DNAZYMES FOR AMINE AND PEPTIDE LYSINE ACYLATION

BY

TIANJIONG YAO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Scott K. Silverman, Chair
Professor Yi Lu
Assistant Professor Erik Procko
Professor Wilfred van der Donk

Abstract

The typical enzymatic biopolymers evolved by nature are proteins and RNA. These biopolymers can fold into complex secondary and tertiary structures capable of performing catalysis. In contrast, no natural DNAzymes (deoxyribozymes) have been identified. Given the structural similarities between DNA and RNA, theoretically DNA is also capable of forming complex structures thus functioning as enzymes. Engineered DNAzymes are identified and developed in laboratories by in vitro selection, and there are several reasons to develop artificial DNAzymes to act as comparable enzymes to proteins and RNA. First, the fundamental base for selection is that candidate pool needs to be amplified after each selection round, otherwise selection cannot be done. Oligonucleotides can be directly amplified by natural polymerases, whereas proteins cannot be amplified directly by any means. Second, oligonucleotide synthesis does not involve many practical challenges associated with protein expression and purification. Third, random oligonucleotide sequences can readily fold into secondary, then tertiary structures, whereas most random peptide sequences will misfold and aggregate. Last but not least, the total number of possible sequences is much smaller for oligonucleotides (4^n, where n is the length of the biopolymer) than for proteins (20^n). Therefore, by combining last two reasons, selections for oligonucleotide enzymes can cover a larger meaningful fraction of total sequence space than for proteins. Furthermore, between the two types of oligonucleotides, DNA has several practical advantages over RNA: DNA is cheaper, more stable, and easier to synthesize. The field of DNAzymes research is relatively unexplored, thus it is more likely to discover novel DNAzymes that perform chemical reactions for which protein enzymes are limited.

My research focus is on lysine acylation, a common post-translational modification (PTM) important in gene expression and regulation, control of protein function, and primary metabolism. Besides the most well-studied acylation type, acetylation, many other acylation types such as malonylation, succinylation, and glutarylation have also been discovered yet are poorly understood.

Engineering of natural acetyltransferases to alter their parent substrates to other longer chain acyl donors is an exciting prospect, but this process usually leads to relaxed substrate selectivity, rather than a true alteration in substrate specificity. Because DNAzymes are identified from pools of random DNA sequences, no inherent peptide sequence biases must be overcome during the selection process, and thus the prospect of truly selective artificial acyltransferases is reasonable. Previous efforts by our lab identified DNAzymes that can catalyze lysine modification by using 5'-phosphorimidazolide (5'-Imp), resulting in the formation of a Lys-phosphoramidite bond. Then, the efforts to investigate DNAzymes (with canonical nucleotides) that catalyze the lysine acylation by using thioesters as electrophile were performed, however no DNAzymes were identified. Directly following the last effort seeking DNAzymes with canonical nucleotides, Chapter 2 describes in vitro selection efforts seeking lysine acylation that employed chemically modified nucleotides to expand the functionality of DNA and facilitate catalysis. The in vitro selection in Chapter 2 also used thioesters as the acyl donor electrophile; meanwhile two amino substrates 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA were used as the nucleophiles. Ten rounds of in vitro selection were performed, but no DNAzymes were identified. We concluded that a thioester is insufficiently reactive as an electrophile to allow the identification of amine-acylating DNAzymes, and a more reactive acyl donor is required.

Therefore, aryl esters were considered as more appropriate acyl donor electrophiles than thioesters. Chapter 3 describes the in vitro selection efforts using highly reactive aryl ester (DMT ester and TFP ester) as acyl donor electrophiles. Acyl donor oligonucleotides were activated *in situ* from their 5'-carboxylic acid (5'-CO_2H) precursors using two common amide-forming coupling reagents, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM) or the combination of the water-soluble 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 2,3,5,6-tetrafluorophenol (TFP), resulting in the formation of DMT ester and TFP ester in the selection process. The purpose of selection in this stage is to identify DNAzymes that catalyze the amine acylation using DMT or TFP ester as electrophiles and two amino substrates as nucleophiles.

The random DNA pool sequences were composed of canonical nucleotides, and same two amino substrates 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA in Chapter 2 were again used as the nucleophiles. After four rounds of selection, several DNA sequences were identified by cloning and sequencing. However, the emergent DNA sequences had no rate enhancement above the uncatalyzed, splinted background reaction under the same incubation conditions. We concluded that each individual DNA sequence likely adopts a combination of secondary and tertiary structure that merely recapitulates a complementary splint. Apparently, rate enhancement beyond the splinting effect cannot be achieved because the DMT or TFP ester electrophile is too reactive.

The lessons learned by the two previous efforts are that neither too low nor too high reactive electrophiles can be considered as appropriate acyl donors; then, the attention was turned to acyl donors with intermediate reactivity. Chapter 4 describes the in vitro selection efforts using intermediate reactive aryl ester acyl donors which are phenyl ester and 4-fluorophenyl ester. The random DNA pool sequences were still composed of canonical nucleotides, and two amino substrates 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA were used as the nucleophile as well. After seven/eight or eleven rounds respectively for the two amino substrates, several DNA sequences emerged from each selection and were identified by cloning and sequencing. These DNA sequences were assayed and have much higher rate enhancement above the uncatalyzed, splinted background reaction under the same incubation conditions, the highest rate enhancement is about 10^3 fold. So, DNAzymes for amine or peptide lysine acylation have been identified by using intermediate aryl ester acyl donor electrophiles. These identified DNAzymes were also assayed for activity by using free peptide as the nucleophile; unfortunately, no free peptide activity was observed. Future efforts will seek DNAzymes that can catalyze lysine acylation of modified free peptide such as azido-peptide by adopting appropriate electrophiles.

While DNAzymes that transfer the big acyl oligonucleotide onto the amino group or peptide lysine have been identified, efforts towards investigating DNAzymes that can transfer a small acyl

group onto amino group have also been performed. Chapter 5 describes the efforts seeking DNAzymes that transfer a glutaryl group to the amine group or peptide lysine. Nine rounds of in vitro selection were performed, but no DNAzymes were identified. Interestingly, when assaying DNAzymes discovered from the lysine acylation selections, it turned out that three DNAzymes, among these previously identified lysine acylation DNAzymes, also have the ability to transfer the glutaryl group to the amine group. This result could develop strategy that, when aiming for DNAzymes that catalyze a difficult substrate, we can first select with an easier counterpart molecule with same reaction mechanism as the target substrate. The resulting DNAzymes could also have the ability to catalyze the target substrate, then we can improve the identified DNAzymes using several methods such as reselection.

Acknowledgements

I am grateful for everyone who has helped me and supported me throughout my Ph.D. studies. I would like to thank my research advisor, Professor Scott Silverman for his outstanding guidance and mentorship. His passion for science always inspires me.

I am thankful to my thesis committee members, Professor Wilfred van der Donk, Professor Yi Lu, and Professor Erik Procko. Thank you for all the helpful discussions and suggestions during my student seminar, preliminary exam, and final defense.

I am also grateful for the talented and wonderful coworkers who have provided tremendous support every day. I would like to thank Ben, Jag, Shannon and Jimmy for their great guidance and instruction, especially at my early Ph.D. life; Puzhou and Cong for all the scientific and nonscientific discussions and advice to evolve and improve my thesis work; Peter, Shukun, Robert, Prakriti, and Jack for the great scientific assistance and preparation to help my thesis work moving forward smoothly. To the undergrads who worked with me and in the lab, Austin, Caleb, Caroline and Morgan – I appreciate your hard work and I also learned a lot from each of you.

I am extremely thankful to Jinlong Guo for his love and support. Thank you for being by my side and helping me fight through the tough times in this long journey. I could not have made it without you.

Finally, I would like to thank my family, especially my parents, for their unconditional love, encouragement, and support throughout my life.

Table of Contents

Chapter 1: Introduction to Chemical and Enzymatic Catalysis

1.1 Overview of Catalysis

1.1.1 Chemical Catalysis

In chemistry, catalysis means the acceleration of the rate of a chemical reaction by addition of a substance not consumed during the reaction.[1] In its most fundamental form, a reaction can be catalyzed by acid or base. The acid–base catalysis means the acceleration of a chemical reaction by the addition of an acid or a base, the acid or base itself not being consumed in the reaction.[2] The catalytic reaction may be acid-specific (acid catalysis), as in the case of the Fisher esterification, in which a carboxylic acid and an alcohol react to form an ester[3]; or base-specific (base catalysis), as in the case of the hydrolysis reactions such as ester hydrolysis or the acylation reactions such as amine acylation.[4,5] The mechanism of acid- and base-catalyzed reactions is explained in terms of the concept of acids and bases as one in which there is an initial transfer of proton from an acidic catalyst to the reactant or from the reactant to a basic catalyst.[6] In terms of the Lewis theory of acids and bases, the reaction involves sharing of an electron pair donated by a base catalyst or accepted by an acid catalyst. Acids (including Lewis acids) and bases act as powerful catalysts for a great variety of chemical reactions, in the laboratory, in industry, and in processes occurring in nature.

In addition to traditional chemical catalysts, small-molecule catalysts have been developed for more challenging chemical reactions. Small-molecule catalysts are different than the acid-base catalysts because these catalysts are discrete molecules or molecular complexes, which can be added to the reaction.[7] This is an important distinction because small-molecule catalysts have built in specificity for catalyzing certain chemical reactions. The catalyst, in this instance, accelerates a reaction only between two very specific reactants, and provides the chemist with a large room to predict and control the outcome of the reaction.[8] These developments have had a profound impact

on the field of organic chemistry, enabling the synthesis of extremely complex small molecules.

1.1.2 Enzymatic Catalysis

Enzymatic catalysis refers to catalysis performed by large, macromolecular structures called enzymes. One of the most significant features of enzymes is the substrate selectivity of enzymes. Enzymes are macromolecules capable of adopting very complex and specific three-dimensional structures which can bind substrates selectively, destabilize substrate ground states, and stabilize reaction transition states.[9] Another significant feature of enzymes is that enzymes have very high rate enhancement compared to usual chemical catalysts. For example, the best enzyme has rate enhancement greater than~10^{15}.[10,11] Enzymes can also have catalytic efficiency that is limited by diffusion of substrates into the enzyme and products out of the enzyme.[12] Because of substrate selectivity and high reaction rate enhancement, enzymes are ideal for many applications and tremendously being used in various academic, medical, industrial fields.[13,14]

1.2 Natural Enzymes

1.2.1 Protein Enzymes

Natural enzymes are large biopolymers composed of a variety of small monomers. Depending on the type of monomers, natural enzymes can be categorized into protein enzymes and ribozymes. Protein enzymes are the predominant enzymes in nature, responsible for cellular power, regulation, replication and catalyzing most of the biochemical reactions in biological systems.

Proteins are polymers made of amino acids linked via peptide bonds. 20 natural canonical amino acids with diverse chemical functional groups on their side chains contribute to the wide variety of protein structures and functions (Figure 1.1). Furthermore, the diversity of protein functions and structures are tremendously expanded by introducing unnatural amino acids and

post-translational modifications (PTMs). The massive chemical diversity of proteins due to various combination of amino acids and PTMs makes protein enzymes capable of catalyzing different kinds of chemical reactions.

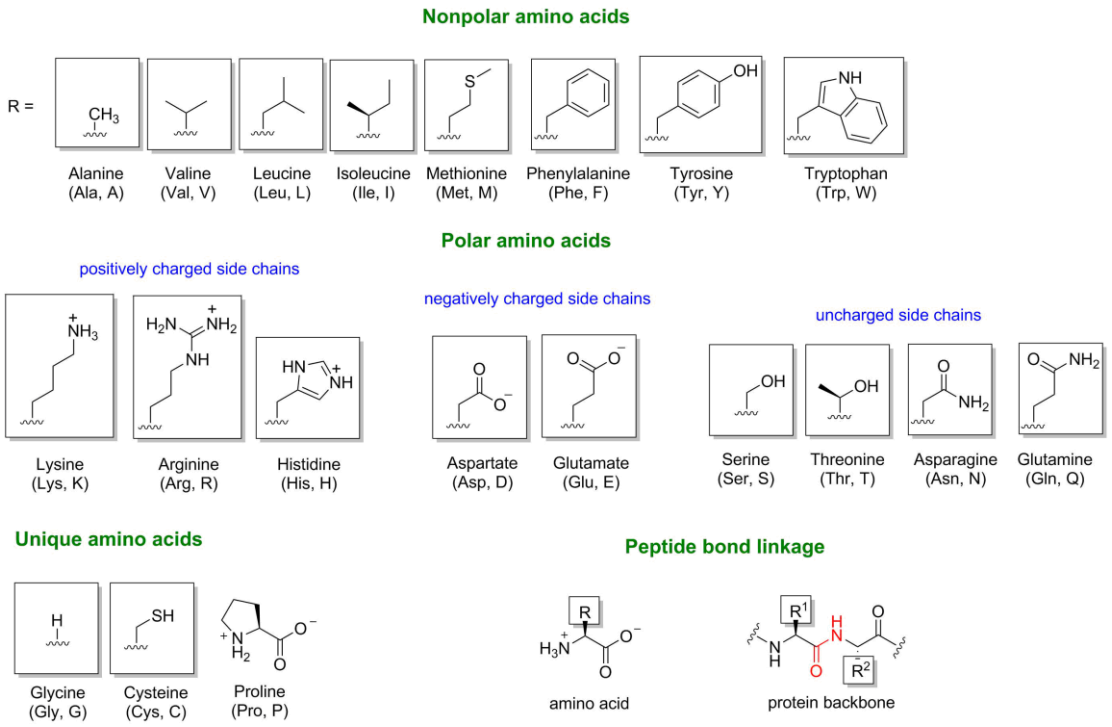


Figure 1.1. Twenty natural amino acids. The diversity of side chain functional groups is illustrated. Proteins are composed of amino acids connected by peptide bonds.

In addition to various chemical functional groups, proteins adopt hierarchical structures which include secondary, tertiary, and quaternary structures. Specific amino acid sequences can fold into secondary structures such as α-helices and β-sheets. Such secondary structures cooperatively form tertiary structures, which can further form quaternary structures by assembly of multiple protein subunits into large protein complexes. The overall structure of a protein enzyme serves as a scaffold to support and position the active site where the substrate can bind and undergo the enzymatic reaction.[15] The active site of a protein enzyme usually consists of several amino acids with desired functional groups that are required to facilitate catalysis. Sequence complexity

allows for proteins to perform enzymatic catalysis using different catalyzing mechanisms, such as transition state stabilization, ground state destabilization, or changing the mechanism via different path.[16]

Acyltransferase is a type of transferase enzyme that acts upon proteins and lipids by utilizing acyl groups. The most well-known and well-studied acyltransferase type is acetyltransferase, which acetylates conserved lysine amino acids on histone proteins by transferring an acetyl group from acetyl-CoA to form ε-N-acetyllysine.[17] Other than acetyltransferase, efforts seeking acyltransferases that transfer longer chain acyl groups such as glutaryl group and crotonyl group have been carried out, yet undiscovered.[18,19] The efforts seeking to identify and expand the scope of acyltransferase other than acetyltransferase and histone have been attempted;[20] alternative approach could be carried out such as by employing nucleic acid enzymes.

1.2.2 Ribozymes

Ribozymes are ribonucleic acid (RNA) molecules that have catalytic functions. Their roles are equivalent to enzymes, which is why they are called ribozymes. RNA is a biopolymer of ribonucleotides linked by phosphodiester bonds (Figure 1.2). RNA was once thought to simply be a passive carrier of genetic information, transferring information from DNA to proteins. RNA was never considered as an enzyme or having similar catalytic functions as protein, because RNA has much fewer functional capabilities than protein due to lack of diverse functional groups. However, many natural RNA molecules have been identified as catalytic RNA since the early 1980s, when the enzymatic abilities of natural RNA were discovered.[21] In spite of the lack of diverse functional groups compared to protein, natural ribozymes can still fold into complex structures and catalyze various biological reactions involving RNA cleavage, RNA ligation, and peptide bond formation.[22]

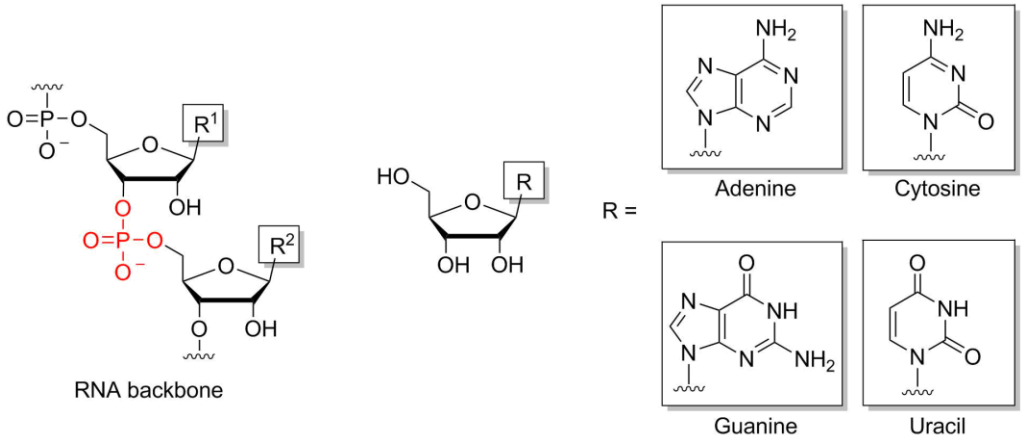


Figure 1.2. Chemical structure of RNA backbone and four nucleobases.

The first natural ribozyme discovered in the early 1980s was the group I intron from *Tetrahymena thermophila*.[21,23] The group I intron is 421 nucleotides long and contains a roughly 200 nucleotides long catalytic core. The group I intron catalyzes the first step in intron splicing, in which the 3'-hydroxyl group of an exogenous guanosine cofactor acts as a nucleophile to perform the self-cleavage during RNA splicing (Figure 1.3A). Another example is RNase P, which was also identified shortly after the discovery of the group I intron in the early 1980s.[24] RNase P is a RNA-protein complex where the RNA component site-specifically catalyzes the hydrolysis of precursor RNA substrates such as tRNA, 4.5S RNA, and viral RNAs, etc.[25,26] Rather than using the 3'-hydroxyl group of a nucleotide, however, RNase P catalyzes the direct nucleophilic attack of a water molecule to cleave the RNA phosphodiester backbone (Figure 1.3B). Other ribozymes exists as large RNA-protein complexes in which both RNA and protein are required for enzymatic ability, but where only the RNA is actually responsible for catalytic function. One representative large RNA-protein complex is the ribosome, in which the rRNA component within the RNA-protein complex catalyzes the peptidyl transfer reaction during protein translation.[27-29]

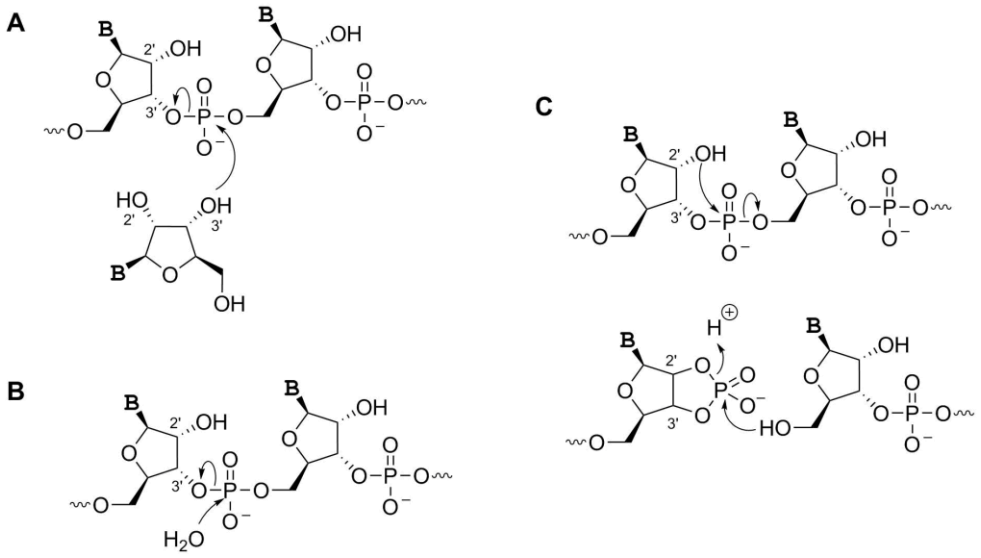


Figure 1.3. Catalytic activities of natural ribozymes. (A) The group I intron uses the 3′-OH of an exogenous guanosine to attack the RNA phosphodiester bond. (B) RNase P catalyzes direct hydrolysis of the RNA phosphodiester bond. (C) RNA cleavage by transesterification, as catalyzed by several naturally occurring small ribozymes. Both enzymes (hammerhead, hairpin) also catalyze the reverse ligation reaction as shown.

