ENZYMATIC SYNTHESIS OF HIGHLY MODIFIED OLIGONUCLEOTIDES: APPLICATION TO THE EVOLUTION OF MODIFIED APTAMERS

by

YI LEI

(Under the Direction of Ryan Hili)

ABSTRACT

Aptamers, single-stranded nucleic acid polymers, are known to bind specific targets with high affinity and specificity. However, functionalities of natural aptamers are limited by four canonical DNA bases compared with 20 amino acids. This deficit of nucleic acids makes it difficult for aptamers to match the performance of antibodies as affinity reagents. Traditional incorporation of chemical functionality has relied on the individual modified nucleobases, which limits the number of unique modifications up to four. T4 DNA ligase-catalyzed oligonucleotide polymerization (LOOPER) has the ability to expand this limitation by catalyzing the ligation of modified 5'-phosphorylated oligonucleotides on a single-stranded DNA template. Thus, the number of unique modifications relies on the codons rather than single nucleotides. Hexylamine modified adenosine is the chemical handle to introduce various functionalities to the natural nucleic acid polymers through amide bond formation. The high fidelity of LOOPER system has been confirmed by 'Duplex Sequencing' method.

Different functional groups such as hydroxyl groups and aromatic rings are successfully incorporated on the hexylamine modified ANNNN building block libraries. These homofunctionalized ANNNN libraries have high polymerization (>60%) and fidelity (>90%), as well as low codon bias. The LOOPER system retains high efficiency and fidelity when applied to

a complex heterofunctionalized system: 16 sub-libraries with different modifications. Thus, LOOPER system enables the generation of diversely functionalized nucleic acid polymers. Current efforts are being directed to the *in vitro* selection of highly modified aptamers against glycosaminoglycans.

INDEX WORDS: Aptamers, T4 DNA ligase, DNA sequencing, LOOPER, SELEX

ENZYMATIC SYNTHESIS OF HIGHLY MODIFIED OLIGONUCLEOTIDES: APPLICATION TO THE EVOLUTION OF MODIFIED APTAMERS

by

YI LEI

BS, University of Science and Technology of China, China, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

ENZYMATIC SYNTHESIS OF HIGHLY MODIFIED OLIGONUCLEOTIDES: APPLICATION TO THE EVOLUTION OF MODIFIED APTAMERS

by

YI LEI

Major Professor:	Ryan Hili
Committee:	Vladimir Popik
	Geert-Jan Boons

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

ACKNOWLEDGEMENTS

I would like to thank Ryan and my committee members first. Ryan has been a great mentor for the past few years. Your passion for science has encouraged me to stay positive when I face with difficulties. You create a research lab without too much pressure, yet I have always proceeded with my projects at a comfortable pace. I am glad that I chose your lab in 2013 while I didn't know too much about your research at that time and I have always enjoyed my research. To my committee members, Dr. Popik and Dr. Boons, thank you for your helpful suggestions and guidance!

I want to thank my fantastic lab members. You not only give me advice about science but also help me with things like seminar presentations. I do enjoy the time we have spent together talking about science and life. You create a wonderful environment, and we support each other along the way. To my family and friends, you have always been a rock behind me. There were some depressing times, but your support helps me get through all those difficulties. I am getting my Ph.D. because of you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

LITERATURE REVIEW AND INTRODUCTION

Nucleic acid structures

Nucleic acids are biopolymers or small biomolecules made from repeating units called nucleotides. The nucleotide monomers are composed of a five-carbon sugar, a phosphate group and a nitrogenous base. The polymer is ribonucleic acid (RNA) if the sugar is a compound ribose; if the sugar is derived from ribose as deoxyribose, the polymer is deoxynucleic acid (DNA).

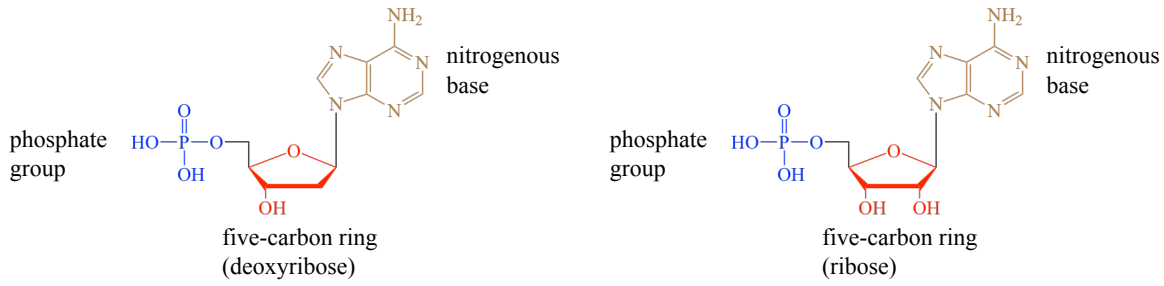


Figure 1.1 Monomer structures of nucleic acids

DNA was first discovered by Johann Friedrich Miescher in 1869[1]. There are four canonical bases to form DNA: adenine (A), guanine (G), thymine (T) and cytosine (C). In RNA, thymine is replaced by uracil (U). The nucleotide monomers linked together by phosphodiester bonds form a linear sequence, known as the primary structure of nucleic acids. It is the order of nucleotides within a DNA or RNA molecule determined by a series of letters. The base-paring interactions within a single nucleic acid polymer or between two polymers determine their secondary structure. In the canonical Watson-Crick base pairing, A forms a base pair with T, G forms another pair with C.

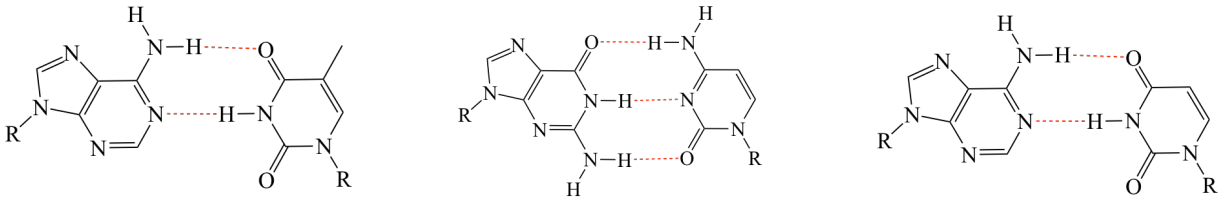


Figure 1.2 Watson-Crick base pairing

The nucleotides are connected by hydrogen bonds and form different shapes, such as the double helix, stem-loop structures, and pseudoknots. The DNA sequences can further fold into a specific three-dimensional shape, which is required in various functions such as molecular recognition and catalysis. The double helix structure of DNA was discovered building on X-ray analysis[2]. The structure of DNA double helix is held together by a number of weak interactions such as hydrogen bonds, stacking interactions and hydrophobic effects. It requires regions of many consecutive base pairs.

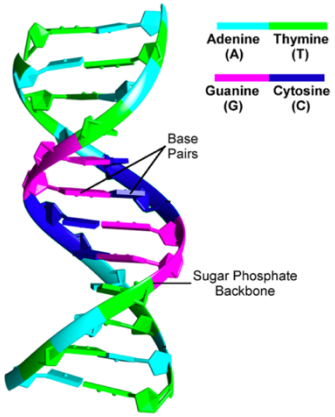


Figure 1.3 Double helix structure of DNA[3]

Nucleic acid functions

It was first suggested that DNA was the carrier of genetic information in 1944 by Oswald. T Avery, Colin Macleod and Maclyn MacCarty[4]. And this function was confirmed by Alfred Hershey and Martha Chase in 1952[5].

RNA has been found to be evolved in various biochemical reactions. It was first suspected that RNA played a role in protein synthesis in 1939[6]. In 1967, Carl Woese hypothesized that RNA might be catalytic, and the earliest forms of life could have relied on RNA to carry genetic information and catalyze biochemical reactions simultaneously – an RNA world[7]. In the early 1970s, it was shown for the first time that enzymes could copy RNA into DNA[8]. David Baltimore, Renato Dulbecco and Howard Temin were awarded Nobel Prize in 1975 for this work. Catalytic RNA molecules (ribozymes) were discovered in the early 1980s[9], and RNA interference was found in 1990[10].

Molecular recognition is central to biological processes, function and specificity. For examples, naturally occurring RNA switches, known as riboswitches, are capable of binding to small-molecule ligands. They have been reported to bind specifically to essential coenzymes and vitamins, amino acids, glucosamine-6-phosphate, and purine bases[11]. Diverse riboswitches are present and able to control the gene expression process[12].

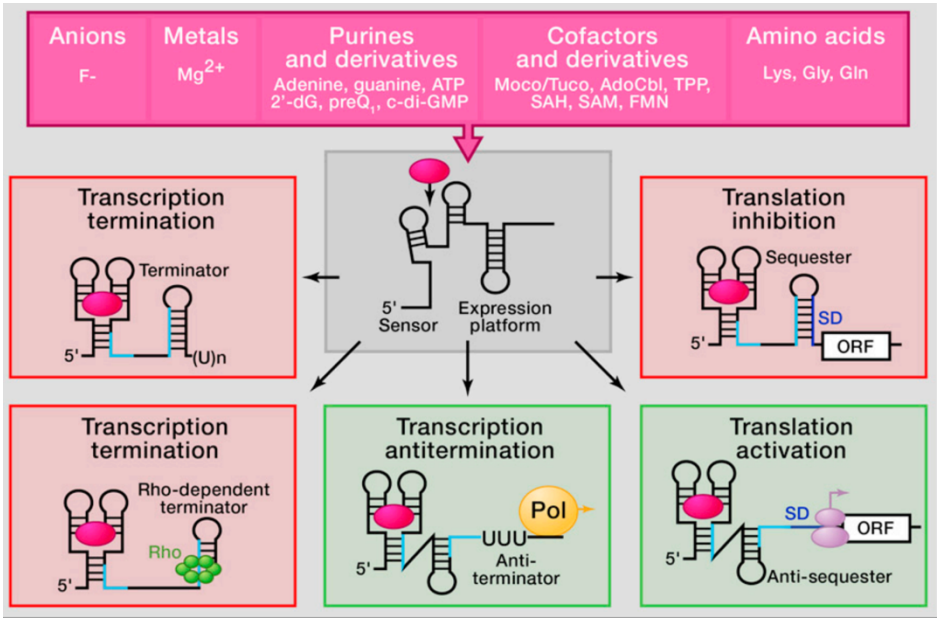


Figure 1.4 Diversity of riboswitches and mechanisms of gene control in bacteria

There are various metabolites present in cells above threshold concentrations so that they can be directly sensed and specifically bound by sensor domains of riboswitches. The binding process typically induces the structural changes in the adjacent expression platform region. The structural elements on the expression platform can either stimulate or repress gene expression. Both translation and transcription processes can be controlled.

Researchers have been working on artificial functional nucleic acids for a variety of purposes such as ligand binding and catalysis[13].

Introduction of nucleic acid aptamers

The word 'aptamer' is derived from Latin 'aptus' ('to fit') and Greek work 'meros' ('part'). They are single-stranded nucleic acid molecules that can fold into discrete 3D structures with ligand-binding sites that are complementary in shape and charge to the desired target[14].

Biological nucleic acids have long been known to form complex shapes and act as scaffolds for molecular interactions. However, methods for the discovery of non-biological nucleic acids with similar functions have been developed only in recent years with a series of technological advances, such as DNA synthetic methods, PCR and high-throughput DNA sequencing.

In 1990, Tuerk and Gold reported their first high-affinity RNA aptamer for the T4 DNA polymerase using an *in vitro* procedure[15]. The RNA ligands were enriched from an 8-nucleotide random region library. The procedure is named *systematic evolution of ligands by exponential enrichment* (**SELEX**). The general process of SELEX is shown in Figure 1.5.

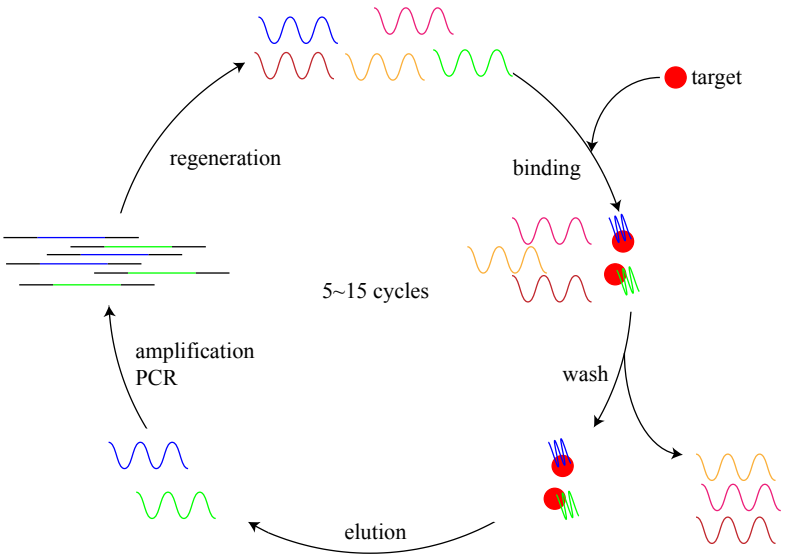


Figure 1.5 Scheme of SELEX rounds

The initial RNA pool consisted of 8 completely randomized nucleotides at specific positions and two primer sites for replication purpose. The RNA pool was subjected to selection against the T4 DNA polymerase target. The weak binders were washed away, while better binding sequences were amplified to dsDNA for further *in vitro* transcription. The newly transcribed RNA pool was subjected to the next cycle. Multiple rounds were performed to create a pool of variant sequences with relatively high affinity for T4 DNA polymerase target. The most enriched sequences could be identified by sequencing, and further used for binding affinity test. The sequencing result revealed two main clusters of binding sequences; one was known wild-type and another novel RNA sequence. The dissociation constant (K_D) was about 5nM, which was about 100 times better than non-specific binding. This was one of first two publications of the SELEX technique. It showed potential to determine the optimal binding sequences for any nucleic acid binding protein.

Another publication was from Ellington and Szostak. They demonstrated the evolutionary power of this power the same year[16]. The initial RNA pool consisted of ~10^{13} different sequences with ~100 bases in length. Subpopulations of RNA molecules that bound specifically to a variety

of organic dyes were successfully isolated using the SELEX technique. They used Cibacron Blue 3GA (CB), Reactive Red 120 (R), Reactive Yellow 86 (Y), Reactive Brown 10 (BR), Reactive Green 19 (GR) and Reactive Blue 4 (B4). These molecules had many possible hydrogen-bond donor and acceptor groups as well as planar surfaces for stack interactions. They also tested the specificity of selected RNA pools by a cross-binding experiment. The results indicated that the binding sites of RNA aptamers were capable of discriminating between functional groups on ligands.

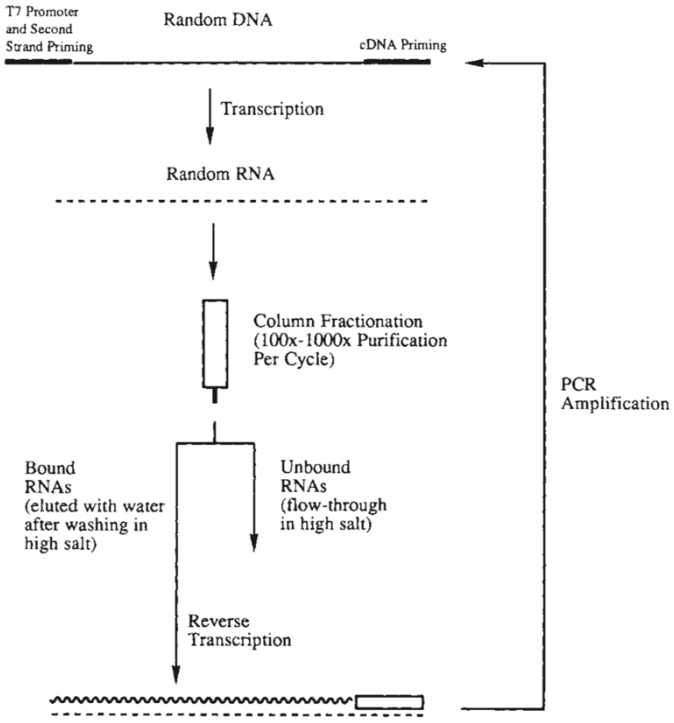


Figure 1.6 *In vitro* selection of RNA aptamers for organic dyes

To date, researchers have identified high-affinity aptamers against a wide range of targets, including proteins, peptides, whole cells, small molecules. Some aptamer examples against different types of targets are summarized in Table 1.1[17].

Target	Estimated K_D (nM)

Small molecules	
Tobramycin (aminoglycoside)	0.8[18]
Arginine	330[19]
NAD	2500[20]
Proteins	
HIV-1 tat protein	0.12[21]
VEGF-165	2.4[22]
Taq DNA polymerase	5.7[23]
Thrombin	102[24]

Table 1.1 Dissociation constants of aptamers and their targets

Aptamers offer several advantages over traditional antibodies[25-26]. Aptamers are produced chemically in a readily scalable process, which is not prone to viral or bacterial contamination. Their smaller sizes (6~30kDa, 2nm in diameter) compared with antibodies (150~180kDa, ~15nm in diameter) allow more efficient entry into a biological compartment and possibly enable small target binding. The phosphodiester backbone is extremely chemically stable, and conjugation chemistries can be readily introduced during synthesis.

Parameters of SELEX

The success of *in vitro* selection depends on factors including the design of nucleic acid library, target choices, selection methods, and experimental conditions.

Nucleic acid types

For many years, researchers preferentially selected RNA over DNA aptamers due to the suggestion that RNA results in higher binding affinity. However, the lack of 2' hydroxyl group of the DNA sugar leads to better chemical and biological stability. It has been calculated that neither DNA nor RNA yield aptamers with better binding affinity[27]. The binding affinities were reported as dissociation constant (K_D values). There was no significant difference between the mean Log10(K_D) for the different aptamer types.

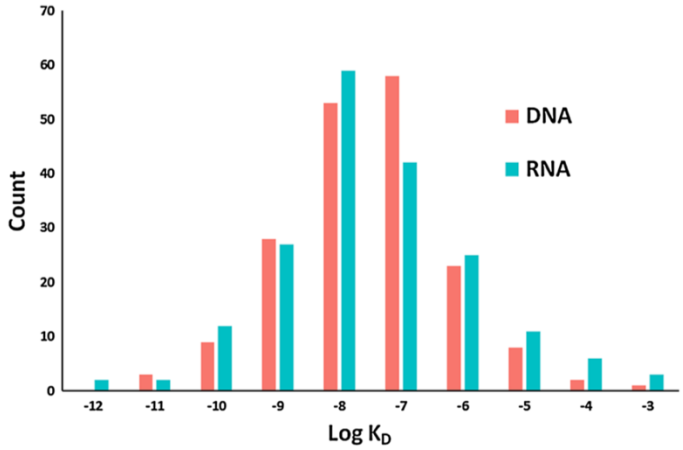


Figure 1.7 Distribution of the highest measured aptamer affinity[27]

Length of the aptamers

The random regions of N40 to N70 are commonly used. The length of random regions helps the balance between appropriate coverage of sequence space and high structural complexity.

Aptamer target

While aptamers are selected against all different kinds of targets, including small molecules, cells, and carbohydrate, the majority of successful selections are against proteins (58%, literature from 1990-2013[27]). Larger targets like proteins provide more functional groups for more possible interactions between the ligand and aptamers. More flexible targets such as carbohydrates can interfere with the process of *in vitro* selection, resulting in lower affinities of aptamers for carbohydrates than those for proteins.

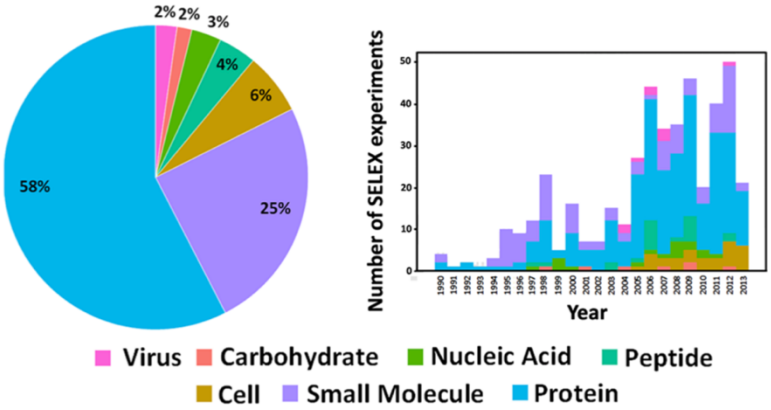


Figure 1.8 Distribution of reported aptamer targets from 1990-2013[27]

Selection conditions

Specific selection conditions such as metal cation concentrations, buffering agents, pH and temperature can all influence the process of SELEX. The secondary structures of aptamers are essential for binding activities, and they are significantly affected by the presence and concentration of metal cations in the buffer. The cations can also shield the negative charges of the DNA backbone, which might interfere with the potential electrostatic binding interactions of aptamers to the target[28]. pH is known to influence the structure and stability of aptamers and targets. It has been reported that binding activities of some aptamers are dependent on pH[28-30].

Selection method

There are different methods to immobilize targets, such as streptavidin magnetic beads capture and classic affinity chromatography method. Heat is the most commonly used method to recover the bound nucleic acids on targets. Other methods such as elution with target, precipitation and change in ionic strength have also been used. The choice of target immobilization and aptamer elution depends on the specific target type. The number of selection rounds depends on the enrichment progress, and usually it requires 5-15 rounds to achieve sufficient enrichment.

Structures of nucleic acid aptamers

Aptamers have a strong tendency to fold into discrete structures following the principals of base pairing. These three-dimensional interactions, including hydrophobic and electrostatic interactions, hydrogen bonding, *van der Waals* forces, shape complementarity and base stacking are essential for aptamer binding affinity and specificity. The success of SELEX results from a large collection of shapes generated from randomized nucleic acid strands. These shapes can recognize other molecules through shape and functional group complementarity[31]. The interactions are similar to the way that antibodies bind to antigens.

The first structure of tRNA was reported in 1974[32-33]. Studies of ribosomal RNA followed by X-ray and NMR structure determinations have revealed numerous structural motifs of natural RNA, which has benefited the first successful SELEX experiment with randomized RNA libraries. Motifs such as hairpin loops, internal loops and pseudoknots as well as structural elements such as base zippers, A-platforms, π-turns and α-loops have been well studied and extensively reviewed[34-35]. Thanks to the development of SELEX method and aptamer characterization technology, it is now clear the single-stranded DNA aptamers are as feasible[36].

The theophylline-binding RNA aptamer (5'-GGCGAUACCAGCCGAAAGGC CCUUGGCAGCGUC-3') was reported to bind theophylline with the dissociation constant K_D of 0.1μM. It was 10,000-fold better than the same RNA molecule affinity for caffeine, which differed from theophylline only by a methyl group at nitrogen atom N-7[37]. The structure of this RNA aptamer-theophylline complex was visualized in 1997 using NMR spectroscopy[38]. The different resonances revealed by ^{15}N, ^{1}H correlation spectrum indicated a large conformational rearrangement for the RNA aptamer upon binding of theophylline ligand.

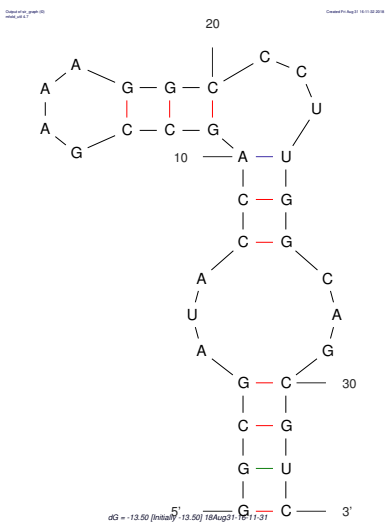


Figure 1.9 Predicted secondary structure of theophylline-binding RNA aptamer

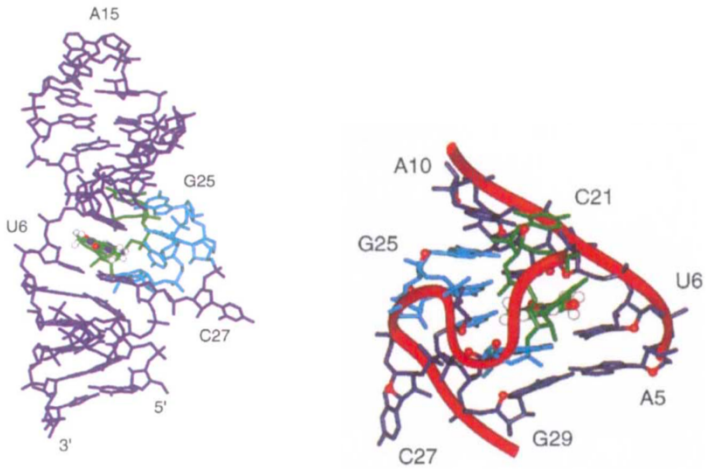


Figure 1.10 Stereoviews of the RNA-theophylline complex and binding pocket

The backbone of the 3' strand forms an unusual S-turn that allows the loops to interact while maintaining an overall linear conformation for the RNA. This S-turn results from the reversed sugar motif of U24, which is the center of the S-turn. These two motifs bring together the conserved residues of upper and lower loops to form the theophylline-binding pocket.

Theophylline binding depends upon the formation of a central base-triple interaction where it forms hydrogen bonds with the bases of C22 and U24. The base-triple composed of U6, U23

and A28 forms the floor of the binding pocket, stacking below the central C22-theophylline-U24 triple. Residues U6, C22, A7 and C21 stack in an interleaved fashion to form a novel 'base-zipper' on one side of the binding pocket. On the other side of the core, the base of G26 intercalates between the bases of U24 and G25 to form a unique 1-3-2 stack, which constitutes the final curve of the S-turn and stabilize the sharp bends in the backbone of globular RNA.

The base C22 is primarily responsible for the molecular recognition of theophylline and the discrimination against caffeine. Hydrogen bonding between C22 base and the purine-like theophylline gives rise to a pseudo-base pair. The additional bulky methyl group would disrupt this pairing alignment[39]. This was a well-studied RNA aptamer enabling high-resolution molecular discrimination.

One of the most famous DNA aptamers is the thrombin-binding aptamer. It was first discovered in 1992[40] (5'-GGTTGGTGTGGTTGG-3'). This 15-mer oligonucleotides were found to inhibit thrombin-catalyzed fibrin clot formation *in vitro*. Two-dimensional ^{1}H NMR revealed the unimolecular quadruplex structure the next year[41]. Different from the theophylline RNA aptamer described above, this DNA aptamer already adopts a define quadruplex structure in solution in the absence of the ligand.

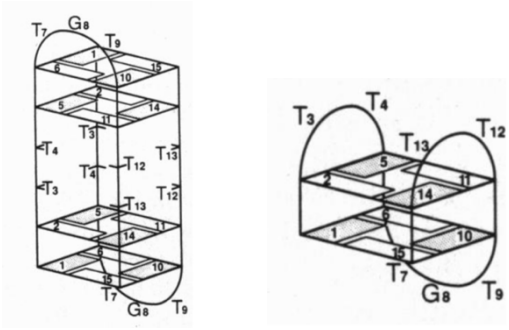


Figure 1.11 Schematic models for the dimer and monomer of the thrombin aptamer

It's of great significance to understand the structural rules that govern specific, high-affinity binding characteristic of aptamer-ligand interactions for many applications. The structures of

binding pockets found in aptamers with those of proteins are distinct despite their similar three-dimensional interactions[42]. The diversity of 20 amino acids allow for a multitude of interactions and precise shape complementarity in open substrate-binding sites. The fit of ligands into binding sites in aptamer folds displays a less-than-perfect shape complementarity due to the more structurally uniform four nucleotides. While the stacking interactions in aptamers are favored by the planarity of the nucleotide bases, intermolecular hydrogen bonds, and general acid-base interactions are preferred by antibodies. Efforts made to improve the characters of nucleic acid aptamers and better mimic antibodies will be described in the following sections.

