CTC	GAG	27	1
CTG	CAG	62	6
CTT	AAG	59	2
GAA	TTC	115	18
GAC	GTC	46	7
GAG	CTC	91	22
GAT	ATC	130	22
GCA	TGC	71	2
GCC	GGC	9	1
GCG	CGC	36	12
GCT	AGC	42	2
GGA	TCC	100	13
GGC	GCC	35	9
GGG	CCC	50	11
GGT	ACC	115	17
GTA	TAC	130	1
GTC	GAC	51	2
GTG	CAC	116	16
GTT	AAC	160	6
TAA	TTA	124	7
TAC	GTA	63	1
TAG	CTA	122	7
TAT	ATA	146	6
TCA	TGA	76	2
TCC	GGA	25	1
TCG	CGA	52	3

TCT	AGA	71	2
TGA	TCA	149	22
TGC	GCA	36	4
TGG	CCA	102	29
TGT	ACA	170	1
TTA	TAA	169	8
TTC	GAA	79	4
TTG	CAA	147	18
TTT	AAA	189	5

Table S7. ESI-MS data confirmation of the synthesized libraries from Chapter 3.

Library	Members	Product Peaks Observed
Cyclopentyl-TNT	TGT	386.75164 (z=3)
	TAT	381.42005 (z=3)
	TTT	378.41628 (z=3)
	TCT	373.41639 (z=3)
Indole-TNG	TGG	616.6246 (z=2)
	TAG	608.6276 (z=2)
	TTG	604.1219 (z=2)
	TCG	596.6219 (z=2)
cyclopropyl-TNC	TGC	559.1143 (z=2)
	TAC	551.1170 (z=2)
	TTC	546.6113 (z=2)
	TCC	539.1115 (z=2)
Amino-TNA	TGA	530.1016 (z=2)
	TAA	522.1039 (z=2)
	TTA	517.5980 (z=2)
	TCA	510.0993 (z=2)
TNA (natural form)	TGA	963.2619
	TAA	947.2348
	TTA	938.3118
	TCA	923.1989
hydroxymethyl-CNT	CGT	969.4062
	CAT	953.5063
	CTT	944.3775
	CCT	929.4103
propyl-CNC	CGC	581.0560 (z=2)

	CAC	573.0023 (z=2)
	CTC	568.3774 (z=2)
	CCC	560.9735 (z=2)
Amino-CNG	CGG	565.6401 (z=2)
	CAG	557.6423 (z=2)
	CTG	553.1369 (z=2)
	CCG	545.6366 (z=2)
Carboxy-CNG	CGG	1008.0973
	CAG	992.4363
	CTG	983.4783
	CCG	968.3273
methoxy acetic acid-CNG	CGG	601.64 (z=2)
	CAG	593.64 (z=2)
	CTG	589.14 (z=2)
	CCG	581.64 (z=2)
hydroxymethyl-CNC	CGC	1092.435
	CAC	1076.4668
	CTC	1067.4282
	CCC	1052.2917
MDPA-CNA	MDPA-CGA	1278.317
	MDPA-CAA	1262.3222
	MDPA-CTA	1253.3103
	MDPA-CCA	1238.3098
ANN	ACC	908.284
	ACT & ATC	923.189
	ACA & AAC	932.4885
	ATT	938.2036
	ATA & AAT	947.3009
	AGC & ACG	948.3376
	AAA	956.3256
	ATG & AGT	963.2671
	AAG & AGA	972.3723
	AGG	988.3848
GNN	GCC	924.2796
	GCT & GTC	939.3158
	GCA & GAC	948.3549
	GTT	954.5196
	GTA & GAT	963.3202
	GGC & GCG	964.3070

	GAA	972.2984
	GTG & GGT	979.4343
	GAG & GGA	988.2438
	GGG	1004.2296