In addition to the ribozymes mentioned above, many other smaller ribozymes have also been identified in nature, such as the hammerhead and hairpin ribozymes. The hammerhead ribozyme catalyzes site-specific RNA self-cleavage by transesterification to form 5'-phosphate and 2',3'-cyclic phosphate (Figure 1.3C),[30,31] and it also catalyzes the reverse RNA ligation reaction.[32] The hairpin ribozymes function similarly as the hammerhead ribozyme that they can catalyze both RNA cleavage and RNA ligation reactions.[33]

Metal ions are very important for ribozyme reactions, and in most cases divalent metal ions are required to be directly involved in catalysis (Figure 1.4A). In fact, the hydrophilicity and negative charge on RNA backbone make ribozymes unstable during catalysis. By introducing metal ions during catalysis reaction, metal ions can stabilize the structure of a ribozyme via positioning substrates, and stabilizing the negative charge on the leaving group in the active

site.[34,35] However, not all ribozymes require metal ions in the active site for catalysis. Other than metal ions, nucleobases can also participate in ribozyme catalysis. For instance, the hepatitis delta virus (HDV) ribozyme catalyzes RNA cleavage by acid-base catalysis, in which a protonated cytosine nucleobase acts as a general acid to donate a proton to the 5'-oxygen atom (Figure 1.4B).[36,37] Studies also suggested a metal ion-independent mechanism for the hairpin ribozyme, likely using nucleobases as general acid and base in catalysis.[38,39]

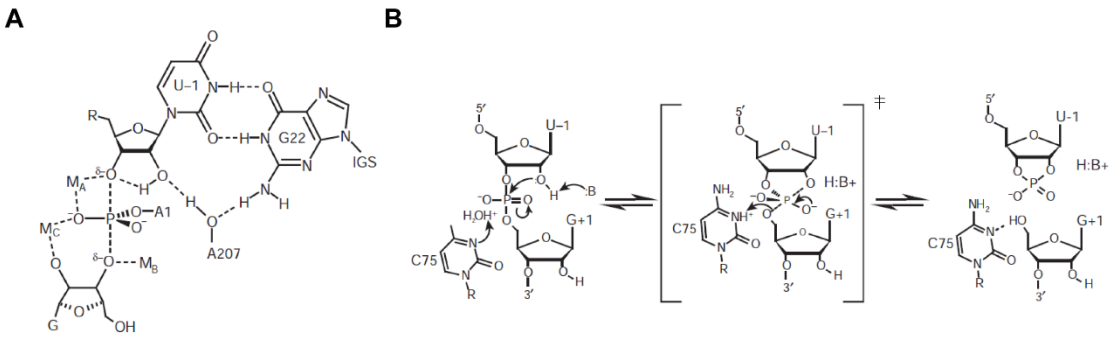


Figure 1.4. Overview of ribozyme reaction mechanisms. (A) Mechanism of RNA cleavage via transesterification by the group I intron. RNA is cleaved by the 3′-hydroxyl group of an exogenous guanosine. The M_A metal ion coordinates the 3′-oxygen atom of U−1, stabilizing the developing negative charge on the leaving 3′-oxygen atom; the M_B metal ion coordinates the 3′-oxygen atom of the exogenous G; the M_C metal ion assists in positioning the substrates and stabilizes the transition state. (B) Mechanism of RNA self-cleavage by the HDV ribozyme. The protonated N3 atom of the C75 nucleobase acts as a general acid, donating a proton to the 5′-oxygen of the leaving ribonucleotide. Figure reprinted with permission from ref. 22.

The discovery of RNA enzymes shows that ribozymes are involved in many important biological processes, such as gene expression and protein synthesis. However, the actual range of chemical reactions for ribozymes is still limited. The significant role played by ribozymes in various biological reactions, yet the limited scope of chemical reactions, urges a development of artificial enzymes to fulfill the needs in biological, chemical, medical studies and applications.

1.3 Engineered Enzymes

Many naturally existing enzymes are used in academic and industrial fields. However, more specific and more efficient enzymes are always desired. Natural enzymes have evolved for billions of years under the evolutionary pressure in nature. It is inefficient and impossible to wait for nature to evolve enzymes with specific and desired function. Therefore, engineered or artificial enzymes have been approached and developed in the laboratory for various purposes, aiming for modified or new substrate specificities or enhanced catalytic abilities. Common approaches for developing protein enzymes and nucleic acid enzymes are described in this section.

1.3.1 Protein Enzymes

Engineered protein enzymes can be derived from either various evolutionary methods to enable catalysis, or from de novo protein design, in which an active site is designed and placed inside of a protein. A method called mRNA display, distinct from protein directed evolution and de novo protein design, has also been developed.

1.3.1.1 Directed Evolution of Proteins

Directed evolution is a robust method and the most common strategy to evolve proteins with desirable functions.[40-44] Directed evolution mimics natural evolution, however much more accelerated, by imposing stringent selection and screening methodologies to identify proteins with optimized functionality, including genetic diversity, binding, catalytic properties, and thermal and environmental stability. A directed evolution experiment encompasses three essential steps: mutagenesis/diversification, screening/selection, and gene amplification (Figure 1.5A). First, the direction evolution requires an existing enzyme, which has a close activity to the desired function, as the starting point. Then, gene diversification techniques are introduced to generate mutated gene variants derived from the parent enzyme gene. Finally, corresponding mutated protein variants are

expressed and screened or selected. Mutants that confer desired activity are subject to subsequent rounds of screening/selection until the final desired function is achieved (Figure 1.5B).

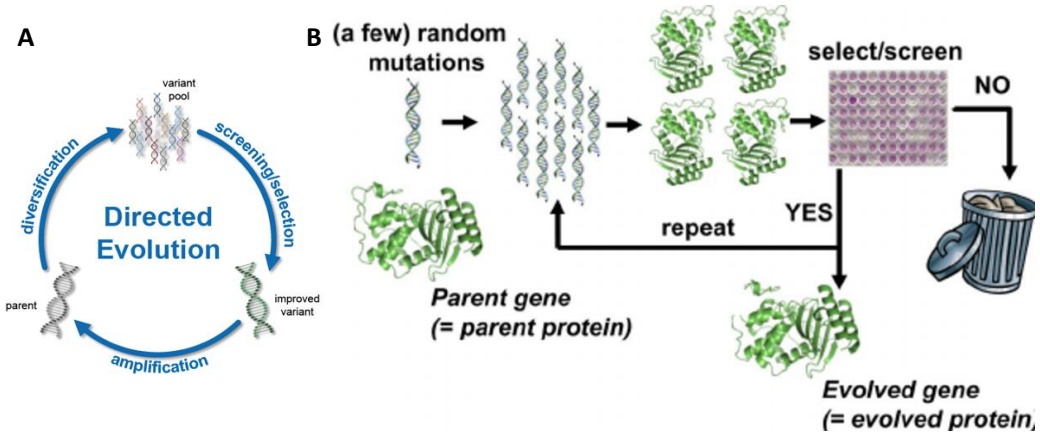


Figure 1.5. (A) Three key steps of directed evolution for proteins. (B) Overview of directed evolution process. Mutations are introduced to diversify the parent gene, generating a library of gene variants. The protein variants are expressed then screened or selected for the desired activity. Undesirable protein variants are discarded, while protein variants with improving desirable phenotype are carried through addition rounds until the desired properties are achieved. Figure reprinted with permission from ref. 40.

The gene diversification step is usually achieved by introducing mutations in the parent enzyme gene. Various mutagenesis techniques such as random mutagenesis, focused mutagenesis, and gene recombination are employed in directed evolution to promote gene mutations.

First, for random mutagenesis, chemical methods can be utilized to introduce random mutations. Chemical mutagens like alkylating compounds such as ethyl methanesulfonate (EMS), or deaminating compounds such as nitrous acid can be used to damage the DNA and lead to errors during replication. However, these chemical methods are not ideal to use for directed evolution due to the biases towards certain types of transitions and transversions.[45,46] Error-prone PCR is a non-chemical method that employs polymerase to generate mutations by imposing nucleotide

incorporation error during DNA replication via using higher concentration of Mg^{2+}, including Mn^{2+}, or using different concentrations of the four dNTPS during the PCR reaction.[47,48] This method is much preferred in gene diversification due to high mutation rate and lower mutation biases.[49,50]

Second, for focused mutagenesis, genetic and structural data are required for the parent enzyme prior to the mutagenesis. Focused mutagenesis is usually accomplished by replacing native DNA with foreign DNA at specific gene sequence region such as the sequence region for enzyme active site, which can likely influence the function of protein enzyme while maintaining the fidelity of the gene at all other locations.[51] The usual method for focused mutagenesis is called site-directed mutagenesis, which enables replication by use of primers with desired bases resulting in specific mutation at a given position.[52] The major advantage of focused mutagenesis is that a smaller number of mutation sites is introduced, which means a higher fraction of possible gene variants from the parent enzyme gene can be explored. The disadvantage of focused mutagenesis is that the understanding of correlation between protein structure and its function is limited, and the residues distant to the active site of protein may have significant impact on its activity.[53] These important residues are impossible to identify based on structural data and therefore cannot be considered when only focused mutations are introduced.

Last, gene recombination includes various techniques such as DNA shuffling, Staggered Extension Protocol (StEP), Incremental Truncation for the Creation of Hybrid Enzymes (ITCHY), and Random Chimeragenesis on Transient Templates (RACHITT). DNA shuffling method includes gene fragmentation using DNase followed by random reassembly by PCR to regenerate gene variants.[54] StEP involves modified annealing and extension steps that can generate staggered fragments.[55] The former two techniques rely on gene libraries of high sequence homology to preserve gene structure among recombinants. ITCHY is a combinatorial methodology for creating libraries of hybrid enzymes which lack high homology,[56] and RACHITT has been used to create

libraries with multiple crossovers events per gene in a single round of gene family shuffling.[57]

Many common methods such as chromatography, mass spectrometry, and colorimetric assays can be used to screen for active protein variants. However, these traditional methods have limited throughput of 10^2-10^4 in a reasonable time. High-throughput flow cytometry screening methods such as fluorescence-activated cell sorting (FACS) can screen up to 10^8 library members in less than 24 hours.[58,59] In contrast to screening, various selection methods have been developed that do not require screening every library member, but rather relate protein desired function to the survival of a host organism.[60] The host organism is subject to the iteration of selection rounds, which will lead to the enrichment of desired protein variants within the host organism. This selection strategy can examine libraries up to 10^8-10^{10} variants, usually limited by the transformation efficiency.[61] In addition to traditional screen/selection methods, a powerful variant method called phage assisted continuous evolution (PACE) has been developed.[62] PACE uses bacteriophage to transfer genes from host cell to host cell, and the desired protein activity is tied to the production of phage protein, which is required for its infection ability. Because successive rounds of evolution are depending on the phage lifecycle, dozens rounds of evolution can happen and hundreds of rounds of selection can be performed in a very short time. PACE can screen libraries up to 10^8-10^{12}, and its library size is limited by titre.[43]

Directed evolution has significant influences on protein engineering. Various protein enzymes with improved catalytic efficiency and desired substrate selectivity have been discovered and developed. However, there are still several drawbacks remaining in direction evolution. The requirement for an existing enzyme as the starting point constrains the scope of the evolved protein. Moreover, evolution aiming for altered substrate specificity usually leads to a relaxed specificity rather than a true alteration.[63] In order to resolve these issues, another protein engineering method has been developed to serve as an alternative approach for designing artificial enzymes—de novo protein design.

1.3.1.2 De Novo Protein Design

There are 20^{200} possible amino acid sequences for a normal 200-residue protein, of which the natural evolutionary process has sampled only a very tiny subset, and it is impossible for directed evolution to assay all sequence variants. However, with the assistance of computational methodologies, de novo protein design is able to explore the full sequence space. Guided by the physical principles that underlie protein folding and chemical principles that interpret catalysis mechanism, unlike directed evolution, de novo protein design is capable of developing novel enzymes without an existing enzyme as the starting point in theory. The overview of de novo protein design process is shown in Figure 1.6.[64] The process is initiated by determining a mechanism for the desired reaction, the amino acids required, and their possible locations within an active site. All possible active-site configurations are then constructed by computational methodology. These three-dimensional active-site models, which are called "theozymes", are placed into larger protein scaffolds at all possible active-site locations and screened for the optimal matches with the best position-related activity. The best "scaffold-active site" matches are expressed and assayed for enzymatic activity. However, de novo protein design is extremely challenging in that it is very difficult to predict which scaffolds will work with certain theozymes, and it is arduous to determine which enzymes will have the best catalytic activity. For enzyme activity is not only affected by best "scaffold-active site" but also by other factors such as temperature and pH when in actual usage. Because of this, de novo protein design is usually paired with directed evolution to improve the resulting enzymes.

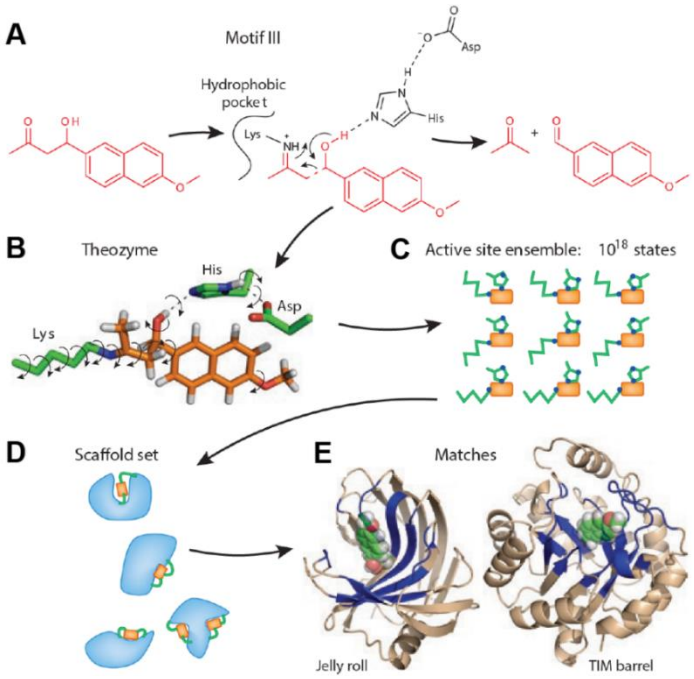


Figure 1.6. Overview of de novo protein design process. (A) The reaction mechanism and the required amino acid residues are determined. (B) Each theozyme is computationally optimized based on physical and chemical principles to accelerate the reaction. (C) Active-site configurations are determined. (D) The active-site configurations are mapped onto existing protein scaffolds to determine the optimal match. (E) Scaffolds that match well with optimized active sites are expressed and assayed. Figure reprinted with permission from ref.63.

Several de novo designed enzymes have been reported to catalyze certain reactions such as Kemp elimination,[65] retro-aldol,[66] Diels-Alder,[67] and ester hydrolysis reaction.[68] Despite these successful cases, de novo protein design is still very challenging. One study of de novo retro-aldolase design reported that the lysine residue designed to be involved in the active site was abandoned and replaced by another lysine residue outside the designed active site after several rounds of directed evolution. An alternative substrate-binding pocket came along with the new catalytic lysine residue rather than using the designed substrate-binding pocket.[69] This case just highlights the huge challenge for de novo protein design and further suggests that de novo protein design could be a good complementary method to serve as a starting point for directed evolution.

1.3.1.3 mRNA Display

mRNA display is a method, which is distinct from directed evolution and de novo protein design, allowing for selection of active protein catalysts.[70,71] The mRNA display method is in many ways similar to the in vitro selection strategies for nucleic acid enzymes that are discussed below.

The overview of mRNA display is shown in Figure 1.7. mRNA, also called messenger RNA, is the single-stranded RNA transcribed from double-stranded DNA. The key step in mRNA display is the modification of mRNA with puromycin which leads to a covalent linkage between the translated protein and coding mRNA. If the translated protein catalyzes the reaction of interest (between A and B in the illustration), then A and B will be covalently linked together with a designed primer (complementary to mRNA) on one end and an anchor group on the other end. Thus, reverse transcription can be performed and cDNA can be isolated by binding to a solid support. cDNA is amplified by PCR and subsequent rounds of selection and amplification can be performed.

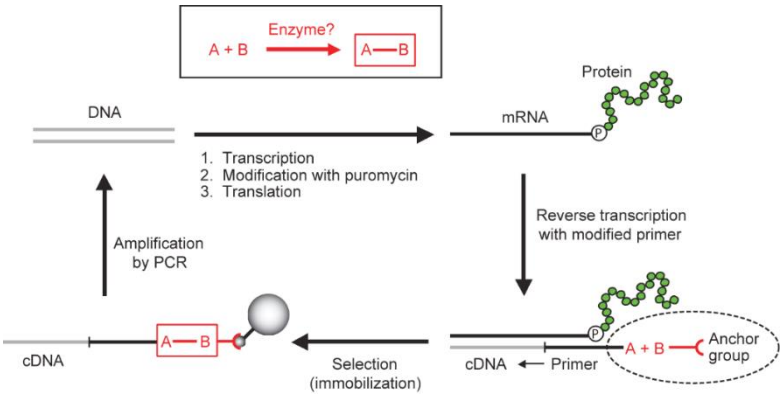


Figure 1.7. Overview of mRNA display process. mRNA is generated from DNA transcription, modified with puromycin, and translated to from the mRNA-protein complex. The complex is subjected to selection conditions in which the protein can catalyze the reaction of interest, and reverse transcription and pull-down of the product separates active proteins which are amplified by PCR. Figure reprinted with permission from ref.70.

1.3.2 Nucleic Acid Enzymes

Plenty discoveries and studies demonstrating the biological significance of natural ribozymes suggests that it is of great interest to engineer artificial nucleic acid enzymes for various reactions.[72] DNA and RNA are two types of nucleic acids, and DNA is structurally very similar to RNA. Thus, DNAzymes can share similar engineering strategies as RNA enzymes have. Development of artificial nucleic acid enzymes was facilitated by a process called in vitro selection, which was first used in the context of RNA to identify RNA aptamers capable of binding to specific molecular targets.[73,74] Since then, many artificial RNA enzymes and DNAzymes have been identified from random sequence pools by in vitro selection.[75-77]

In vitro selection, which is often called systematic evolution of ligands by exponential enrichment (SELEX) for identifying aptamers, is a process in which a large number of sequences are assayed for function at the same time, and those sequences with the desired function are physically separated, PCR-amplified, and subjected to additional rounds of selections (Figure 1.8). One of the primary advantages of in vitro selection is that it can start with a large pool of random sequences, up to 10^{14} molecules, and sequences can be amplified by PCR. The only reason that in vitro selection is possible is because of PCR amplification, which cannot be done for protein. Sequences with desired activity are enriched through the selection step; meanwhile, undesired sequences are eliminated under the selection pressure. However, it is inevitable to have some noncatalytic sequences survive due to the background reaction and the artifacts of the separation method. The surviving population with enriched catalytic sequences undergoes the amplification step to restore the pool size, which is crucial for in vitro selection to ensure a sufficient number of sequences for the next selection round. RNA sequences must be reverse-transcribed to cDNA, amplified, and then transcribed back to RNA. DNA sequences can be directly amplified by PCR. The surviving population with restored pool size is subjected to additional rounds of selection until the desired function is achieved. At that point, the active pool is cloned and sequenced, and

individual nucleic acid enzymes are identified and assayed.