Applications of aptamers

Aptamers are suitable for applications based on molecular recognition as analytical, diagnostic and therapeutic tools.

Analytical applications

Aptamers are capable of molecular recognition with high affinity and specificity, which makes them very promising in the field of affinity chromatography. Immobilized DNA aptamers have been used to resolve enantiomers of small bioactive molecules (D-adenosine and L-tyrosinamide), by HPLC on microbore columns[43]. Adenosine aptamers have been applied to monitor the adenosine level in the brain of live rats. The biotinylated aptamer was mixed with streptavidin-coated porous glass beads, then packed into a fused-silica capillary. The adenosine elution was injected into the column. The detection limit was 30nM with UV absorbance detection, and the column could be used for 200 injections[44].

Aptamers can also assist protein capture and analysis in MALDI-MS. It has been proved that thrombin aptamers can selectively capture thrombin from a complex matrix and then release the thrombin from those aptamer spots for further mass spectrometry analysis. Compared with traditional affinity reagent used in biosensors like monoclonal antibodies, aptamers are able to

extend the range of targets from small molecules to whole cells. Further, aptamers allow efficient immobilization at high density due to their relatively small size. Along with these two advantages, the high chemical stability of aptamers makes them reusable as biosensors. The targets can be released upon washing and the aptamers can fold back to the specific structure under appropriate conditions[45]. Optical bioassays are one of the many examples of using aptamers as bio-recognition elements in the biosensors. Fluorescent detection is widely used due to the ease of labeling aptamers with fluorescent dyes, the various fluorophores and quenchers, and the inherent capability for real-time detection. Among all different design of assay formats, the aptamer-based molecular beacon is most frequently adopted. Figure 1.12 is the scheme of a common design. A fluorophore-labeled aptamer in a duplex structure with a complementary DNA sequence labeled with a quencher is separated after the binding of the target and the aptamer, leading to an increase in fluorescence[46].

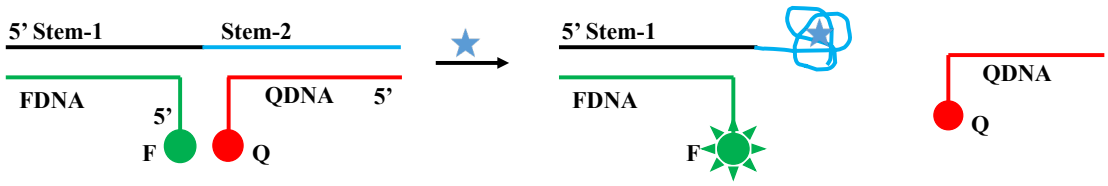


Figure 1.12 Design of an aptamer-based molecular beacon[46]

Aptamers are also used in "proximity ligation assay"[47]. The first assay was applied to the homodimer of the platelet-derived growth factor B-chain (PDGF-BB), a cytokine with growth- and differentiation-promoting effects[48]. The DNA aptamers were used as affinity probes and they were extended at either 5' or 3' end, forming a proximity probe pair. Upon pairs of probes binding to PDGF-BB, the free ends of their sequence extension were brought sufficiently close to hybridize together to a subsequently added connector oligonucleotide, allowing the enzymatic ligation of the free ends. The ligation products were then amplified by PCR, while unreacted probes remained silent. This proximity ligation assay provides a valuable method to investigate

modifications on a specific protein and also the interactions between biomolecules forming macromolecular complexes.

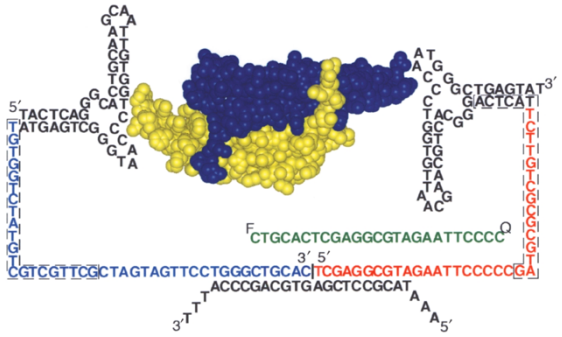


Figure 1.13 Schematic view of the aptamer-based proximity probe

Aptamer-based therapeutics

There are three typical aptamer-based therapeutics strategies: an aptamer can serve as an antagonist to block the interaction of disease-associated targets, such as protein-protein or receptor-ligand interactions; an aptamer can serve as an agonist to activate the function of target receptors; or a cell-type-specific aptamer can serve as a carrier to deliver other therapeutic agents to the target cells or tissue[26]. As the advantages of nucleic acid aptamers mentioned in previous sections, aptamer-based therapeutics show great potential over traditional monoclonal antibodies. Pegaptanib (Macugen; Eyetech Pharmaceuticals/Pfizer) is an RNA aptamer directed against vascular endothelial growth factor VEGF-165. It has been shown in clinical trials to be effective in treating choroidal neovascularization associated with age-related macular degeneration. It was approved by the US FDA in December 2004. Pegaptanib is the first aptamer therapeutic approved for use in humans[22].

However, the development of clinically effective therapeutic aptamers has lagged far behind that of therapeutic antibodies[49]. Their inherent physicochemical characteristics limit the *in vivo* therapeutic potency of aptamers. First of all, wild-type RNA and DNA aptamers are very

susceptible to nuclease degradation. The half-life of unmodified nucleotide aptamers in blood can be as short as two minutes[50]. Besides, aptamers are susceptible to rapid elimination from the blood by renal filtration due to the small size of aptamers (about 5-15kDa). Toxicities of aptamers can arise from polyanionic effects, unexpected tissue accumulation and nonspecific immune activation.

Modified nucleic acid polymers

Researchers have been working on modifications of natural nucleic acid to overcome the limitations mentioned above. The common strategies include modifications on the terminals of nucleic acid polymers, on the phosphodiester linkage, on the sugar rings and the bases. XNA (*xeno*-nucleic acid) was first prosed by Herdewijn and Marliere to describe any such synthetic genetic polymer with a focus on those that have shown potential for either chemical and/or enzymatic replication[51].

Resistance to nuclease degradation

Natural nucleic acid aptamers are poor candidates for diagnostic and therapeutic applications due to their limited stability in the biological environment.

The 3'-end capping with inverted thymidine has been a common strategy. It could increase the stability and resistance of aptamers to 3'-exonuclease in human serum[52]. The modification can be incorporated using solid-phase synthesis. 2'-hydroxyl groups of ribonucleotides are fundamental for phosphodiester bond cleavage by nucleases. 2'-substitutions such as 2'-fluoro, 2-amino or 2'-methoxy groups can confer resistance.

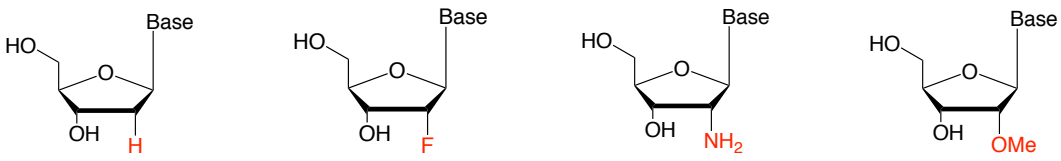


Figure 1.14 Common 2'-substitutions

An aptamer binding to VEGF composed entirely of 2'-OMe nucleotides was reported in 2005[53]. It was necessary to find conditions under which a template-dependent polymerase would accept 2'-OMe NTPs as substrates. After screening various conditions, they found a T7 RNA mutant that enabled transcription of 2'-OMe NTPs (0.5mM each) with only a spike of 2'-OH GTP (30μM). This method was applied to a direct selection of fully 2'-OMe-modified aptamers against VEGF. Two 23nt long aptamers were identified with K_D values of 2nM (ARC245) and 1nM (ARC224) respectively. The aptamers were highly resistant to chemical, physical thermal and enzymatic insults. The clearance half-life of the 5'-PEGylated ARC245 administrated to CD-1 mice was 23h.

Locked nucleic acid is an analog of ribonucleotide with a methylene linkage between 2'-O and 4'-C of the sugar ring. An aptamer against lymphoma Ramos cells was used as a model to test the improvement of stability in serum upon incorporation of LNA[54].

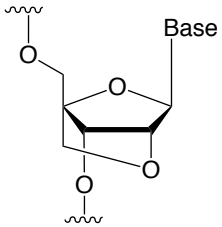


Figure 1.15 Structure of LNA

The modified aptamer with 7-base pair-LNA substitution and supplemented 3'-3'-thymidine (3'-3'-T) capping exhibited the significantly elevated detection stability half-life from ~0.5h of natural DNA aptamer to 5~6h of modified version for target cells in serum. Moreover, a much slower clearance rate in tumor-bearing mice was also observed, thus leading to the extended tumor imaging window from <150min of the natural aptamer to >600min. This discovery supported the feasibility and effectiveness of the LNA modification strategy to improve the aptamer-based tumor detection efficacy and imaging quality *in vivo*.

Replaced phosphodiester linkage of aptamers is another modification strategy for better biological stability. Methylphosphonate and phosphorothioate analogs are commonly used. Their impact on thermodynamic stability was studied[55-57]. Another replacement of phosphodiester linkage is the triazole linkage. There are several types of triazole linkage[58].

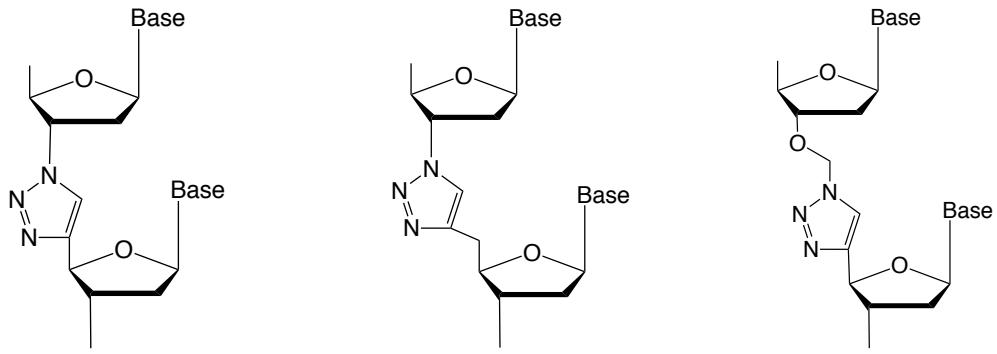
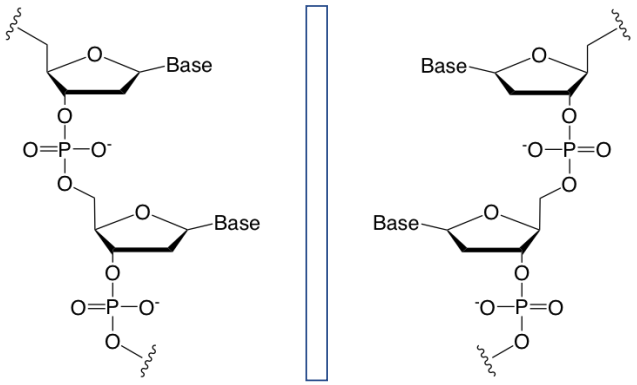


Figure 1.16 Fragments of oligonucleotides analogs with triazole linkages

A series of triazole-modified DNA aptamers with the similar structure to thrombin-binding aptamer TBA15 was synthesized and tested in 2013[59]. The rigid triazole linkage on central loop successfully protected the aptamers from nuclease digestion and retained the anticoagulant activity.

Natural DNAs are in D-form. The mirror image L-DNA (also named Spiegelmers) may display high resistance to the degradation of nucleases. A Spiegelmer with high affinity (K_D = 20 nM) for the peptide hormone, gonadotropin-releasing hormone (GnRH) was reported in 2002[60].



D-DNA
natural form

L-DNA
unnatural form

Figure 1.17 Structure of L-DNA

Recently, a synthetic molecular system capable of mirror-image genetic replication and transcription was reported[61]. This discovery introduced the potential of direct *in vitro* selection of this mirror image aptamers.

Resistance to renal clearance

Even the modifications can protect the aptamers from nuclease degradation, the small size of aptamers are subject to rapid excretion through renal clearance. Enlarging the size of aptamers using bulky moiety formulation can overcome the renal filtration and extend circulation time. Cholesterol[62], dialkyl lipids[63] and polyethylene glycol (PEG)[64] have been conjugated to 5'-end of aptamers for improved half-time in plasma. N-hydroxysuccinimide (NHS)-ester-activated PEG is the most widely used for manufacturing PEGylated oligonucleotides. It can react with the intermediate amino-modified oligonucleotides. The unmodified aptamers exhibit a very short *in vivo* half-life (<1h) while the conjugation of high molecular weight PEG can extend half-life to 24-48h[64-65].

Expansion of chemical functionalities

Despite the large initial libraries of aptamers to begin with for SELEX, many targets have proved to be resistance to selections. This can be explained by the limited chemical functionalities of nucleic acids. Aptamers have fewer monomer units compared with natural amino acids (4 natural bases vs 20 amino acids).

One approach to expand the chemical functionality is to introduce unnatural base pairs that achieve base-paring through an alternative arrangement of hydrogen bonding. DNA aptamers against VEGF-165 and IFN-γ with K_D of 0.65pM and 0.038nM respectively were successfully selected from DNA aptamer pool containing an unnatural nucleotide with hydrophobic base 7-(2-thienyl)imidazo[4,5-b]pyridine (Ds)[66]. The Ds base was proposed to enhance the interaction with hydrophobic cavities in target proteins compared with hydrophilic natural nucleic acids. The same group previously reported the optimized PCR conditions involving the Ds-Px pair as a third base pair with high efficiency and fidelity[67], which enables the application of this unnatural base pair to *in vitro* selection.

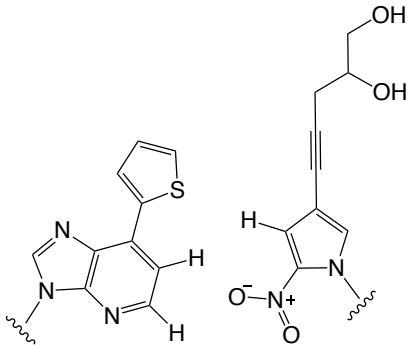


Figure 1.18 Chemical structures of the unnatural Ds–Px pair

The initial library of selection was constructed with unnatural hydrophobic base Ds at several predefined positions. The affinities of selected aptamers were <100-fold higher over those aptamers containing only four natural bases against the same targets.

Another conceptually similar approach was termed as click-SELEX[68]. The DNA library pool was alkyne-modified by replacing thymidine triphosphate (TTP) with C5-ethynyl-2'-

deoxyuridine (EdU), which could be further modified with an azide-containing compound using click chemistry.

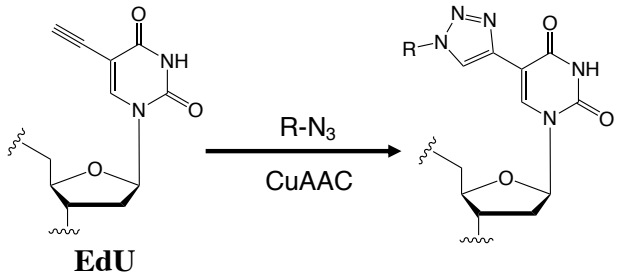


Figure 1.19 CuAAC functionalization of EdU-containing DNA

The DNA library was flanked by two 21nt thymidine-free primer binding sites and a Cy5 molecule at the 5'-end. The EdU positions would be modified with an azide-bearing molecule by CuAAC. After binding with the target molecule, the bound sequences were eluted and amplified by PCR with EdU triphosphate instead of thymidine. The single-stranded DNA was prepared by λ-exonuclease digestion of the 5'-phosphorylated antisense strands. A modified aptamer against cycle 3 green fluorescent protein (C3-GFP) with the K_D value of 18.4nM was selected using this approach.

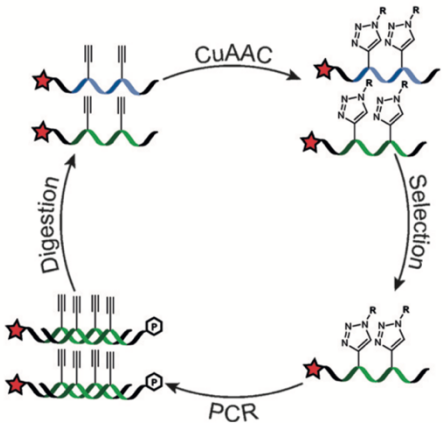


Figure 1.20 Schematic representation of click-SELEX

SomaLogic has developed a class of aptamers called SOMAmers[69] (slow off-rate modified aptamers). The nucleotide triphosphate analogs modified at the 5-position (R) of uridine (dUTP)

are incorporated into the nucleic acid polymers through polymerase-mediated synthesis. DNA molecules in which each thymine residue in uniformly replaced by a specific type of functional group are produced this way. Functional groups that resemble the side chains of hydrophobic amino acid residues have proved to be successful against a wide range of protein targets[70-71].

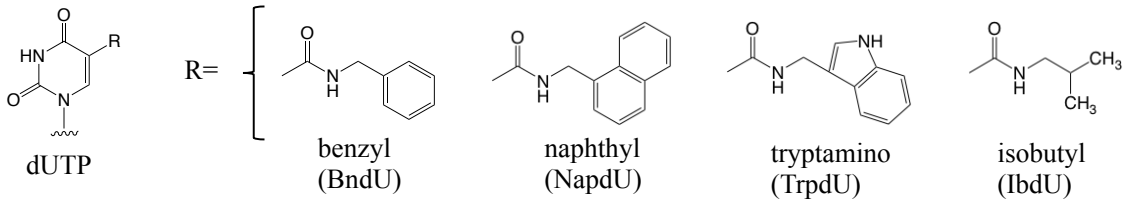


Figure 1.21 Nucleotide triphosphate analogs modified at 5-position of dUTP

Several SOMAmer-protein crystal structures indicate that the functional groups have a role in target binding and aptamer binding[31]. The modified bases not only improve the binding affinities and binding kinetics but also significantly increase the rate of successful selection[72]. According to SomaLogic, SOMAmers have been generated against over 3,000 proteins and these reagents are currently used as a point-of-care technology for monitoring protein levels in human serum[73].

Among functional groups tested, hydrophobic, aromatic side chains consistently produce the best ligands, especially when selection pressure includes a demand for slow dissociation rates[31]. It's not surprising because amino acids like tyrosine, tryptophan, and phenylalanine are about five times more common in paratopes that on the rest of antibody surfaces[74].

Some SOMAmer-protein co-crystal structures have been obtained and researchers are beginning to gain insight into the role modifications play in folding and binding. Here described is the PDGF (Platelet-derived growth factor) SOMAmer example. The SOMAmer is composed of a stem-loop and a short pseudoknot (Figure 1.22). PDGF is shown as electrostatic surface

renderings with the binding interface in orange. Atoms within 4Å of the target are colored orange.

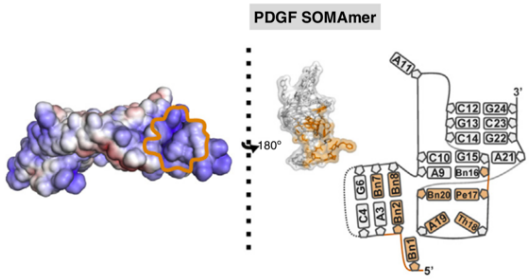


Figure 1.22 Open-book views of PDGF-SOMAmer binding interfaces

The eight modified nucleotides are distributed evenly between the domains. All eight are in direct contact with the protein forming an interaction surface with complementary hydrophobic pockets on the protein. The modified nucleotides also form an extensive network of face-to-face and edge-to-face π-π interactions that serve as a bridge between the two domains (Figure 1.23). Replacement of benzyl residue at position 8 with an aliphatic isobutyl side chain (equivalent to a Phe to Leu mutation) disrupts the network and reduces the binding affinity[75].

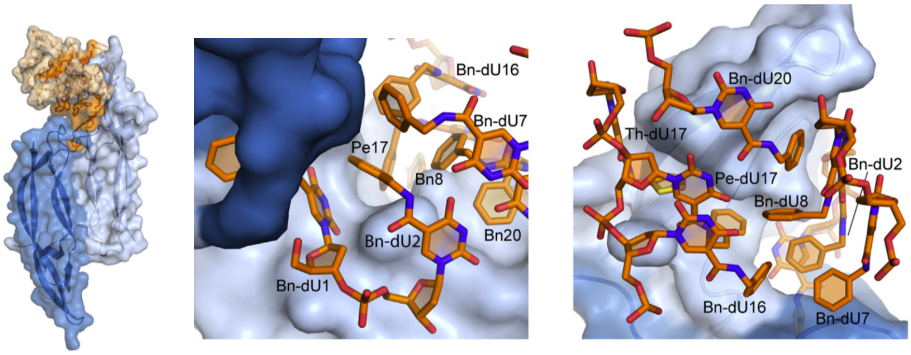


Figure 1.23 Intramolecular contacts of PDGF and the SOMAmer

Ligase-mediated DNA-templated polymerization

All the modifications described above are limited to one or two types of chemical modifications per library. A ligase-mediated translation system developed by Liu and co-workers

makes it possible to synthesize nucleic acid polymers that carry a variety of functional groups in the same library[76].

T4 DNA ligase has been known to effectively join two adjacent, short 5'-phosphorylated oligonucleotides. It catalyzed the ATP-dependent phosphodiester bond formation between the 5' phosphoryl and the 3' hydroxyl groups as guided by DNA templates[77].

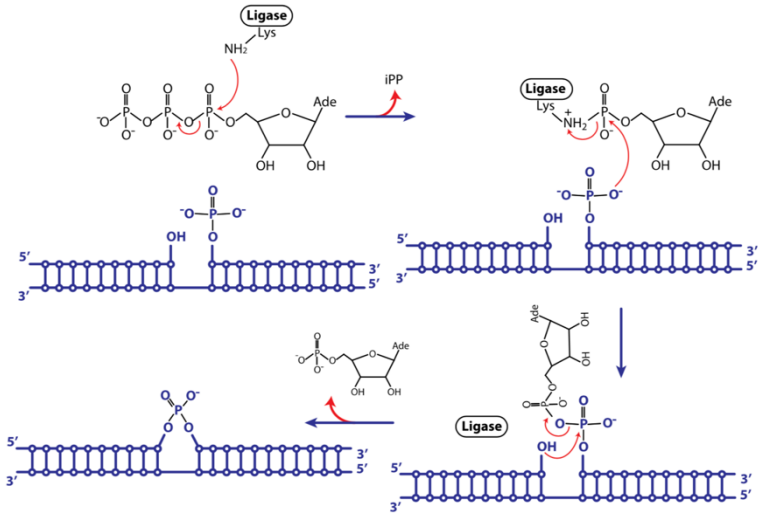


Figure 1.24 Mechanism of T4 DNA ligase-mediated templated ligation[78]

The covalent phosphodiester bond is formed in a three-step process[78-80]. In the first step, the ATP cofactor is locked into the enzyme's active pocket where the ε-amino group of the residue attacks the α-phosphorus atom of ATP to form the AMP-NH_2-ligase intermediate, and a pyrophosphate anion is released simultaneously. In the second step, the adenylyl residue from the intermediate is transferred to the 5'-phosphate of a short oligonucleotide, forming the AMP-activated 5'-phosphate the nick site. Finally, the 3'-hydroxyl group of another short oligonucleotide acts as a nucleophile and attacks the phosphorus atom of the activated 5'-phosphate group. The new phosphodiester bond is formed, and the AMP molecule is released.

The shortest substrate for T4 DNA ligase polymerization reported was a pentamer[81]. It has been reported that T4 DNA ligase can mediate construction of branched DNA strands[82] as well as its ability to tolerate modified DNA template[83].

Based on these previous discoveries, Liu and co-workers developed a new system using T4 DNA ligase to translate the DNA template into sequence-defined functionalized nucleic acid polymers with trinucleotides[76].

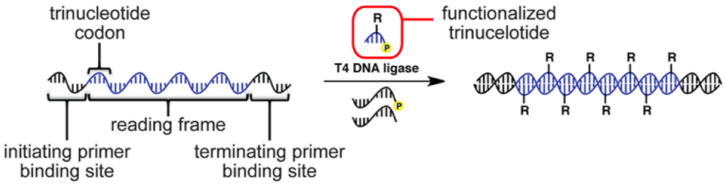


Figure 1.25 T4 DNA ligase-mediated polymerization of modified oligonucleotides

The various functionalities were introduced through amine-linked bases. The amine groups were served as a chemical handle to install functional groups using well-established bioconjugation chemistries such as activated NHS ester.

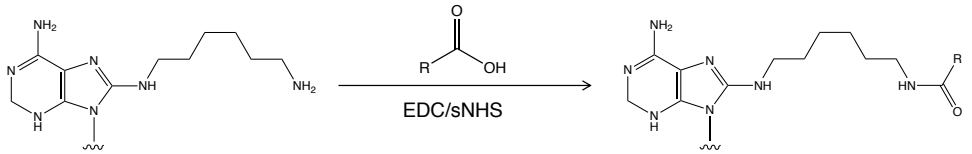


Figure 1.26 Coupling reaction of amino-modified adenosine

Eight different functional groups were successfully incorporated throughout a polymer product, with the possibility of expanding the substrate set up to 64 (4^3)[76]. The polymerization exhibited high efficiency as well as sequence specificity. The highly functionalized nucleic acid polymers survived iterative cycles of translation, selection, template regeneration and PCR amplification (the cycle of SELEX).

This method has proved the potential of highly functionalized nucleic acid polymers and their ready application to *in vitro* selection.

References

1. Dahm, R., Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **2008,** *122* (6), 565-581.

2. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171* (4356), 737-738.

3. Ali, R. H. From genomes to post-processing of Bayesian inference of phylogeny. Doctoral thesis, comprehensive summary, KTH Royal Institute of Technology, Stockholm, 2016.

4. Avery, O. T.; MacLeod, C. M.; McCarty, M., Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J. Exp. Med.* **1944,** *79* (2), 137-158.

5. Hershey, A. D.; Chase, M., Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *J. Gen. Physiol.* **1952,** *36* (1), 39-56.

6. Caspersson, T.; Schultz, J., Pentose nucleotides in the cytoplasm of growing tissues. *Nature* **1939,** *143*, 602-603.

7. King, L.; Carl, J., Lithium Ion Effects on Brain in Vivo. *Fed Proc* **1967,** *26* (2), 813-&.

8. Baltimore, D., Viral Rna-Dependent DNA Polymerase - Rna-Dependent DNA Polymerase in Virions of Rna Tumour Viruses. *Nature* **1970,** *226* (5252), 1209-+.

9. Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R., Self-Splicing Rna - Auto-Excision and Auto-Cyclization of the Ribosomal-Rna Intervening Sequence of Tetrahymena. *Cell* **1982,** *31* (1), 147-157.

10. Napoli, C.; Lemieux, C.; Jorgensen, R., Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in Trans. *Plant Cell* **1990,** *2* (4), 279-289.

11. Schwalbe, H.; Buck, J.; Fürtig, B.; Noeske, J.; Wöhnert, J., Structures of RNA Switches: Insight into Molecular Recognition and Tertiary Structure. *Angew. Chem. Int. Edit.* **2007,** *46* (8), 1212-1219.