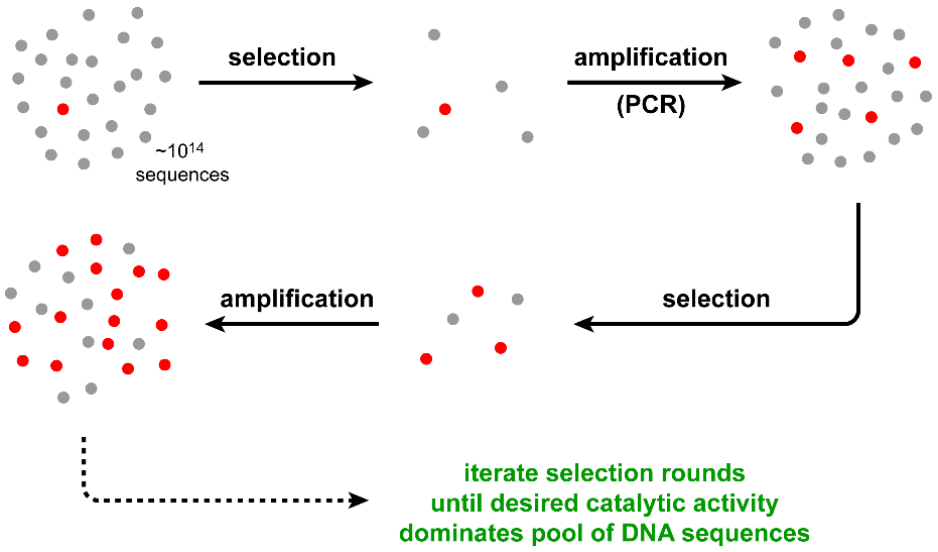


Figure 1.8. Overview of in vitro selection process. The initial pool contains ~10^{14} sequences. The catalytically active sequences (red dots) are enriched in the selection step, and the noncatalytic sequences (gray dots) are minimized under selection pressure. All the surviving catalytic sequences and noncatalytic sequences undergo the amplification step to restore the pool size and then are subjected to iteration of selection rounds.

1.3.2.1 Ribozymes

Engineered ribozymes have been identified for a variety of reactions involving RNA as substrate, such as RNA cleavage via transesterification,[78,79] ligation,[80,81] phosphorylation,[82,83] and polymerization.[84,85] Artificial ribozymes have also been discovered for a number of reactions not using RNA as substrates, such as Diels-Alder reaction,[86,87] aldol reaction,[88] alcohol oxidation reaction,[89] and acyl transfer reactions.[90-92] The scope of chemical reactions catalyzed by engineered ribozymes is extremely diverse, demonstrating the excellent catalytic potential of ribozymes.

Despite the large number of engineered ribozymes that have been discovered for various

chemical reactions, the understanding and information regarding how engineered ribozymes work are still limited due to very few structural and mechanistic data of artificial ribozymes,[93] such as a lead-dependent ribozyme catalyzing RNA cleavage[94,95] and a ribozyme that catalyzes the Diels-Alder reaction.[96] Next-generation sequencing has been employed to analyze the enrichment process of active sequences during the in vitro selection rounds,[97] however, it does not lead to the structural or mechanistic understanding of engineered ribozymes.

1.3.2.2 Deoxyribozymes

As mentioned above, DNA shares structural similarities with RNA, and the only differences are that DNA lacks the 2'-hydroxyl group of RNA and replaces the uridine ribonucleotide with thymidine deoxyribonucleotide (Figure 1.9). Because of these similarities, single-stranded DNA can also fold into complex secondary thus tertiary structures and be capable of catalyzing reactions. In 1994, the very first DNAzyme was discovered,[98] which can catalyze the cleavage of a single RNA nucleotide linkage in a DNA substrate. Since then, a large number of DNAzymes have been identified.[99] However, in similar cases to ribozymes, structural or mechanistic studies of DNAzymes are poorly explored, due to extreme difficulty of obtaining crystal structures for DNAzymes.[100] The first high-resolution crystal structure of a DNAzyme, which is a RNA-ligating DNAzyme called 9DB1, was reported in 2016.[101] This first crystal structure contributes to the understanding of DNAzymes mechanism such as metal ions participation during the reaction. Metal ions are very important and required for the activity of nucleic acid enzymes.[34,35] However, the crystal structure revealed no metal ion at the active site, but instead, an internucleotide phosphodiester oxygen atom showed significant interaction with the newly formed phosphodiester linkage. This observation indicates that DNAzymes are not necessarily metalloenzymes. But still, exploring the reaction mechanisms of DNAzymes is extremely challenging.

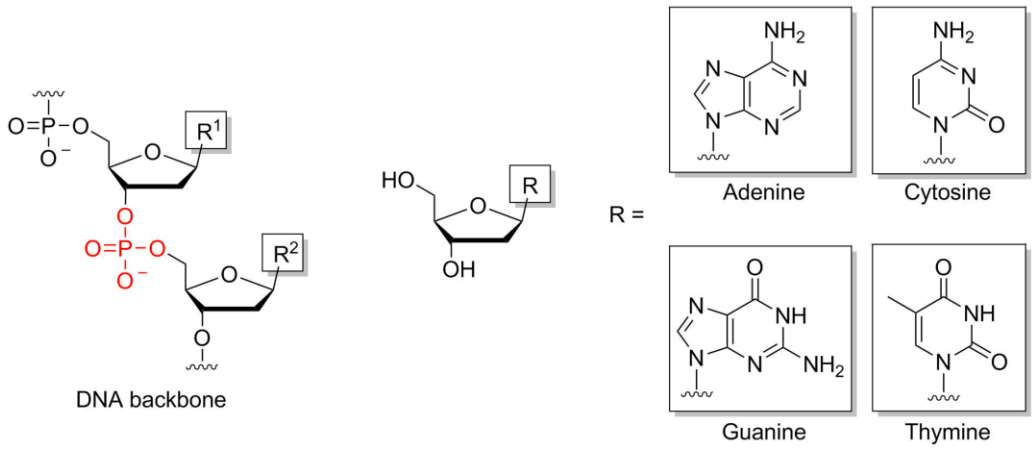


Figure 1.9. Chemical Structure of DNA backbone and four nucleobases.

DNAzymes have been identified by in vitro selection for various biochemical reactions involving oligonucleotides as substrates, including RNA cleavage by transesterification[102,103] or hydrolysis,[104] RNA ligation,[105-107] DNA cleavage by hydrolysis[108,109] or deglycosylation,[110] DNA ligation,[111] and DNA phosphorylation.[112,113] A large number of DNAzymes have also been discovered for reactions using non-oligonucleotide substrates, such as conjugation of oligonucleotides to tyrosine,[114-116] tyrosine phosphorylation,[117,118] phosphotyrosine (pTyr) and phosphoserine (pSer) dephosphorylation,[119] formation of dehydroalanine by elimination of phosphate from pSer,[120] tyrosine modification via azido-adenylylation,[121] lysine modification by phosphorimidazolidated oligonucleotides,[122] and amide-bond hydrolysis.[123] DNAzymes that catalyzes lysine acylation have been identified as part of this thesis work and will be described in later chapters.

The use of nucleic acids as catalysts has several advantages over protein in the context of identifying artificial enzymes using selection or evolution-based methods.[76] First, the in vitro selection of nucleic acids enzymes can start with entirely random sequence libraries, which allows the identification of novel enzymes without the requirement for a specific starting point, such as an existing enzyme required by directed evolution or a known reaction mechanism required by de novo protein design. Second, nucleic acid sequences can be amplified directly by natural

polymerases, unlike the protein enzymes that cannot be directly amplified. The ability to be directly amplified is the fundamental base for in vitro selection. Third, considering there are only 4 building blocks for nucleic acid enzymes relative to 20 building blocks for protein enzymes, nucleic acid selections can cover a much larger fraction of sequence space (4^n, where n is the random sequence length) compared to protein selections (20^n). Last but not least, random nucleic acid sequences can readily fold into secondary thus tertiary structures while most random protein sequences will misfold and aggregate. Combined with the third reason, these two facts further limit the actual coverage of protein sequence space. When comparing between DNA and RNA, DNA has further practical advantages over RNA, including its chemical stability, cost, and ease of synthesis. Thus, DNA is a strong biopolymer candidate in terms of discovering novel macromolecular catalysts.

1.4 Thesis Research Focus

The research described in this thesis is focused on identifying DNAzymes for amine and peptide lysine acylation. The DNA-anchored single amino substrate and tethered peptide lysine were used as nucleophiles in all chapters, while different electrophiles were developed and employed in individual chapter. Chapter 2 describes the efforts towards DNAzymes for amine acylation via thioester acyl donor electrophiles. This part of the work was largely based on prior graduate student work. In this chapter's work, the DNA included modified nucleotides. Chapter 3 describes the efforts towards DNAzymes for amine acylation via highly reactive aryl ester acyl donor electrophiles. Coupling reagents were used as activators and additives for generating acyl donors. In this chapter's work, the DNA was composed of four standard nucleotides, and same for the remainder chapters. Chapter 4 describes the identification of DNAzymes for amine and peptide lysine acylation via intermediate-reactivity aryl ester acyl donor electrophiles. Acyl donors such as phenyl ester and 4-fluoro-phenyl ester were pre-formed, purified, and then utilized in selections. Chapter 5 describes the efforts towards DNAzymes for amine and peptide lysine glutarylation.

1.5 References

(1) Masel, R. I. *Chemical kinetics and catalysis*; Wiley-Interscience New York, 2001; Vol. 10.

(2) Jencks, W. P. General acid-base catalysis of complex reactions in water. *Chem. Rev.* **1972**, *72*, 705-718.

(3) Loening, K. L.; Garrett, A.; Newman, M. S. Steric effects. II. Acid-catalyzed esterification. *J. Am. Chem. Soc.* **1952**, *74*, 3929-3931.

(4) Fritz, K. S. Chemical Acetylation and Deacetylation. In *Sirtuins: Methods and Protocols*; Methods in Molecular Biology, 2013; Vol. 1077, pp. 191-201.

(5) Jencks, W. P.; Carriuolo, J. General Base Catalysis of Ester Hydrolysis. *J. Am. Chem. Soc.* **1961**, *83*, 1743-1750.

(6) Eigen, M. Proton transfer, acid‐base catalysis, and enzymatic hydrolysis. Part I: elementary processes. *Angew. Chem. Int. Ed.* **1964**, *3*, 1-19.

(7) Knowles, R. R.; Jacobsen, E. N. Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20678-20685.

(8) Chen, M. S.; White, M. C. A predictably selective aliphatic C–H oxidation reaction for complex molecule synthesis. *Science* **2007**, *318*, 783-787.

(9) Larsen, C. N.; Krantz, B. A.; Wilkinson, K. D. Substrate specificity of deubiquitinating enzymes: ubiquitin C-terminal hydrolases. *Biochemistry* **1998**, *37*, 3358-3368.

(10) Lipscomb, W. N. Acceleration of reactions by enzymes. *Acc. Chem. Res.* **1982**, *15*, 232-238.

(11) Wolfenden, R.; Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* **2001**, *34*, 938-945.

(12) Zhou, G. Q.; Zhong, W. Z. Diffusion‐Controlled Reactions of Enzymes: A Comparison between Chou's Model and Alberty‐Hammes‐Eigen's Model. *Eur. J. Biochem.* **1982**, *128*, 383-387.

(13) Kirk, O.; Borchert, T. V.; Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **2002**, *13*, 345-351.

(14) Liang, J. F.; Li, Y. T.; Yang, V. C. Biomedical application of immobilized enzymes. *J. Pharm. Sci.* **2000**, *89*, 979-990.

(15) Park, H.-S.; Nam, S.-H.; Lee, J. K.; Yoon, C. N.; Mannervik, B.; Benkovic, S. J.; Kim, H.-S. Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **2006**, *311*, 535-538.

(16) Alberty, R. A. A short history of the thermodynamics of enzyme-catalyzed reactions. *J. Biol. Chem.* **2004**, *279*, 27831-27836.

(17) Carrozza, M. J.; Utley, R. T.; Workman, J. L.; Côté, J. The diverse functions of histone acetyltransferase complexes. *Trends Genet.* **2003**, *19*, 321-329.

(18) Sabari, B. R.; Tang, Z.; Huang, H.; Yong-Gonzalez, V.; Molina, H.; Kong, H. E.; Dai, L.; Shimada, M.; Cross, J. R.; Zhao, Y. Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol. Cell* **2015**, *58*, 203-215.

(19) Tan, M.; Peng, C.; Anderson, K. A.; Chhoy, P.; Xie, Z.; Dai, L.; Park, J.; Chen, Y.; Huang, H.; Zhang, Y. Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab.* **2014**, *19*, 605-617.

(20) Yanagisawa, T.; Umehara, T.; Sakamoto, K.; Yokoyama, S. Expanded genetic code technologies for incorporating modified lysine at multiple sites. *Chembiochem* **2014**, *15*, 2181-2187.

(21) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **1981**, *27*, 487-496.

(22) Doudna, J. A.; Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **2002**, *418*, 222-228.

(23) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147-157.

(24) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849-857.

(25) Coughlin, D. J.; Pleiss, J. A.; Walker, S. C.; Whitworth, G. B.; Engelke, D. R. Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12218-12223.

(26) Evans, D.; Marquez, S. M.; Pace, N. R. RNase P: interface of the RNA and protein worlds. *Trends Biochem. Sci.* **2006**, *31*, 333-341.

(27) Hiller, D. A.; Singh, V.; Zhong, M.; Strobel, S. A. A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* **2011**, *476*, 236-239.

(28) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000**, *289*, 920-930.

(29) Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242-253.

(30) Forster, A. C.; Symons, R. H. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell* **1987**, *49*, 211-220.

(31) Pley, H. W.; Flaherty, K. M.; McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **1994**, *372*, 68-74.

(32) Canny, M. D.; Jucker, F. M.; Pardi, A. Efficient ligation of the Schistosoma hammerhead ribozyme. *Biochemistry* **2007**, *46*, 3826-3834.

(33) Hampel, A.; Tritz, R. RNA catalytic properties of the minimum (-) sTRSV sequence. *Biochemistry* **1989**, *28*, 4929-4933.

(34) Shan, S.-o.; Kravchuk, A. V.; Piccirilli, J. A.; Herschlag, D. Defining the catalytic metal ion interactions in the Tetrahymena ribozyme reaction. *Biochemistry* **2001**, *40*, 5161-5171.

(35) Stahley, M. R.; Strobel, S. A. RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr. Opin. Struct. Biol.* **2006**, *16*, 319-326.

(36) Chen, J.-H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and

general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry* **2010**, *49*, 6508-6518.

(37) Ferré-D'Amaré, A. R.; Zhou, K.; Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **1998**, *395*, 567-574.

(38) Nesbitt, S.; Hegg, L.; Fedor, M. An unusual mechanism for hairpin ribozyme catalysis: pH and metal ion independence. *Chem. Biol.* **1997**, *4*, 619-630.

(39) Rupert, P. B.; Ferré-D'Amaré, A. R. Crystal structure of a hairpin ribozyme–inhibitor complex with implications for catalysis. *Nature* **2001**, *410*, 780-786.

(40) Arnold, F. H.; Volkov, A. A. Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **1999**, *3*, 54-59.

(41) Bloom, J. D.; Arnold, F. H. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9995-10000.

(42) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **2015**, *44*, 1172-1239.

(43) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015**, *16*, 379-394.

(44) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the enzyme universe: accessing non‐natural reactions by mechanism‐guided directed evolution. *Angew. Chem. Int. Ed.* **2015**, *54*, 3351-3367.

(45) Lai, Y. P.; Huang, J.; Wang, L. F.; Li, J.; Wu, Z. R. A new approach to random mutagenesis in vitro. *Biotechnol. Bioeng.* **2004**, *86*, 622-627.

(46) Myers, R. M.; Lerman, L. S.; Maniatis, T. A general method for saturation mutagenesis of cloned DNA fragments. *Science* **1985**, *229*, 242-247.

(47) Cadwell, R. C.; Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **1992**, *2*, 28-33.

(48) Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **1996**, *255*, 589-603.

(49) Mccullum, E.; Williams, B.; Zhang, J.; Chaput, J. In Vitro Mutagenesis Protocols. *Methods Mol. Biol.* **2010**, *634*, 103-109.

(50) Wong, T. S.; Tee, K. L.; Hauer, B.; Schwaneberg, U. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* **2004**, *32*, e26-e26.

(51) Wells, J. A.; Vasser, M.; Powers, D. B. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* **1985**, *34*, 315-323.

(52) Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **1989**, *77*, 51-59.

(53) Marshall, N. M.; Garner, D. K.; Wilson, T. D.; Gao, Y.-G.; Robinson, H.; Nilges, M. J.; Lu, Y. Rationally tuning the reduction potential of a single cupredoxin beyond the natural range. *Nature* **2009**, *462*, 113-116.

(54) Stemmer, W. P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994**, *370*, 389-391.

(55) Aguinaldo, A. M.; Arnold, F. H. Staggered extension process (StEP) in vitro recombination. In *Directed Evolution Library Creation*; Springer, 2003, pp. 105-110.

(56) Ostermeier, M.; Nixon, A. E.; Benkovic, S. J. Incremental truncation as a strategy in the engineering of novel biocatalysts. *Bioorgan. Med. Chem.* **1999**, *7*, 2139-2144.

(57) Coco, W. M. Rachitt. In *Directed Evolution Library Creation*; Springer, 2003, pp. 111-127.

(58) Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997**, *15*, 553-557.

(59) Fulwyler, M. J. Electronic separation of biological cells by volume. *Science* **1965**, *150*, 910-911.

(60) Santoro, S. W.; Schultz, P. G. Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 4185-4190.

(61) Dower, W. J.; Miller, J. F.; Ragsdale, C. W. High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res.* **1988**, *16*, 6127-6145.

(62) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **2011**, *472*, 499-503.

(63) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **2006**, *10*, 498-508.

(64) Nanda, V. Do-it-yourself enzymes. *Nat. Chem. Biol.* **2008**, *4*, 273-275.

(65) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190-195.

(66) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F. De novo computational design of retro-aldol enzymes. *Science* **2008**, *319*, 1387-1391.

(67) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010**, *329*, 309-313.

(68) Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **2012**, *134*, 16197-16206.

(69) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494-498.

(70) Keefe, A. D.; Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **2001**, *410*, 715-718.

(71) Seelig, B.; Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **2007**, *448*, 828-831.

(72) Jaeger, L. The new world of ribozymes. *Curr. Opin. Struct. Biol.* **1997**, *7*, 324-335.

(73) Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818-822.

(74) Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505-510.

(75) Breaker, R. R. DNA aptamers and DNA enzymes. *Curr. Opin. Chem. Biol.* **1997**, *1*, 26-31.

(76) Silverman, S. K. Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.* **2016**, *41*, 595-609.

(77) Wilson, D. S.; Szostak, J. W. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(78) Jayasena, V. K.; Gold, L. In vitro selection of self-cleaving RNAs with a low pH optimum. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10612-10617.

(79) Pan, T.; Uhlenbeck, O. C. A small metalloribozyme with a two-step mechanism. *Nature* **1992**, *358*, 560-563.

(80) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **1995**, *269*, 364-370.

(81) Hager, A. J.; Szostak, J. W. Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem. Biol.* **1997**, *4*, 607-617.

(82) Curtis, E. A.; Bartel, D. P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **2005**, *12*, 994-1000.

(83) Lorsch, J. R.; Szostak, J. W. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994**, *371*, 31-36.

(84) Ekland, E. H.; Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **1996**, *382*, 373-376.

(85) Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **2001**, *292*, 1319-1325.

(86) Seelig, B.; Jäschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **1999**, *6*, 167-176.

(87) Seelig, B.; Keiper, S.; Stuhlmann, F.; Jäschke, A. Enantioselective ribozyme catalysis of a bimolecular cycloaddition reaction. *Angew. Chem. Int. Ed.* **2000**, *39*, 4576-4579.

(88) Fusz, S.; Eisenführ, A.; Srivatsan, S. G.; Heckel, A.; Famulok, M. A ribozyme for the aldol reaction. *Chem. Biol.* **2005**, *12*, 941-950.

(89) Tsukiji, S.; Pattnaik, S. B.; Suga, H. An alcohol dehydrogenase ribozyme. *Nat. Struct. Mol. Biol.* **2003**, *10*, 713-717.

(90) Jenne, A.; Famulok, M. A novel ribozyme with ester transferase activity. *Chem. Biol.* **1998**, *5*, 23-34.

(91) Lohse, P. A.; Szostak, J. W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **1996**, *381*, 442-444.

(92) Suga, H.; Lohse, P. A.; Szostak, J. W. Structural and kinetic characterization of an acyl

transferase ribozyme. *J. Am. Chem. Soc.* **1998**, *120*, 1151-1156.

(93) Shechner, D. M.; Bartel, D. P. The structural basis of RNA-catalyzed RNA polymerization. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1036.

(94) Wedekind, J. E.; McKay, D. B. Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis. *Nat. Struct. Biol.* **1999**, *6*, 261-268.

(95) Wedekind, J. E.; McKay, D. B. Crystal structure of the leadzyme at 1.8 Å resolution: Metal ion binding and the implications for catalytic mechanism and allo site ion regulation. *Biochemistry* **2003**, *42*, 9554-9563.

(96) Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Höbartner, C.; Polonskaia, A.; Phan, A. T.; Wombacher, R.; Micura, R. Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat. Struct. Mol. Biol.* **2005**, *12*, 218-224.

(97) Ameta, S.; Winz, M.-L.; Previti, C.; Jäschke, A. Next-generation sequencing reveals how RNA catalysts evolve from random space. *Nucleic Acids Res.* **2014**, *42*, 1303-1310.

(98) Breaker, R. R.; Joyce, G. F. A DNA enzyme that cleaves RNA. *Chem. Biol.* **1994**, *1*, 223-229.

(99) Silverman, S. K. Pursuing DNA catalysts for protein modification. *Acc. Chem. Res.* **2015**, *48*, 1369-1379.

(100) Nowakowski, J.; Shim, P. J.; Prasad, G. S.; Stout, C. D.; Joyce, G. F. Crystal structure of an 82-nucleotide RNA–DNA complex formed by the 10-23 DNA enzyme. *Nat. Struct. Biol.* **1999**, *6*, 151-156.

(101) Ponce-Salvatierra, A.; Wawrzyniak-Turek, K.; Steuerwald, U.; Höbartner, C.; Pena, V.

Crystal structure of a DNA catalyst. *Nature* **2016**, *529*, 231-234.

(102) Breaker, R. R.; Joyce, G. F. A DNA enzyme with Mg2+-dependent RNA phosphoesterase activity. *Chem. Biol.* **1995**, *2*, 655-660.

(103) Santoro, S. W.; Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4262-4266.

(104) Parker, D. J.; Xiao, Y.; Aguilar, J. M.; Silverman, S. K. DNA catalysis of a normally disfavored RNA hydrolysis reaction. *J. Am. Chem. Soc.* **2013**, *135*, 8472-8475.

(105) Coppins, R. L.; Silverman, S. K. A DNA enzyme that mimics the first step of RNA splicing. *Nat. Struct. Mol. Biol.* **2004**, *11*, 270-274.

(106) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. Deoxyribozymes with 2 '− 5 'RNA ligase activity. *J. Am. Chem. Soc.* **2003**, *125*, 2444-2454.

(107) Purtha, W. E.; Coppins, R. L.; Smalley, M. K.; Silverman, S. K. General Deoxyribozyme-Catalyzed Synthesis of Native 3 '− 5 'RNA Linkages. *J. Am. Chem. Soc.* **2005**, *127*, 13124-13125.

(108) Chandra, M.; Sachdeva, A.; Silverman, S. K. DNA-catalyzed sequence-specific hydrolysis of DNA. *Nat. Chem. Biol.* **2009**, *5*, 718-720.

(109) Xiao, Y.; Wehrmann, R. J.; Ibrahim, N. A.; Silverman, S. K. Establishing broad generality of DNA catalysts for site-specific hydrolysis of single-stranded DNA. *Nucleic Acids Res.* **2012**, *40*, 1778-1786.

(110) Dokukin, V.; Silverman, S. K. Lanthanide ions as required cofactors for DNA catalysts. *Chem. Sci.* **2012**, *3*, 1707-1714.

(111) Cuenoud, B.; Szostak, J. W. A DNA metalloenzyme with DNA ligase activity. *Nature* **1995**, *375*, 611-614.

(112) Camden, A. J.; Walsh, S. M.; Suk, S. H.; Silverman, S. K. DNA oligonucleotide 3′-phosphorylation by a DNA enzyme. *Biochemistry* **2016**, *55*, 2671-2676.

(113) Li, Y.; Breaker, R. R. Phosphorylating DNA with DNA. *Proc. Nat. Acad. Sci.* **1999**, *96*, 2746-2751.

(114) Chu, C. C.; Wong, O. Y.; Silverman, S. K. A Generalizable DNA‐Catalyzed Approach to Peptide–Nucleic Acid Conjugation. *ChemBioChem* **2014**, *15*, 1905-1910.

(115) Pradeepkumar, P. e. I.; Höbartner, C.; Baum, D. A.; Silverman, S. K. DNA-Catalyzed Formation of Nucleopeptide Linkages. *Angew. Chem. Int. Ed.* **2008**, *120*, 1777-1781.

(116) Wong, O. Y.; Pradeepkumar, P.; Silverman, S. K. DNA-catalyzed covalent modification of amino acid side chains in tethered and free peptide substrates. *Biochemistry* **2011**, *50*, 4741-4749.

(117) Walsh, S. M.; Konecki, S. N.; Silverman, S. K. Identification of Sequence-Selective Tyrosine Kinase Deoxyribozymes. *J. Mol. Evol.* **2015**, *81*, 218-224.

(118) Walsh, S. M.; Konecki, S. N.; Silverman, S. K. Identification of sequence-selective tyrosine kinase deoxyribozymes. *J. Mol. Evol.* **2015**, *81*, 218-224.

(119) Chandrasekar, J.; Silverman, S. K. Catalytic DNA with phosphatase activity. *Proc. Natl.*

Acad. Sci. USA **2013**, *110*, 5315-5320.

(120) Chandrasekar, J.; Wylder, A. C.; Silverman, S. K. Phosphoserine lyase deoxyribozymes: DNA-catalyzed formation of dehydroalanine residues in peptides. *J. Am. Chem. Soc.* **2015**, *137*, 9575-9578.

(121) Wang, P.; Silverman, S. K. DNA‐Catalyzed Introduction of Azide at Tyrosine for Peptide Modification. *Angew. Chem. Int. Ed.* **2016**, *55*, 10052-10056.

(122) Brandsen, B. M.; Velez, T. E.; Sachdeva, A.; Ibrahim, N. A.; Silverman, S. K. DNA‐Catalyzed Lysine Side Chain Modification. *Angew. Chem. Int. Ed.* **2014**, *53*, 9045-9050.

(123) 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*, 2106-2109.

Chapter 2: Efforts towards DNAzymes for Amine Acylation via Thioester Acyl Donor Electrophiles

2.1 Introduction

2.1.1 Lysine Acylation

Lysine acylation is a common and important post-translational modification. The most well-known and well-studied acylation type is lysine acetylation, which is involved in gene expression, lipid and amino acid metabolism, energy production, oxidative stress response, and intracellular signaling.[1] However, protein acylation is not limited to histone lysine acetylation,[2] and its application scale has been expanded for different acyl donors and different protein substrates. Other types of lysine acylation, such as malonylation, glutarylation, succinylation and crotonylation, have revealed significant functions in metabolism, gene expression, immunology and diseases.[3] Also, protein acetylation is an important target to study not only in histone proteins but also in many other biologically significant proteins such as tumor protein p53[4] and the inducible transcription protein NF-ƙB, which plays a central role in the immune response.[5,6]

Despite the broad biological scope of lysine acylation, robust methods to prepare peptides and proteins with specific acylation type is still limited. Chemical methods and enzymatic methods are two general ways to create peptides or proteins with acylated lysine. For chemical methods, amine acylation do not provide the expected site specificity for selective protein modification. Synthesis of site-specifically acetylated histone proteins has been reported using native chemical ligation to join N-terminal peptide fragments prepared by solid-phase synthesis to recombinant globular histone domains.[7] However, this method requires the expensive chemical synthesis of large quantities of peptide thioesters. Furthermore, the ligation sites of the peptide thioesters are quite limited, because they must be next to a Cys (or Ala) residue. For enzymatic methods, lysine acetyltransferases have been reconstituted and utilized in vitro for incorporating the natural acetyl

modification,[8] as well as for incorporating unnatural acyl modifications for protein labeling.[9,10] In many cases, however, heterogeneously modified forms are generated by the enzymatic methods, because it is difficult to control the sites and degrees of modifications when using these enzymes in vitro. Genetic methods have also been developed that incorporate an unnatural amino acid which is subsequently converted to dehydrolalanine,[11,12] forming an electrophile within the protein. This dehydroalanine is then reacted with derivatized thiols, including acetyl amines, to yield site-specifically modified proteins. However, there are drawbacks of dehydroalanine method. One drawback is that this method requires lots of cysteine site in protein, and it is not so useful for protein that does not process a lot of cysteine site. Another major drawback is that modification site is restricted by cysteine site, and this method cannot modify specific site as desire. In general, robust methods to generate peptides and proteins with site-specifically acylated lysine are still highly desired.

2.1.2 DNA-Catalyzed Lysine Side Chain Modification

DNAzymes that can catalyze the nucleophilic attack of water to cleave the amide bond have been identified,[13] meanwhile the amino group on lysine side chain is more nucleophilic than water. DNAzymes has also been found to catalyze the modification of lysine side chain using phosphorimidazolide as electrophile, resulting in formation of a phosphoramidate linkage.[14] Thus, it is logic to assume that DNAzymes are also capable of catalyzing the lysine acylation reaction using acyl donor electrophile. Previous efforts showed that the choice of electrophile has significant impact on selection design and potential selection outcome. Thioester was then the first acyl donor chosen and used in selections as electrophile, for which has been employed in many biological reactions naturally.

2.1.3 Prior Work towards DNAzymes with Unmodified Nucleotides for Amine Acylation

Prior selections seeking DNAzymes with regular nucleotides has been carried out by previous

graduate student Benjamin Brandsen. These selections were designed around two thioester substrates as electrophiles: 5'-glutaryl-DNA thioester and glutaryl-coenzyme A (Figure 2.1). The DNA-anchored thioester substrate is closely tethered to the DNAzyme by Watson-Crick base pairing, and thus the DNAzyme can directly use the substrate in reaction. In contrast, glutaryl-CoA is present in solution at a high concentration, and the DNAzyme is required to bind to the glutaryl-CoA first, and then use it in reaction.

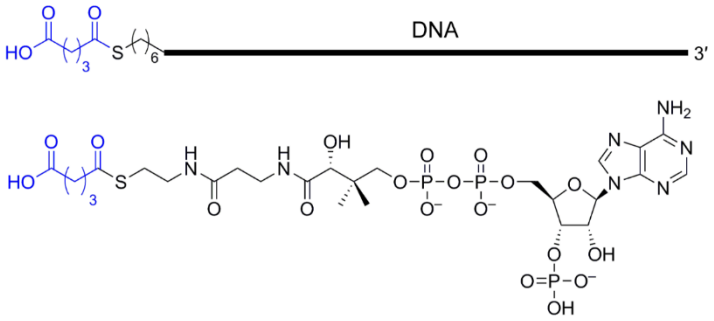


Figure 2.1. Two thioester substrates used for in vitro selection employing random DNA pool sequence with unmodified nucleotides.

Two amine substrates were chosen as nucleophiles: 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA. 5'-DNA-C_3-NH_2 is closely tethered to DNAzymes, which allows DNAzymes to directly use this substrate during reaction. 5'-DNA-HEG-AAAKAA has the hexapeptide substrate and is loosely tethered to DNAzymes, which requires DNAzymes to bind it first and then catalyze the acylation reaction.

Two random region lengths were chosen: N_{40} and N_{80}. N_{40} has worked well for identifying deoxyribozymes for a large number of chemical reactions, while N_{80} should be able to adopt more complex catalytic structures than N_{40}, although it has a smaller fraction of sequence space covered compared to N_{40}. Six selections were performed (Figure 2.2) under the same incubation condition. The incubation condition was 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl. The samples were incubated at 37 °C for 12-14 h.

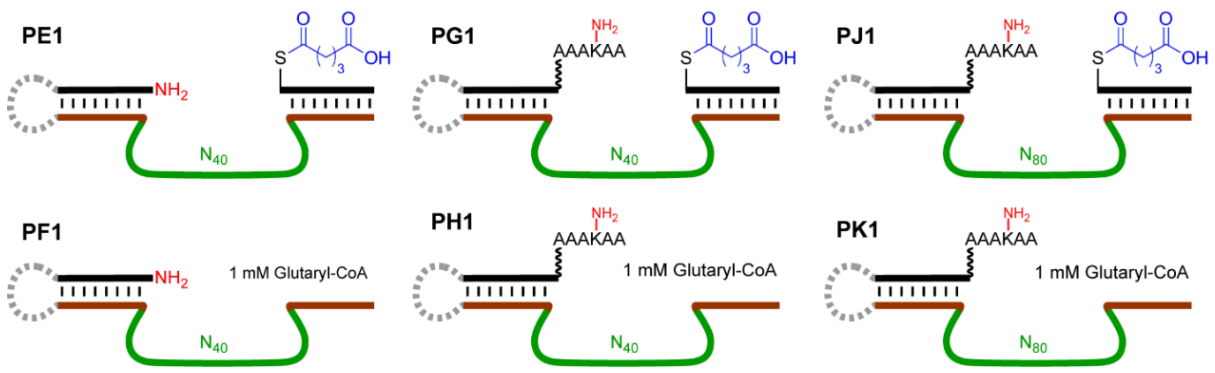


Figure 2.2. Overview of selections performed for amine acylation via DNA pools with unmodified nucleotides. Selection variables of amine nucleophile, thioester acyl donor electrophile, and random region length are illustrated.

The key selection and capture steps are illustrated in Figure 2.3. Provided that the acylation reaction is catalyzed by DNAzymes, glutaric acid is transferred from the thioester acyl donor electrophile to the amine nucleophile. The transfer of glutaric acid does not provide sufficient mass to enable a direct gel shift selection. Instead, the additional carboxylic acid on glutaric acid is used as an extra chemical handle for a subsequent capture reaction with a DNA splint and 5'-amino DNA capture oligo. This additional capture reaction adds sufficient mass to enable selection by gel shift.

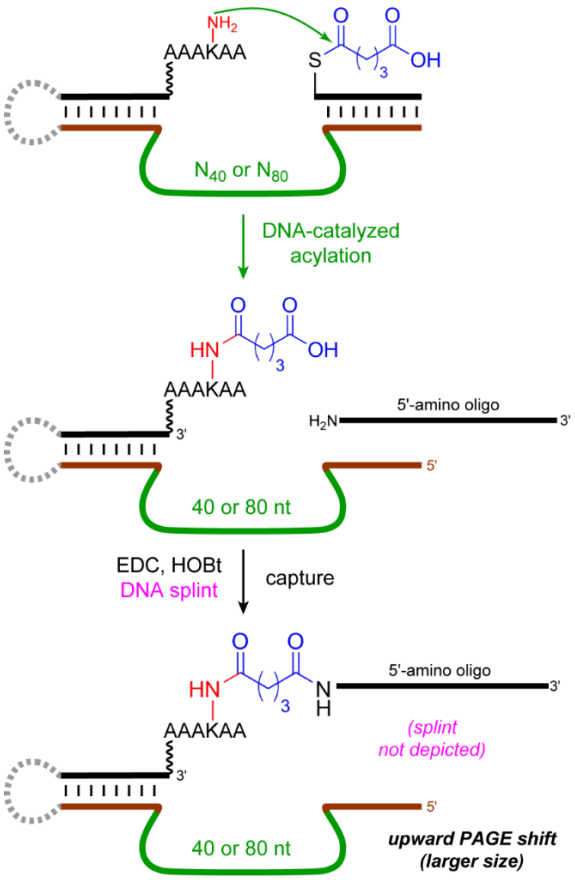


Figure 2.3. Key selection and capture steps for in vitro selection using DNA pools with regular nucleotides. DNA-catalyzed acylation using the glutaryl-DNA thioester (as shown) or untethered glutaryl-CoA results in transfer of glutaric acid to the amine nucleophile. The acylated product is captured in a DNA splinted chemical reaction using an amino-modified DNA, EDC and HOBt. Captured products are separated by PAGE and amplified by PCR.

In vitro selections were performed for 11 rounds (Figure 2.4) by previous graduate student Benjamin Brandsen. In each round, a standard capture reaction capturing either the glutarylated DNA-C_3-NH_2 or the glutarylated DNA-HEG-AAAKAA was performed. The capture yield of glutarylated DNA-C_3-NH_2 was 40-60%, and the capture yield of glutarylated DNA-HEG-AAAKAA was between 25-50%. After 11 rounds of selection, selections using the glutaryl thioester donor were showing 5-10% activity. Each selection was assayed in trans, without ligating the DNA-anchored amine substrate to the PCR product. However, in these assays, no activity in

trans was observed. This finding means that the deoxyribozymes are only active in cis, when the DNA-anchored amine is ligated. One possible explanation is that deoxyribozyme sequence requires the DNA loop being present during selection for activity, and it loses activity when the loop is not present as in trans assays. Considering the highest yield after 11 rounds was only close to 10%, other possible explanation is that the acylation reaction is too difficult to be catalyzed at first place due to unmodified DNA sequences are functional limited.

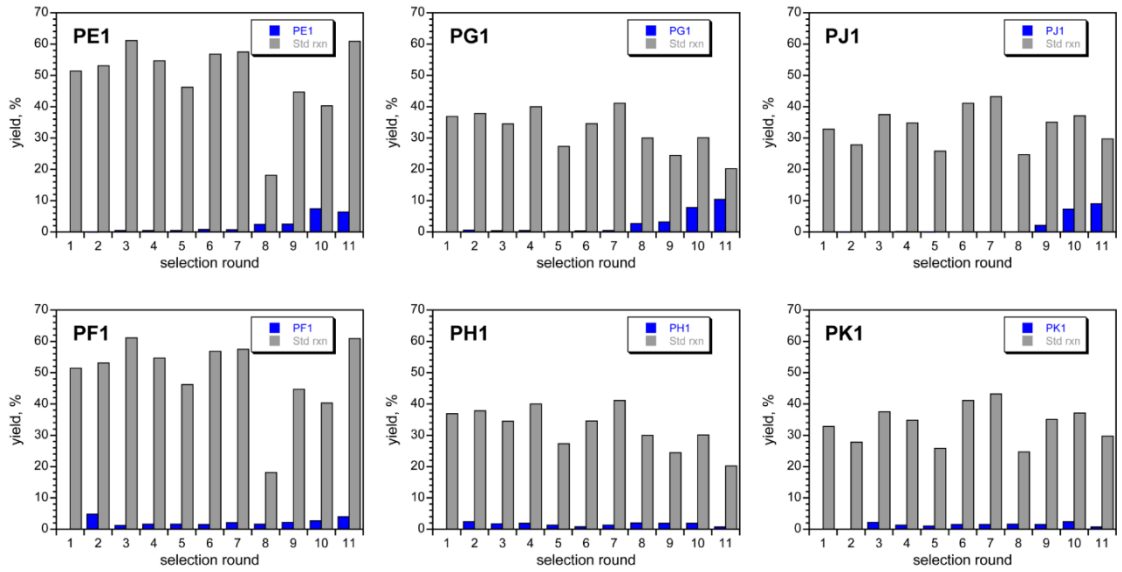


Figure 2.4. Progression data of selections with regular nucleotides DNA. Blue bars show selection activity, and gray bars show the standard capture yield in each round of selection.