12. Serganov, A.; Nudler, E., A Decade of Riboswitches. *Cell* **2013,** *152* (1), 17-24.

13. Silverman, S. K., Artificial Functional Nucleic Acids: Aptamers, Ribozymes, and Deoxyribozymes Identified by In Vitro Selection. In *Functional Nucleic Acids for Analytical Applications*, Yingfu, L.; Yi, L., Eds. Springer New York: New York, NY, 2009; pp 47-108.

14. Dunn, M. R.; Jimenez, R. M.; Chaput, J. C., Analysis of aptamer discovery and technology. *Nat. Rev. Chem.* **2017,** *1*, 0076.

15. Tuerk, C.; Gold, L., Systematic Evolution of Ligands by Exponential Enrichment - Rna Ligands to Bacteriophage-T4 DNA-Polymerase. *Science* **1990,** *249* (4968), 505-510.

16. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346*, 818.

17. Wilson, D. S.; Szostak, J. W., In Vitro Selection of Functional Nucleic Acids. *Annu. Rev. Biochem.* **1999,** *68* (1), 611-647.

18. Wang, Y.; Rando, R. R., Specific binding of aminoglycoside antibiotics to RNA. *Chem. Biol.* **1995,** *2* (5), 281-290.

19. Geiger, A.; Burgstaller, P.; von der Eltz, H.; Roeder, A.; Famulok, M., RNA Aptamers That Bind l-Arginine with Sub-Micromolar Dissociation Constants and High Enantioselectivity. *Nucleic Acids Res.* **1996,** *24* (6), 1029-1036.

20. Lauhon, C. T.; Szostak, J. W., RNA aptamers that bind flavin and nicotinamide redox cofactors. *J. Am. Chem. Soc.* **1995,** *117* (4), 1246-1257.

21. Yamamoto, R.; Katahira, M.; Nishikawa, S.; Baba, T.; Taira, K.; Kumar, P. K. R., A novel RNA motif that binds efficiently and specifically to the Tat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo. *Genes to Cells* **2001,** *5* (5), 371-388.

22. Ng, E. W. M.; Shima, D. T.; Calias, P.; Cunningham Jr, E. T.; Guyer, D. R.; Adamis, A. P., Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug Discovery* **2006,** *5*, 123.

23. Noma, T.; Ikebukuro, K., Aptamer selection based on inhibitory activity using an evolution-mimicking algorithm. *Biochem Bioph Res Co* **2006,** *347* (1), 226-231.

24. Pasternak, A.; Hernandez, F. J.; Rasmussen, L. M.; Vester, B.; Wengel, J., Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer. *Nucleic Acids Res.* **2011,** *39* (3), 1155-1164.

25. Keefe, A. D.; Pai, S.; Ellington, A., Aptamers as therapeutics. *Nat. Rev. Drug Discovery* **2010,** *9* (7), 537-550.

26. Zhou, J. H.; Rossi, J., Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug Discovery* **2017,** *16* (3), 181-202.

27. McKeague, M.; McConnell, E. M.; Cruz-Toledo, J.; Bernard, E. D.; Pach, A.; Mastronardi, E.; Zhang, X. R.; Beking, M.; Francis, T.; Giamberardino, A.; Cabecinha, A.; Ruscito, A.; Aranda-Rodriguez, R.; Dumontier, M.; DeRosa, M. C., Analysis of In Vitro Aptamer Selection Parameters. *J. Mol. Evol.* **2015,** *81* (5-6), 150-161.

28. Hianik, T.; Ostatná, V.; Sonlajtnerova, M.; Grman, I., Influence of ionic strength, pH and aptamer configuration for binding affinity to thrombin. *Bioelectrochemistry* **2007,** *70* (1), 127-133.

29. Deng, Q.; German, I.; Buchanan, D.; Kennedy, R. T., Retention and separation of adenosine and analogues by affinity chromatography with an aptamer stationary phase. *Anal. Chem.* **2001,** *73* (22), 5415-5421.

30. Neves, M. A. D.; Reinstein, O.; Saad, M.; Johnson, P. E., Defining the secondary structural requirements of a cocaine-binding aptamer by a thermodynamic and mutation study. *Biophys. Chem.* **2010,** *153* (1), 9-16.

31. Gelinas, A. D.; Davies, D. R.; Janjic, N., Embracing proteins: structural themes in aptamer-protein complexes. *Curr. Opin. Struct. Biol.* **2016,** *36*, 122-32.

32. Kim, S. H.; Sussman, J. L.; Suddath, F. L.; Quigley, G. J.; McPherson, A.; Wang, A. H. J.; Seeman, N. C.; Rich, A., The General Structure of Transfer RNA Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1974,** *71* (12), 4970-4974.

33. Robertus, J. D.; Ladner, J. E.; Finch, J. T.; Rhodes, D.; Brown, R. S.; Clark, B. F. C.; Klug, A., Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **1974,** *250*, 546.

34. Hendrix, D. K.; Brenner, S. E.; Holbrook, S. R., RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys* **2005,** *38* (3), 221-243.

35. Holbrook, S. R., RNA structure: the long and the short of it. *Curr. Opin. Struct. Biol.* **2005,** *15* (3), 302-308.

36. Ozer, A.; Pagano, J. M.; Lis, J. T., New Technologies Provide Quantum Changes in the Scale, Speed, and Success of SELEX Methods and Aptamer Characterization. *Mol. Ther. - Nucleic Acids* **2014,** *3*, e183.

37. Jenison, R.; Gill, S.; Pardi, A.; Polisky, B., High-resolution molecular discrimination by RNA. *Science* **1994,** *263* (5152), 1425-1429.

38. Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J.-P.; Pardi, A., Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat. Struct. Biol.* **1997,** *4*, 644.

39. Hermann, T.; Patel, D. J., Biochemistry - Adaptive recognition by nucleic acid aptamers. *Science* **2000,** *287* (5454), 820-825.

40. 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.

41. Macaya, R. F.; Schultze, P.; Smith, F. W.; Roe, J. A.; Feigon, J., Thrombin-Binding DNA Aptamer Forms a Unimolecular Quadruplex Structure in Solution. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (8), 3745-3749.

42. Marshall, K. A.; Robertson, M. P.; Ellington, A. D., A biopolymer by any other name would bind as well: a comparison of the ligand-binding pockets of nucleic acids and proteins. *Structure* **1997,** *5* (6), 729-734.

43. Michaud, M.; Jourdan, E.; Ravelet, C.; Villet, A.; Ravel, A.; Grosset, C.; Peyrin, E., Immobilized DNA aptamers as target-specific chiral stationary phases for resolution of nucleoside and amino acid derivative enantiomers. *Anal. Chem.* **2004,** *76* (4), 1015-1020.

44. Deng, Q.; Watson, C. J.; Kennedy, R. T., Aptamer affinity chromatography for rapid assay of adenosine in microdialysis samples collected in vivo. *J. Chromatogr. A* **2003,** *1005* (1), 123-130.

45. Song, S.; Wang, L.; Li, J.; Fan, C.; Zhao, J., Aptamer-based biosensors. *TrAC, Trends Anal. Chem.* **2008,** *27* (2), 108-117.

46. Nutiu, R.; Li, Y., Structure-Switching Signaling Aptamers. *J. Am. Chem. Soc.* **2003,** *125* (16), 4771-4778.

47. Fredriksson, S.; Gullberg, M.; Jarvius, J.; Olsson, C.; Pietras, K.; Gustafsdottir, S. M.; Ostman, A.; Landegren, U., Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* **2002,** *20* (5), 473-7.

48. Heldin, C.-H.; Westermark, B., Mechanism of Action and In Vivo Role of Platelet-Derived Growth Factor. *Physiol. Rev.* **1999,** *79* (4), 1283-1316.

49. Sundaram, P.; Kurniawan, H.; Byrne, M. E.; Wower, J., Therapeutic RNA aptamers in clinical trials. *Eur. J. Pharm. Sci.* **2013,** *48* (1-2), 259-271.

50. Griffin, L. C.; Tidmarsh, G. F.; Bock, L. C.; Toole, J. J.; Leung, L. L., In vivo anticoagulant properties of a novel nucleotide-based thrombin inhibitor and demonstration of regional anticoagulation in extracorporeal circuits. *Blood* **1993,** *81* (12), 3271-6.

51. Herdewijn, P.; Marlière, P., Toward Safe Genetically Modified Organisms through the Chemical Diversification of Nucleic Acids. *Chemistry & Biodiversity* **2009,** *6* (6), 791-808.

52. Shaw, J.-P.; Kent, K.; Bird, J.; Fishback, J.; Froehler, B., Modified deoxyoligonucleotides stable to exonuclease degradation in serum. *Nucleic Acids Res.* **1991,** *19* (4), 747-750.

53. Burmeister, P. E.; Lewis, S. D.; Silva, R. F.; Preiss, J. R.; Horwitz, L. R.; Pendergrast, P. S.; McCauley, T. G.; Kurz, J. C.; Epstein, D. M.; Wilson, C.; Keefe, A. D., Direct In Vitro Selection of a 2′ -O-Methyl Aptamer to VEGF. *Chem. Biol.* **2005,** *12* (1), 25-33.

54. Shi, H.; He, X.; Cui, W.; Wang, K.; Deng, K.; Li, D.; Xu, F., Locked nucleic acid/DNA chimeric aptamer probe for tumor diagnosis with improved serum stability and extended imaging window in vivo. *Anal. Chim. Acta* **2014,** *812*, 138-144.

55. Zaitseva, M.; Kaluzhny, D.; Shchyolkina, A.; Borisova, O.; Smirnov, I.; Pozmogova, G., Conformation and thermostability of oligonucleotide d(GGTTGGTGTGGTTGG) containing thiophosphoryl internucleotide bonds at different positions. *Biophys. Chem.* **2010,** *146* (1), 1-6.

56. Pozmogova, G. E.; Zaitseva, M. A.; Smirnov, I. P.; Shvachko, A. G.; Murina, M. A.; Sergeenko, V. I., Anticoagulant Effects of Thioanalogs of Thrombin-Binding DNA-Aptamer and Their Stability in the Plasma. *Bull. Exp. Biol. Med.* **2010,** *150* (2), 180-184.

57. Saccà, B.; Lacroix, L.; Mergny, J.-L., The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **2005,** *33* (4), 1182-1192.

58. Varizhuk, A. M.; Kaluzhny, D. N.; Novikov, R. A.; Chizhov, A. O.; Smirnov, I. P.; Chuvilin, A. N.; Tatarinova, O. N.; Fisunov, G. Y.; Pozmogova, G. E.; Florentiev, V. L., Synthesis of triazole-linked oligonucleotides with high affinity to DNA complements and an analysis of their compatibility with biosystems. *The Journal of organic chemistry* **2013,** *78* (12), 5964-9.

59. Varizhuk, A. M.; Tsvetkov, V. B.; Tatarinova, O. N.; Kaluzhny, D. N.; Florentiev, V. L.; Timofeev, E. N.; Shchyolkina, A. K.; Borisova, O. F.; Smirnov, I. P.; Grokhovsky, S. L.; Aseychev, A. V.; Pozmogova, G. E., Synthesis, characterization and in vitro activity of thrombin-binding DNA aptamers with triazole internucleotide linkages. *Eur. J. Med. Chem.* **2013,** *67*, 90-97.

60. Wlotzka, B.; Leva, S.; Eschgfaller, B.; Burmeister, J.; Kleinjung, F.; Kaduk, C.; Muhn, P.; Hess-Stumpp, H.; Klussmann, S., In vivo properties of an anti-GnRH Spiegelmer: an example of an oligonucleotide-based therapeutic substance class. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (13), 8898-902.

61. Wang, Z.; Xu, W.; Liu, L.; Zhu, T. F., A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nat. Chem.* **2016,** *8*, 698.

62. Lee, C. H.; Lee, S.-H.; Kim, J. H.; Noh, Y.-H.; Noh, G.-J.; Lee, S.-W., Pharmacokinetics of a Cholesterol-conjugated Aptamer Against the Hepatitis C Virus (HCV) NS5B Protein. *Mol. Ther. - Nucleic Acids* **2015,** *4*, e254.

63. Willis, M. C.; Collins, B. D.; Zhang, T.; Green, L. S.; Sebesta, D. P.; Bell, C.; Kellogg, E.; Gill, S. C.; Magallanez, A.; Knauer, S.; Bendele, R. A.; Gill, P. S.; Janjic, N., Liposome-anchored vascular endothelial growth factor aptamers. *Bioconjugate Chem.* **1998,** *9* (5), 573-82.

64. Da Pieve, C.; Blackshaw, E.; Missailidis, S.; Perkins, A. C., PEGylation and biodistribution of an anti-MUC1 aptamer in MCF-7 tumor-bearing mice. *Bioconjugate Chem.* **2012,** *23* (7), 1377-81.

65. Healy, J. M.; Lewis, S. D.; Kurz, M.; Boomer, R. M.; Thompson, K. M.; Wilson, C.; McCauley, T. G., Pharmacokinetics and Biodistribution of Novel Aptamer Compositions. *Pharm. Res.* **2004,** *21* (12), 2234-2246.

66. Kimoto, M.; Yamashige, R.; Matsunaga, K.-i.; Yokoyama, S.; Hirao, I., Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **2013,** *31*, 453.

67. Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I., Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **2012,** *40* (6), 2793-2806.

68. Tolle, F.; Brändle, G. M.; Matzner, D.; Mayer, G., A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew. Chem. Int. Edit.* **2015,** *54* (37), 10971-10974.

69. Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T.; Flather, D.; Forbes, A.; Foreman, T.; Fowler, C.; Gawande, B.; Goss, M.; Gunn, M.; Gupta, S.; Halladay, D.; Heil, J.; Heilig, J.; Hicke, B.; Husar, G.; Janjic, N.; Jarvis, T.; Jennings, S.; Katilius, E.; Keeney, T. R.; Kim, N.; Koch, T. H.; Kraemer, S.; Kroiss, L.; Le, N.; Levine, D.; Lindsey, W.; Lollo, B.; Mayfield, W.; Mehan, M.; Mehler, R.; Nelson, S.

K.; Nelson, M.; Nieuwlandt, D.; Nikrad, M.; Ochsner, U.; Ostroff, R. M.; Otis, M.; Parker, T.; Pietrasiewicz, S.; Resnicow, D. I.; Rohloff, J.; Sanders, G.; Sattin, S.; Schneider, D.; Singer, B.; Stanton, M.; Sterkel, A.; Stewart, A.; Stratford, S.; Vaught, J. D.; Vrkljan, M.; Walker, J. J.; Watrobka, M.; Waugh, S.; Weiss, A.; Wilcox, S. K.; Wolfson, A.; Wolk, S. K.; Zhang, C.; Zichi, D., Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLOS ONE* **2010,** *5* (12), e15004.

70. Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E., Expanding the Chemistry of DNA for in Vitro Selection. *J. Am. Chem. Soc.* **2010,** *132* (12), 4141-4151.

71. Gupta, S.; Hirota, M.; Waugh, S. M.; Murakami, I.; Suzuki, T.; Muraguchi, M.; Shibamori, M.; Ishikawa, Y.; Jarvis, T. C.; Carter, J. D.; Zhang, C.; Gawande, B.; Vrkljan, M.; Janjic, N.; Schneider, D. J., Chemically modified DNA aptamers bind interleukin-6 with high affinity and inhibit signaling by blocking its interaction with interleukin-6 receptor. *J. Biol. Chem.* **2014,** *289* (12), 8706-19.

72. Davies, D. R.; Gelinas, A. D.; Zhang, C.; Rohloff, J. C.; Carter, J. D.; O'Connell, D.; Waugh, S. M.; Wolk, S. K.; Mayfield, W. S.; Burgin, A. B.; Edwards, T. E.; Stewart, L. J.; Gold, L.; Janjic, N.; Jarvis, T. C., Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (49), 19971-6.

73. Ostroff, R. M.; Bigbee, W. L.; Franklin, W.; Gold, L.; Mehan, M.; Miller, Y. E.; Pass, H. I.; Rom, W. N.; Siegfried, J. M.; Stewart, A.; Walker, J. J.; Weissfeld, J. L.; Williams, S.; Zichi, D.; Brody, E. N., Unlocking Biomarker Discovery: Large Scale Application of Aptamer Proteomic Technology for Early Detection of Lung Cancer. *PLOS ONE* **2010,** *5* (12), e15003.

74. Ramaraj, T.; Angel, T.; Dratz, E. A.; Jesaitis, A. J.; Mumey, B., Antigen–antibody interface properties: Composition, residue interactions, and features of 53 non-redundant

structures. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2012,** *1824* (3), 520-532.

75. Davies, D. R.; Gelinas, A. D.; Zhang, C.; Rohloff, J. C.; Carter, J. D.; O'Connell, D.; Waugh, S. M.; Wolk, S. K.; Mayfield, W. S.; Burgin, A. B.; Edwards, T. E.; Stewart, L. J.; Gold, L.; Janjic, N.; Jarvis, T. C., Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (49), 19971-19976.

76. 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.

77. Wu, D. Y.; Wallace, R. B., Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene* **1989,** *76* (2), 245-254.

78. Pawlowska, R.; Korczynski, D.; Nawrot, B.; Stec, W. J.; Chworos, A., The α-thio and/or β-γ-hypophosphate analogs of ATP as cofactors of T4 DNA ligase. *Bioorg. Chem.* **2016,** *67*, 110-115.

79. Rossi, R.; Montecucco, A.; Ciarrocchi, G.; Biamonti, G., Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic Acids Res.* **1997,** *25* (11), 2106-2113.

80. Shuman, S., Vaccinia virus DNA ligase: specificity, fidelity, and inhibition. *Biochemistry* **1995,** *34* (49), 16138-16147.

81. Kotler, L. E.; Zevin-Sonkin, D.; Sobolev, I. A.; Beskin, A. D.; Ulanovsky, L. E., DNA sequencing: modular primers assembled from a library of hexamers or pentamers. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (9), 4241-4245.

82. Mendel‐Hartvig, M.; Kumar, A.; Landegren, U., Ligase‐mediated construction of branched DNA strands: a novel DNA joining activity catalyzed by T4 DNA ligase. *Nucleic Acids Res.* **2004,** *32* (1), e2-e2.

83. Liang, X.; Fujioka, K.; Asanuma, H., Nick Sealing by T4 DNA Ligase on a Modified DNA Template: Tethering a Functional Molecule on D-Threoninol. *Chemistry – A European Journal* **2011,** *17* (37), 10388-10396.

CHAPTER 2

A HIGH-FIDELITY CODON SET FOR THE T4 DNA LIGASE-CATALYZED POLYMERIZATION OF MODIFIED OLIGONUCLEOTIDES[1]

[1]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.

Abstract

In vitro selection of nucleic acid polymers can readily deliver highly specific receptors and catalysts for a variety of applications; however, it is suspected that the functional group deficit of nucleic acids has limited their potential with respect to proteinogenic polymers. This has stimulated research toward expanding their chemical diversity to bridge the functional gap between nucleic acids and proteins to develop a superior biopolymer. In this study, we investigate the effect of codon library size and composition on the sequence specificity of T4 DNA ligase in the DNA-templated polymerization of both unmodified and modified oligonucleotides. Using high-throughput DNA sequencing of duplex pairs, we have uncovered a 256-membered codon set that yields sequence-defined modified ssDNA polymers in high yield and with high fidelity.

Introduction

In the rapidly growing fields of proteomics and glycomics, in-depth analysis of protein expression, modification, and interaction on a genomic scale requires the development of specific and high-affinity reagents[2-3]. Traditionally, proteinogenic polymers, especially antibodies, have served in this role; however, they suffer from a number of shortcomings, ranging from poor stability and variable production quality to limited target availability and issues with specificity profiles[4-6]. Despite over 500,000 commercially available antibodies[5], researchers continue to express concerns over accessibility to high-quality antibodies for biomedical research[7-9]. These issues have stimulated the development of alternative technologies to generate high-affinity reagents for proteins that rival the performance of traditional antibodies[10-14].

To this end, the use of nucleic acid aptamers as high-affinity reagents has come to the fore[15], as they provide numerous advantages over traditional antibodies, such as (i) their generation

through *in vitro* selection enables ready tuning of binding and specificity properties; (ii) their active structure can be reversibly formed by thermal denaturation and cooling; (iii) they exhibit excellent chemical stability and shelf life; and (iv) their chemical synthesis is predictable and scalable. Unfortunately, the functional group deficit of nucleic acids limits their potential to match the performance of proteinogenic affinity reagents. It is anticipated that expanding the functional group repertoire of DNA will increase the heteromultivalent interactions with their molecular targets and better resemble the binding events that take place at protein-protein interfaces[16]. Thus, methods to increase the chemical functionality present in nucleic acid polymers are of significant interest[17].

The sequence-specific incorporation of chemical functionality throughout a nucleic acid polymer has traditionally relied on polymerase-catalyzed DNA-templated primer extension using nucleobase-modified dNTPs[18]. When using a four-base genetic code, this approach enables up to four different functional groups to be incorporated throughout ssDNA and also provides flexibility over the sugar backbone structure[18-19]. While this has produced nucleic acid polymers with superior function when compared with their unmodified counterparts[20], the ability to incorporate greater functionality remained elusive. Recently, the T4 DNA ligase-catalyzed DNA-templated polymerization of modified 5′-phosphorylated oligonucleotides has expanded the number of different modifications on a ssDNA to eight and relaxed limitations on the size of the modification[21](Figure 2.1). As the method relies on codons, rather than single nucleotides, the theoretical number of unique modifications that can be incorporated increases with increasing codon length. Thus, a trinucleotide codon set enables the incorporation of up to 64 unique modifications.

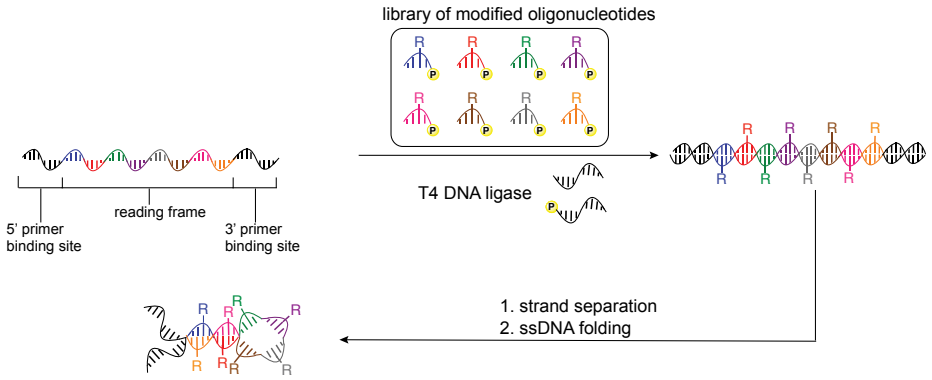


Figure 2.1 Synthesis of modified ssDNA polymers using T4 DNA ligase

While the specificity of the reported ligase-catalyzed polymerization using an eight-membered trinucleotide codon library was high, there is little known about the specificity of larger libraries that can access greater sequence space. One early report documents the filling of a 12nt gapped DNA duplex with either a library of 64 trinucleotides or library of 4096 hexanucleotides using Escherichia coli ligase[22]. The yield of the full-length product was very low; however, the fidelity of the system at the single nucleotide level was high, with hexanucleotides providing the best fidelity of 0.8 single-nucleotide errors per 12nt sequence. Unfortunately, when the error rate is considered at the level of hexanucleotide incorporation, this corresponds to 60% fidelity. For longer sequences, it is not clear if such a low level of fidelity could support iterative rounds of *in vitro* selection, and the ability for E. *coli* ligase to accept chemical modifications was not studied.

We sought to identify a codon set for the T4 DNA ligase-catalyzed DNA-templated polymerization of modified 5'-phosphorylated oligonucleotides that satisfies the following requirements: (i) highly efficient polymerization; (ii) high fidelity DNA-templated polymerization; (iii) broad coverage of sequence space; (iv) tolerant of small modifications on polymerized oligonucleotides; and (v) readily accessible with standard phosphoramidite mixtures. Herein, we report a codon set for the ligase-catalyzed polymerization of modified oligonucleotides that satisfies all of these requirements and should enable the *in vitro* selection of

modified nucleic acids as receptors and catalysts with larger repertoires of chemical functionality.

Results

The initial report on the T4 DNA ligase-catalyzed polymerization of modified oligonucleotides was based on trimers[21]. We pursued a tetranucleotide codon set with the aim of expanding the number of possible modifications and increasing the coverage of sequence space. First, we evaluated the efficiency of polymerization using a PEGylated DNA template and PAGE analysis. Unfortunately, the polymerization efficiency of modified tetranucleotides dropped dramatically when expanding the codon set to 64 members as shown in Figure 2.2.

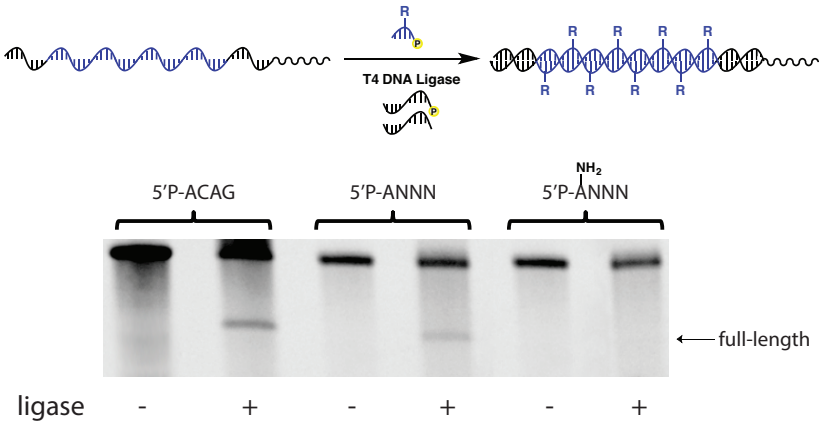


Figure 2.2 Polymerization efficiency of modified tetranucleotides

As our objective was to identify a codon set capable of exploring a greater portion of nucleic acid sequence space, we next examined the efficiency of T4 DNA ligase to polymerize a library of 5′-phosphorylated pentanucleotides along a library of DNA templates[23]. The pentanucleotide library consisted of all 256 possible sequences derived from 5′P-ANNNN, where A was either unmodified or modified as an N8-hexylamine derivative (Figure 2.3).

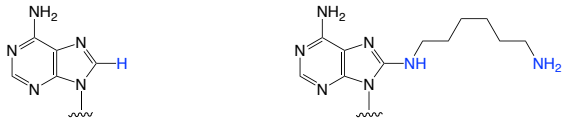


Figure 2.3 Structures of modified adenosine

The corresponding template library comprised two primer sites that flanked eight repeats of the codon NNNNT. The primers used during the polymerization were fluorescently labeled with CY5 and 6FAM fluorophores, which enabled identification of full- length products by 2-channel fluorescent PAGE imaging (Figure 2.4).

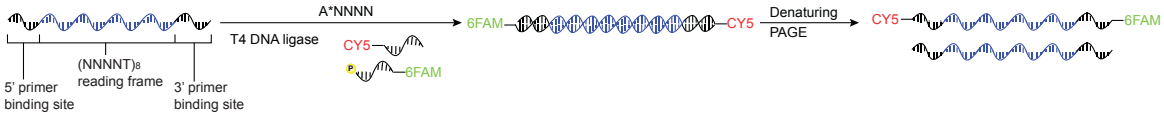


Figure 2.4 Dual fluorophore approach for evaluating polymerization efficiency

In contrast with the tetranucleotide systems, efficient polymerization of both the modified and unmodified pentanucleotide libraries was observed as judged by the presence of a heavy dual fluorescently labeled product band (Figure 2.5).