2.1.4 In Vitro Selection Design towards DNAzymes with Modified Nucleotides for Amine Acylation

Based on the previous negative results that no DNAzymes composed of unmodified nucleotides were identified for amine acylation, meanwhile DNA sequences with chemically modified nucleotides have extended chemical functionality and are able to catalyze more difficulty reactions.[13] The next logic step was incorporating chemically modified nucleotides in selection experiments to enhance the catalytic ability of DNAzymes for amine acylation, which was also the starting point of whole my amine and peptide lysine acylation project. These new in vitro selections

were simplified by using only one thioester acyl donor electrophile (5'-glutaryl-DNA thioester) and only one length of random DNA pool (N_{40}), due to the introduction of modified nucleotides and extra primer extension step.

The in vitro selections were designed on the same substrates as mentioned above: one thioester substrate (5'-glutaryl-DNA thioester) as electrophile, and two amine substrates (5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA) as nucleophiles (Figure 2.5). The 5'-glutaryl-DNA thioester was utilized rather than glutaryl-CoA in these selections, because previous selection progression data showed DNA-anchored glutaryl thioester had a greater reaction potential than glutaryl-CoA, which will increase the chance of finding active DNAzymes.

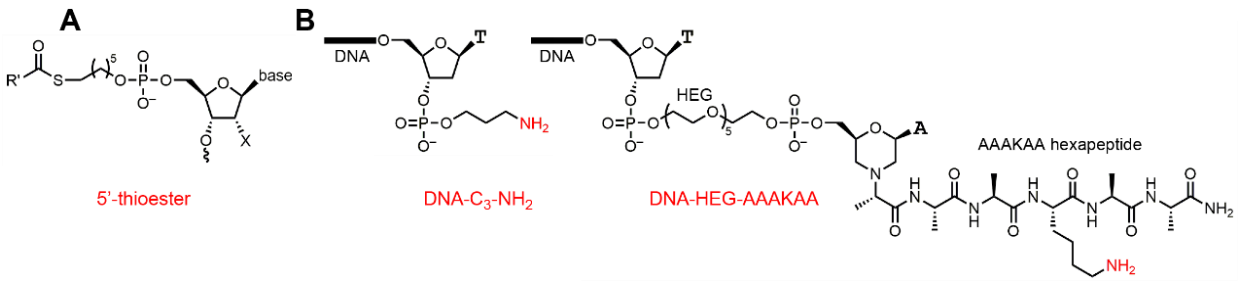


Figure 2.5. Substrates used in selections. (A) 5'-thioester as electrophile. (B) DNA-C_3-NH_2 and DNA-HEG-AAAKAA as nucleophiles.

Four types of modified nucleotide with extended chemical functionality were developed. These groups of primary amino, primary hydroxyl, carboxyl, or imidazolyl are attached to the 5-position of thymidine, thereby forming 5-substituted 2′-deoxyuridine derivatives ^{Am}dU, ^{HO}dU, ^{COOH}dU, or ^{Im}dU (Figure 2.6). These functional groups could potentially function as nucleophiles, general acids/bases, or metal-binding ligands in plausible mechanisms of amide hydrolysis.

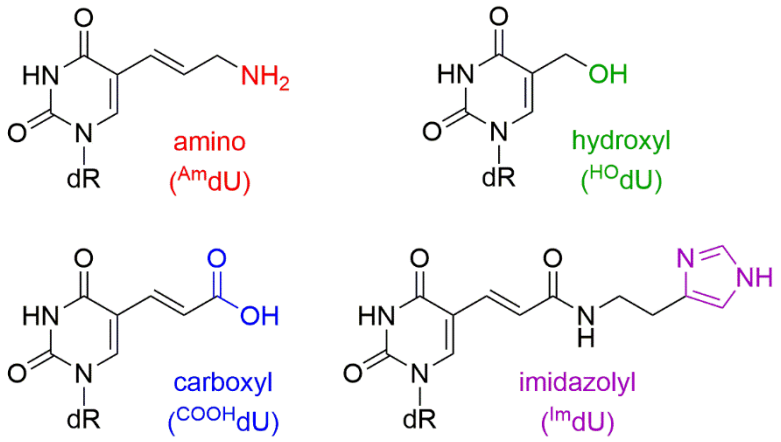


Figure 2.6. Structures of four modified dU with extended chemical functionality (dR = deoxyribose).

To prepare the N_{40} pools for selections using ^{Am}dU, ^{HO}dU, and ^{COOH}dU, the corresponding modified nucleoside phosphoramidites were used in solid-phase oligonucleotide synthesis. For selection using ^{Im}dU, the pool was prepared by primer extension on a reverse complement template N_{40} pool using ^{Im}dUTP and KOD XL DNA polymerase. ^{Am}dU, ^{HO}dU, and ^{COOH}dU phosphoramidite were purchased, and ^{Im}dUTP was synthesized in our lab by previous graduate student Cong Zhou.

The key selection and capture steps for in vitro selection process via using thioester acyl donor electrophile and modified DNA pool are shown in Figure 2.7. After synthesizing the N_{40} DNA pool with modified nucleotides, the ligation step that covalently links the amine substrate to the modified pool was carried out. The subsequent selection step for amine acylation was performed in presence of 5'-glutaryl thioester in 70 mM HEPES, pH 7.5, 1 mM $ZnCl_2$, 20 mM $MnCl_2$, 40 mM $MgCl_2$, 150 mM NaCl at 37 °C for 12-14 h, where Zn^{2+}, Mn^{2+}, and Mg^{2+} are metal ion cofactors required by most deoxyribozymes reactions. These selection experiments used a capture reaction in which the carboxyl group of the glutaryl moiety transferred to the amine in the selection step was subsequently captured using a 5'-amino oligonucleotide, DMT-MM, and a DNA splint. At the end of each selection round, the sequence population was restored by PCR amplification, and the reverse complement DNA strand was isolated. The new modified DNA pool

with enriched catalytic activity was generated by primer extension from the reverse complement strand population using the appropriate modified dUTP and KOD XL DNA polymerase.

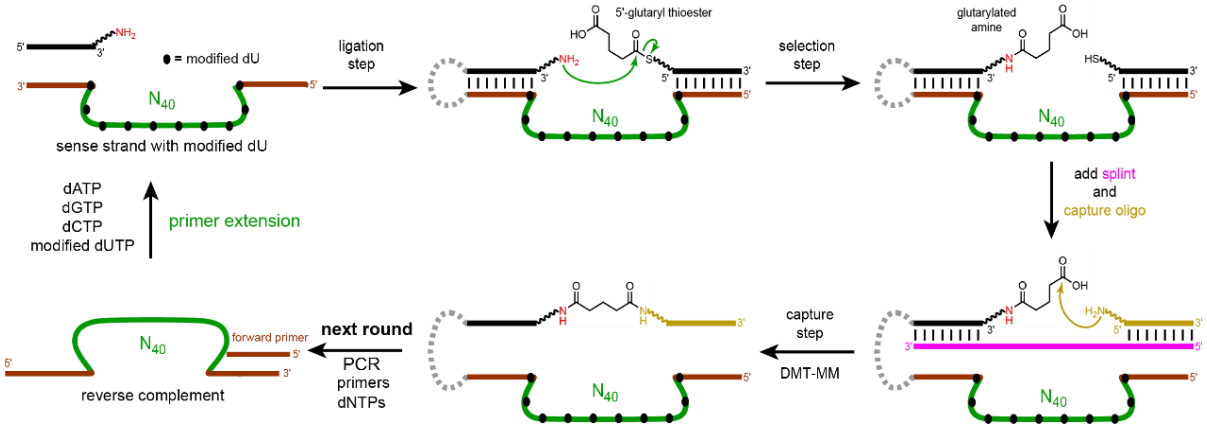


Figure 2.7. Overview of in vitro selection using a thioester acyl donor electrophile.

2.2 Results and Discussion

Eight in vitro selections were performed for 10 rounds (Figure 2.8). Each selection used N_{40} random region DNA pool which contained three canonical nucleotides (dATP, dGTP, dCTP) and only one specific modified nucleotide (modified dUTP), and these eight selections can be divided into two sets. The first set selection (VT1, VV1, VW1, VX1) used DNA-C_3-NH_2 as their nucleophile, meanwhile the second set selections (WA1, WB1, WC1, WD1) used DNA-HEG-AAAKAA as their nucleophile. Both sets used 5'-glutaryl thioester oligonucleotide as acyl donor electrophile.

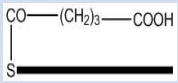
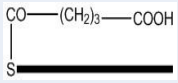
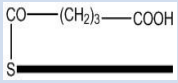
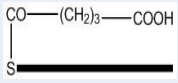
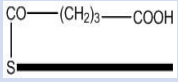
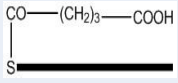
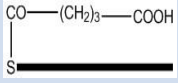
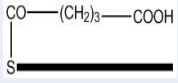
Selection	Modified Pool	1st Substrate	2nd Substrate	Rounds
VT1	^{Am}-dU			10
VV1	^{Im}-dU			10
VW1	^{HOOC}-dU			10
VX1	^{HO}-dU			10
WA1	^{Am}-dU			10
WB1	^{Im}-dU			10
WC1	^{HOOC}-dU			10
WD1	^{HO}-dU			10

Figure 2.8. Overview of selections performed for amine acylation via DNA pools with modified nucleotides. Selection variables of amine nucleophile, thioester acyl donor electrophile, and choice of modified nucleotides are illustrated.

After the selection step, an extra 5'-amino oligonucleotide that serves as mass addition to the acylation product, for which enables a visible gel band shift, was introduced in the following capture reaction. The capture reaction was performed to capture the carboxyl handle within the transferred glutaryl moiety; meanwhile, glutarylated DNA-C_3-NH_2 and glutarylated DNA-HEG-AAAKAA were also assayed as capture standards. Their yields were calculated and shown in Figure 2.9. The capture yield of glutarylated DNA-C_3-NH_2 was 35%-60%, and the capture yield of glutarylated DNA-HEG-AAAKAA was 40%-50%. The capture yield of selections using DNA-C_3-NH_2 was 1%-7%, and the capture yield of selections using DNA-HEG-AAAKAA was 1%-6%.

All data were internally consistent except the yields for VW1 and WC1 selections, which were around 15% and 20%. The explanation for high yield of VW1 and WC1 were: PhosphorImager counts were very low after exposure, and there were massive smear in the gel images, which caused the calculated yield not accurate. Despite the discrepancy between VW1/WC1 and other selections, all selections were subject to in trans assays, without covalently ligating 5'-DNA-C_3-NH_2 or 5'-DNA-HEG-AAAKAA to the DNA sequence pool. Provided that the glutaryl moiety was successfully transferred onto DNA sequence pool, the additional glutaryl moiety changed the DNA sequence charge from +1 (NH^{3+}) to -1 (COO^-). Since DNA runs from positive end to negative end, despite the trivial mass addition, a visible downward band shift (towards negative end) can be seen directly in the gel exposure assay using 20% gel and radiolabeled amine substrates. Thus, no capture reaction were performed for the in trans assays. However, in each instance, no in trans activity was observed.

Figure 2.9. (cont.)

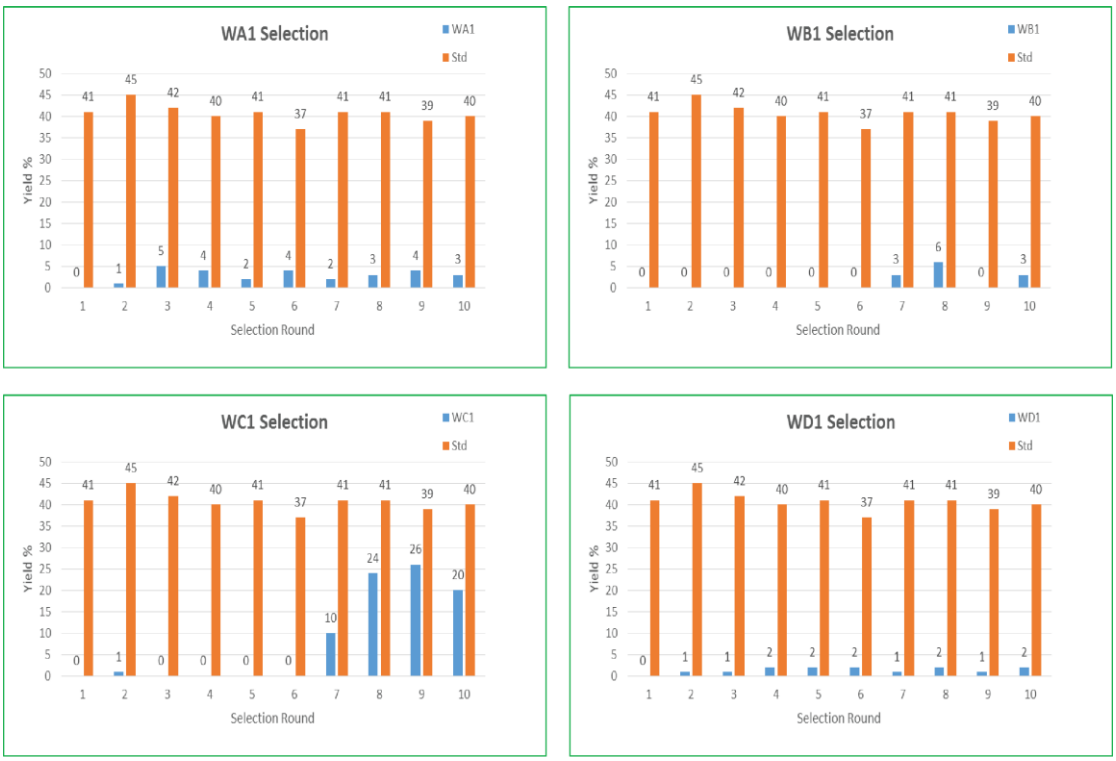


Figure 2.9. Selection progression data with modified DNA pool. Blue bars show selection activity, and orange bars show the standard capture yield in each round of selection.

There are two possible explanations for this outcome. One explanation is that the deoxyribozymes are only active in cis where the DNA-anchored amine is covalently ligated to DNA pool. Perhaps any deoxyribozyme sequence requires the DNA loop that is present during selection for activity, and it loses activity when the loop is not present, such as when the deoxyribozyme is assayed in trans. The other explanation is based on the fact that, with the progression of selection rounds, there was never a significant increase in the capture yield excluding the weirdly high yields of VW1/WC1 selections due to the artificial factors. We concluded that a thioester is insufficiently reactive as an electrophile to allow the emergence of DNAzymes showing activity surpassing the detection threshold. In such case, a more reactive acyl donor electrophile is required.

2.3 Summary

Lysine acylation is involved in many cellular processes including gene expression, cellular metabolism, and direct control of protein function. There is always a desire to identify robust methods to perform site-specific lysine acylation. The investigation of deoxyribozymes that catalyze acylation would provide a valuable approach to studying different acylation states. Acylation selections, either with regular or modified nucleotides, did not show catalytic activity. This is surprising because thioesters are good electrophiles employed by nature in many spontaneous biological reactions. In addition, a small amount of background activity was detected during selections, suggesting that this reaction can occur readily even without catalyst. Despite this, no in trans activity was observed. One possible reason is that the enzymes can work only in cis. This issue could be probed by performing the selections with the loop region base paired by a complementary oligo, which would greatly reduce the possibility of the loop being involved in catalysis. The other possible reason is that the thioester acyl donor electrophile is too unreactive that doesn't allow extent of acylation reaction to surpass the detection threshold. To solve this issue, more reactive acyl donor electrophiles will be investigated, which leads to the next chapter.

2.4 Materials and Methods

2.4.1 Preparation of Oligonucleotides

DNA oligonucleotides and oligonucleotide conjugates were obtained from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an ABI 394 instrument using reagents from Glen Research. All oligonucleotides and conjugates were purified by 7 M urea denaturing PAGE with running buffer 1X TBE (89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3), extracted from the polyacrylamide with TEN buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaCl), and precipitated with ethanol as described previously.[15,16] Hexapeptide AAAK(tfa)AA was prepared by solid-phase synthesis using Fmoc Rink amide MBHA resin as

described. The trifluoroacetyl protecting group was removed after conjugation to the DNA anchor oligonucleotide.

2′-Deoxynucleotide 5′-Triphosphates and 2′-Deoxynucleoside 3′-Phosphoramidites preparation. The ^{Am}dU 5′-triphosphate (^{Am}dUTP) was from TriLink BioTechnologies (cat. no. N-0249). The ^{Am}dU phosphoramidite was from Berry & Associates (cat. no. BA0311). The ^{HO}dU 5′-triphosphate (^{HO}dUTP) was from TriLink BioTechnologies (cat. no. N-0259). The ^{HO}dU phosphoramidite was from Glen Research (cat. no. 10-1093). The ^{COOH}dU phosphoramidite was from Glen Research (cat. no. 10-1035). The ^{COOH}dU 5′-triphosphate (^{COOH}dUTP) and ^{Im}dU 5′-triphosphate (^{Im}dUTP) were synthesized by previous graduate student Cong Zhou as described in previous report.[13]

Conjugation of AAAKAA to DNA anchor oligonucleotide. The DNA anchor oligonucleotide was 5'-GGATAATACGACTCACTAT-HEG-rA-3', where the 3'-terminal ribonucleotide was oxidized by $NaIO_4$ and used for conjugation to the peptide N-terminus. A 100 µL sample containing 1.0 nmol of DNA anchor oligonucleotide in 100 mM HEPES, pH 7.5, and 10 mM $NaIO_4$ was incubated at room temperature for 1 h. The oxidized product was precipitated to remove excess $NaIO_4$ by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The precipitated product was dissolved in 65 µL of water and used directly in the next step. A 100 µL sample containing the $NaIO_4$- oxidized DNA anchor oligonucleotide and 100 nmol (100 equiv) of AAAK(tfa)AA hexapeptide in 100 mM NaOAc, pH 5.2, 50 mM $NiCl_2$, and 10 mM $NaCNBH_3$ was incubated at 37 °C for 14 h. The DNA-anchored hexapeptide was precipitated by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The trifluoroacetyl protecting group was removed by incubation in 100 uL of 30% NH_4OH for 1 h at room temperature. NH_4OH was removed in a speedvac and the conjugation product was purified by 20% PAGE.

Synthesis of DNA-Glutaryl Thioester. The DNA anchor oligonucleotide was 5'-

GGATAATACGACTCACTAT-C3-SSC3-OH-3', where the 5'-disulfide linker was introduced via standard solid-phase DNA synthesis and unmasked to a 5'-thiol by DTT treatment. A 50 µL sample containing 2 nmol of DNA anchor oligonucleotide in 50 mM HEPES, pH 7.5, and 50 mM DTT was incubated at 37 °C for 2 h. The reduced product was precipitated to remove excess DTT by addition of 50 µL of water, 10 µL of 3 M NaCl, and 330 µL of ethanol. The precipitated product (5'-HS-C6-DNA) was dissolved in 35 µL of water and used directly in the next reaction. A 50 µL reaction containing 100 mM NaP buffer pH 7.4 and 60 mM glutaric anhydride was incubated at room temperature for 2 h. Excess salts were removed using a 3kD Amicon column, and the DNA thioester product was purified by reverse-phase HPLC. Collected HPLC peaks were pooled, dried in a speedvac, redissolved in water, and quantified by UV absorbance (A_{260}).