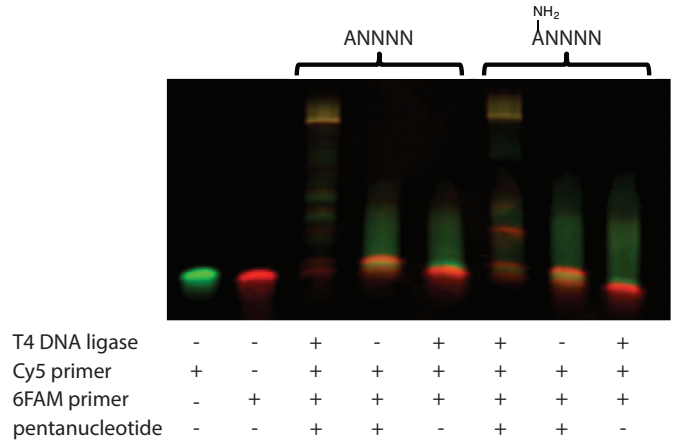


Figure 2.5 Dual-channel fluorescent image of a denaturing PAGE

Encouraged by the efficiency of polymerization, we next examined the fidelity of the process. We first sought to determine if T4 DNA ligase was able to efficiently incorporate a pentanucleotide containing a single-nucleotide mismatch and extend from the misincorporation site (i.e., read through). Using a set of 5'-phosphorylated hairpin templates (Figure 2.6a), we assessed the ability of T4 DNA ligase to incorporate a pentanucleotide (5′P-AGAGA) across from a codon with a single-nucleotide containing a single-nucleotide error; a 3'-cytidine was

added to the template to preclude blunt-end ligation of the full-length products. Following analysis by denaturing PAGE, full-length products were observed for all cases under the examined polymerization conditions. These experiments demonstrate that misincorporation can readily occur when the matched pentanucleotide sequence is not present in solution and that such misincorporations can be extended to full-length products (Figure 2.6b).

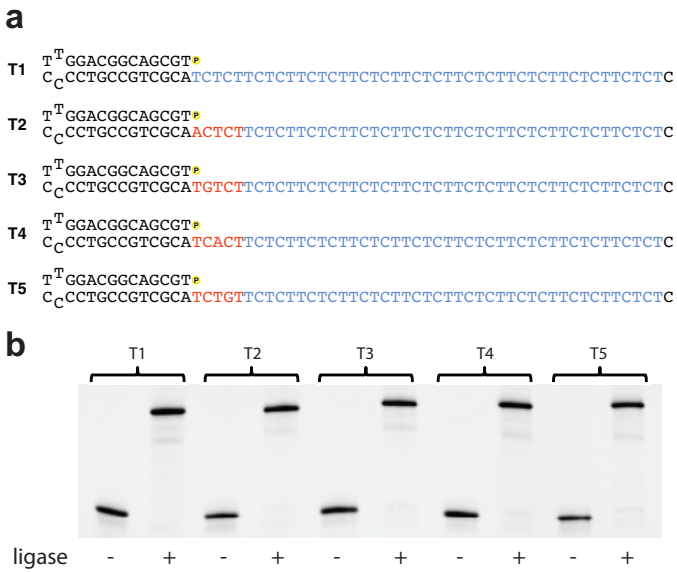


Figure 2.6 Incorporation of pentanucleotides with a single-nucleotide mismatch

While potentially serving as an issue with respect to fidelity, this feature could serve an important role during in vitro evolution of modified nucleic acids, as mutations will not terminate the nucleic acid polymer synthesis, thus enabling novel fit phenotypes to survive and be replicated to generate novel genotypes for subsequent rounds of selection.

We next designed a simple chain-termination[24-25] competition experiment to provide an estimation of single- nucleotide discrimination during pentanucleotide incorporation, as we anticipated single-nucleotide mismatches to be the greatest threat to fidelity. Thus, a 5'-phosphorylated hairpin template was synthesized containing a TCTCT codon followed by seven repeats of AGAGT and capped with a 3'-cytidine (Figure 2.7a). The template was used to direct the polymerization of 5'P-AGAGA and 5'P-ACTCT in the presence of increasing amounts of a

mixture of single-nucleotide mismatch terminators of 5'P-AGAGA (mixture contained equal parts: 5'-TGAGA, 5'-ACAGA, 5'-AGTGA, 5'-AGACA, 5'-AGAGT). As the terminator mixture was increased from 0 to 15-fold with respect to 5'P-AGAGA, chain termination at codon position one of the templates increased from 0-5% (Figure 2.7b), suggesting that high single-nucleotide discrimination occurred during the templated polymerization.

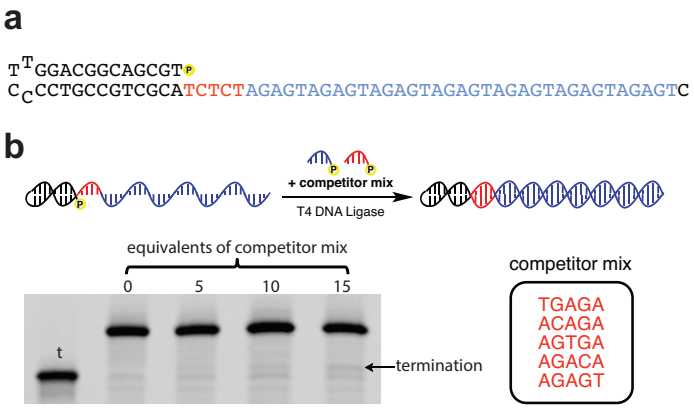


Figure 2.7 Single-nucleotide mismatch competition experiments

We further analyzed the sequence specificity by restriction enzyme digestion (Figure 2.8) using non-denaturing PAGE analysis. We designed a DNA template that contained a *BcoDI* digest site, which would allow us to evaluate the fidelity of polymerization around the digestion site. T4 DNA ligase was challenged to correctly incorporate the correct pentanucleotides from the 5'P-ANNNN library along the DNA template. As a positive control, Klenow fragment was used to generate the correct DNA duplex prior to restriction digest. We found that the product of the T4 DNA ligase-catalyzed polymerization reaction was efficiently cleaved by *BcoDI*, suggesting that highly specific incorporation of pentanucleotides was occurring at the digest site.

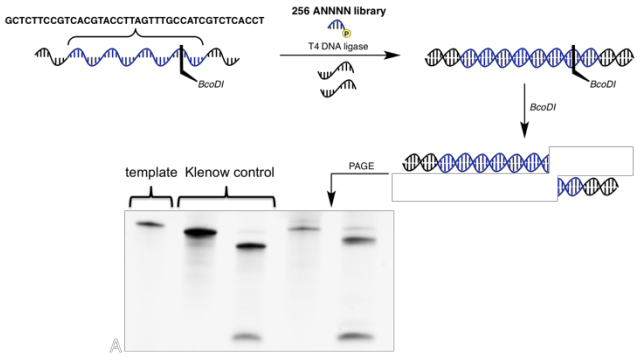


Figure 2.8 Restriction enzyme digestion of polymerized products

The high fidelity of T4 DNA ligase was further supported by Sanger sequencing of the polymerized strand following PAGE purification, which revealed a consensus sequence that was the reverse complement of the starting template strand (Figure 2.9). Klenow fragment-catalyzed extension was used as a positive control.

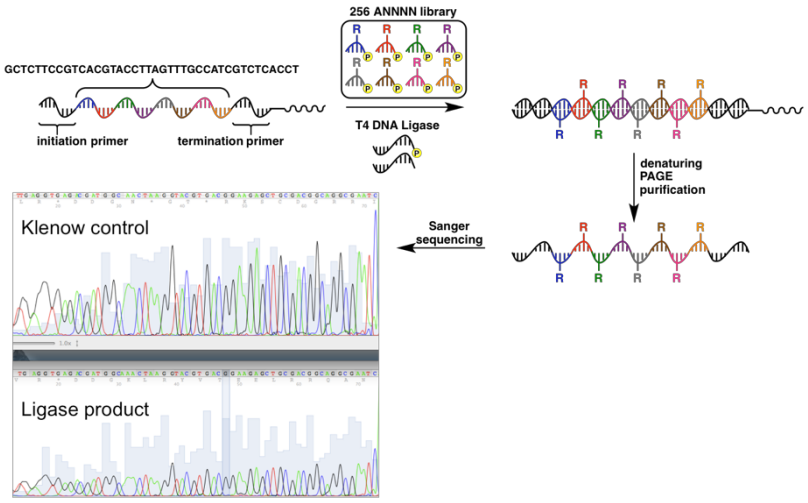


Figure 2.9 Sanger sequencing workflow for analyzing fidelity of polymerization

Following a series of preliminary sequence specificity experiments, we observed sufficiently high fidelity of polymerization to warrant a high-throughput evaluation of the system. The measure of fidelity in XNA systems[19], which typically comprise a four-nucleotide code, relies on the sequencing analysis of the product of polymerization from a single template. Due to the large number of codons used during ligase-catalyzed pentanucleotide polymerization, a high

throughput approach was needed to provide adequate sampling of each codon to determine the fidelity of the process. Since the large size of the pentanucleotide codon set prohibits the ready analysis of sequence specificity by a conventional chain-termination approach, we looked to high-throughput DNA sequencing. Inspired by the "Duplex Sequencing" approach[26], which was developed to increase accuracy in high-throughput DNA sequencing, we reasoned that the barcoding of duplex pairs would enable post-sequencing association of template and polymerized strands (Figure 2.10). Thus, the sequences of the template and polymerized strands could be directly compared to reveal the fidelity of the ligase system for various codon sets in a high-throughput manner.

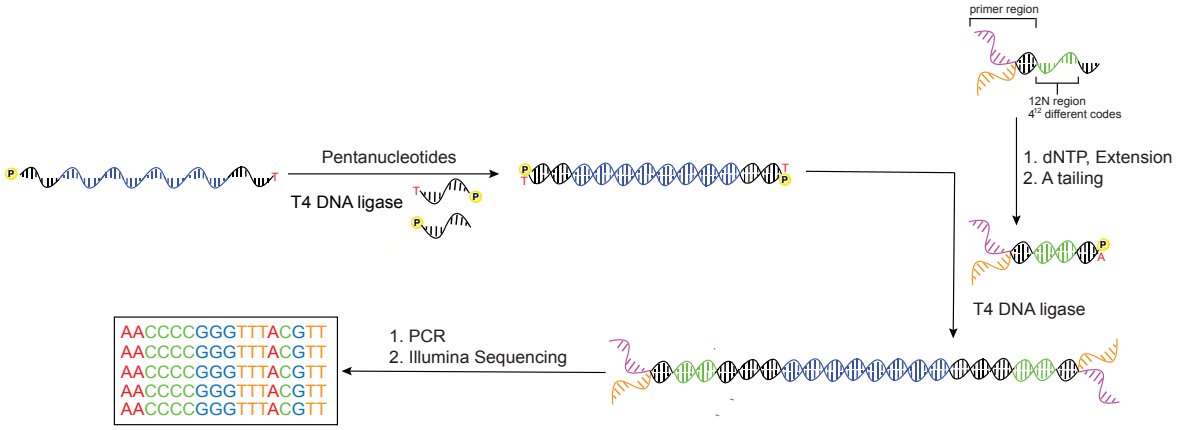


Figure 2.10 Workflow of duplex DNA sequencing

DNA template libraries were designed with two primer binding sites flanking a reading frame comprising eight consecutive repeats of various codon sets (Table 2.1). The DNA templates were modified with a 5'-phosphate and a 3'-T to facilitate downstream ligations of sequencing adapters. Two primers were then annealed to the template library; the initiation primer was modified with a 5'-phosphate and a 3'-T overhang, and the termination primer was modified with a 5'-phosphate. Following T4 DNA ligase-catalyzed polymerization of the corresponding pentanucleotide library, the duplex barcoding tag, which contained standard Illumina sequencing adapters and a 12bp randomized region, was ligated directly to both ends of

the duplex product and subsequently purified by PAGE. The adapter-ligated products were PCR amplified with flow-cell adapter sequences and subjected to paired-end DNA sequencing.

After DNA sequencing and processing, duplex pairs were grouped by their matching 24nt barcode tags and the template and polymerized strands were identified. The reading frames of the polymerized strands were then parsed into pentanucleotide sequences and compared against their template codons to identify errors. The frequency of each 5'-phosphorylated pentanucleotide sequence was determined, and their misincorporation rate was calculated. Readouts of errant codon sites for each pentanucleotide were generated to determine trends in fidelity. Overall fidelity of the codon set was calculated by the aggregated error frequency of all codons. Since PCR amplification (20-30 cycles) was performed after duplex tag ligation, the fidelity is a combined assessment of ligase-catalyzed polymerization and PCR amplification of the modified dsDNA, which is required for iterative rounds of *in vitro* selection.

Sequencing results for the polymerization of unmodified pentanucleotide libraries are summarized in Table 2.1. Surprisingly, polymerization with a 1024-membered NNNNN pentanucleotide library resulted in sequence-specific incorporation with 81.1% fidelity (entry 1, Table 2.1). Considering the complexity of such a large codon library, and the presence of single-nucleotide mismatch competitors, this level of fidelity is remarkable. As expected, with decreasing library complexity, the fidelity of the polymerization increases; when using a 64-membered codon set, 91.6% fidelity was observed (Table 2.1, entry 2). We also surveyed the fidelity of a 256-membered codon set, where adenosine was held constant at one position along the pentanucleotide (Table 2.1, entries 3-8). The fidelities for these codon sets were also high, ranging from 83.7-88.5%. All these pentanucleotides were unmodified and the fidelity was calculated for pentanucleotide incorporations.

Entry	Pentanucleotide	Codon set	Reads	Fidelity

1	5’P-NNNNN	NNNNN	3376	81.1%
2	5’P-ANWNW	WNWNT	1800	91.6%
3	5’P-ANNNN	NNNNT	1536	86.7%
4	5’P-NANNN	NNNTN	3968	87.5%
5	5’P-NNANN	NNTNN	288	88.5%
6	5’P-NNNAN	NTNNN	176	85.2%
7	5’P-NNNAN	NTNNN	$2.1x10^6$	83.7%
8	5'P-NNNNA	TNNNN	288	84.7%

Table 2.1 Fidelity of unmodified pentanucleotide polymerization

We next evaluated the sequence-specificity for ligase-catalyzed polymerization of pentanucleotides modified with chemical functionality. Using the 256-membered codon system, pentanucleotides containing a hexylamine group on the adenine nucleobase were polymerized along their corresponding templates and the fidelity of the process was assessed by duplex DNA sequencing (Table 2.2, entries 1-5, 7). Remarkably, the addition of a modification on the pentanucleotide decreased error rates >5-fold, resulting in fidelities of >95%. Error rates were the highest when the modification was at the 5’-end, but this level of fidelity should not preclude *in vitro* selection, as 65% of octacodon templates were generated faithfully without error. Notwithstanding, we focused on the NTNNN codon set, which provided 98.4 % fidelity. This level of fidelity was confirmed with a second higher throughout sequencing analysis, which used a lower amount of DNA template (35attmols) for PCR prior to DNA sequencing (Table 2.2, entry 5). We were concerned that the amine functionality, which would be protonated under the ligation conditions (pH7.6), might have enabled the observed higher sequence-specificity. This would pose problems when pentanucleotides containing various uncharged functional groups are used. To investigate this possible undesired effect, we synthesized a pentanucleotide library that was modified with an uncharged N-hexylacetamide group. Gratifyingly, the level of fidelity

remained high (Table 2.2, entry 6) suggesting that various small functional groups on the adenosine nucleobase could be accommodated by this polymerization method and 256-membered NTNNN codon set.

Entry	Pentanucleotide	Codon set	Reads	Fidelity
1	5'P-ANNNN	NNNNT	31936	95.1%
2	5'P-NANNN	NNNTN	2912	97.8%
3	5'P-NNANN	NNTNN	5232	98.0%
4	5'P-NNNAN	NTNNN	3032	98.4%
5	5'P-NNNAN	NTNNN	$2.4x10^5$	98.1%
6	5'P-NNNAN	NTNNN	$2.1x10^5$	97.6%
7	5'P-NNNNA	TNNNN	600	98.0%

Table 2.2 Fidelity of modified pentanucleotide polymerization

Although the reasons for the striking increase in fidelity that arise upon nucleobase modification are not entirely understood, one possibility is that they result from the perturbation in melting temperature upon nucleobase modification. It is known that modifications at C-8 of adenine results in the adenosine adopting a *syn* conformation about the N-glycoside bond in solution, which inhibits the kinetics of annealing and decreases the thermal stability of the DNA duplex[27-28]. The resulting drop in thermal duplex stability and slower annealing kinetics should give rise to a more discriminating ligation process. However, sequencing data indicate that GC-content, and thus the T_m of the pentanucleotide has little influence on fidelity within a codon set. For example, when polymerizing the NNNAN pentanucleotide library modified with N-hexylacetamide, codons grouped by %GC-content showed little variability with respect to error rate (80%GC = 2.4 %; 60%GC = 2.4 %; 40%GC = 2.4%; 20%GC = 2.6%; 0%GC = 2.7%). Notwithstanding, the decrease in thermal stability caused by the C-8 modification could have a similar destabilizing effect on each member within the codon set.

The observed high level of fidelity should allow for *in vitro* selection by DNA display methods[29], and in *vitro* evolution since the error rate is low enough to avoid error catastrophe during iterated cycles of selection for lengths typically implemented in SELEX experiments[30]. To highlight the application of this polymerization to in *vitro* selection, we performed oligonucleotide polymerization, selection, and amplification on a model DNA library. We prepared a DNA library, LIB1, which contained two primer-binding sites flanking eight consecutive repeats of the NTNNN codon set. We also prepared a biotinylated DNA template, POS1, which contained a *BcoDI* restriction site (Figure 2.11a). We diluted the POS1 template 1000-fold into LIB1 and performed a single round of in *vitro* selection. The round involved polymerization with 5'P-NNNAN library that was modified with a N-hexylacetamide group; a selection pressure, which involved binding to streptavidin-coated magnetic particles, followed by thorough washing; elution of the modified DNA strand upon incubation in 100mM NaOH; and then amplification of the eluted strand by PCR. To evaluate enrichment over the single round, the PCR product was digested with *BcoDI* and the ratio was compared against digestion of the amplified products of LIB1 and POS1 (Figure 2.11b).

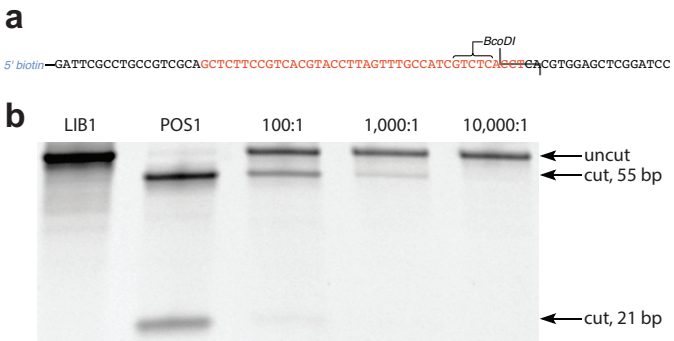


Figure 2.11 Single round of mock selection

This single round of mock selection resulted in approximately 100-fold enrichment, suggesting that the codon set could work effectively in an *in vitro* selection against molecular targets.

Conclusions

In summary, we have described a strategy to determine the fidelity of the T4 DNA ligase-catalyzed DNA-templated polymerization of modified oligonucleotides using high-throughput DNA duplex sequencing. The method allowed the evaluation of fidelity for each pentanucleotide within a large codon set. We applied this strategy to discover a 256-member codon set that enables the incorporation of small molecule functionality throughout a DNA polymer at >97 % fidelity. This high-fidelity polymerization strategy should find immediate application in DNA nanotechnology, DNA computing, and in *vitro* selection of functional nucleic acids.

Materials and methods

General Information

Unless otherwise noted, water was purified with Milli-Q purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies (IDT®). DNA oligonucleotides with amine modification were synthesized on a Bioautomation MerMade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent1260) using a C18 stationary phase (Eclipse-XDB C18, 5µm, 9.4 x 200mm) and an acetonitrile/100mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop™ ND2000 spectrophotometer. DNA sequences are written from 5'→3'. <Aam> = Amino-modifer C6 dA, <N>=A/T/C/G, <W>=A/T

Synthesis of amino-modified pentanucleotides

Pentanucleotides were synthesized on a MerMade 12 DNA synthesizer using a DMT-ON protocol on a 1µmol scale (1000Å CPG column). Amine-modifier C6 dA (Glen Research 10-

1089), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65°C in 500µl of a 1:1 mixture of ammonium hydroxide and methylamine for 15min. The cleaved resin was filtered away by filtration, and the oligonucleotide was concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The residue was then taken up into 100µl of H_2O and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 500µl of 80% aqueous acetic acid for 1h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 500µl 30% ammonium hydroxide at room temperature for 15min to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The dried product was dissolved into 100µl H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

T4-DNA Ligase-mediate polymerization protocols

DNA sequences

TP(NNNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNC ACG TGG AGC TCG GAT CCT

TP(WNWNT)8P /5Phos/GA TTC GCC TGC CGT CGC AWN WNT WNW NTW NWN TWN WNT WNW NTW NWN TWN WNT WNW NTC ACG TGG AGC TCG GAT CCT

TP(NNNNT)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTC ACG TGG AGC TCG GAT CCT

TP(NNNTN)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNC ACG TGG AGC TCG GAT CCT

TP(NNTNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNC ACG TGG AGC TCG GAT CCT

TP(NTNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNC ACG TGG AGC TCG GAT CCT

TP(TNNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ATN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNC ACG TGG AGC TCG GAT CCT

PR5 /5Phos/GG ATC CGA GCT CCA CGT G

PR6 /5Phos/TG CGA CGG CAG GCG AAT CT

NNNNN /5Phos/NNNNN

ANWNW /5Phos/ANWNW

ANNNN /5Phos/ANNNN

NH_2-ANNNN /5Phos/<Ama>NNNN

NANNN /5Phos/NANNN

NH_2-NANNN /5Phos/N<Ama>NNN

NNANN /5Phos/NNANN

NH_2-NNANN /5Phos/NN<Ama>NN

NNNAN /5Phos/NNNAN

NH_2-NNNAN /5Phos/NNN<Ama>N

NNNNA /5Phos/NNNNA

NH_2-NNNNA /5Phos/NNNN<Ama>

Protocols

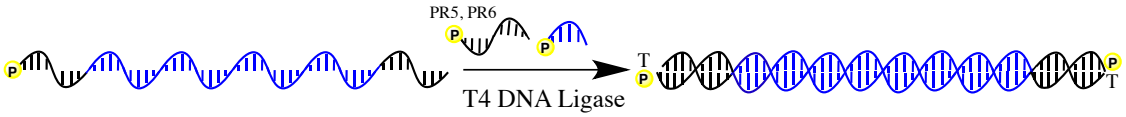


Figure 2.12 T4 DNA ligase-mediated polymerization

In a PCR tube was added DNA template (15pmol in 1.5µl water), PR5 (22.5pmol in 2.25µl water), PR6 (22.5pmol in 2.25µl water), 4µl NEBNext® Quick Ligation Reaction Buffer 5X, 7µl water. The mix was heated to 90°C for 2min and then cooled to 25°C at the rate of 0.1°C/s. In this PCR tube was then added pentanuclotides (480pmol in 1µl water), BSA (2µg in 1µl water) and 400U T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25°C for 24 hours. The products were then purified with MinElute® PCR Purification Kit for adapter ligation.

Polymerization efficiency with fluorescently-labeled primers

In a PCR tube was added DNA template TP(NNNNT)8P (15pmol in 1.5µl water), 5'-Cy5 modified PR5 (22.5pmol in 2.25µl water), 3'-6FAM modified PR6 (22.5pmol in 2.25µl water), 4µl NEBNext® Quick Ligation Reaction Buffer 5×, 7µl water for the experimental reaction, and 8µl water for the two control reactions. The mix was heated to 90°C for 2min and then cooled to 25°C at the rate of 0.1°C/s. In experimental reaction PCR tube was then added pentanucleotides ANNNN (480pmol in 1µl water), BSA (2µg in 1µl water) and 400U T4 DNA ligase (New England Biolabs, M0202L). In one of the control reactions was added pentanucleotides ANNNN (480pmol in 1µl water) and BSA (2µg in 1µl water). The other control reaction was added 400 U T4 DNA ligase (New England Biolabs, M0202L) and BSA (2µg in 1µl water). The polymerization was performed at 25°C for 24h. The products were then purified with MinElute® PCR Purification Kit. The image was observed by Typhoon™ imager.

Forced misincorporation polymerization

DNA sequences

matchedB-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT C

mismatchedB1-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAA CTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT C

mismatchedB2-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT GTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT C

mismatchedB3-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CAC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT C

mismatchedB4-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CTG TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT C

mismatchedB5-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CTC ATC TCT TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT C

P-AGAGA /5Phos/AG AGA

Protocol

In the PCR tube was added DNA template (15pmol in 1.5µl water), 4µl NEBNext® Quick Ligation Reaction Buffer 5X, 11.5µl water. The mix was heated to 90°C for 2min and then cooled to 25°C at the rate of 0.1°C/s. In all tubes were added BSA (2µg in 1µl water), 400 U T4 DNA ligase (New England Biolabs, M0202L) and P-AGAGA (480pmol in 1µl water). The products were then purified with MinElute® PCR Purification Kit. PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55°C for 90min.

Single-nucleotide chain termination polymerization

DNA sequences

idelitymixA1-c-5' /5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CTC TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT C

P-ACTCT /5Phos/ACTCT

P-AGAGA /5Phos/AGAGA

TGAGA TGAGA

ACAGA ACAGA

AGTGA AGTGA

AGAGT AGAGT

Protocol

In 5 PCR tubes were added DNA template (15pmol in 1.5µl water), 4µl NEBNext® Quick Ligation Reaction Buffer 5X, 11.5µl water. The mix was heated to 90°C for 2min and then cooled to 25°C at the rate of 0.1°C/s. In all tubes were added BSA (2µg in 1µl water), 400 U T4 DNA ligase (New England Biolabs, M0202L) and P-ACTCT (420pmol in 1µl water). In the first tube was added P-AGAGA (60pmol in 1µl water), the second tube was added TGAGA (60pmol in 1µl water), the third tube was added TGAGA/ACAGA/AGTGA/AGACA/AGAGT mixture (60pmol each, 300pmol total in 5µl water), the fourth tube was added mixture (120pmol each, 600pmol total in 5µl water), and the fifth tube was added mixture (180pmol each, 900pmol total in 5µl water). The polymerization was performed at 25°C for 24h. The products were then purified with MinElute® PCR Purification Kit. PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55°C for 90min.

Restriction digestion of polymerization product

DNA sequences

TP(rand1)8P GAT TCG CCT GCC GTC GCA TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TCA CGT GGA GCT CGG ATC C

ANNNN /5Phos/ANNNN

PR1 GGA TCC GAG CTC CAC GTG

PR2 /5Phos/TG CGA CGG CAG GCG AAT C

Protocol

In a PCR tube polymerization was performed with TP(rand1)8P and ANNNN following the T4-DNA ligase-mediate polymerization protocol with PR1 and PR2.

In another PCR tube was added DNA template TP(rand1)8P (15pmol in 1.5µl water), PR1 (22.5pmol in 2.25µl water), 5µl 10× NEBuffer™2 (New England Biolabs, M0212L), 37.5µl water. The mix was heated to 90°C for 2min and then cooled to 37°C at the rate of 0.1°C/s. The tube was then added 5U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L) and dNTP (1.25µl 10mM stock solution, Thermo Scientific, R0192). The extension was then performed at 37°C for 1h. The products were then purified with MinElute® PCR Purification Kit for further digestion.

In the PCR tubes were added 10µl polymerized dsDNA and extended dsDNA separately, 5µl 10× CutSmart® Buffer (New England Biolabs, R0542S), 10U *BcoDI* (New England Biolabs, R0542S) and 34µl water. The mix was then incubated at 37°C overnight. The digestion products were purified with Centri-Sep columns (Princeton Separations, CS-900). PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55°C for 90min.

Sanger sequencing of polymerized product

DNA sequence

TP(rand1)8PS GAT TCG CCT GCC GTC GCA TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TCA CGT GGA GCT CGG ATC C/iSp18/AACA ACA ACA ACA A

PR1 GGA TCC GAG CTC CAC GTG

PR2 /5Phos/TG CGA CGG CAG GCG AAT C

Protocol

In a PCR tube polymerization was performed with TP(rand1)8PS and ANNNN following the T4-DNA ligase-mediate polymerization protocol with PR1 and PR2.

In another PCR tube was added DNA template TP(rand1)8PS (15pmol in 1.5µl water), PR1 (22.5pmol in 2.25µl water), 5µl 10× NEBuffer™2 (New England Biolabs, M0212L), 37.5µl water. The mix was heated to 90°C for 2min and then cooled to 37°C at the rate of 0.1°C/s. The tube was then added 5U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L) and dNTP (1.25µl 10mM stock solution, Thermo Scientific, R0192). The extension was then performed at 37°C for 1h. The products were then purified with MinElute® PCR Purification Kit.

The polymerized strand and extended strand were then purified by gel purification (15% Mini-PROTEAN® TBE-Urea Gel). After stained with 0.5× SYBR® safe DNA gel stain (Life Technologies, S33100) for 15min, the product bands were cut out and the nucleotides were eluted with 0.3M NaCl at room temperature overnight. The products were then purified with Centri-Sep columns (Princeton Separations, CS-900). The nucleotides were then amplified with 16-cycle and 25-cycle PCR for extended strand and polymerized strand separately. The PCR products were sent to the Georgia Genomics and Bioinformatics Core (GGBC)for SANGER sequencing.

T4-DNA Ligase-mediate adapter ligation protocols

DNA sequences

iTruS_i7_D701 CAA GCA GAA GAC GGC ATA CGA GAT ATT ACT CGG TGA CTG GAG TTC AG

iTruS_i7_D702 CAA GCA GAA GAC GGC ATA CGA GAT TCC GGA GAG TGA CTG GAG TTC AG

iTruS_i7_D703 CAA GCA GAA GAC GGC ATA CGA GAT CGC TCA TTG TGA CTG GAG TTC AG

iTruS_i7_D704 CAA GCA GAA GAC GGC ATA CGA GAT GAG ATT CCG TGA CTG GAG TTC AG

iTruS_i7_D705 CAA GCA GAA GAC GGC ATA CGA GAT ATT CAG AAG TGA CTG GAG TTC AG

iTruS_i7_D706 CAA GCA GAA GAC GGC ATA CGA GAT GAA TTC GTG TGA CTG GAG TTC AG

iTruS_i7_D707 CAA GCA GAA GAC GGC ATA CGA GAT CTG AAG CTG TGA CTG GAG TTC AG

iTruS_i7_D708 CAA GCA GAA GAC GGC ATA CGA GAT TAA TGC GCG TGA CTG GAG TTC AG

iTruS_i7_D709 CAA GCA GAA GAC GGC ATA CGA GAT CGG CTA TGG TGA CTG GAG TTC AG

iTruS_i7_D710 CAA GCA GAA GAC GGC ATA CGA GAT TCC GCG AAG TGA CTG GAG TTC AG

iTruS_i7_D711 CAA GCA GAA GAC GGC ATA CGA GAT TCT CGC GCG TGA CTG GAG TTC AG

iTruS_i7_D712 CAA GCA GAA GAC GGC ATA CGA GAT AGC GAT AGG TGA CTG GAG TTC AG

PRIMER B AAT GAT ACG GCG ACC ACC GAG

AdapterA AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

AdapterB /5phos/AC TGN NNN NNN NNN NNA GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC

Protocol

Adapter duplex synthesis

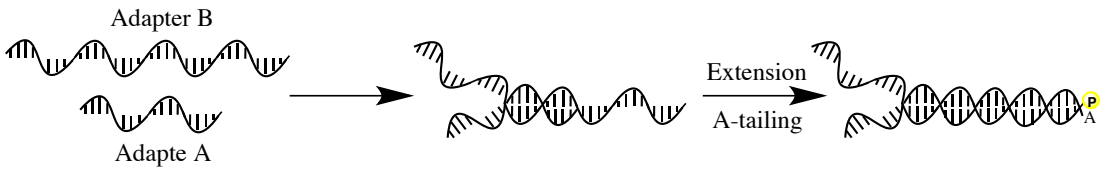


Figure 2.13 Adapter duplex synthesis

In a PCR tube was added 15µl of 100µM adapter A and 15µl of 100µM adapter B, then the tube was heated to 95°C for 5min and cooled to room temperature over 1 hour. Then in this PCR tube was added 4µl NEBuffer^{TM}2 10× (New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo^{-}, New England Biolabs, M0212L), 1µl dNTP Mix (Thermo Scientific, 10mM each). The extension was performed at 37°C for 1h. The adapter duplex was purified with QIAquick® Nucleotide Removal Kit, and then diluted in 30µl water.

In a PCR tube was added 30µl purified adapter duplex, 5µl NEBuffer^{TM}2 10× (New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo^{-}, New England Biolabs, M0212L), 5µl dATP (Thermo Scientific, 10mM), 5µl water. This PCR tube was incubated at 37°C for 1h for A-tailing. Then product was purified with QIAquick® Nucleotide Removal Kit, and then diluted in 30µl water.

Adapter ligation

Figure 2.14 Adapter ligation process

In a PCR tube was added polymerization products (12pmol in 10µl water), A-tailing adapter duplex (240pmol), 10µl NEBNext® Quick Ligation Reaction Buffer 5x, BSA (5µg in 2.5µl water), 1000U T4 DNA ligase (New England Biolabs, M0202L), water (make total volume 50µl). Then the ligation was performed at 16°C for 16h followed by gel. The products were diluted to 10µl water.

PCR amplification

Each purified adapter ligation product was amplified with a different primer from iTrus_D701 to iTrus_D712. The combinations were shown below:

iTrus_D701+ TP(NNNNN)8P+ NNNNN

iTrus_D702+ TP(WNWNT)8P+ ANWNW

iTrus_D703+ TP(NNNNT)8P+ ANNNN

iTrus_D704+ TP(NNNNT)8P+ NH_2-ANNNN

iTrus_D705+ TP(NNNTN)8P+ NANNN

iTrus_D706+ TP(NNNTN)8P+ NH_2-NANNN

iTrus_D707+ TP(NNTNN)8P+ NNANN

iTrus_D708+ TP(NNTNN)8P+ NH_2-NNANN

iTrus_D709+ TP(NTNNN)8P+ NNNAN

iTrus_D710+ TP(NTNNN)8P+ NH_2-NNNAN

iTrus_D711+ TP(TNNNN)8P+ NNNNA

iTrus_D712+ TP(TNNNN)8P+ NH_2-NNNNA

In a PCR tube was added 1µl purified adapter ligation product, 1.25µl 10µM Primer B, 1.25µl 10µM corresponding iTrus_D7XX primer, 9µl water and 12.5µl Q5® High-Fidelity 2× Master Mix (New England Biolabs, M0492S). The tube was then transferred to a preheated

thermocycler (98°C). The annealing temperature was 55°C for the first two cycles and remained 71°C for the rest of the cycles.

The PCR cycles were 20 for iTrus_D704 and iTrus_D712 since the corresponding adapter ligation yields were lower, and all the others were amplified for 10 cycles. The PCR products were then purified by gel purification.

High-Throughput DNA sequencing Protocol

The concentrations of gel purified samples were determined with Kapa library quantification kit for Illumina® libraries (KK4845) on Roche LightCycler 480. Paired-end Illumina® sequencing was performed on an Illumina® MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing) at the Georgia Genomics and Bioinformatics Core (GGBC) in the University of Georgia, Athens, GA, USA.

In vitro selection

DNA sequences

POS2(rand1) /5BiosG/GA TTC GCC TGC CGT CGC AGC TCT TCC GTC ACG TAC CTT AGT TTG CCA TCG TCT CAC CTC ACG TGG AGC TCG GAT

TP(NTNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNC ACG TGG AGC TCG GAT CCT

PR1 GGA TCC GAG CTC CAC GTG

PR2 /5Phos/TG CGA CGG CAG GCG AAT C

NH_2-NNNAN /5Phos/NNN<Ama>N

Protocol

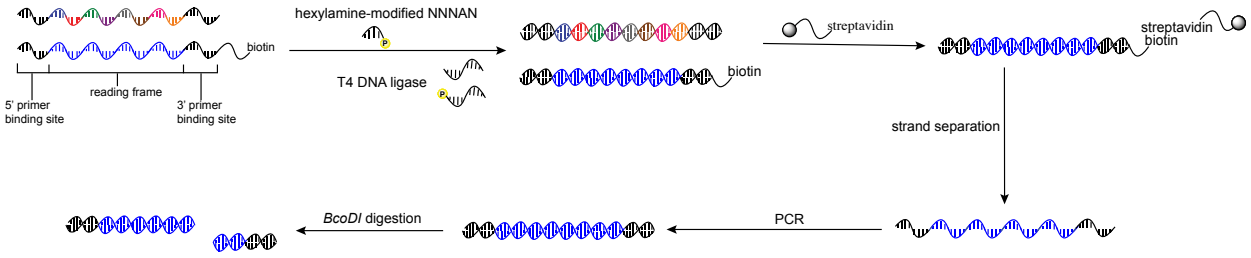


Figure 2.15 Single round of in *vitro* selection using hexylamine modified NNNAN

Polymerization was performed with the template mixture of POS2(rand1):TP(NTNNN)8P = 1:100, following the following the T4 DNA ligase-mediate polymerization protocol with NH_2-NNNAN pentanucleotides. The polymerized products were then incubated with 30µl streptavidin magnetic beads (Life Technologies, Dynabeads® MyOne™ Streptavidin C1), which were washed with 200µl binding buffer (0.5M NaCl, 200mM Tris-HCl pH7.5, 1mM EDTA) 3 times before binding. After 30min incubation at room temperature, TP(NTNNN)8P products were washed away with 200µl binding buffer 3 times. The dsDNA was then separated with 40µl freshly made 150mM NaOH. After incubation at room temperature for 15min, 4µl 1.5M HCl quenched the reaction. The liberated strand was then purified with Centri-Sep columns (Princeton Separations, CS-900). The products were then PCR amplified with 10 cycles. The products were then purified with MinElute® PCR Purification Kit for further digestion. In the PCR tubes were added 10µl PCR amplified dsDNA, 5µl 10× CutSmart® Buffer (New England Biolabs, R0542S), 10U *BcoDI* (New England Biolabs, R0542S) and 34µl water. The mix was then incubated at 37°C overnight. The digestion products were purified with Centri-Sep columns (Princeton Separations, CS-900). PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55°C for 90min.

References

1. 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.

2. Blow, N., The generation game. *Nature* **2007,** *447*, 741.

3. Cummings, Richard D.; Pierce, J. M., The Challenge and Promise of Glycomics. *Chem. Biol.* **2014,** *21* (1), 1-15.

4. Berglund, L.; Björling, E.; Oksvold, P.; Fagerberg, L.; Asplund, A.; Al-Khalili Szigyarto, C.; Persson, A.; Ottosson, J.; Wernérus, H.; Nilsson, P.; Lundberg, E.; Sivertsson, Å.; Navani, S.; Wester, K.; Kampf, C.; Hober, S.; Pontén, F.; Uhlén, M., A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies. *Mol. Cell. Proteomics* **2008,** *7* (10), 2019-2027.

5. Colwill, K.; Renewable Protein Binder Working, G.; Persson, H.; Jarvik, N. E.; Wyrzucki, A.; Wojcik, J.; Koide, A.; Kossiakoff, A. A.; Koide, S.; Sidhu, S.; Dyson, M. R.; Pershad, K.; Pavlovic, J. D.; Karatt-Vellatt, A.; Schofield, D. J.; Kay, B. K.; McCafferty, J.; Mersmann, M.; Meier, D.; Mersmann, J.; Helmsing, S.; Hust, M.; Dübel, S.; Berkowicz, S.; Freemantle, A.; Spiegel, M.; Sawyer, A.; Layton, D.; Nice, E.; Dai, A.; Rocks, O.; Williton, K.; Fellouse, F. A.; Hersi, K.; Pawson, T.; Nilsson, P.; Sundberg, M.; Sjöberg, R.; Sivertsson, Å.; Schwenk, J. M.; Takanen, J. O.; Hober, S.; Uhlén, M.; Dahlgren, L.-G.; Flores, A.; Johansson, I.; Weigelt, J.; Crombet, L.; Loppnau, P.; Kozieradzki, I.; Cossar, D.; Arrowsmith, C. H.; Edwards, A. M.; Gräslund, S., A roadmap to generate renewable protein binders to the human proteome. *Nat. Methods* **2011,** *8*, 551.

6. Nollau, P., Generating high-quality protein binders: a large screening effort pays off. *Nat. Methods* **2011,** *8*, 545.

7. Taussig, M. J.; Stoevesandt, O.; Borrebaeck, C. A. K.; Bradbury, A. R.; Cahill, D.; Cambillau, C.; de Daruvar, A.; Dübel, S.; Eichler, J.; Frank, R.; Gibson, T. J.; Gloriam, D.; Gold, L.; Herberg, F. W.; Hermjakob, H.; Hoheisel, J. D.; Joos, T. O.; Kallioniemi, O.; Koegl, M.; Konthur, Z.; Korn, B.; Kremmer, E.; Krobitsch, S.; Landegren, U.; van der Maarel, S.; McCafferty, J.; Muyldermans, S.; Nygren, P.-Å.; Palcy, S.; Plückthun, A.; Polic, B.; Przybylski,

M.; Saviranta, P.; Sawyer, A.; Sherman, D. J.; Skerra, A.; Templin, M.; Ueffing, M.; Uhlén, M., ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods* **2007,** *4*, 13.

8. Bordeaux, J.; Welsh, A. W.; Agarwal, S.; Killiam, E.; Baquero, M. T.; Hanna, J. A.; Anagnostou, V. K.; Rimm, D. L., Antibody validation. *BioTechniques* **2010,** *48* (3), 197-209.

9. Uhlén, M.; Björling, E.; Agaton, C.; Szigyarto, C. A.-K.; Amini, B.; Andersen, E.; Andersson, A.-C.; Angelidou, P.; Asplund, A.; Asplund, C.; Berglund, L.; Bergström, K.; Brumer, H.; Cerjan, D.; Ekström, M.; Elobeid, A.; Eriksson, C.; Fagerberg, L.; Falk, R.; Fall, J.; Forsberg, M.; Björklund, M. G.; Gumbel, K.; Halimi, A.; Hallin, I.; Hamsten, C.; Hansson, M.; Hedhammar, M.; Hercules, G.; Kampf, C.; Larsson, K.; Lindskog, M.; Lodewyckx, W.; Lund, J.; Lundeberg, J.; Magnusson, K.; Malm, E.; Nilsson, P.; Ödling, J.; Oksvold, P.; Olsson, I.; Öster, E.; Ottosson, J.; Paavilainen, L.; Persson, A.; Rimini, R.; Rockberg, J.; Runeson, M.; Sivertsson, Å.; Sköllermo, A.; Steen, J.; Stenvall, M.; Sterky, F.; Strömberg, S.; Sundberg, M.; Tegel, H.; Tourle, S.; Wahlund, E.; Waldén, A.; Wan, J.; Wernérus, H.; Westberg, J.; Wester, K.; Wrethagen, U.; Xu, L. L.; Hober, S.; Pontén, F., A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Mol. Cell. Proteomics* **2005,** *4* (12), 1920-1932.

10. Mahon, C. S.; Fulton, D. A., Mimicking nature with synthetic macromolecules capable of recognition. *Nat. Chem.* **2014,** *6*, 665.

11. Gebauer, M.; Skerra, A., Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* **2009,** *13* (3), 245-255.

12. Mayer, G., The Chemical Biology of Aptamers. *Angew. Chem. Int. Edit.* **2009,** *48* (15), 2672-2689.

13. Lollo, B.; Steele, F.; Gold, L., Beyond antibodies: New affinity reagents to unlock the proteome. *PROTEOMICS* **2014,** *14* (6), 638-644.

14. Marx, V., Calling the next generation of affinity reagents. *Nat. Methods* **2013,** *10*, 829.

15. Wilson, D. S.; Szostak, J. W., In Vitro Selection of Functional Nucleic Acids. *Annu. Rev. Biochem.* **1999,** *68* (1), 611-647.

16. Rohloff, J. C.; Gelinas, A. D.; Jarvis, T. C.; Ochsner, U. A.; Schneider, D. J.; Gold, L.; Janjic, N., Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol. Ther. - Nucleic Acids* **2014,** *3*, e201.

17. Taylor, A. I.; Arangundy-Franklin, S.; Holliger, P., Towards applications of synthetic genetic polymers in diagnosis and therapy. *Curr. Opin. Chem. Biol.* **2014,** *22*, 79-84.

18. Chaput, John C.; Yu, H.; Zhang, S., The Emerging World of Synthetic Genetics. *Chem. Biol.* **2012,** *19* (11), 1360-1371.

19. Pinheiro, V. B.; Holliger, P., The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **2012,** *16* (3), 245-252.

20. Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M., Toward the Combinatorial Selection of Chemically Modified DNAzyme RNase A Mimics Active Against all-RNA Substrates. *ACS Combinatorial Science* **2013,** *15* (4), 174-182.

21. 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.

22. 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), 5203-5211.

23. 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.

24. Niu, J.; Hili, R.; Liu, D. R., Enzyme-Free Translation of DNA into Sequence-Defined Synthetic Polymers Structurally Unrelated to Nucleic Acids. *Nat. Chem.* **2013,** *5* (4), 282-292.

25. Schmitt, M. W.; Kennedy, S. R.; Salk, J. J.; Fox, E. J.; Hiatt, J. B.; Loeb, L. A., Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* **2012,** *109* (36), 14508-13.

26. Schmitt, M. W.; Kennedy, S. R.; Salk, J. J.; Fox, E. J.; Hiatt, J. B.; Loeb, L. A., Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (36), 14508-14513.

27. Dierckx, A.; Dinér, P.; El-Sagheer, A. H.; Kumar, J. D.; Brown, T.; Grøtli, M.; Wilhelmsson, L. M., Characterization of photophysical and base-mimicking properties of a novel fluorescent adenine analogue in DNA. *Nucleic Acids Res.* **2011,** *39* (10), 4513-4524.

28. Luyten, I.; Herdewijn, P., Hybridization properties of base-modified oligonucleotides within the double and triple helix motif. *Eur. J. Med. Chem.* **1998,** *33* (7), 515-576.

29. Li, G.; Zheng, W.; Liu, Y.; Li, X., Novel encoding methods for DNA-templated chemical libraries. *Curr. Opin. Chem. Biol.* **2015,** *26*, 25-33.

30. Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A., Effect of Stalling after Mismatches on the Error Catastrophe in Nonenzymatic Nucleic Acid Replication. *J. Am. Chem. Soc.* **2010,** *132* (16), 5880-5885.

CHAPTER 3

ENZYMATIC SYNTHESIS OF SEQUENCE-DEFINED SYNTHETIC NUCLEIC ACID POLYMERS WITH DIVERSE FUNCTIONAL GROUPS[1]

[1]Kong, D.; Lei, Y.; Yeung, W.; Hili, R., Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angew. Chem. Int. Edit.* **2016,** *55* (42), 13164-13168.

Abstract

The development and in-depth analysis of T4 DNA ligase-catalyzed DNA-templated oligonucleotides polymerization is described here. NNNNT codon set is proved to enable high efficiency, high fidelity and low codon bias for the modified ANNNN libraries. The 256-membered library comprises 16 modified sub-libraries with different functional groups. The fidelity of polymerization remains high despite the complexity of the system. This discovery should prove its ready application in DNA nanotechnology, DNA computing and in *vitro* evolution of functional nucleic acid polymers.

Introduction

Proteins play the most diverse roles and functions in nature among all the sequence-defined biopolymers. The hegemony of proteinogenic biopolymers as biological receptors and catalysts arises from their broad structural and chemical diversity, which enables them to effectively engage their molecular targets with high affinity and specificity. Despite the ability of nucleic acid polymers to fold into complex three-dimensional structures, their functional group deficit has limited their ability to match the diverse activities of proteins[2]. Thus, technologies that enable the sequence-defined synthesis of nucleic acid polymers with diverse chemical functionality has received significant attention[3-4]. Indeed, DNA and RNA polymerases have been used to achieve the DNA-templated incorporation of modified nucleotides to facilitate the evolution of functionalized nucleic acid polymers with significantly improved activities over their unmodified counterparts[5-6]. Homomultivalent display of a hydrophobic functional group has been demonstrated to increase the binding affinity and lower kinetic off-rates of nucleic acid aptamers raised against various protein targets[7-8]. Furthermore, the homomultivalent display of various functional groups has increased the catalytic potential of nucleic acid polymers for ribonuclease activity[9] and protease activity[10]. Expanding DNA into a heteromultivalent polymer

comprising two or three functional groups has enabled the evolution of divalent metal-independent nuclease activity, thus opening the door to their application *in vivo*[11-14].

Methods that permit the sequence-defined incorporation of a diverse library of functional groups along a library of ssDNA will allow the concomitant *in vitro* evolution of both the DNA architecture and the identities of the displayed functional groups. Therefore, the spatial optimization of the heteromultivalent ensemble of weak interactions between the functional groups and their molecular target can be achieved[15].

Our group has explored enzymatic methods to incorporate multiple instances of a diverse set of functional groups throughout a DNA polymer. Recently, we developed the Ligase-catalyzed *OligOnucleotide PolymERization* (LOOPER) method to access this class of sequence-defined synthetic biopolymers[16-18]. The method relies upon the T4 DNA ligase-catalyzed DNA-templated copolymerization of a library of modified 5'-phosphorylated pentanucleotides. As the method employs codons, rather than single nucleotides, the theoretical maximum number of unique modifications that can be incorporated for a pentanucleotide system is 1024.

DNA duplex sequencing was used to analyze the fidelity of LOOPER when polymerizing pentanucleotide libraries of various nucleic acid composition along large libraries of DNA templates[17]. Our initial analysis with amine-modified building blocks revealed high fidelity of the system, thus we sought to apply this to the simultaneous incorporation of different modifications.

Results

First, we analyzed the efficiency of polymerization with a fixed amine-modified adenosine base (Figure 3.1) at different positions on the pentanucleotide building blocks as shown in Table 3.1.

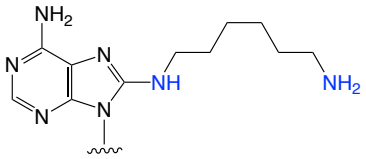


Figure 3.1 Amine-modified adenosine base

Pentanucleotide	Codon Set	Yield	Fidelity
A*NNNN	NNNNT	68%	95.1%
NA*NNN	NNNTN	62%	97.8%
NNA*NN	NNTNN	53%	98.0%
NNNA*N	NTNNN	30%	98.4%
NNNNA*	TNNNN	10%	98.0%

Table 3.1 Yield and fidelity of pentanucleotide building blocks

The highest fidelity was achieved with NTNNN codon set. However, the yield was only 30%. This could potentially result in skewing of the codon distribution and convolute *in vitro* evolution efforts of this class of biopolymer. We compared the codon bias of NTNNN codon set with the highest yield codon set NNNNT by comparing the frequency of a codon in the polymers and frequency of the same codon in the templates, which was defined as the enrichment of that codon. Bias was the standard deviation of the enrichment. The results were shown in Figure 3.2.

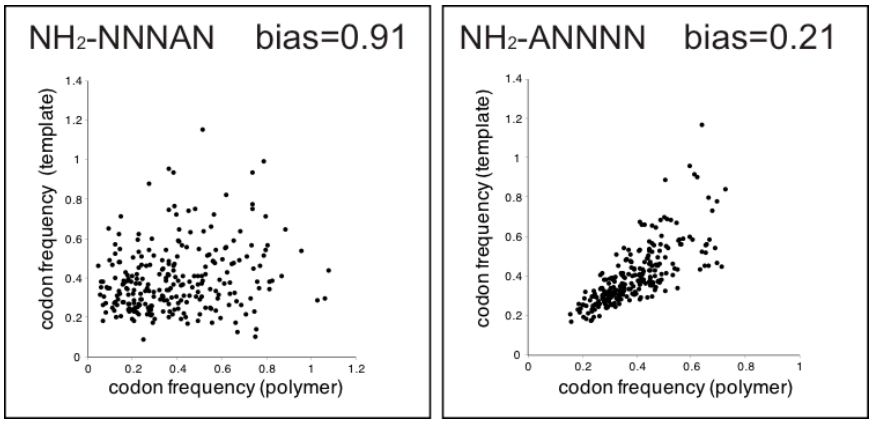


Figure 3.2 Codon bias analysis of NTNNN and NNNNT

It was obvious that NTNNN had a high codon bias compared with NNNNT. Considering the application of *in vitro* selection, we decided to further pursue NNNNT codon set instead of NTNNN due to its low bias and high polymerization efficiency with high fidelity level.

Optimizations with various incubation temperatures and ATP concentrations were performed with NNNNT codon set. Temperatures had little effect on the fidelity and yield of polymerization. Lower ATP concentrations resulted in higher yield, which might be the result of decreased inhibition of DNA binding[19] and the minimization of an over-adenylated system[18].

Modified 5'-phosphorylated ANNNN pentanucleotides were prepared to study the capacity of LOOPER to tolerate various functional groups. The modifications were incorporated through amide bond-formation between the hexylamine appended to the C8 position of adenosine and various carboxylic acid derivatives. This modification site has been previously shown to be the most permissive for this system[18].

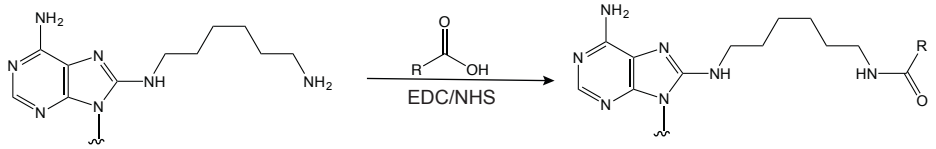


Figure 3.3 Amide bond formation reaction

Duplex DNA sequencing was performed on the products of LOOPER for each monofunctionalized library, and the yields, biases, and fidelities for each library were reported in Table 3.2.

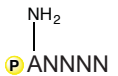
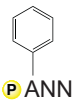
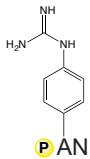
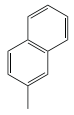
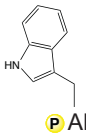
	pANNNN 1	 pANNNN 2	 pANNNN 3	 pANNNN 4	 pANNNN 5	 pANNNN 6	 pANNNN 7
Yield	84%	75%	73%	63%	66%	62%	60%
Fidelity	86.7%	95.1%	93.9%	94%	91.3%	95.1%	93.2%
Bias	0.16	0.21	0.34	0.36	0.36	0.25	0.37

Table 3.2 Characterization of homofunctionalized ANNNN libraries

On the whole, polymerization fidelities were not governed by the molecular size of the functional group modification. This is not surprising, as pentanucleotides modified with peptide fragments have been show to polymerize efficiency and with high fidelity[18]. While codon biases were greater than either the unmodified or amine-modified ANNNN libraries, the biases remained considerably lower than the amine-modified NNNAN library, suggesting that this codon set could be applied to *in vitro* selection systems where strong levels of codon bias would be problematic, such as selections operating with poor enrichment levels.