2.4.2 In Vitro Selection Procedure Steps

The selection procedure, cloning, and initial screening of individual clones were performed essentially as described previously,[17] with additional steps to enable incorporation of the modified nucleotides. An overview of the key selection and capture steps of each round is shown in Figure 2.7. The random deoxyribozyme pool was 5'-CGAAGTCGCCATCTCTTC-N_{40}-ATAGTGAGTCGTATTAAGCTGATCCTGATGG -3', where the initially random N_{40} region includes modified nucleotides. Each of the ^{Am}dU, ^{HO}dU and ^{COOH}dU random pools was prepared by solid-phase synthesis, replacing all dT in the N_{40} region with one of ^{Am}dU, ^{HO}dU, or ^{COOH}dU using the corresponding 2'-deoxynucleoside 3'-phosphoramidite. The ^{Im}dU random pool was prepared by primer extension from a reverse complement template pool using synthesized ^{Im}dUTP and KOD XL polymerase. The reverse complement template pool was5'-CCATCAGGATCAGCTTAATACGACTCACTAT-N_{40}-GAAGAGATGGC-GACTTCGAGATCACGTCGATAACAACAACAACAACAACAACAACAAC-3', where the underlined nucleotides are arbitrary sequence included to allow PAGE separation of the primer extension product from the template. Primers were 5'-CGAAGTCGCCATCTCTTC-3' (forward

primer) and 5'-$(AAC)_{18}$XCCATCAGGATCAGCT-3', where X is the HEG spacer to stop Taq polymerase (reverse primer). The $(AAC)_{18}$ tail was especially long to assure PAGE separation of template and ^{Am}dU/ ^{Im}dU-modified product strands. For convenience, the same tail length was used with the ^{HO}dU and ^{COOH}dU modifications, although the more standard $(AAC)_4$ tail can be used instead. In each round, the ligation step to attach the deoxyribozyme pool at its 3'-end with the 5'-end of the amine substrate was performed using a DNA splint and T4 DNA ligase. The DNA splint sequence for ligation was 5'-ATAGTGAGTCGTATTATCCCCATCAGGATCAGCTTAATACGACTCACTAT-3'. T4 DNA ligase was from Thermo Fisher. KOD XL polymerase was made by previous graduate student Shannon Walsh.

Procedure for primer extension to prepare the ^{Im}dU random pool in round 1. A 50 µL sample was prepared containing 250 pmol of the reverse complement template pool (which contains a 3′-segment of arbitrary sequence to allow PAGE separation of the primer extension product from the template), 500 pmol of forward primer, 15 nmol each of dATP, dGTP, dCTP, and ImdUTP, 5 µL of 10X KOD XL polymerase buffer, and 1 µL of 2.5 U/µL KOD XL polymerase. Primer extension was performed by cycling 4 times in a PCR thermocycler according to the following program: 94 °C for 2 min, 4x (94 °C for 2 min, 47 °C for 2 min, 72 °C for 30 min), 72 °C for 30 min. The sample was separated by 8% PAGE to provide the ^{Im}dU random pool for the selection.

Procedure for primer extension to generate the modified DNA pool in subsequent rounds. A 25 µL sample was prepared containing the reverse-complement single strand (which contains a nonamplifiable spacer that stops KOD XL polymerase), 50 pmol of forward primer, 7.5 nmol each of dATP, dGTP, dCTP, and modified ^{X}dUTP (X = Am, HO, COOH, or Im), 20 µCi of α-32P-dCTP (800 Ci/mmol), 2.5 µL of 10x KOD XL polymerase buffer, and 0.5 µL of 2.5 U/µL KOD XL polymerase. Primer extension was performed in a PCR thermocycler according to the following program: 94 °C for 2 min, 47 °C for 2 min, 72 °C for 1 h. The sample was separated by 8% PAGE

to provide the modified DNA pool for the next selection round.

Procedure for ligation step in round 1. A 25 µL sample containing 600 pmol of DNA pool, 750 pmol of DNA splint, and 900 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 3 µL of 10x T4 DNA ligase buffer and 2 µL of 5 U/µL T4 DNA ligase. The sample was incubated at 37 °C for 12 h and purified by 8% PAGE.

Procedure for ligation step in subsequent rounds. A 17 µL sample containing the PCR-amplified DNA pool (~5–10 pmol), 30 pmol of DNA splint, and 50 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 2 µL of 10x T4 DNA ligase buffer and 1 µL of 1 U/µL T4 DNA ligase. The sample was incubated at 37 °C for 12 h and purified by 8% PAGE.

Procedure for selection step in round 1. Each selection experiment was initiated with 200 pmol of the ligated pool. A 20 µL sample containing 200 pmol of ligated pool and 300 pmol of 5'-glutaryl thioester DNA was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 40 µL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl. The Mn^{2+} was added from a 10x stock solution containing 200 mM $MnCl_2$. The Zn^{2+} was added from a 10X stock solution containing 10 mM $ZnCl_2$, 20 mM HNO_3, and 200 mM HEPES at pH 7.5; this stock solution was freshly prepared from a 100x stock of 100 mM $ZnCl_2$ in 200 mM HNO_3. The metal ion stocks were added last to the final sample. The sample was incubated at 37 °C for 12-14 h and separated by 8% PAGE.

Procedure for selection step in subsequent rounds. A 10 µL sample containing the ligated

pool and 30 pmol of 5'-glutaryl thioester DNA was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 20 μL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl. The Mn^{2+} was added from a 10X stock solution containing 200 mM $MnCl_2$. The Zn^{2+} was added from a 10x stock solution containing 10 mM $ZnCl_2$, 20 mM HNO_3, and 200 mM HEPES at pH 7.5; this stock solution was freshly prepared from a 100x stock of 100 mM $ZnCl_2$ in 200 mM HNO_3. The metal ion stocks were added last to the final sample. The sample was incubated at 37 °C for 12-14 h and separated by 8% PAGE.

Procedure for capture step in round 1. A 90 μL sample containing the selection product, 300 pmol of DNA splint, and 400 pmol of 5'-amino capture oligonucleotide was annealed in 100 mM MOPS, pH 7.0, and 1 M NaCl by heating at 95 °C for 3 min and cooling on ice for 5 min. The capture reaction was initiated by bringing the sample to 100 μL total volume containing 100 mM MOPS, pH 7.0, 1 M NaCl, and 50 mM DMT-MM. The sample was incubated at 37 °C for 12 h.

Procedure for capture step in subsequent rounds. A 20 μL sample containing the selection product, 50 pmol of DNA splint, and 100 pmol of 5'-amino capture oligonucleotide was annealed in 100 mM MOPS, pH 7.0, and 1 M NaCl by heating at 95 °C for 3 min and cooling on ice for 5 min. The capture reaction was initiated by bringing the sample to 25 μL total volume containing 100 mM MOPS, pH 7.0, 1 M NaCl, and 50 mM DMT-MM. The sample was incubated at 37 °C for 12 h.

Procedure for PCR. In each selection round, two PCR reactions were performed, 10-cycle PCR followed by 30-cycle PCR. First, a 100 μL sample was prepared containing the PAGE-purified selection product, 200 pmol of forward primer, 50 pmol of reverse primer, 20 nmol of each dNTP, and 10 μL of 10x Taq polymerase buffer [1x = 20 mM Tris-HCl, pH 8.8, 10 mM

$(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, and 0.1% Triton X-100]. This sample was cycled 10 times according to the following PCR program: 94 °C for 2 min, 10x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Taq polymerase was removed by phenol/chloroform extraction. Second, a 50 µL sample was prepared containing 1 µL of the 10-cycle PCR product, 100 pmol of forward primer, 25 pmol of reverse primer, 10 nmol of each dNTP, 20 µCi of α-^{32}P-dCTP (800 Ci/mmol), and 5 µL of 10x Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Samples were separated by 8% PAGE.

2.5 References

(1) Fritz, K. S. Chemical acetylation and deacetylation. In *Sirtuins*; Springer, 2013, pp. 191-201.

(2) Spange, S.; Wagner, T.; Heinzel, T.; Krämer, O. H. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 185-198.

(3) Hirschey, M. D.; Zhao, Y. Metabolic regulation by lysine malonylation, succinylation, and glutarylation. *Mol. Cell. Proteomics* **2015**, *14*, 2308-2315.

(4) Brooks, C. L.; Gu, W. The impact of acetylation and deacetylation on the p53 pathway. *Protein Cell* **2011**, *2*, 456-462.

(5) Ghosh, S.; Dass, J. F. P. Study of pathway cross-talk interactions with NF-κB leading to its activation via ubiquitination or phosphorylation: A brief review. *Gene* **2016**, *584*, 97-109.

(6) Zhang, Q.; Lenardo, M. J.; Baltimore, D. 30 years of NF-κB: a blossoming of relevance to human pathobiology. *Cell* **2017**, *168*, 37-57.

(7) He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.;

Meng, F.; Kelleher, N.; McCafferty, D. G. Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12033-12038.

(8) Dormeyer, W.; Ott, M.; Schnölzer, M. Probing lysine acetylation in proteins: strategies, limitations, and pitfalls of in vitro acetyltransferase assays. *Mol. Cell. Proteomics* **2005**, *4*, 1226-1239.

(9) Yang, Y.-Y.; Ascano, J. M.; Hang, H. C. Bioorthogonal chemical reporters for monitoring protein acetylation. *J. Am. Chem. Soc.* **2010**, *132*, 3640-3641.

(10) Yu, M.; De Carvalho, L. P. S.; Sun, G.; Blanchard, J. S. Activity-based substrate profiling for Gcn5-related N-acetyltransferases: the use of chloroacetyl-coenzyme A to identify protein substrates. *J. Am. Chem. Soc.* **2006**, *128*, 15356-15357.

(11) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. Site‐specific incorporation of methyl‐and acetyl‐lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed.* **2008**, *47*, 6399-6401.

(12) Wang, Z. U.; Wang, Y.-S.; Pai, P.-J.; Russell, W. K.; Russell, D. H.; Liu, W. R. A facile method to synthesize histones with posttranslational modification mimics. *Biochemistry* **2012**, *51*, 5232-5234.

(13) 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*, 2106-2109.

(14) Brandsen, B. M.; Velez, T. E.; Sachdeva, A.; Ibrahim, N. A.; Silverman, S. K. DNA‐Catalyzed Lysine Side Chain Modification. *Angew. Chem. Int. Ed.* **2014**, *53*, 9045-9050.

(15) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. Deoxyribozymes with 2 ‘− 5 ‘RNA ligase activity. *J. Am. Chem. Soc.* **2003**, *125*, 2444-2454.

(16) Wang, Y.; Silverman, S. K. Characterization of deoxyribozymes that synthesize branched RNA. *Biochemistry* **2003**, *42*, 15252-15263.

(17) Langner, J.; Klussmann, S. PCR primers containing stop codons reduce the number of false-negatives during blue-white screening. *BioTechniques* **2003**, *34*, 950-954.

Chapter 3: Efforts towards DNAzymes for Amine Acylation via Highly Reactive Aryl Ester Acyl Donor Electrophiles

3.1 Introduction

As described in Chapter 2, the thioester acyl donor electrophile was too unreactive to allow the activity of amine acylation reaction to surpass the detection threshold. Also inspired by the capture step described in previous chapter, the intermediate ester formed during the capture reaction, which was the DMT ester, was a good candidate as highly reactive acyl donor electrophile. Other suitable acyl donor such as phenol derivative ester was sought and tested as well.

3.1.1 DMT-MM Mechanism and DMT Ester

One common method to activate carboxylic acids forming its corresponding aryl esters is using an amide-forming coupling reagent, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM).[1] The negative charge on the 5'-carboxylic acid oligonucleotide attacks the 2-carbon on the triazine ring and releases the methylmorpholinium group, thus forming the DMT ester. The DMT ester is attacked by the amino oligonucleotide as the nucleophile, and then an amide bond is formed via amine acylation. The overall mechanism is illustrated in Figure 3.1.

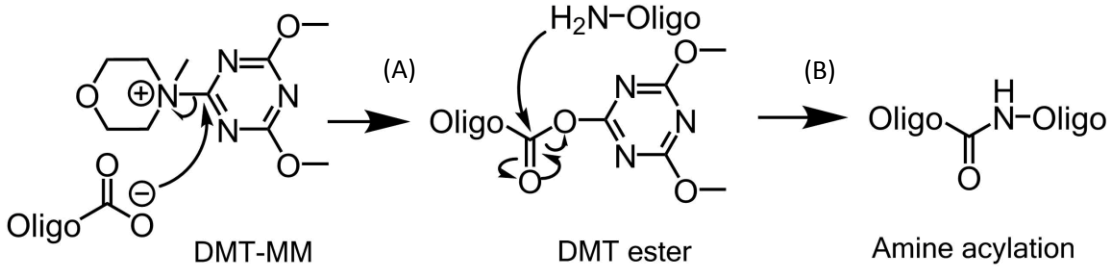


Figure 3.1. (A) The mechanism of DMT-MM activation and formation of DMT ester. (B) The amine acylation reaction using DMT ester.

3.1.2 EDC/TFP Mechanism and TFP Ester

Other common method to activate carboxylic acids forming its corresponding acyl donor oligonucleotide is using the coupling reagents combination of the water-soluble 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 2,3,5,6-tetrafluorophenol (TFP)[2-5]. The negative charge on the 5'-carboxylic acid attacks the carbodiimide group on EDC and forms the intermediate EDC ester that is unstable due to rearrangement into urea byproduct. In the presence of TFP, the TFP attacks the carbonyl group on the unstable EDC ester and replaces the EDC moiety with TFP moiety, thus forming the TFP ester. The TFP ester is a much more stable aryl ester comparing to EDC ester and also a highly reactive acyl donor electrophile, which reacts with amino group readily and forms the amide bond. The overall reaction mechanism is illustrated in Figure 3.2.

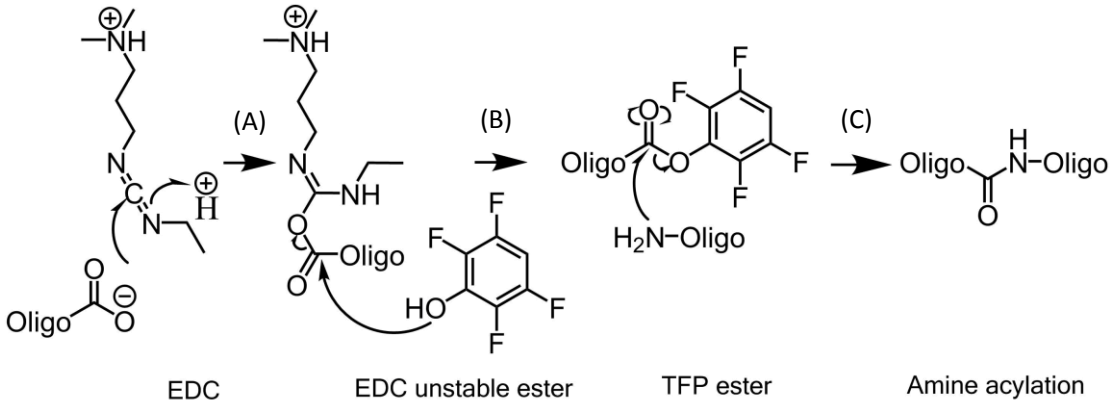


Figure 3.2. (A) The mechanism of EDC activation. (B) The formation of TFP ester. (C) The amine acylation reaction by using TFP ester.

3.1.3 In vitro Selection Design towards DNAzymes for Amine Acylation with in situ Activated Acyl Donors

Eight selections BV2-CD2 were designed based on the two highly reactive aryl ester acyl donors (DMT ester and TFP ester) as described above. Same two amine substrates (5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA) as described in chapter 2 were used as nucleophiles. These in vitro selection experiments used N_{40} initially random regions with unmodified nucleotides rather than modified nucleotides. The key selection steps for in vitro selection BV2-CD2 are shown in Figure 3.3. After synthesizing the N_{40} randomized DNA pool with unmodified nucleotides via solid phase oligo synthesizer, the ligation step that covalently links the amine substrate to the DNA pool was carried out. The 5'-carboxylic acid oligonucleotide was then added into the selection step and activated in situ by DMT-MM or EDC/TFP rather than directly using purified and isolated DMT ester or TFP ester. This key in situ activation and selection step for DMT ester was performed in presence of 50 mM DMT-MM, 100mM MES for pH 6.0 or 100 mM MOPS for pH 7.0, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 2 h; the in situ activation and selection step for TFP ester was performed in presence of 50 mM EDC and 50 mM TFP, 100 mM MES for pH 6.0 or 100 mM MOPS for pH 7.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 2 h. In the selection step, the oligonucleotide part (transferred from the 5'-carboxylic acid oligonucleotide) was conjugated to the amino group on the pool sequence that provided enough mass difference for PAGE isolation, so no additional capture step was needed. At the end of each selection round, the sequence population was restored by PCR amplification using Taq polymerase, and then subjected to subsequent iteration of rounds.

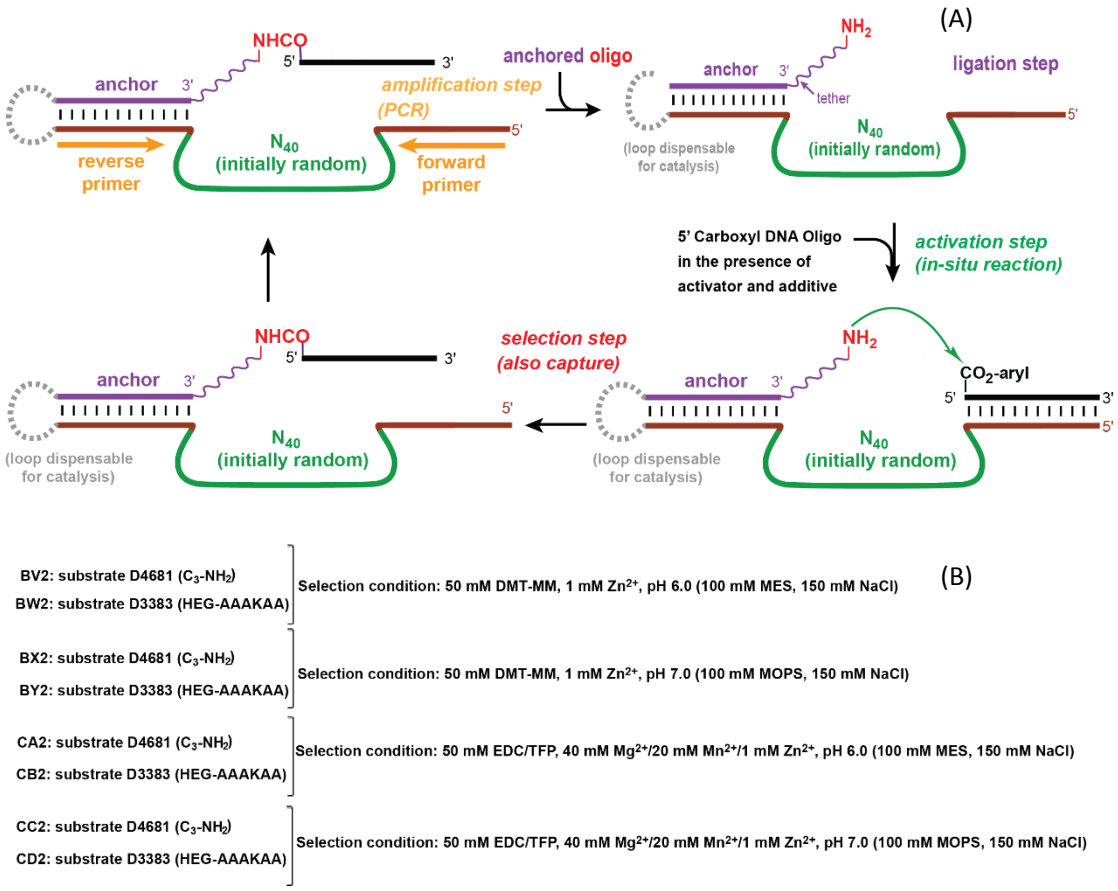


Figure 3.3. (A) Overview of key steps for selections BV2-CD2. (B) Individual incubation condition for selections BV2-CD2.