Further analysis of the NNNNT codon sets revealed several trends with respect to codon sequence and fidelity. We analyzed the consensus sequence for the misincorporations of each codon during LOOPER with amine-modified and phenyl-modified ANNNN libraries. One of the most striking trends is that single-nucleotide errors arising from misincorporation most frequently occur at the first two nucleotides at the 5'-end of the codon (3'-end of the pentanucleotide building block). The only exception is when the 5'-nucleotide of the codon is a dC, which is largely conserved during misincorporations. Figure 3.4 is an example of the sequence logos indicating the mismatched pentanucleotides (codon sequence written from

5’→3’) in amine-modified ANNNN libraries. This trend is witnessed across the entire 256-membered codon set.

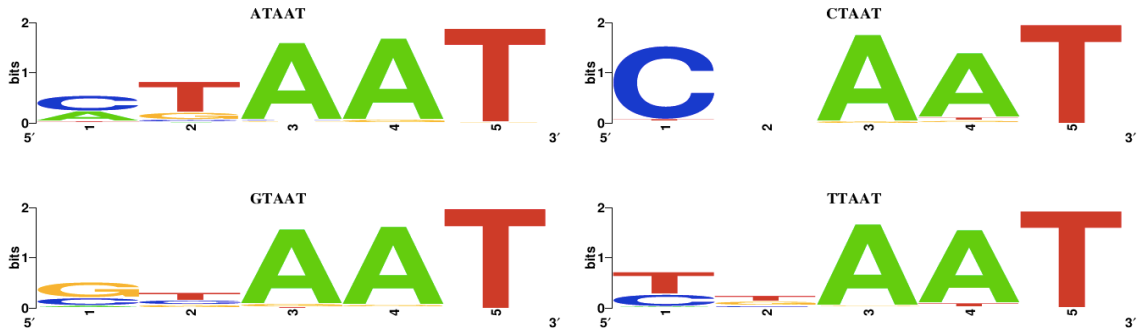
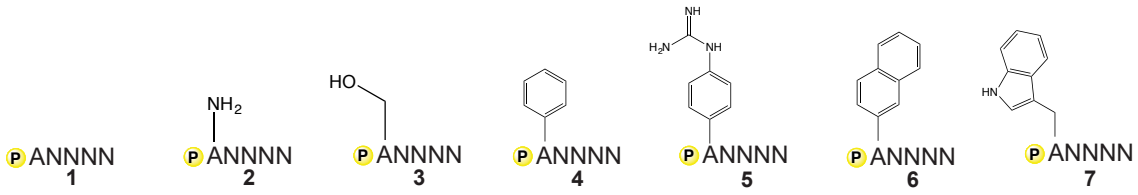


Figure 3.4 DNA sequence logo for misincoportation at NTAAT codons

The monofunctionalized ANNNN libraries were then analyzed by nucleotide identity at each position. The results would indicate whether specific nucleotides would influence the fidelity (Figure 3.5).

All the error rates were normalized so the sum of four sub-libraries was set to 1.0. The figures showed the ratios of these sub-libraries for every individual functional group. As the results indicated, different functional groups had little influence on the distribution of the error rate ratios. This proved the similar tolerance of LOOPER system to a variety of functionalization. On the other hand, we observed higher error rates for pentanucleotide building blocks that contained either a dA or dT at the 3’-position. Other positions were less sensitive to the nucleotide identity.



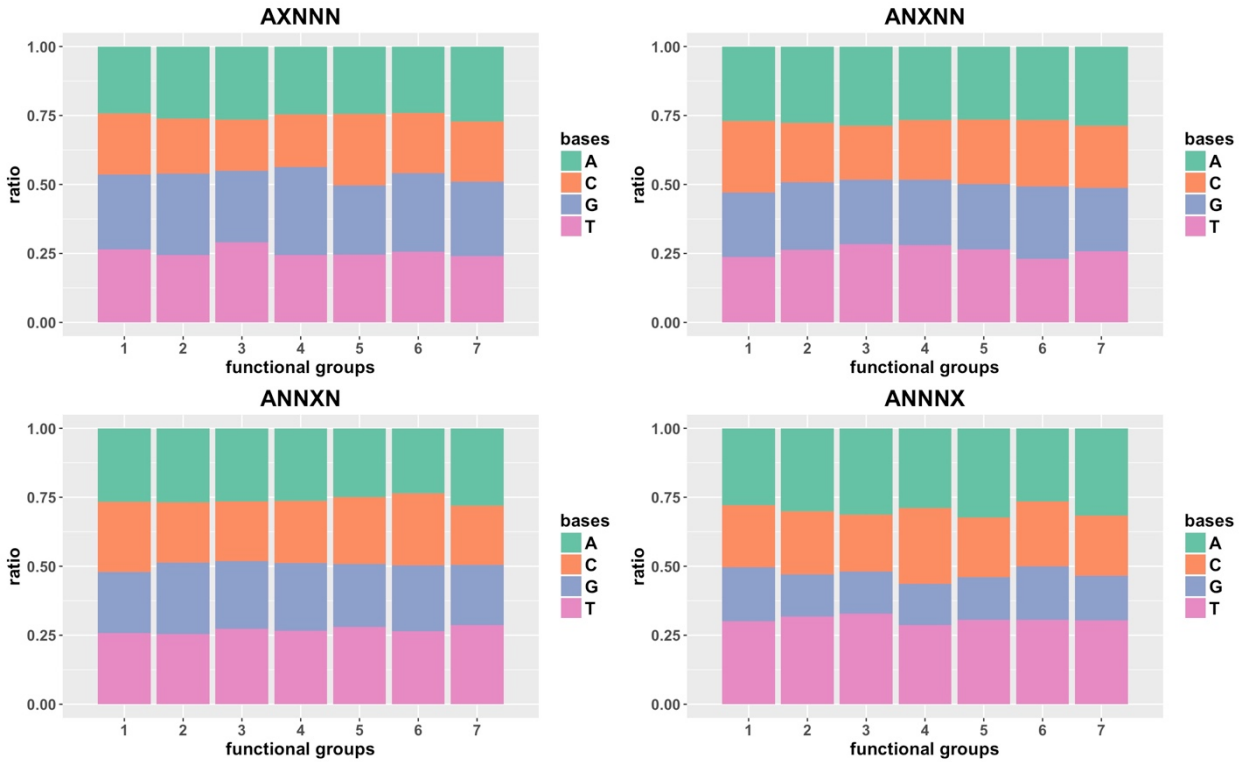


Figure 3.5 Analysis of error rate influenced by nucleotide identity

It is possible that a 128-membered codon set lacking a dA or dT at the 5'-position, namely SNNNT, could result in higher fidelity, albeit at the cost of sequence diversity. A somewhat similar approach has been used in the successful molecular evolution of TNA aptamers with a reduced genetic code[20]. Alternatively, noncanonical bases such as 2,6-diaminopurines, could be used in place of dA to modulate the fidelity[21].

Inspired by the trinucleotide code used during ribosomal translation of mRNA into proteins, we sought to organize the 256-membered NNNNT codon set into degenerate sub-libraries, where each sub-library would encode a unique modification. Thus, evolution of the DNA scaffold and functional group can occur concomitantly. Namely, errors that arise during LOOPER could either result in synonymous mutations that change the ssDNA scaffold, while maintaining the displayed functional group identity, or missense mutations that concomitantly change the identity of the displayed functional group and the ssDNA scaffold. The comparison of different sub-libraries design is shown in Figure 3.6.

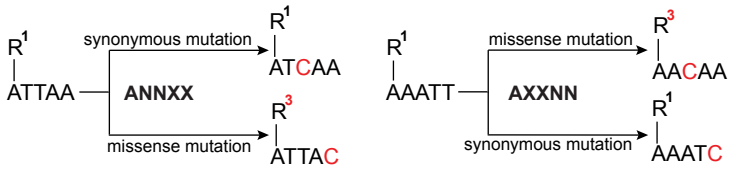


Figure 3.6 Comparison of ANNXX and AXXNN sub-libraries

It has been previously analyzed that more mismatches would happen at the 3'-end of the ANNNN pentanucleotides. Thus, ANNXX sub-libraries would have a higher probability of missense mutation compared with AXXNN. To achieve better evolutionary results, we decided to pursue a codon set derived from XXNNT, where XX represents a dinucleotide sequence that specifies the functional group modification on the pentanucleotide (Figure 3.7).

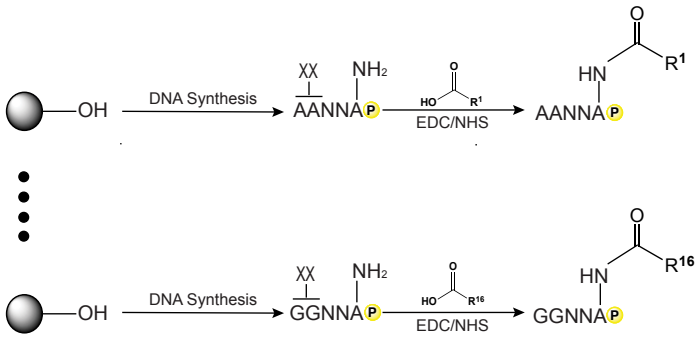


Figure 3.7 Synthesis of modified sub-libraries

Functional groups and their corresponding genotype building blocks are specified as shown in Figure 3.8. The 16 sub-libraries were mixed in equimolar ratios to generate the 256-membered heterofunctionalized pentanucleotide library, which was used in LOOPER with a library of templates comprising eight repeats of the NNNNT codon set within the reading frame. The success of all coupling reactions was confirmed by ESI MS analysis.

Diverse functionalities were incorporated, and such complex system retained high fidelity level about 91.2%. Further analysis on the heterofunctionalized library was performed by my lab mate using similar strategy as the homofunctionalized libraries.

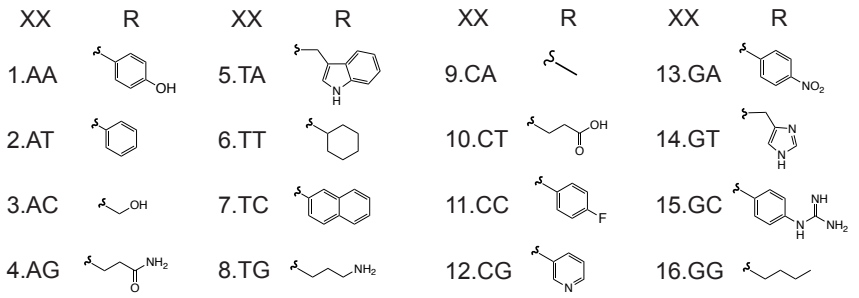


Figure 3.8 Functional groups of 16 sub-libraries

Conclusions

Our findings reported herein demonstrate that the LOOPER process represents a viable approach toward the library generation of sequence-defined nucleic acid polymers with diverse chemical functionality. We show that LOOPER can operate with high fidelity and low sequence bias along an NNNNT codon set, enabling the sequence-defined copolymerization of a 256-membered ANNNN pentanucleotide library. Furthermore, we demonstrate that this library can be divided into 16 sub-libraries, each comprising 16 pentanucleotides adorned with a unique functional group. This heterofunctionalized pentanucleotide library was shown to polymerize with excellent fidelity and minimal bias. The high fidelity of the process and customizability of the functional groups gives promise to its application to modified nucleic acid polymer synthesis for DNA nanotechnology; *in vitro* selection of nucleic acid polymers for desired molecular function; and DNA computing.

Materials and methods

General Information

Unless otherwise noted, water was purified with Milli-Q purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies (IDT®). DNA oligonucleotides with amine modification were synthesized on a Bioautomation MerMade 12 synthesizer. All materials and reagents used for oligonucleotide

synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent1260) using a C18 stationary phase (Eclipse-XDB C18, 5µm, 9.4 x 200mm) and an acetonitrile/100mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop™ ND2000 spectrophotometer. DNA sequences are written from 5'→3'. <Aam> = Amino-modifer C6 dA, <N>=A/T/C/G.

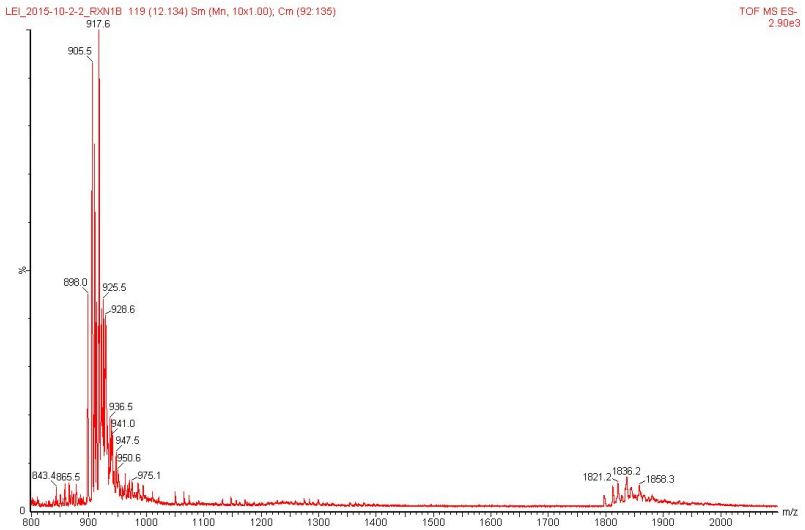
Synthesis of amino-modified pentanucleotides

Pentanucleotides were synthesized on a Mermaid12 DNA synthesizer using a DMT-ON protocol on a 1µmol scale (1000Å CPG column). Amine-modifier C6 dA (Glen Research 10-1089), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 25°C in 400µl of a 1:1 mixture of ammonium hydroxide and methylamine (AMA) for 25mins. The cleaved resin was filtered away by filtration, and the oligonucleotide was further incubated in the same AMA solution at 60°C for 30min. The cleaved oligonucleotides concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The residue was then taken up into 100µl of H_2O and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 1ml of 40% aqueous acetic acid for 1h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 400µl 30% ammonium hydroxide at room temperature for 15min to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using the CentriVap DNA Vacuum

Concentrators (LABCONCO). The dried product was dissolved into 100µl H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

Functionalization of amino-modified pentanucleotides with carboxylic acids

A mixture of 25µL carboxylic acid (100mM in DMSO), 25µl N-hydroxysulfosuccinimide (sNHS, 100mM in 1:1 mixture of DMSO and H_2O), 5µl 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 100mM in DMSO) and 7.5µl DMSO was incubated at room temperature for 30min. Followed by addition 7.5µl of amino modified pentanucleotides (10nmol in H_2O) and 30µl Na_2CO_3 buffer (500mM, pH9.0). The mixture was incubated at room temperature overnight with vortex. The reaction was then quenched by addition of 15µl Tris buffer (500mM, pH8.0) at room temperature for one hour. Functionalized pentanucleotide was then purified with HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified pentanucleotide was dissolved in water and characterized with mass spectrometry. Figure 3.9 showed an example of MS result (ANNTA coupling reaction).



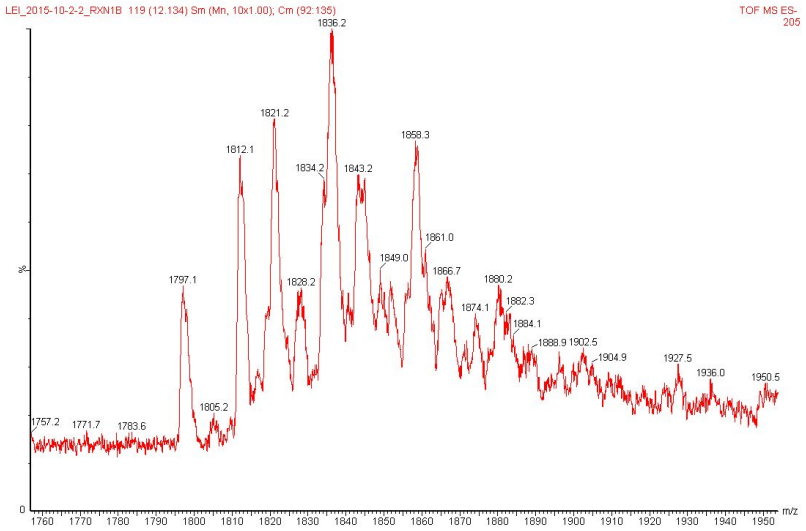


Figure 3.9 Mass spectrometry result of ANNTA sub-library

The theoretical molecular weight range after coupling reaction was 1797~1877. The lowest molecular weight was from ACCTA, and the next peak should be 1812(ACTTA/ATCTA). All 16 members in this sub-library could be confirmed by MS even though some of signals were not that strong. The other coupling reactions were all confirmed the same way.

T4-DNA Ligase-mediate polymerization protocols

DNA sequences

TP(NNNNT)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTC ACG TGG AGC TCG GAT CCT

PR5 /5Phos/GG ATC CGA GCT CCA CGT G

PR6 /5Phos/TG CGA CGG CAG GCG AAT CT

NH_2-ANNNN /5Phos/<Ama>NNNN

Protocols

In a PCR tube was added DNA template (15pmol in 1.5µl water), PR5 (22.5pmol in 2.25µl water), PR6 (22.5pmol in 2.25µl water), 5µl reaction buffer 4X (1X buffer: 66mM Tris-HCl, 10mM $MgCl_2$, 10mM DTT, 6% PEG6000, pH7.6@25°C), 2µl modified ATP cofactor (0.25mM

in water) and 4µl water. The mix was heated to 90°C for 2min and then cooled to 25°C at the rate of 0.1°C/s. In this PCR tube was then added pentanuclotides (480pmol in 1µl water), BSA (2µg in 1µl water) and 400U T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25°C for 24h. The products were then purified with MinElute® PCR Purification Kit for adapter ligation.

T4-DNA Ligase-mediate adapter ligation protocols

DNA sequences

iTruS_i7_D701 CAA GCA GAA GAC GGC ATA CGA GAT ATT ACT CGG TGA CTG GAG TTC AG

iTruS_i7_D702 CAA GCA GAA GAC GGC ATA CGA GAT TCC GGA GAG TGA CTG GAG TTC AG

iTruS_i7_D703 CAA GCA GAA GAC GGC ATA CGA GAT CGC TCA TTG TGA CTG GAG TTC AG

iTruS_i7_D704 CAA GCA GAA GAC GGC ATA CGA GAT GAG ATT CCG TGA CTG GAG TTC AG

iTruS_i7_D705 CAA GCA GAA GAC GGC ATA CGA GAT ATT CAG AAG TGA CTG GAG TTC AG

iTruS_i7_D706 CAA GCA GAA GAC GGC ATA CGA GAT GAA TTC GTG TGA CTG GAG TTC AG

iTruS_i7_D707 CAA GCA GAA GAC GGC ATA CGA GAT CTG AAG CTG TGA CTG GAG TTC AG

iTruS_i7_D708 CAA GCA GAA GAC GGC ATA CGA GAT TAA TGC GCG TGA CTG GAG TTC AG

PRIMER B AAT GAT ACG GCG ACC ACC GAG

AdapterA AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

AdapterB /5phos/AC TGN NNN NNN NNN NNA GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC

Protocol

Adapter duplex synthesis

In a PCR tube was added 15µl of 100µM adapter A and 15µl of 100µM adapter B, then the tube was heated to 95°C for 5min and cooled to room temperature over 1 hour. Then in this PCR tube was added 4µl NEBuffer™2 10× (New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 1µl dNTP Mix (Thermo Scientific, 10mM each). The extension was performed at 37°C for 1h. The adapter duplex was purified with QIAquick® Nucleotide Removal Kit, and then diluted in 30µl water.

In a PCR tube was added 30µl purified adapter duplex, 5µl NEBuffer™2 10× (New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 5µl dATP (Thermo Scientific, 10mM), 5µl water. This PCR tube was incubated at 37°C for 1h for A-tailing. Then product was purified with QIAquick® Nucleotide Removal Kit, and then diluted in 30µl water.

Adapter ligation

In a PCR tube was added polymerization products (12pmol in 10µl water), A-tailing adapter duplex (240pmol), 10µl NEBNext® Quick Ligation Reaction Buffer 5x, BSA (5µg in 2.5µl water), 1000U T4 DNA ligase (New England Biolabs, M0202L), water (make total volume 50µl). Then the ligation was performed at 16°C for 16h followed by gel purification. The products were diluted to 10µl water.

PCR amplification

Each purified adapter ligation product was amplified with a different primer from iTrus_D701 to iTrus_D708. Each iTrus_D70X was paired with a specific functional group.

In a PCR tube was added 1µl purified adapter ligation product, 1.25µl 10µM Primer B, 1.25µl 10µM corresponding iTrus_D70X primer, 9µl water and 12.5µl Q5® High-Fidelity 2× Master Mix (New England Biolabs, M0492S). The tube was then transferred to a preheated thermocycler (98°C). The annealing temperature was 55°C for the first two cycles and remained 71°C for the rest of the cycles.

High-Throughput DNA sequencing Protocol

The concentrations of gel purified samples were determined with Kapa library quantification kit for Illumina® libraries (KK4845) on Roche LightCycler 480. Paired-end Illumina® sequencing was performed on an Illumina® MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing) at the Georgia Genomics and Bioinformatics Core (GGBC) in the University of Georgia, Athens, GA, USA.

References

1. Kong, D.; Lei, Y.; Yeung, W.; Hili, R., Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angew. Chem. Int. Edit.* **2016,** *55* (42), 13164-13168.

2. Wilson, D. S.; Szostak, J. W., In Vitro Selection of Functional Nucleic Acids. *Annu. Rev. Biochem.* **1999,** *68* (1), 611-647.

3. Badi, N.; Lutz, J.-F., Sequence control in polymer synthesis. *Chem. Soc. Rev.* **2009,** *38* (12), 3383-3390.

4. Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M., Sequence-Controlled Polymers. *Science* **2013,** *341* (6146).

5. Meek, K. N.; Rangel, A. E.; Heemstra, J. M., Enhancing aptamer function and stability via in vitro selection using modified nucleic acids. *Methods* **2016,** *106*, 29-36.

6. Kong, D.; Yeung, W.; Hili, R., Generation of Synthetic Copolymer Libraries by Combinatorial Assembly on Nucleic Acid Templates. *ACS Comb. Sci.* **2016,** *18* (7), 355-370.

7. Rohloff, J. C.; Gelinas, A. D.; Jarvis, T. C.; Ochsner, U. A.; Schneider, D. J.; Gold, L.; Janjic, N., Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol. Ther. - Nucleic Acids* **2014,** *3*, e201.

8. Davies, D. R.; Gelinas, A. D.; Zhang, C.; Rohloff, J. C.; Carter, J. D.; O'Connell, D.; Waugh, S. M.; Wolk, S. K.; Mayfield, W. S.; Burgin, A. B.; Edwards, T. E.; Stewart, L. J.; Gold, L.; Janjic, N.; Jarvis, T. C., Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (49), 19971-19976.

9. Santoro, S. W.; Joyce, G. F.; Sakthivel, K.; Gramatikova, S.; Barbas, C. F., RNA Cleavage by a DNA Enzyme with Extended Chemical Functionality. *J. Am. Chem. Soc.* **2000,** *122* (11), 2433-2439.

10. Zhou, C.; Avins, J. L.; Klauser, P. C.; Brandsen, B. M.; Lee, Y.; Silverman, S. K., DNA-Catalyzed Amide Hydrolysis. *J. Am. Chem. Soc.* **2016,** *138* (7), 2106-2109.

11. Perrin, D. M.; Garestier, T.; Hélène, C., Bridging the Gap between Proteins and Nucleic Acids: A Metal-Independent RNAseA Mimic with Two Protein-Like Functionalities. *J. Am. Chem. Soc.* **2001,** *123* (8), 1556-1563.

12. Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M., A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M 2+). *Nucleic Acids Res.* **2009,** *37* (5), 1638-1649.

13. Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M., Toward the Combinatorial Selection of Chemically Modified DNAzyme RNase A Mimics Active Against all-RNA Substrates. *ACS Comb. Sci.* **2013,** *15* (4), 174-182.

14. Thomas, J. M.; Yoon, J.-K.; Perrin, D. M., Investigation of the Catalytic Mechanism of a Synthetic DNAzyme with Protein-like Functionality: An RNaseA Mimic? *J. Am. Chem. Soc.* **2009,** *131* (15), 5648-5658.

15. Mammen, M.; Choi, S. K.; Whitesides, G. M., Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chem. Int. Edit.* **1998,** *37* (20), 2755-2794.

16. 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.

17. 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.

18. 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.

19. Cherepanov, A. V.; de Vries, S., Kinetics and thermodynamics of nick sealing by T4 DNA ligase. *Eur. J. Biochem.* **2003,** *270* (21), 4315-4325.

20. Yu, H.; Zhang, S.; Chaput, J. C., Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat. Chem.* **2012,** *4*, 183.

21. Wu, X.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A., 2,6-Diaminopurine in TNA: Effect on Duplex Stabilities and on the Efficiency of Template-Controlled Ligations1. *Org. Lett.* **2002,** *4* (8), 1283-1286.

CHAPTER 4

STRUCTURE-ACTIVITY RELATIONSHIPS OF THE ATP COFACTOR IN LIGASE-CATALYZED OLIGONUCLEOTIDE POLYMERIZATIONS[1]

[1]Lei, Y.; Hili, R., Structure–activity Relationships of the ATP Cofactor in Ligase-catalysed Oligonucleotide Polymerisations. *Org. Biomol. Chem.* **2017,** *15* (11), 2349-2352.

Abstract

The T4 DNA ligase-catalyzed oligonucleotide polymerization process has been recently developed to enable the incorporation of multiple functional groups throughout a nucleic acid polymer. T4 DNA ligase requires ATP as a cofactor to catalyze phosphodiester bond formation during the polymer process. Herein, we describe the structure-activity relationship of ATP within the context of T4 DNA ligase-catalyzed oligonucleotide polymerization. Using high-throughput sequencing, we study not only the influence of ATP modification on polymerization efficiency, but also on the fidelity and sequence bias of the polymerization process.

Introduction

Aptamers are sequence-defined, mono-dispersed nucleic acid polymers that can bind molecular targets. The molecular recognition properties of this class of polymer can be evolved and tuned using *in vitro* selection technologies[2-3]. Aptamers with high affinities and selectivity have been evolved for myriad targets, ranging from metal ions[4] to mammalian cells[5]. Notwithstanding their success as high affinity reagents, the dearth of chemical functionality present in natural nucleic acid polymers has limited their potential. This contrasts with the diverse side chains present in proteinogenic polymers. To address the functional group deficit of aptamers, researchers have explored ways to sequence-specifically incorporate diverse functionality either through nucleobase or sugar backbone modifications[6-7]. Indeed, enhancing the functional group repertoire of nucleic acid polymers has yielded more favorable binding properties[8], and also expanded the utility of this class of polymer within the realm of catalysis[9-11].

The traditional method to incorporate modifications throughout a nucleic acid polymer has been to use a polymerase to incorporate modified nucleotide triphosphates[12-13]. This approach has been successful at increasing the functional properties of nucleic acid polymers; however, it

is inherently limited to a maximum of four unique modifications using a standard genetic code[14] ; current efforts to expand the genetic code can be used to expand the chemical diversity of the polymer[15]. Our lab has developed Ligase-catalyzed OligOnucleotide PolymERisation (LOOPER), which enables the sequence-defined incorporation of multiple unique instances of functional groups along a single-stranded DNA (ssDNA) polymer (Figure 4.1)[16-18]. The method relies upon T4 DNA ligase to catalyze the DNA-templated polymerization of modified 5′-phosphorylated oligonucleotides. LOOPER has been shown to proceed with excellent fidelities in a library context with codon sets as large as 256 members (codon set = NNNNT). Furthermore, the system can also tolerate a host of chemical modifications on the pentanucleotide building blocks, ranging from small molecules[17] to peptide fragments[16].