3.2 Results and Discussion

3.2.1 Splinted Background Reaction for Selections Using in situ Activated DMT Ester and TFP Ester

The overall reaction rate is depending on two steps: first the activation reaction of carboxyl oligo becoming aryl ester, and second the amine acylation reaction. Our interest is focused on the rate enhancement of second reaction, thus the rate of first reaction has to be constant and maximized in order to minimize its contribution to overall rate enhancement. Therefore, activators/additives were assayed under different concentration since the activation reaction rate

was maximized with the saturated activator/additive concentration. The activators/additives were assayed under 2 mM, 10 mM, 50 mM, and 100 mM concentration. The results showed that 50 mM was the saturation concentration for each activator and additive, and this concentration was applied for all the following experiments described in this chapter.

Before performing in vitro selection, the DMT and TFP esters were assayed in situ for their uncatalyzed background reactivities, using a DNA splint complementary to the DNA-anchored acyl donor and the DNA-anchored amine nucleophiles (Figure 3.4). The 5'-DNA-C_3-NH_2 nucleophile was assayed first with different metal ion combination to give a general idea to determine the selection condition (Table 3.1). The results indicated that for 2 hour incubation at 37 °C, the amide formation yield by DMT-MM activation was 55-60% at pH 6.0, and 75-80% at pH 7.0; the amide formation yield by EDC/TFP activation was 15-20% at pH 6.0, and 30-35% at pH 7.0. The 5'-DNA-HEG-AAAKAA nucleophile substrate was then assayed under specific selection condition (Table 3.2). The results indicated that for 2 hour incubation at 37 °C, the amide formation yield by DMT-MM activation was around 40% at pH 6.0, and 65% at pH 7.0; the amide formation yield by EDC/TFP activation was about 15% at pH 6.0, and 40% at pH 7.0.

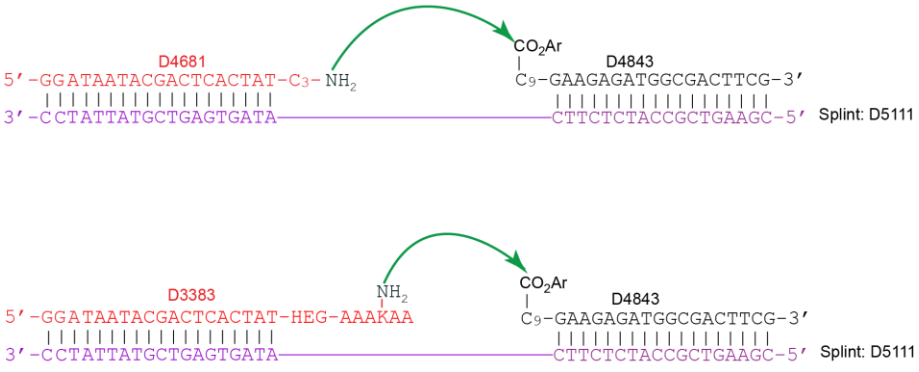


Figure 3.4. Splinted background reaction using 5'-DNA-C_3-NH_2 or 5'-DNA-HEG-AAAKAA substrates with aryl ester acyl donors.

Substrates:5'-DNA-C_3-NH_2, 5'-CO_2H-C_9-DNA Incubation time: 2 h Incubation temp: 37 °C	pH 6.0 (100 mM MES, 150 mM NaCl)				pH 7.0 (100 mM MOPS, 150 mM NaCl)			
Metal ion and activator concentration (mM)	Mg (40)	Mn (20)	Zn (1)	Mg (40) Mn (20) Zn (1)	Mg (40)	Mn (20)	Zn (1)	Mg (40) Mn (20) Zn (1)
DMT-MM (50) activation	58.6%	59.9%	56.3%	ND	76.4%	77.4%	76.3%	ND
EDC/TFP (50) activation	16.8%	16.4%	17.3%	ND	29.4%	33.0%	31.1%	ND

Table 3.1. Splinted background reaction yield using 5'-DNA-C_3-NH_2 and 5'-CO_2H-C_9-DNA. ND means "Not Determined".

Substrates: 5'-DNA-HEG-AAAKAA, 5'-CO_2H-C_9-DNA Incubation time: 2 h Incubation temp: 37 °C	pH 6.0 (100 mM MES, 150 mM NaCl)				pH 7.0 (100 mM MOPS, 150 mM NaCl)			
Metal ion and activator concentration (mM)	Mg (40)	Mn (20)	Zn (1)	Mg (40) Mn (20) Zn (1)	Mg (40)	Mn (20)	Zn (1)	Mg (40) Mn (20) Zn (1)
DMT-MM (50) activation	ND	ND	38.3%	ND	ND	ND	64.5%	ND
EDC/TFP (50) activation	ND	ND	ND	14.7%	ND	ND	ND	39.6%

Table 3.2. Splinted background reaction yield using 5'-DNA-HEG-AAAKAA and 5'-CO_2H-C_9-DNA. ND means "Not Determined".

Both DMT ester and TFP ester led to relatively high uncatalyzed background reactivity, with substantial formation of acylation product. Nevertheless, in vitro selection still had the potential to lead to DNAzymes with rate enhancement above this uncatalyzed background reaction. We therefore proceeded to perform in vitro selection experiments using the in situ activated DMT ester and TFP ester acyl donors with both the 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA substrates.

3.2.2 Selection Progression Data

In vitro selections were performed for 6 rounds (Figure 3.5). The splinted background reaction, which uses both amino substrate nucleophiles to react with in situ activated DMT ester and TFP ester, was employed as position indicator for acylation product. Unlike the capture yield of positive control (capture standard) described in Chapter 2 that caps the maximum yield for the actual selection product, however, the yield of splinted background reaction was totally independent on the actual selection reaction. Since no capture step was involved in selection rounds, the acylation yield in each round was directly correlated to the catalytic activity tendency. Thus, in the selection progression data figures, only the actual selection yields were presented with no comparison to the yields of splinted background reaction. After 4 rounds, all selections reached maximum yield plateau except CB2 selection that reached maximum yield at round 5. For BV2-BY2 selections that used DMT-MM as activator under pH 6.0 or pH 7.0, the highest yield was 37% of BX2 selection at round 4, and 34% of BY2 selection at round 4. For CA2-CD2 selections that use EDC/TFP as activator/additive under pH 6.0 or pH 7.0, the highest yield was 13% of CC2 selections at round 4, and this was the only yield higher than 10% among all four selections using EDC/TFP as activator/additive.

Figure 3.5. Selection progression data for selections BV2-CD2 using in situ activated DMT ester and TFP ester.

3.2.3 DNA Sequences for Selections BX2, BY2, and CC2

For BX2 selection using 5'-DNA-C_3-NH_2 substrate and in situ activated DMT ester at pH 7.0, a maximum 37% yield was reached at round 4. For BY2 selection using 5'-DNA-HEG-AAAKAA substrate and in situ activated DMT ester at pH 7.0, a maximum 34% yield was reached at round 4. For CC2 selection using 5'-DNA-C_3-NH_2 substrate and in situ activated TFP ester at pH 7.0, a maximum 13% yield was reached at round 4. Thus, BX2, BY2, CC2 selections were cloned after round 4 and sequenced. 7 DNA sequences were identified for BX2 selection with 7 unique N_{40} regions; 5 DNA sequences were identified for BY2 selection with 4 unique N_{40} regions and 1 N_{39} region; 6 DNA sequences were identified for CC2 selection with 6 unique N_{40} regions (Figure 3.6).

```
4BX205:   CGAAGTCGCCATCTCTTCCGGCAAGAAGGGTTGTGTTGGGGTATGCCGCCAGCTCGTCATAGTGAGTCGTATTA
4BX206:   CGAAGTCGCCATCTCTTCCGGCGGCAAAGTTATAGCTAGATATAGGAGGCTTCTCTGCATAGTGAGTCGTATTA
4BX214:   CGAAGTCGCCATCTCTTCCTAGGCTTAGTAGGAGTCGCAGAGCTTACAATGGCCGTGAATAGTGAGTCGTATTA
4BX216:   CGAAGTCGCCATCTCTTCGGGAGGAGTATCAGAAGGTCAGGCGGAAATAAAGCGGTTGATAGTGAGTCGTATTA
4BX225:   CGAAGTCGCCATCTCTTCGGCACATTGTTCAACCTACTTCATATATGGGTTACATGGCATAGTGAGTCGTATTA
4BX232:   CGAAGTCGCCATCTCTTCGGCAGGGGGGGCGAAGGTTATAGAACGGGACGATACTGACATAGTGAGTCGTATTA
4BX236:   CGAAGTCGCCATCTCTTCAGGCTGCGAGTGAGGGCTATAGGGAAAGGACACGAGCCCTATAGTGAGTCGTATTA

4BY204:   CGAAGTCGCCATCTCTTCCGGCACAGGGGGGACTCGATAGGGTCGGCACGTCCGTCT ATAGTGAGTCGTATTA
4BY207:   CGAAGTCGCCATCTCTTCCCTAGCGCTCATAAGGCGAAGACATATCGCGGTTGGTAGTATAGTGAGTCGTATTA
4BY210:   CGAAGTCGCCATCTCTTCACCACCGGCATCTAAATCCTGGTAGGTGCGAGTGTTCGTAATAGTGAGTCGTATTA
4BY217:   CGAAGTCGCCATCTCTTCGGGCGCACCAGTAAAAGGTTAGCCGTAAACGGGGGAGACCATAGTGAGTCGTATTA
4BY223:   CGAAGTCGCCATCTCTTCCAGCGCGATACCAGGAGAGAGATGAAGTTGCCGAGCCTTCATAGTGAGTCGTATTA

4CC210:   CGAAGTCGCCATCTCTTCTCACACGAATTAATATGCTTAGCGTTGAACGCCATCTCGCATAGTGAGTCGTATTA
4CC211:   CGAAGTCGCCATCTCTTCCGCGGTTGAAGACGTGTGCGAGGAGAGGGAGCCAACTCTGATAGTGAGTCGTATTA
4CC212:   CGAAGTCGCCATCTCTTCTCAGGACAACGCTCTTGTGACTATCCGAGGGGGGGACTGAATAGTGAGTCGTATTA
4CC214:   CGAAGTCGCCATCTCTTCCGGACCACGGCGCAAAACTGTCTGTTATAGATGGTGGACGATAGTGAGTCGTATTA
4CC216:   CGAAGTCGCCATCTCTTCCGCAATAAGGTTAGACACTGGTAGGTTCTATATTTTAGCGATAGTGAGTCGTATTA
4CC217:   CGAAGTCGCCATCTCTTCGGGCATAGTTCTAAGCATTCGCCATAACGAAGGTGAGCCCATAGTGAGTCGTATTA
```

Figure 3.6. Sequences identified for BX2, BY2 and CC2 selections. Initially randomized N_{40} region are highlighted.

3.2.4 In Trans Partial Kinetic Assays Using 5'-DNA-C_3-NH_2 Amine Substrate with Both in situ Activated Aryl Esters

Among all the 18 sequences identified from BX2, BY2 and CC2 selections, all sequences from BX2 and CC2 selection use 5'-DNA-C_3-NH_2 as parent substrate in reaction, whereas all sequences from BY2 selection use 5'-DNA-HEG-AAAKAA as nucleophile in reaction. Furthermore, certain sequences among all BX2 and CC2 DNAzymes have better acylation yields

according to the preliminary test results when their corresponding PCR products were assayed. Thus, only 6 DNA sequences with the best acylation yield: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, and 4CC214 were assayed with 5'-DNA-C_3-NH_2 substrate and both in situ activated aryl esters by partial kinetic experiments, in which only two time points (0.5 min and 2 h) were checked, to measure their in trans catalytic activity. The splinted background reaction was also carried out using the same amino substrate and aryl ester acyl donors under the same incubation condition.

First, for using DMT ester, at 0.5 min incubation time point, the splinted background reaction had an amine acylation yield around 34%, meanwhile all 6 DNA sequences had an amine acylation yield between 3-15%; at 2 h incubation time point, the splinted background reaction had an amine acylation yield around 77%, meanwhile all 6 DNA sequences had an amine acylation yield between 65-80% (Figure 3.7).

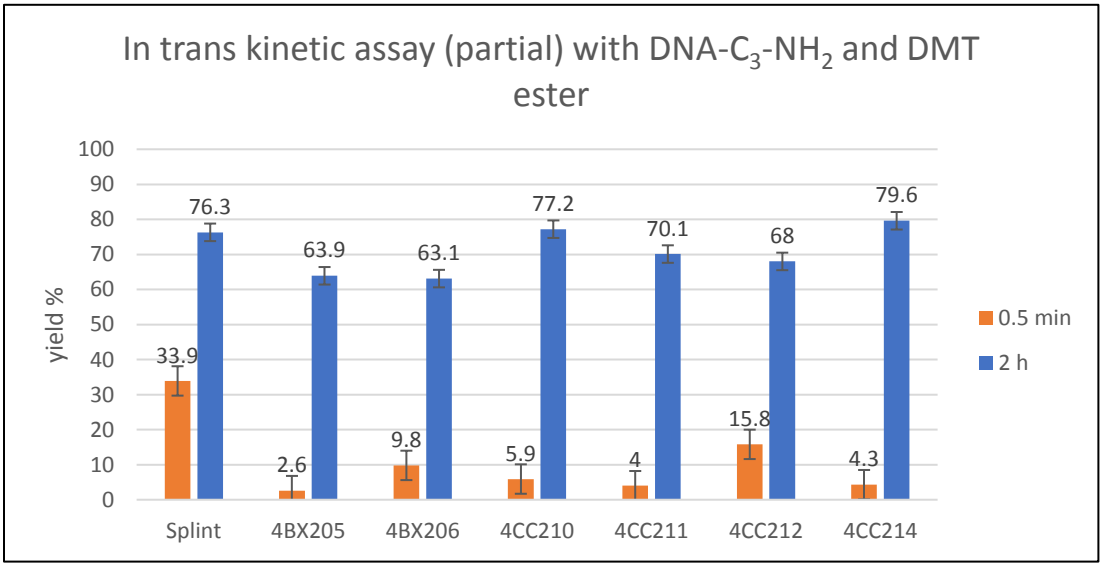


Figure 3.7. In trans kinetic assays for 6 DNA sequences: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, and 4CC214 using 5'-DNA-C_3-NH_2 substrate and in situ activated DMT ester at two time points 0.5 min and 2 h.

Similarly, for using TFP ester, at 0.5 min incubation time point, the splinted background reaction had an amine acylation yield around 15%, meanwhile all 6 DNA sequences had an amine acylation yield between 1-6%; at 2 h incubation time point, the splinted background reaction had an amine acylation yield around 40%, meanwhile all 6 DNA sequences had an amine acylation yield between 3-25% (Figure 3.8).

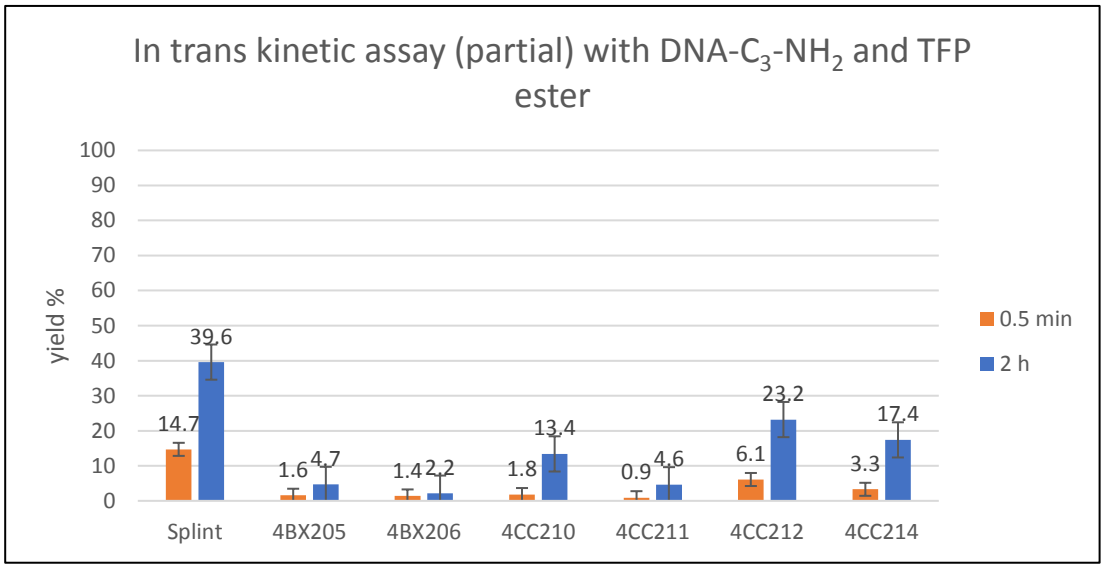


Figure 3.8. In trans kinetic assays for 6 DNA sequences: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, and 4CC214 using 5'-DNA-C_3-NH_2 substrate and in situ activated TFP ester at two time points 0.5 min and 2 h.

To summarize the results that use 5'-DNA-C_3-NH_2 as amino substrate nucleophile and DMT ester or TFP ester as aryl ester acyl donor electrophile, the yields from all 6 DNA sequences cannot exceed the yield of splinted background reaction at either time points (0.5 min or 2 h). These results strongly imply that there is no rate enhancement of identified DNA sequences above the splinted background reaction when using 5'-DNA-C_3-NH_2 as the amino substrate and DMT or TFP ester as acyl donor.

3.2.5 In Trans Partial Kinetic Assays Using 5'-DNA-HEG-AAAKAA Amine Substrate with both in situ Activated Aryl Esters

Based on the results shown above, with addition to the 6 DNAzymes from BX2 and CC2 selections that were already assayed with 5'-DNA-C_3-NH_2 substrates, all 5 DNA sequences from BY2 selections were assayed with their parent amino substrate 5'-DNA-HEG-AAAKAA. Thus, among all the 18 sequences identified from BX2, BY2 and CC2 selections, 11 DNA sequences: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, 4CC214, 4BY204, 4BY207, 4BY210, 4BY217, and 4BY223 were assayed with 5'-DNA-HEG-AAAKAA substrate and both in situ activated aryl esters by partial kinetic experiments, in which only two time points (0.5 min and 2 h) were checked, to measure their in trans catalytic activity. For comparison, the splinted background reaction was performed using the same amino substrate and aryl esters acyl donor.