Pentanucleotides are ideal for this LOOPER, as they polymerize efficiently, yet require a primer to initiate the polymerization process, thus defining the reading frame. Longer oligonucleotides also decrease the density of functional group modifications on the polymer product. Despite the efficiency and high fidelity of LOOPER, increased polymerization fidelities are likely to benefit its application to the in vitro selection of modified nucleic acids by increasing enrichment factors and allowing access to larger reading frames.

We have previously shown that adjustment of certain LOOPER reaction parameters had no observable effect on the polymerization fidelity[18]. For example, varying temperatures ranging from 10–30°C did not produce significantly different polymerization fidelities, presumably due to the temperatures being higher than the annealing temperatures of the penta- nucleotide library. We sought to modify and evaluate other factors that might result in a more discerning polymerization. To this end, focused on the T4 DNA ligase cofactor, ATP. Since the ATP cofactor is used by T4 DNA ligase to covalently activate the 5′-phosphate, it is thus intimately involved in the transition state of the rate-limiting phosphodiester bond formation step[19]. We

hypothesized that modifications of the ATP cofactor may disrupt ligation kinetics, resulting in a more stringent and higher fidelity ligation process.

Previous reports have shown that specific modifications can be tolerated on the ATP cofactor during the T4 DNA ligase-cata- lysed splint ligation. For example, 2'-deoxy modification on the ribose ring or 2-amino modifications on the adenine ring are accommodated by the T4 DNA ligase, albeit with decreased levels of efficiency[20]. Interestingly, the 1-thiotriphosphate derivative of ATP was recently shown to modestly enhance substrate specificity during a single-ligation process[21].

In order to explore the potential of enhancing LOOPER fidelity using different cofactors, we chose to examine the influence of chemical modifications at four specific regions of the ATP cofactor, namely modification at the triphosphate region; the 2' and 3' positions of the ribose ring; the Hoogsteen face of the adenine nucleobase; and the Watson–Crick face of the adenine nucleobase (Figure 4.1). Critical to exploring the hypothesis that a modified adenylated oligonucleotide is ligated with the perturbed fidelity, we explored only those ATP modifications that are not lost upon 5'-adenylation of the oligonucleotide substrate.

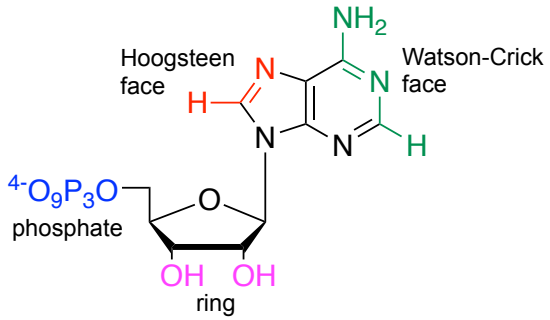


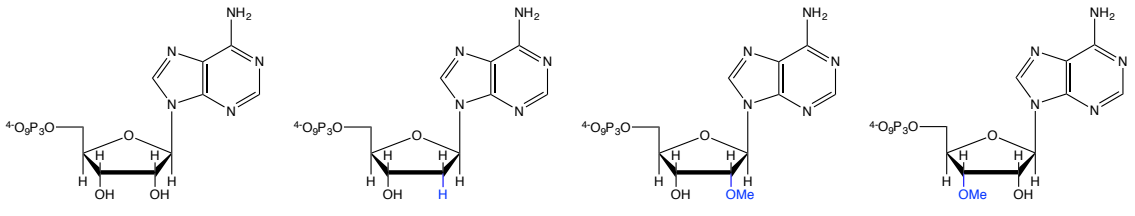
Figure 4.1 Modification sites of ATP cofactor

Results

The selected library of modified ATP cofactors is shown in Figure 4.2. Their efficiency during LOOPER system was examined first. This was performed using a 3'-hairpin template

library comprising eight-codon reading frame of the NNNNT codon set. The modified 5'-phosphorylated pentanucleotide building block library was A*NNNN, where A* is the 5'-dA nucleotide modified with an aminohexylamine at position 8 of the adenine ring. This modification has been shown to be well-behaved during previous LOOPER and sequencing studies[17-18]. While our primary goal was to increase the fidelity of LOOPER, the efficiency of the process would need to remain sufficiently high to ensure generation of full-length polymers.

Our previous findings have shown that high cofactor concentration can have a detrimental impact of yield during LOOPER, due to re-adenylation of dissociated ligase during the process. Since the ligase must be in its un-adenylated state to catalyze phosphodiester-bond formation, this effectively shuts down the polymerization[16]. However, T4 DNA ligase is likely to have an increased Km for modified ATP cofactors. Thus, LOOPER was performed at both high (1mM) and low (25μM) cofactor concentrations and the yields determined by denaturing PAGE analysis.



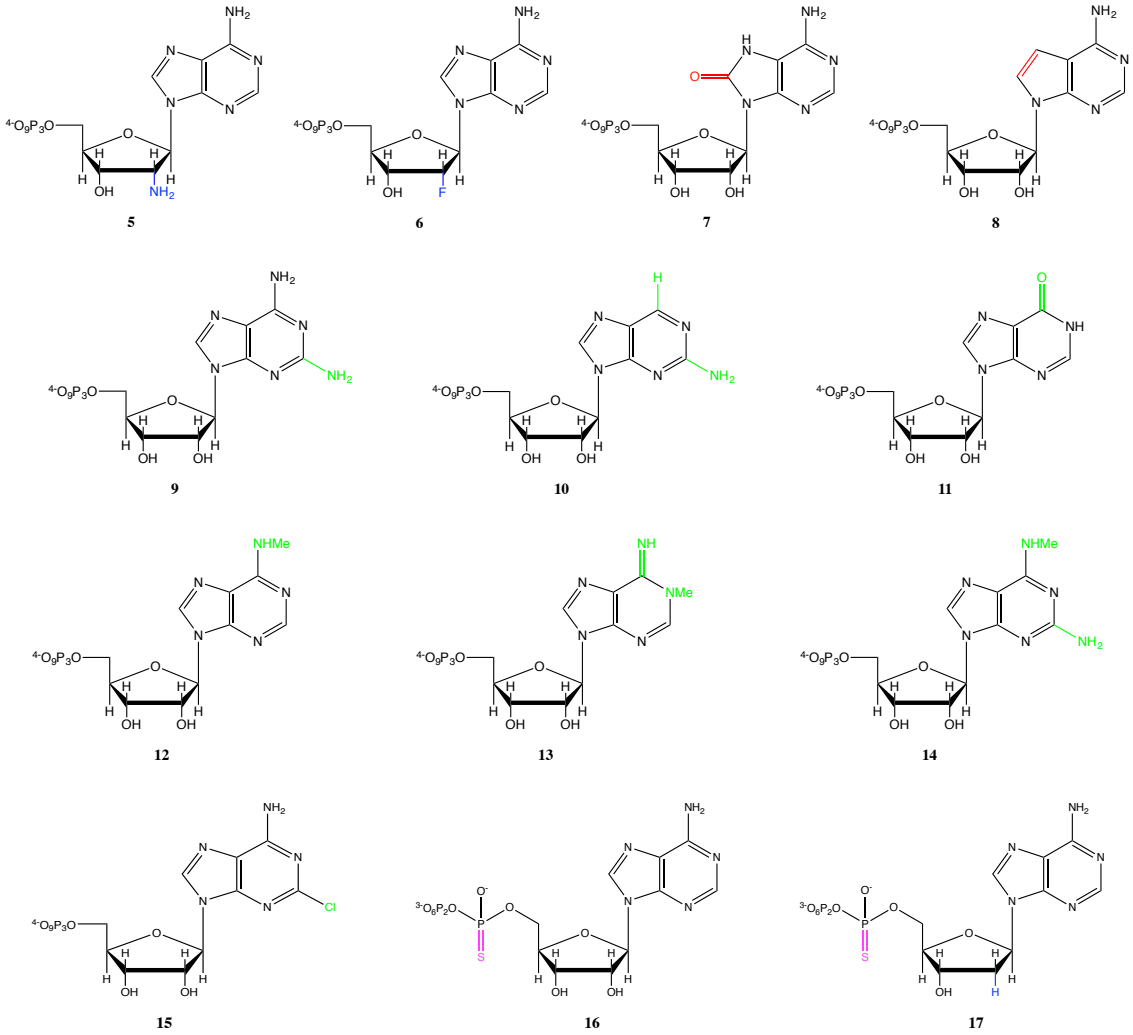


Figure 4.2 Library of ATP cofactor derivatives surveyed in LOOPER

The efficiency results were summarized in Table 4.1. The majority of modified ATP cofactors were ineffective at promoting the LOOPER process. Contrary to the splint-ligation study, which showed that dATP was an effective cofactor for T4 DNA ligase, this was not true during LOOPER. In fact, we observed that any modification to the ribose ring at either the 2' or 3' positions (Structure **2-6** in Figure 4.2) resulted in very poor polymerization efficiency. All the yields were analyzed by denaturing PAGE using hairpin template (Figure 4.3).

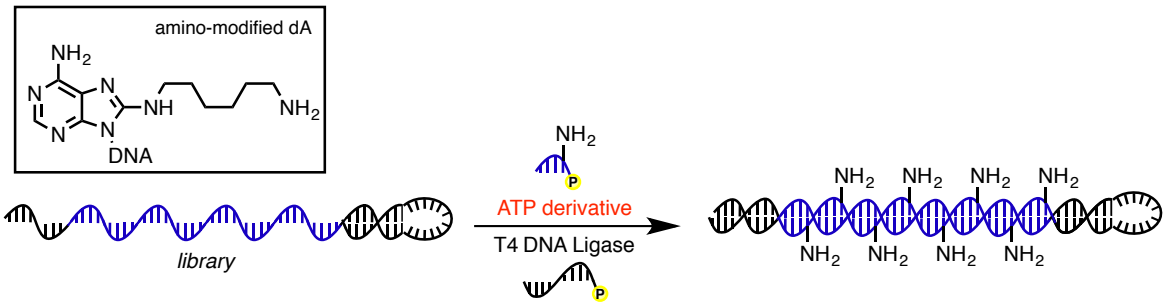


Figure 4.3 Hairpin polymerization for polymerization efficiency analysis

Cofactor	Concentration (μM)	Yield	Modification site
1	25	90%	N/A
	1000	60%	
8	25	5%	Hoogsteen face
	1000	0	
9	25	87%	Watson-Crick face
	1000	48%	
12	25	0	
	1000	70%	
15	25	64%	
	1000	37%	
16	25	10%	Phosphate
	1000	0	

Table 4.1 Efficiency of polymerization using modified ATP cofactors

Oxidation of ATP at position 8 of adenine results in shut-down of the polymerization (Cofactor **7**, Figure 4.2). Removal or substitution of the amine with an oxo group at position 6 of adenine was also not tolerated (Cofactor **10** and **11**, Figure 4.2). Of the tested cofactors, only five yielded measurable amounts of full-length product during LOOPER. 7-deaza ATP cofactor 8 resulted in only 5% yield of full-length product when used at 25μM, with higher concentrations failing to improve yield. 2-Amino ATP cofactor **9** gave a comparable yield (87% with 25μM concentration) to ATP. Cofactor **12**, which has a methylation of the amine at position 6 of ATP was also well-tolerated and gave good yield (42% with 1000μM concentration). Interestingly, highly efficient polymerization when using cofactor **12** was only observed at high cofactor

concentration, suggesting that T4 DNA ligase exhibits an increased K_m with this cofactor. Noteworthy is that when the modifications of cofactors **9** and **12** are combined to yield cofactor **14**, the LOOPER process become very inefficient. 2-Cl ATP cofactor **15** was also effective at promoting the polymerization; however, unidentified additional higher weight products were also produced when analyzed by PAGE. Lastly, the 1-thiotriphosphate derivative **16** gave measurable yield of the full-length product.

Errors in the LOOPER process can then be readily identified allowing the fidelity and bias of each codon within the entire codon set to be calculated (Table 4.2). Fidelity was calculated as the percentage by which the correct pentanucleotide building block was incorporated across from its cognate codon. Codon bias was calculated as the standard deviation of codon enrichment, whereby codon enrichment is the frequency of a codon in the polymer product divided by its corresponding frequency in the original template library. A codon bias closer to zero denotes lower codon bias during the LOOPER process.

We decided to analyze the fidelity of LOOPER using the cofactors that enabled sufficient yield of pure full-length polymer (cofactors **1, 8, 9, 12, 16**). To do this, we employed the high-throughput duplex DNA sequencing method that was recently developed in our lab.17 This method implements a randomized barcoding strategy of the LOOPER products in a library format, which enables the reading of both the template and the corresponding product polymer strand using next-generation sequencing.

Errors in the LOOPER process can then be readily identified allowing the fidelity and bias of each codon within the entire codon set to be calculated (Table 4.2). Fidelity was calculated as the percentage by which the correct pentanucleotide building block was incorporated across from its cognate codon. Codon bias was calculated as the standard deviation of codon enrichment, whereby codon enrichment is the frequency of a codon in the polymer product divided by its

corresponding frequency in the original template library. A codon bias closer to zero denotes lower codon bias during the LOOPER process.

Cofactor	Reads	Bias	Fidelity
1	31,936	0.21	95.1 %
8	10,023	0.51	98.2 %
9	12,565	0.32	96.6 %
12	12,278	0.28	96.2 %
16	21,962	0.42	98.6 %

Table 4.2 Fidelity and bias of polymerization using modified ATP

Using ATP as a cofactor, the original LOOPER process with an amine-modified ANNNN library proceeded at 95.1 % fidelity with a codon bias of 0.21. This level of fidelity is sufficient for short reading frames, but for extended reading frames, the portion of the polymer products with at least one error can become a significant fraction of the total population. Consistent with our hypothesis, we observed that all modified cofactors examined improved the fidelity of LOOPER. While cofactors **8** and **16** provided the greatest fidelities, their low polymerization efficiencies may preclude their use in preparative applications. Furthermore, both of these cofactors suffered from greater codon bias compared to the other tested cofactors; indeed, the codon bias observed with cofactor **16** led to strong de-enrichment of several codons, primarily codons with low GC-content (Figure 4.4). All other cofactors gave low to moderate degrees of codon bias (Figure 4.3), with individual codon enrichments/de-enrichments being suitable for applications such as in vitro selection.

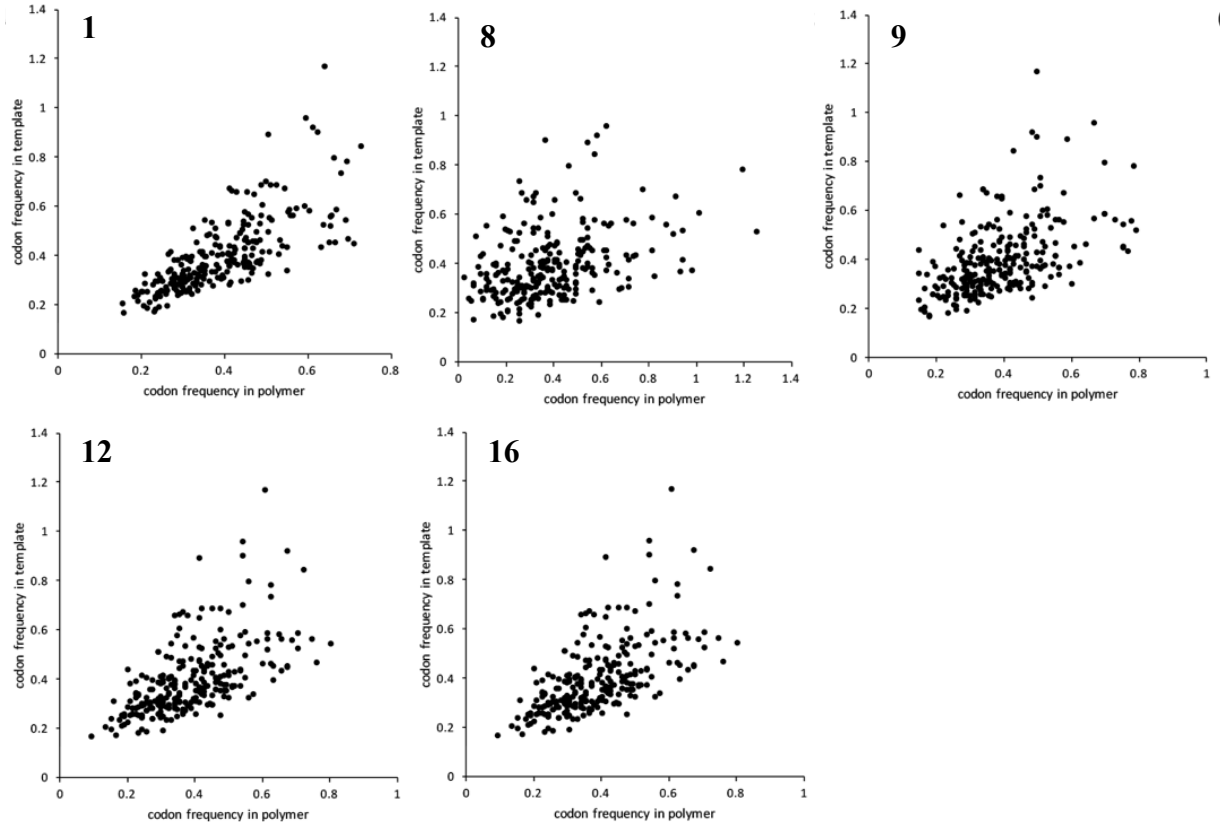


Figure 4.4 Bias analysis of NNNNT codon set used during LOOPER with various cofactors

We also examined the ability of LOOPER to accommodate larger hydrophobic modifications (benzylated amino-modified dA) using the optimal set of modified ATP cofactors. While ATP (25μM) generated full-length product in 56% yield, cofactors **9** (25μM) and **12** (1mM) gave 35% and 39% yield analyzed by hairpin polymerization.

Conclusions

We have shown that specific ATP modifications can enhance the fidelity of T4 DNA ligase during LOOPER. Of those tested, cofactors **9** and **12** are the most promising ATP substitutes, as they gave the highest polymerization yield with nucleobase modifications, lowest codon bias, and higher fidelity than unmodified ATP. Due to T4 DNA ligase catalyzing self-adenylation, substrate adenylation, and phosphodiester formation, cofactor modifications are likely to have a complex effect on this catalytic process. Thus, further studies are needed to unravel the

molecular basis behind the observed increase in fidelity using modified ATP cofactors. Notwithstanding, this improvement to fidelity is likely to have a marked impact on enrichment factors during the *in vitro* selection of modified nucleic acid polymers generated via LOOPER. Moreover, access to longer reading frames will enable longer polymers to be surveyed during selection experiments.

Materials and methods

General Information

Unless otherwise noted, water was purified with Milli-Q purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies (IDT®). DNA oligonucleotides with amine modification were synthesized on a Bioautomation Mermade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent1260) using a C18 stationary phase (Eclipse-XDB C18, 5µm, 9.4 x 200mm) and an acetonitrile/100mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop™ ND2000 spectrophotometer. DNA sequences are written from 5'→3'. <Aam> = Amino-modifer C6 dA, <N>=A/T/C/G.

Synthesis of amino-modified pentanucleotides

Pentanucleotides were synthesized on a Mermaid12 DNA synthesizer using a DMT-ON protocol on a 1µmol scale (1000Å CPG column). Amine-modifier C6 dA (Glen Research 10-1089), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation

at 25°C in 400µl of a 1:1 mixture of ammonium hydroxide and methylamine (AMA) for 25min. The cleaved resin was filtered away by filtration, and the oligonucleotide was further incubated in the same AMA solution at 60°C for 30min. The cleaved oligonucleotides concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The residue was then taken up into 100µl of H_2O and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 1ml of 40% aqueous acetic acid for 1h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 400µl 30% ammonium hydroxide at room temperature for 15min to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The dried product was dissolved into 100µl H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

Synthesis of benzylated amino-modified pentanucleotides

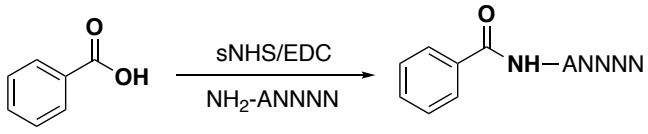


Figure 4.5 Synthesis of benzylated amino-modified pentanucleotides

A mixture of 25µl benzoic acid (100mM in DMSO), 25µl N-hydroxysulfosuccinimide (sNHS, 100mM in 1:1 mixture of DMSO and H_2O), 5µl 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 100mM in DMSO) and 7.5µl DMSO was incubated at room temperature for 30min. Followed by addition 7.5µl of amino-modified pentanucleotides (10nmol in H_2O) and 30µl Na_2CO_3 buffer (500mM in H_2O, pH9.0). The mixture was incubated

at room temperature overnight with vortex. The reaction was then quenched by addition of 15µl Tris buffer (500mM in H_2O, pH8.0) at room temperature for one hour and functionalized pentanucleotide was then purified with HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified modified pentanucleotides were dissolved in water and characterized with mass spectrometry.

T4-DNA Ligase-mediate polymerization protocols

DNA sequences

TP(NNNNT)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTC ACG TGG AGC TCG GAT CCT

PR5 /5Phos/GG ATC CGA GCT CCA CGT G

PR6 /5Phos/TG CGA CGG CAG GCG AAT CT

NH_2-ANNNN /5Phos/<Ama>NNNN

Protocols

In a PCR tube was added 1.5µl DNA template TH(NNNNT)8P (10µM in water), 2.25µl PR5 and PR6 (both 10µM in water), 5µl reaction buffer 4X (1X buffer: 66mM Tris-HCl, 10mM MgCl2, 1mM DTT, 6% PEG6000, pH7.6@25°C), 2µl modified ATP cofactor (0.25mM or 10mM in water) and 4µl water. The mix was heated to 90 °C for 2 minutes and then cooled to 25°C at the rate of 0.1°C/s. In this PCR tube was then added 1µl pentanucleotides NH_2-ANNNN (480µM in water), 1µl BSA (2 mg/mL in water) and 1µl T4 DNA ligase (400U/µl, New England Biolabs, M0202L). The polymerization was performed at 25°C for 24 hours. The products were then purified with MinElute® PCR Purification Kit for adapter ligation step.

T4-DNA Ligase-mediate adapter ligation protocols

DNA sequences

iTruS_i7_D709 CAA GCA GAA GAC GGC ATA CGA GAT CGG CTA TGG TGA CTG GAG TTC AG

iTruS_i7_D710 CAA GCA GAA GAC GGC ATA CGA GAT TCC GCG AAG TGA CTG GAG TTC AG

iTruS_i7_D711 CAA GCA GAA GAC GGC ATA CGA GAT TCT CGC GCG TGA CTG GAG TTC AG

iTruS_i7_D712 CAA GCA GAA GAC GGC ATA CGA GAT AGC GAT AGG TGA CTG GAG TTC AG

PRIMER B AAT GAT ACG GCG ACC ACC GAG

AdapterA AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

AdapterB /5phos/AC TGN NNN NNN NNN NNA GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC

Protocol

Adapter duplex synthesis

In a PCR tube was added 15µl of 100µM adapter A and 15µl of 100µM adapter B, then the tube was heated to 95°C for 5min and cooled to room temperature over 1h. Then in this PCR tube was added 4µl NEBuffer™2 10× (New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo-, New England Biolabs, M0212L), 1µl dNTP Mix (Thermo Scientific, 10mM each). The extension was performed at 37°C for 1h. The adapter duplex was purified with QIAquick® Nucleotide Removal Kit, and then diluted in 30µl water.

In a PCR tube was added 30µl purified adapter duplex, 5µl NEBuffer™2 10× (New England Biolabs, M0212L), 25U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 5µl dATP (Thermo Scientific, 10mM), 5µl water. This PCR tube was incubated at 37°C for 1h for

A-tailing. Then product was purified with QIAquick® Nucleotide Removal Kit, and then diluted in 30µl water.

Adapter ligation

In a PCR tube was added polymerization products (12pmol in 10µl water), A-tailing adapter duplex (240pmol), 10µl NEBNext® Quick Ligation Reaction Buffer 5x, BSA (5µg in 2.5µl water), 1000U T4 DNA ligase (New England Biolabs, M0202L), water (make total volume 50µl). Then the ligation was performed at 16°C for 16h followed by gel purification. The products were diluted to 10µl water.

PCR amplification

Each purified adapter ligation product was amplified with a different primer from iTrus_D709 to iTrus_D712. Each iTrus_D7XX was paired with a different ATP cofactor.

In a PCR tube was added 1µl purified adapter ligation product, 1.25µl 10µM Primer B, 1.25µl 10µM corresponding iTrus_D7XX primer, 9µl water and 12.5µl Q5® High-Fidelity 2× Master Mix (New England Biolabs, M0492S). The tube was then transferred to a preheated thermocycler (98°C). The annealing temperature was 55°C for the first two cycles and remained 71°C for the rest of the cycles.

High-Throughput DNA sequencing Protocol

The concentrations of gel purified samples were determined with Kapa library quantification kit for Illumina® libraries (KK4845) on Roche LightCycler 480. Paired-end Illumina® sequencing was performed on an Illumina® MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing) at the Georgia Genomics and Bioinformatics Core (GGBC) in the University of Georgia, Athens, GA, USA.

References

1. Structure–activity Relationships of the ATP Cofactor in Ligase-catalysed Oligonucleotide Polymerisations. *Org. Biomol. Chem.* **2017,** *15* (11), 2349-2352. -Reproduced by permission of The Royal Society of Chemistry

2. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346*, 818.

3. Tuerk, C.; Gold, L., Systematic Evolution of Ligands by Exponential Enrichment - Rna Ligands to Bacteriophage-T4 DNA-Polymerase. *Science* **1990,** *249* (4968), 505-510.

4. Qu, H.; Csordas, A. T.; Wang, J.; Oh, S. S.; Eisenstein, M. S.; Soh, H. T., Rapid and Label-Free Strategy to Isolate Aptamers for Metal Ions. *ACS Nano* **2016,** *10* (8), 7558-7565.

5. Fang, X.; Tan, W., Aptamers Generated from Cell-SELEX for Molecular Medicine: A Chemical Biology Approach. *Accounts of Chemical Research* **2010,** *43* (1), 48-57.

6. Meek, K. N.; Rangel, A. E.; Heemstra, J. M., Enhancing aptamer function and stability via in vitro selection using modified nucleic acids. *Methods* **2016,** *106*, 29-36.

7. Kong, D.; Yeung, W.; Hili, R., Generation of Synthetic Copolymer Libraries by Combinatorial Assembly on Nucleic Acid Templates. *ACS Comb. Sci.* **2016,** *18* (7), 355-370.

8. Rohloff, J. C.; Gelinas, A. D.; Jarvis, T. C.; Ochsner, U. A.; Schneider, D. J.; Gold, L.; Janjic, N., Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol. Ther. - Nucleic Acids* **2014,** *3*, e201.

9. Santoro, S. W.; Joyce, G. F.; Sakthivel, K.; Gramatikova, S.; Barbas, C. F., RNA Cleavage by a DNA Enzyme with Extended Chemical Functionality. *J. Am. Chem. Soc.* **2000,** *122* (11), 2433-2439.

10. Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M., Toward the Combinatorial Selection of Chemically Modified DNAzyme RNase A Mimics Active Against all-RNA Substrates. *ACS Comb. Sci.* **2013,** *15* (4), 174-182.

11. Zhou, C.; Avins, J. L.; Klauser, P. C.; Brandsen, B. M.; Lee, Y.; Silverman, S. K., DNA-Catalyzed Amide Hydrolysis. *J. Am. Chem. Soc.* **2016,** *138* (7), 2106-2109.

12. Hocek, M., Synthesis of Base-Modified 2′-Deoxyribonucleoside Triphosphates and Their Use in Enzymatic Synthesis of Modified DNA for Applications in Bioanalysis and Chemical Biology. *The Journal of organic chemistry* **2014,** *79* (21), 9914-9921.

13. Hottin, A.; Marx, A., Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Accounts of Chemical Research* **2016,** *49* (3), 418-427.

14. 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.

15. Malyshev, D. A.; Romesberg, F. E., The Expanded Genetic Alphabet. *Angew. Chem. Int. Edit.* **2015,** *54* (41), 11930-11944.

16. 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.

17. Kong, D.; Lei, Y.; Yeung, W.; Hili, R., Enzymatic Synthesis of Sequence-Defined Synthetic Nucleic Acid Polymers with Diverse Functional Groups. *Angewandte Chemie* **2016,** *128* (42), 13358-13362.

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.

19. Cherepanov, A. V.; de Vries, S., Kinetics and thermodynamics of nick sealing by T4 DNA ligase. *Eur. J. Biochem.* **2003,** *270* (21), 4315-4325.

20. Kinoshita, Y.; Nishigaki, K., Unexpectedly General Replaceability of ATP in ATP-Requiring Enzymes. *The Journal of Biochemistry* **1997,** *122* (1), 205-211.

21. Kim, J.; Mrksich, M., Profiling the selectivity of DNA ligases in an array format with mass spectrometry. *Nucleic Acids Res.* **2010,** *38* (1), e2-e2.

CHAPTER 5

IN VITRO SELECTION OF DNA APTAMERS AGAINST A HEPARAN SULFATE EPTITOPE

Introduction

We are applying the T4 DNA ligase-catalyzed DNA templated oligonucleotide polymerization toward the evolution of a functionalized aptamer against a heparan sulfate epitope. A 256-membered ANNNN library containing 16 functionalized sub-libraries is used in the *in vitro* selection. Certain sequences have been highly enriched; the binding affinity and specificity are being explored. The success of the selection should demonstrate improved binding activities of functionalized aptamers generated by LOOPER system compared with natural DNA aptamers against carbohydrate molecule targets.

Glycans are covalent assembles of sugars. They exist in either free form or in covalent complexes with proteins or lipids[1]. The glycan molecules are involved in many cellular interactions, including cell adhesion, cell-matrix interactions, cellular signaling and glycoprotein folding. It is not surprising given the fact that all cells in nature are covered by a dense and complex array of carbohydrates[2].

Monosaccharides surpass nucleotides or amino acids in coding capacity due to the chemically equivalent hydroxyl groups, which facilitate the formation of the glycosidic bond via various linkages[3]. Possibilities for two anomeric configurations, as well as introduction of branching and additional site-specific acetylation, phosphorylation or sulfation further underlie the structural diversity of glycan molecules[4].

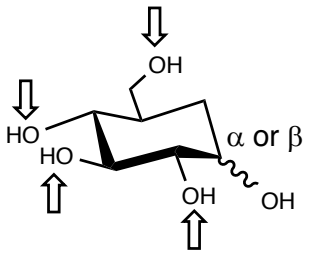


Figure 5.1 Illustration of the linkage points of oligomer formation of glycans

The study of structure and functions of glycans in biological systems has promises to enhance the understanding of human health and disease. Thus, glycan recognition molecules (GRMs), including glycan-binding proteins (GBPs), are essential regarding the studies of specific glycan structure and function. However, the repository of GRMs is extremely limited. Lectins are the most widely used receptor for glycan identifications so far. Hundreds of different lectins have been isolated from natural sources. They show high selectivity towards defined glycosidic linkages despite their modest binding affinity (10^{-4}-10^{-7}M) compared with antibodies. Natural sources also result in the purity issue for lectins. High-quality binders for glycan molecules are still in demand.

Aptamers are being explored as a promising class of GRMs. It is possible to use the SELEX method for the selection against glycan molecules[5]. Aptamers have shown the ability to distinguish between forms of the same protein with slightly different post-translational modifications[6]. However, successful selections against glycan molecule remain challenging due to the lack of aromatic ring structures and the intrinsic flexible conformation of glycan molecules. Typically, the aptamers selected for glycan molecules are weak binders. An aptamer against cellobiose, a disaccharide repeating unit of the cellulose, was reported to have the dissociation constant between 10^{-5} to 10^{-7}M[7].

The boronic acid-modified DNA has been explored to facilitate the binding activities toward glycan molecules. It was successfully applied to a selection against fibrinogen glycoprotein with dissociation constant at nanomolar range (6nM)[8]. This study only incorporated one modification

on the thymine nucleobase and resulted in a high-affinity aptamer. Chemical modifications serve as a reasonable solution for poor binding activity of aptamers against glycan molecules.

LOOPER system is able to incorporate multiple modifications simultaneously. Various modifications are chosen based on the known lectin-glycan binding sites as well as some unnatural functionalities such as boronic acid. An epitope of heparan sulfate (HS) is used as a model target in this study for its biological importance and commercial availability.

HS is linear polysaccharide, a type of glycosaminoglycan (GAG). It is attached to cell surface or extracellular matrix proteins, namely proteoglycan (HSPG). This form of HS is able to bind to a variety of protein ligands and receptors involved in diverse biological processes, such as signal transduction, cell adhesion and lipid metabolism.

HS consists of repeating disaccharide units each composed of an uronic acid and a derivative of glucosamine, N-sulfated glucosamine or unsubstituted glucosamine that is variably O sulfated. The disaccharide subunits of HS are shown in Figure 5.2[9]. Large sections of HS chains are unmodified units (Figure 5.2a), while the short modified segments (Figure 5.2b) are interspersed among them. Biochemical studies have shown that HS binding depends on the arrangement of the NA and NS segments and the modified sugar residues in the N-sulfated regions[10-11]. However, the structure of the binding sequence is only available in a few cases.

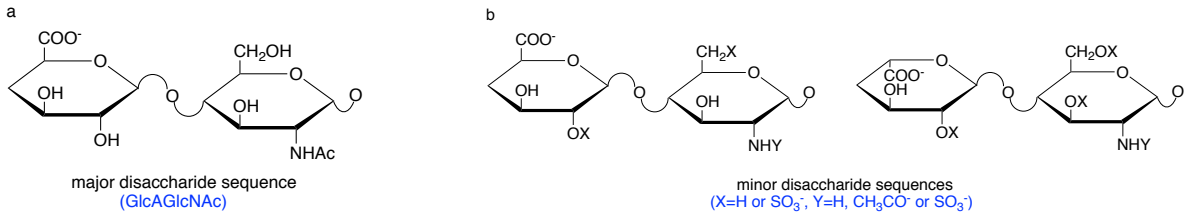


Figure 5.2 Major and minor disaccharide sequences of heparan sulfate (HS)

In this study, a commercially available hexasaccharide has been chosen (Figure 5.3). The sequence is Glc_{NS6S}-GlcA-Glc_{NS6S}-GlcA-Glc_{NS6S}-GlcA-Biotin. It is biotinylated for immobilization purpose during SELEX.

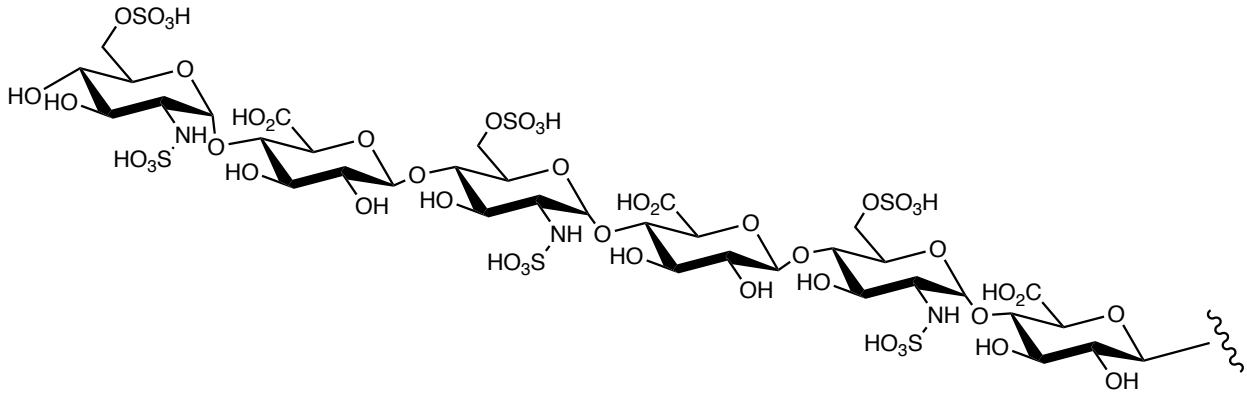


Figure 5.3 Structure of the HS epitope target

Materials and methods

General Information

Unless otherwise noted, water was purified with Milli-Q purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies (IDT®). DNA oligonucleotides with amine modification were synthesized on a Bioautomation MerMade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent1260) using a C18 stationary phase (Eclipse-XDB C18, 5µm, 9.4 x 200mm) and an acetonitrile/100mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop™ ND2000 spectrophotometer. DNA sequences are written from 5'→3'. <Aam> = Amino-modifer C6 dA, <N>=A/T/C/G.

Synthesis of amino-modified pentanucleotides

Pentanucleotides were synthesized on a Mermaid12 DNA synthesizer using a DMT-ON protocol on a 1µmol scale (1000Å CPG column). Amine-modifier C6 dA (Glen Research 10-1089), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 25°C in 400µl of a 1:1 mixture of ammonium hydroxide and methylamine (AMA) for 25mins. The cleaved resin was filtered away by filtration, and the oligonucleotide was further incubated in the same AMA solution at 60°C for 30mins. The cleaved oligonucleotides concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The residue was then taken up into 100µL of H_2O and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 1ml of 40% aqueous acetic acid for 1h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 400µl 30% ammonium hydroxide at room temperature for 15min to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using the CentriVap DNA Vacuum Concentrators (LABCONCO). The dried product was dissolved into 100µl H_2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

Functionalization of amino-modified pentanucleotides with carboxylic acids

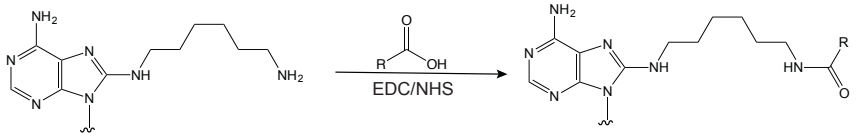


Figure 5.4 Amide bond formation reaction

A mixture of 25µl carboxylic acid (100mM in DMSO), 25µl N-hydroxysulfosuccinimide (sNHS, 100mM in 1:1 mixture of DMSO and H_2O), 5µl 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 100mM in DMSO) and 7.5µl DMSO was incubated at room temperature for 30mins. Followed by addition 7.5µL of amino modified pentanucleotides (10nmol in H_2O) and 30 µl Na_2CO_3 buffer (500mM, pH9.0). The mixture was incubated at room temperature overnight with vortex. The reaction was then quenched by addition of 15µl Tris buffer (500mM, pH8.0) at room temperature for one hour. Functionalized pentanucleotide was then purified with HPLC purification using a [10% acetonitrile in 0.1M TEAA, pH7.0] to [80% acetonitrile in 0.1M TEAA, pH7.0] solvent gradient with a column temperature of 45°C. The purified pentanucleotide was dissolved in water and characterized with mass spectrometry.

Design of diversely functionalized sub-libraries

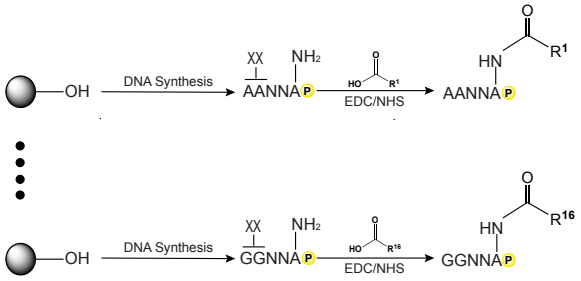
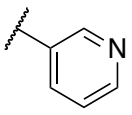
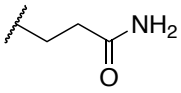
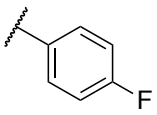


Figure 5.5 Design and synthesis of modified sub-libraries

XX	R	XX	R	XX	R	XX	R
AA		CA		TA	-NH_2	GA	

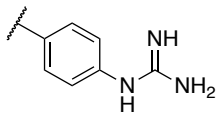
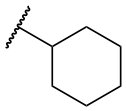
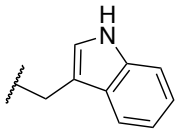
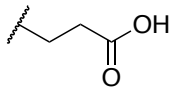
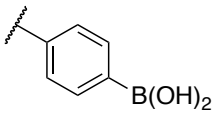
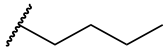
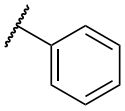
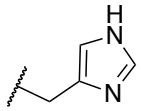
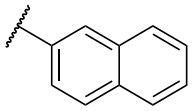
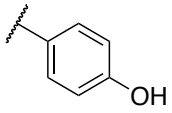
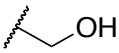
AC		CC		TC		GC	
AT		CT		TT		GT	
AG		CG		TG		GG	

Table 5.1 Modifications of 16 sub-libraries (ANNXX)

16 sub-libraries were mixed in equimolar ratios to generate the 256-membered heterofunctionalized pentanucleotide library, referred to as mix libraries.

In vitro selection cycles

DNA sequences

Bt_Temp /5BioTinTEG/GA TTC GCC TGC CGT CGC ANN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTC ACG TGG AGC TCG GAT CC

Poly_Prim1 /5Phos/TG CGA CGG CAG GCG AAT C

Poly_Prim2 GGA TCC GAG CTC CAC GTG

Bt_PrimA /5BioTinTEG/GA TTC GCC TGC CGT CGC A

PrimB GGA TCC GAG CTC CAC GTG

Protocols

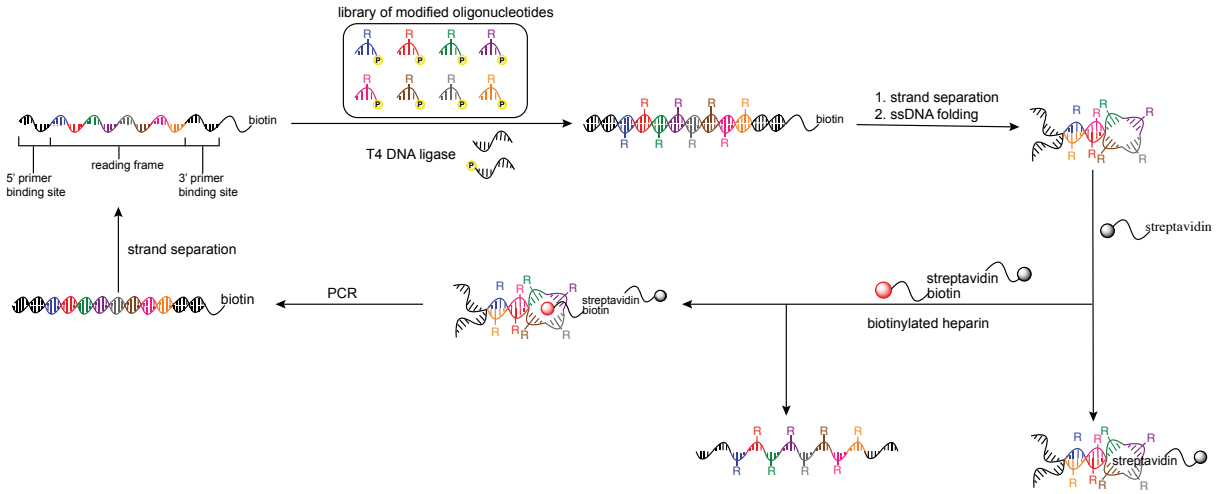


Figure 5.6 Schematic representation of *in vitro* selection cycles

In a PCR tube was added 1.5µl DNA template Bt_Temp (10µM in water), 2.25µl Poly_Prim1 and 2.5µl Poly_Prim2 (10µM in water, 1.5equiv. of the template), 5µl reaction buffer 4X (1X buffer: 66mM Tris-HCl, 10mM $MgCl_2$, 10mM DTT, 6% PEG6000, pH7.6@25°C), 2µl ATP (0.25mM in water) and 4µl water. The mix was heated to 90°C for 2min and then cooled to 25°C at the rate of 0.1°C/s. In this PCR tube was then added 1µl mix libraries (480µM in water), 1µl BSA (2mg/ml in water) and 1µl T4 DNA ligase (400U/µl, New England Biolabs, M0202L). The polymerization was performed at 25°C for 24h. 12 polymerization experiments were set up to generate the initial ssDNA library pool. The products were then purified with DNA Clean & Concentrator™-5 (ZYMO Research, D4013).

The amount of polymerized dsDNA was then measured by Nanodrop ND2000™ spectrophotometer. Dynabeads® MyOne™ Streptavidin C1 beads were washed with 100µl binding buffer (5mM $MgCl_2$, 250mM NaCl, 50mM Tris-HCl, pH7.6@25°C) three times before using. The volume of streptavidin was calculated based on the theoretical capacity for dsDNA (0.2µg/µl), and twice of the theoretical volume was used to capture all biotinylated dsDNA. The purified dsDNA was resuspended in binding buffer, and the volume was 4 times of the

streptavidin beads used. The dsDNA was then incubated with prewashed streptavidin beads at room temperature for 1h. The bound dsDNA was washed with 100µl binding buffer three times. The strand separation was performed with 40µl fresh 150mM NaOH at 37°C for 1min. The modified non-biotinylated strand was collected, neutralized with 4µl 1.5M HCl, and then purified using CENTRI-SEP™ columns (Princeton Separations).

The biotinylated target was immobilized on Dynabeads® MyOne™ Streptavidin C1 beads using the same binding buffer as dsDNA. Excess amount of target molecules was used, twice the amount of theoretical capacity (4pmol/µl). The beads were incubated with target at room temperature for 2h, washed with 100µl binding buffer 3 times and then resuspended in binding buffer.

The purified modified ssDNA pool was heated to 95°C for 5min then cool to room temperature for 30min. Then the folded ssDNA was incubated with immobilized target beads at room temperature for 1h. Negative selection step (folded ssDNA incubation with streptavidin beads at room temperature for 30min then collected the non-binders) was performed before this binding step starting from round 2 selection. The non-binders were then removed after incubation, and the beads were washed with binding buffers (Round 2-3, 2*40µl; Round 4-5, 2*50µl; Round 6, 3*50µl). The bound ssDNA was eluted with 80µl water at 95°C for 5min.

The bound fraction was first amplified with qPCR using KOD Hot Start polymerase. 10µl modified ssDNA was added to a PCR tube, 5µl KOD buffer (10X), 5µl dNTP (2mM each), 2µl Bt_PrimA and 2µl PrimB (both 10µM), 0.5µl SYBR green (100X), 1µl KOD Hot Start polymerase (1.0U/µl), and 21.5µl water. The total reaction volume was 50µl. qPCR was performed with MiniOpticon™ Real-Time PCR System.

The amplified dsDNA was then incubated with prewashed Dynabeads® MyOne™ Streptavidin C1 beads in the binding buffer at room temperature for 1h. The non-biotinylated strands were removed by incubation using 150mM NaOH at 37°C for 1min. The biotinylated templates were then recovered with elution buffer (95% formamide, 10mM EDTA, pH8.2) at 90°C for 2min. The templates were purified with CENTRI-SEP columns followed by gel purification then subjected to the next round selection. Total of 6 rounds selection were performed.

High-throughput sequencing

DNA sequences

iTruS_i7_D704C CAA GCA GAA GAC GGC ATA CGA GAT GAG ATT CCG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT GAT TCG CCT GCC GTC GCA

iTruS_i7_D705C CAA GCA GAA GAC GGC ATA CGA GAT ATT CAG AAG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT GAT TCG CCT GCC GTC GCA

iTruS_i7_D706C CAA GCA GAA GAC GGC ATA CGA GAT GAA TTC GTG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT GAT TCG CCT GCC GTC GCA

iTruS_i7_D707C CAA GCA GAA GAC GGC ATA CGA GAT CTG AAG CTG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT GAT TCG CCT GCC GTC GCA

PRIMEC AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TGG ATC CGA GCT CCA CGT G

Protocol

In a PCR tube was added 1µl amplified dsDNA after round 4, corresponding iTrus_i7_D70XC 2µl (10µM, D704 to round4, D705 to round5, D706 to round6, and D707 to initial template pool) and 2µl PRIMEC (10µM), 25µl Q5® Hot Start High-Fidelity 2X Master Mix and 20µl water. The polymerase was initially activated at 98°C for 30s. The first two PCR

cycles were 10s at 98°C (denaturation), 30s at 55°C (annealing), and 30s at 72°C (extension). Starting from third cycle, annealing was changed to 71°C for 30s. The PCR products were then gel purified and sent to Georgia Genomics and Bioinformatics Core (GGBC) for further sequencing analysis.

The concentrations of gel purified samples were determined with Kapa library quantification kit for Illumina® libraries (KK4845) on Roche LightCycler 480. Paired-end Illumina® sequencing was performed on an Illumina® MiSeq system using the kit v2 with 300 cycles (150bp PE sequencing).

References

1. Fuster, M. M.; Esko, J. D., The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer* **2005,** *5*, 526.

2. Varki, A., Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **2007,** *446*, 1023.

3. Gabius, H.-J.; Roth, J., An introduction to the sugar code. *Histochem. Cell Biol.* **2017,** *147* (2), 111-117.

4. Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D., From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **2011,** *36* (6), 298-313.

5. Svarovsky, S. A.; Joshi, L., Biocombinatorial Selection of Carbohydrate Binding Agents of Therapeutic Significance. *Curr. Drug Discovery Technol.* **2008,** *5* (1), 20-28.

6. Partha, R.; A., S. B.; R., W. R., Further Characterization of the Target of a Potential Aptamer Biomarker for Pancreatic Cancer: Cyclophilin B and Its Posttranslational Modifications. *Nucleic Acid Ther.* **2013,** *23* (6), 435-442.

7. Yang, Q.; Goldstein, I. J.; Mei, H.-Y.; Engelke, D. R., DNA ligands that bind tightly and selectively to cellobiose. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (10), 5462-5467.

8. Li, M.; Lin, N.; Huang, Z.; Du, L.; Altier, C.; Fang, H.; Wang, B., Selecting Aptamers for a Glycoprotein through the Incorporation of the Boronic Acid Moiety. *J. Am. Chem. Soc.* **2008,** *130* (38), 12636-12638.

9. Cummings, R. D.; Pierce, J. M., *Handbook of Glycomics*. Elsevier Science & Technology: San Diego, UNITED STATES, 2009.

10. Spillmann, D.; Lindahl, U., Glycosaminoglycan-protein interactions: a question of specificity. *Curr. Opin. Struct. Biol.* **1994,** *4* (5), 677-682.

11. Esko, J. D.; Selleck, S. B., Order Out of Chaos: Assembly of Ligand Binding Sites in Heparan Sulfate. *Annu. Rev. Biochem.* **2002,** *71* (1), 435-471.

CHAPTER 6

CONCLUSIONS

There has been tremendous growth in the aptamer field since the introduction of *in vitro* selection method in the 1990s. The nucleic acid aptamers mimic and extend the features of monoclonal antibody reagents, thus serve as a powerful class of synthetic affinity reagents. Nucleic acid aptamers can be chemically synthesized without any biological system involved. The quality and scale of production can then be ensured. The nucleic acid aptamers can be engineered into sensor, actuators and other devices for their molecular recognition abilities. As therapeutics, aptamers exhibit a lower level of immune response compared with proteins. However, antibodies still dominate the global medicine marketplace. The limitations of aptamers mentioned in the previous chapters have delayed their widespread distribution like antibodies. Further, their pharmacokinetic and other systemic properties are variable and hard to predict[2]. The lack of chemical functionalities has limited the broad application of nucleic acid aptamers as affinity reagents compared with antibodies. This is supported by the fact that proteins remain the dominant target category of aptamers even though aptamers are thought to bind various target categories.

Modifications on nucleic acid aptamers are essential for their practical applications. Several strategies mentioned in the previous chapters, such as unnatural base pairs and SOMAmers, have resulted in some aptamers with enhanced activity and stability. SOMAmers have been proved to increase the success rate of *in vitro* selection with the incorporation of hydrophobic functional groups. Modified-aptamers are promising to mimic the binding behavior of polypeptides and

proteins better as some structural studies revealed. These studies are of great significance for a better understanding of the binding activities and also aptamer design.

LOOPER system has been developed based on DNA-templated polymerization, which enables the heterofunctionalization on one nucleic acid polymer chain. It has been successfully applied to the selection against thrombin protein with 1.6nM dissociation constant and high level of selectivity. As LOOPER enables simultaneous incorporation of various functionalities, the aptamers can mimic peptides with functionalized side chains. This unique character can lead to aptamers against more 'difficult' targets in traditional SELEX such as carbohydrates. The system expands the diversity of functionalities to a comparative level to the natural proteins, which are composed of 20 amino acids.

On the other hand, the LOOPER system can be improved in several aspects. First, systematic studies of possible functional groups are beneficial to the design of modified building blocks. Even though all functionalities can be incorporated through the polymerization process, it is essential to carefully choose the linker and the right derivatives of carboxylic acids. For examples, both 4-nitrobenzoic acid and nitroacetic acid can be coupled on the hexylamine modified adenosine through amide bond formation. There are several protocols of this coupling reaction, and not all carboxylic acids can be successfully incorporated on the oligonucleotides. The following purification process and characterization methods are also important. After the successful incorporation and purification, the distance between the functional groups and nucleic acid polymers as well as the rigidity should be considered.

Second, the scale of modified aptamers is lower compared with the traditional *in vitro* selection pool. As mentioned previously, LOOPER depends on the T4 DNA ligase-mediated polymerization, which is a two-edged sword. On the one hand, various functionalities are incorporated at the same time in a sequence-defined manner. On the other hand, it cannot be

amplified efficiently like traditional PCR, thus decrease the amount of the modified aptamer generated from one reaction. The excess amount of modified building blocks in the polymerization reaction also makes it not easy to increase the scale to a comparable level of PCR. Thus, the initial library pool of aptamers is not as large. Third, the generation of a selected aptamer still depends on the T4 DNA ligase-mediated polymerization. Unlike traditional *in vitro* selection, where you can chemically synthesize the whole aptamer sequence, the modified aptamers still need the DNA-templated polymerization process. Specific building blocks are synthesized separately and subject to the reaction with the corresponding DNA template. The process can be time-consuming, and the efficiency is not as high as pure chemical synthesis.

The structures of these highly functionalized aptamers are not fully understood, and further studies need to be done. These studies will reveal the importance of modifications on the side chains. Further, optimizations can be performed based on the interactions between the aptamers and the targets. Better design of the initial pool, such as the length of the random region and the choices of functional groups, is also possible. This might compensate the effect of smaller initial aptamer library compared with traditional *in vitro* selection. Chemical synthesis of selected aptamers is being explored. Solid-phase synthesis of modified aptamers is possible.

LOOPER overcomes the limitation of one or two modifications on one nucleic acid strand using the DNA-templated polymerization with oligonucleotide building blocks. It is inspired by the peptide synthesis process in biological systems. The generated aptamers might adopt unique folding patterns due to the unnatural functional groups, thus bind to the targets at various sites through different kinds of non-covalent interactions. The mechanism of the polymerization is fully revealed, and there's very little structural information of these modified aptamers. In-depth studies are still in need for us to understand this new class of aptamers.

Aptamers are still relatively new technology compared with antibodies, and some challenges remain to be addressed. However, they have proved to be a promising class of affinity reagents as well as therapeutics over the past two decades.