First, for using DMT ester, at 0.5 min incubation time point, the splinted background reaction had an amine acylation yield around 14%, meanwhile all 11 DNA sequences had an amine acylation yield between 1-7%; at 2 h incubation time point, the splinted background reaction had an amine acylation yield around 57%, meanwhile all 11 DNA sequences had an amine acylation yield between 13-45% (Figure 3.9)

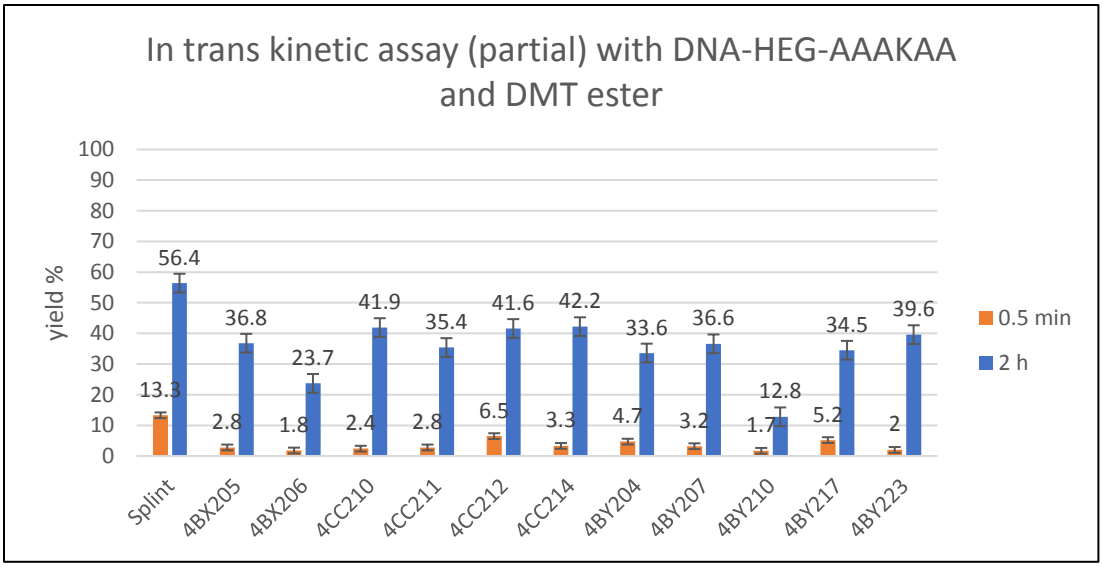


Figure 3.9. In trans kinetic assays for 11 DNA sequences: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, 4CC214, 4BY204, 4BY207, 4BY210, 4BY217, and 4BY223 using 5'-DNA-HEG-AAAKAA substrate and in situ activated DMT ester at two time points 0.5 min and 2 h.

Similarly, for using TFP ester, at 0.5 min incubation time point, the splinted background reaction had an amine acylation yield around 3.4%, meanwhile all 11 DNA sequences had an amine acylation yield about 0%; at 2 h incubation time point, the splinted background reaction had an amine acylation yield around 15%, meanwhile all 11 DNA sequences had an amine acylation yield between 1-10% (Figure 3.10).

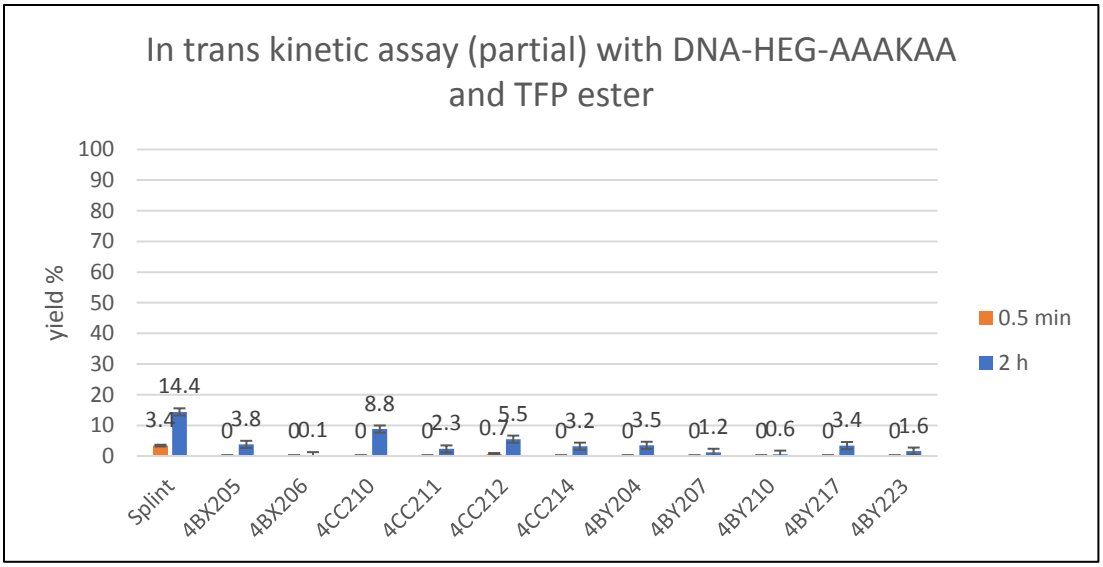


Figure 3.10. In trans kinetic assays for 11 DNA sequences: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, 4CC214, 4BY204, 4BY207, 4BY210, 4BY217, and 4BY223 using 5'-DNA-HEG-AAAKAA substrate and in situ activated TFP ester at two time points 0.5 min and 2 h.

In summary, the outcomes of these assays, which used 5'-DNA-HEG-AAAKAA as amino substrate nucleophile and DMT ester or TFP ester as aryl ester acyl donor electrophile, showed that none of the 11 DNA sequences has a yield surpassing the yield of splinted background reaction at either time points (0.5 min or 2 h). These results again strongly indicate that there is no rate enhancement of identified DNA sequences above the splinted background reaction when using 5'-DNA-HEG-AAAKAA as the amino substrate and DMT or TFP ester as acyl donor.

3.2.6 In Trans Partial Kinetic Assays Using Free Peptide Amine Substrate with Both in situ Activated Aryl Esters

Free peptide AAAKAA is a small molecule amine substrate. Our ultimate goal in this thesis work is trying to select DNAzymes that can acylate small molecule substrate. 11 identified DNA sequences were already assayed with tethered 5'-DNA-HEG-AAAKAA substrate. Even though those DNA sequences did not have a rate enhancement beyond the splinted background reaction, however, they still processed a fairly decent reaction yield when acylating tethered peptide lysine under the DMT-MM activation in 2 hours (yield 13-45%). It was still possible that those DNAzymes could acylate free peptide and have a high rate enhancement beyond the splinted background reaction. Based on this assumption, free peptide substrate was used in partial kinetic assays for 11 identified DNA sequences that were already with tethered peptide. Those 11 DNA sequences: 4BX205, 4BX206, 4CC210, 4CC211, 4CC212, 4CC214, 4BY204, 4BY207, 4BY210, 4BY217, and 4BY223 were assayed with free peptide substrate (AAAKAA) and both in situ activated esters by partial kinetic experiments, in which only two time points (0.5 min and 2 h) were checked, to measure their in trans catalytic activity. The splinted background reaction was also carried out using the same amino substrate and aryl esters acyl donor.

First, for using DMT ester, at 0.5 min incubation time point, the splinted background reaction had an amine acylation yield around 7.4%, meanwhile all 11 DNA sequences had an amine acylation yield between 5-7%; at 2 h incubation time point, the splinted background reaction had an amine acylation yield around 55%, meanwhile all 11 DNA sequences had an amine acylation yield between 45-55%.

Similarly, for using TFP ester, at 0.5 min incubation time point, the splinted background reaction had an amine acylation yield around 1.9%, meanwhile all 11 DNA sequences had an amine acylation yield about 0.5-1.2%; at 2 h incubation time point, the splinted background reaction had

an amine acylation yield around 4.4%, meanwhile all 11 DNA sequences had an amine acylation yield between 2.5-4.1%.

All the results above showed that, for any combination of either amino substrates (5'-DNA-C_3-NH_2 substrate, 5'-DNA-HEG-AAAKAA substrate, or free peptide) paired with either in situ activated aryl esters (DMT ester or TFP ester), the splinted background reaction always presented the highest amine acylation yield at both time points 0.5 min and 2 h. It was logical to assume that the splint background reaction also had a higher acylation yield than any DNA sequence identified at any time point during the incubation period. Thus, we concluded that all DNA sequences identified from selections BX2, BY2 and CC2 have no rate enhancement above the splinted background reaction. One possible explanation is that, the secondary/tertiary structures of DNA sequences when folding during the reaction mimics "the splint", thus the identified DNA sequences from the selections BV2-CD2 basically serve the same role as the splint.

3.3 Summary

For each selection using 5'-DNA-C_3-NH_2 amine substrate, the emergent DNA sequences had no rate enhancement above the uncatalyzed splinted background reaction under the same incubation conditions. We concluded that each individual DNA sequence likely adopts a combination of secondary and tertiary structure that merely recapitulates a complementary splint. Apparently, rate enhancement beyond the splinting effect cannot be achieved because the DMT ester or TFP ester electrophile is too reactive. Similarly, using the considerably less preorganized and presumably less reactive 5'-DNA-HEG-AAAKAA amine substrate did not suppress the too-high background reactivity of the DMT ester and TFP ester acyl donors. The same outcome of finding DNA sequences that have no rate enhancement was found for the DMT ester and TFP ester acyl donors when the amine substrate was instead a free AAAKAA hexapeptide that included a single Lys residue.

The lesson we learned here from Chapter 2 and Chapter 3 is that neither an insufficiently reactive electrophile nor a too reactive electrophile is good for selecting appropriate DNAzymes. Thus, we have to investigate appropriate aryl ester acyl donors with intermediate reactive and optimize their incubation condition. This issue will be discussed and resolved in Chapter 4.

3.4 Methods and Materials

3.4.1 Preparation of Oligonucleotides

DNA oligonucleotides and oligonucleotide conjugates were obtained from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an ABI 394 instrument using reagents from Glen Research. All oligonucleotides and conjugates were purified by 7 M urea denaturing PAGE with running buffer 1X TBE (89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3), extracted from the polyacrylamide with TEN buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaCl), and precipitated with ethanol as described previously.[6,7] Hexapeptide AAAK(tfa)AA was prepared by solid-phase synthesis using Fmoc Rink amide MBHA resin as described. The trifluoroacetyl protecting group was removed after conjugation to the DNA anchor oligonucleotide. The coupling reagents DMT-MM and EDC/TFP were commercially available (Fisher Scientific).

Conjugation of AAAKAA to DNA anchor oligonucleotide. The DNA anchor oligonucleotide was 5'-GGATAATACGACTCACTAT-HEG-rA-3', where the 3'-terminal ribonucleotide was oxidized by $NaIO_4$ and used for conjugation to the peptide N-terminus. A 100 µL sample containing 1.0 nmol of DNA anchor oligonucleotide in 100 mM HEPES, pH 7.5, and 10 mM $NaIO_4$ was incubated at room temperature for 1 h. The oxidized product was precipitated to remove excess $NaIO_4$ by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The precipitated product was dissolved in 65 µL of water and used directly in the next step. A 100 µL sample containing the $NaIO_4$- oxidized DNA anchor oligonucleotide and 100 nmol (100 equiv) of AAAK(tfa)AA

hexapeptide in 100 mM NaOAc, pH 5.2, 50 mM $NiCl_2$, and 10 mM $NaCNBH_3$ was incubated at 37 °C for 14 h. The DNA-anchored hexapeptide was precipitated by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The trifluoroacetyl protecting group was removed by incubation in 100 µL of 30% NH_4OH for 1 h at room temperature. NH_4OH was removed in a speedvac and the conjugation product was purified by 20% PAGE.

Synthesis of 5'-CO_2H-C_9-DNA oligonucleotide. The carboxylic acid DNA oligonucleotide sequence was 5'-CO_2H-C_9-GAAGAGATGGCGACTTCG-3'. The 5'-GAAGAGATGGCGACTTCG-3' was the precursor oligonucleotide, which was prepared by ABI 394 solid-phase synthesizer, for the eventual 5'-CO_2H-C_9-DNA oligonucleotide. The precursor column was dried and stored in -20 °C condition before usage. 5'-Carboxyl-C_{10}-Modifier (Glen Research, 100 µmol) was dissolved in 1 mL anhydrous acetonitrile under the protection of argon. 500 µL of dissolved 5'-Carboxyl-C_{10}-Modifier was mixed with 500 µL activator, and this mixture was applied to the precursor column for 12 min coupling reaction under argon. The mixture was removed, and the precursor column was washed with 1 mL anhydrous acetonitrile 3 times. 1 mL oxidizing reagent was applied to the precursor column for 15 min oxidization reaction under argon. The oxidizing reagent was removed, and the precursor column was washed with 1 mL anhydrous acetonitrile 3 times. 500 µL Cap A reagent was mixed with 500 µL Cap B reagent, and this mixture was applied to the precursor column for 5 min capping reaction under argon. The mixture was removed, and the precursor column was washed with 1 mL anhydrous acetonitrile 3 times, and the column was dried with argon for 5 min. The CPG from precursor column was transferred to a back ring tube, and then 400 µL 0.4 M NaOH in H_2O/MeOH (1:4, v/v) was added into the tube and incubated for 48 hours at room temperature. The black ring tube (with CPG and NaOH mixture inside) was sonicated for 5 min, and the CPG was spun down. The supernatant was transferred to a new clean tube, the CPG was rinsed with 200 µL water, and the rinsed water was combined with the supernatant. The supernatant was ethanol precipitated and finally dissolved in 200 µL water. 50 µL of unpurified 5'-CO_2H-C_9-DNA oligonucleotide solution was taken for HPLC purification

each time. The HPLC is using a Shimadzu Prominence instrument with a Phenomenex Gemini-NX C_{18} column (5 µm, 10 × 250 mm), solvent A (20 mM triethylammonium acetate in 50% acetonitrile/50% water, pH 7.0), solvent B (20 mM triethylammonium acetate in water, pH 7.0), and flow rate of 3.5 mL/min. For 5'-CO_2H-C_9-DNA oligonucleotide, the gradient was 15% solvent A/85% solvent B at 0 min to 30% solvent A/70% solvent B at 45 min, and the elution time was between 24-25 min.

3.4.2 Mass Spectrometry of Oligonucleotides

Data were acquired on a Bruker UltrafleXtreme MALDI-TOF mass spectrometer with matrix 3-hydroxypicolinic acid in positive ion mode at the UIUC School of Chemical Sciences Mass Spectrometry Laboratory. All *m*/*z* values are for $[M+H]^+$. Samples were desalted by Millipore C_{18} ZipTip before analysis.

Data for oligonucleotides and conjugates were as follows. For those that bind to the left-hand DNAzyme binding arm, the DNA sequence was 5'-GGATAATACGACTCACTAT-3'. For those that bind to the right-hand DNA binding arm, the DNA sequence was 5'-GAAGAGATGGCGACTTCG-3'.

Oligonucleotides and conjugates that bind to left-hand DNAzyme binding arm

DNA-C_3-NH_2 *m*/*z* calcd. 5934.9, found 5931.1, Δ= –0.06%

DNA-HEG-AAAKAA *m*/*z* calcd. 6934.6, found 6938.9, Δ= +0.06%

Oligonucleotides and conjugates that bind to right-hand DNAzyme binding arm

5'-CO_2H-C_9-DNA *m*/*z* calcd. 5839.9, found 5840.3, Δ= +0.007%

3.4.3 In vitro selection procedure steps

The key selection step of each round using a 5'-aryl ester oligonucleotide substrate is shown in Figure 3.11. The DNA anchor oligonucleotide sequence was 5'-GGATAATACGACTCACTAT-3'. The 5'-aryl ester oligonucleotide sequence was 5'-GAAGAGATGGCGACTTCG-3'. The random DNAzyme pool was 5'-CGAAGTCGCCATCTCTTC-N_{40}-ATAGTGAGTCGTATTAAGCTGATCCTGATGG -3'. PCR primers were 5'-CGAAGT-CGCCATCTCTTC-3' (forward primer) and 5'-$(AAC)_{18}$XCCATCAGGATCAGCT-3', where X is the HEG spacer to stop Taq polymerase (reverse primer). In each round, the ligation step to attach the DNAzyme pool at its 3'-end with the 5'-end of the amine-containing substrate was performed using a DNA splint and T4 DNA ligase (Thermo Fisher). The splint sequence was 5'-ATAGTGAGTCGTATTATCCCCATCAGGATCAGCTTAATACGACTCACTAT-3'.

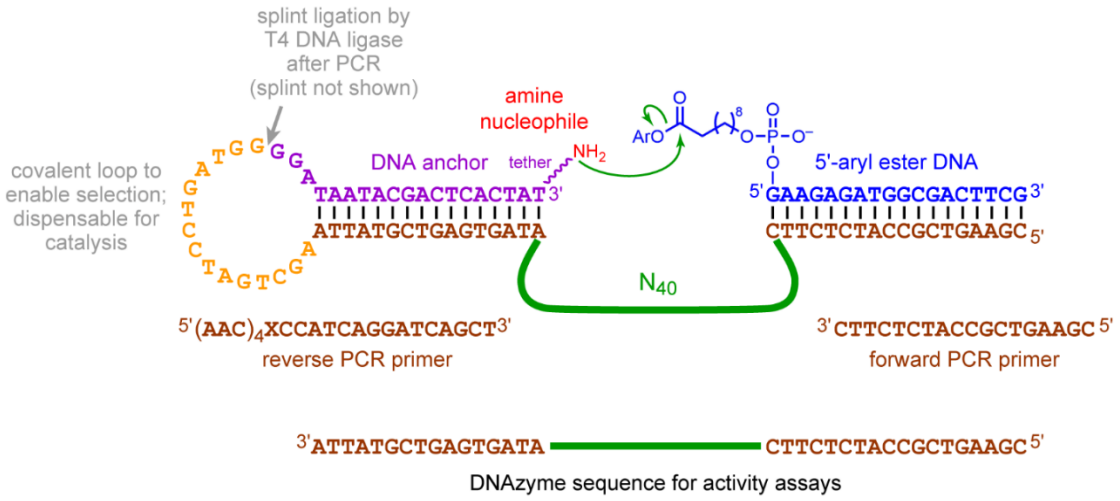


Figure 3.11. Nucleotide details of the in vitro selection experiments with a 5'-aryl ester substrate.

Procedure for ligation step in round 1. A 25 µL sample containing 600 pmol of DNA pool, 750 pmol of DNA splint, and 900 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 3 µL of 10X T4 DNA ligase buffer (400

mM Tris, pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, and 5 mM ATP) and 2 μL of 5 U/μL T4 DNA ligase. The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for ligation step in subsequent rounds. A 17 μL sample containing the PCR-amplified DNA pool (~5–10 pmol), 30 pmol of DNA splint, and 50 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 2 μL of 10x T4 DNA ligase buffer and 1 μL of 1 U/μL T4 DNA ligase. The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for selection step in round 1. Each selection experiment was initiated with 300 pmol of the ligated pool. A 32 μL sample containing 300 pmol of ligated pool and 600 pmol of 5'-CO_2H-C_9-DNA was annealed in 5 mM MES, pH 6.0, 15 mM NaCl, and 0.1 mM EDTA or 5 mM MOPS, pH 7.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 40 μL total volume containing (a) 50 mM DMT-MM, 100 mM MES at pH 6.0 or 100 mM MOPS at pH 7.0, 1 mM $ZnCl_2$, and 150 mM NaCl; (b) 50 mM EDC/TFP, 100 mM MES at pH 6.0 or 100 mM MOPS at pH 7.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$ and 150 mM NaCl. The Mn^{2+} was added from a 10x stock solution containing 200 mM $MnCl_2$. The Zn^{2+} was added from a 10x stock solution containing 10 mM $ZnCl_2$, 20 mM HNO_3, and 200 mM HEPES at pH 7.5; this stock solution was freshly prepared from a 100x stock of 100 mM $ZnCl_2$ in 200 mM HNO_3. The metal ion stocks were added last to the final sample. The sample was incubated at 37 °C for 2 h and separated by 8% PAGE.

Procedure for selection step in subsequent rounds. A 16 μL sample containing the ligated pool and 30 pmol of 5'-CO_2H-C_9-DNA was annealed in 5 mM MES, pH 6.0, 15 mM NaCl, and 0.1 mM EDTA or 5 mM MOPS, pH 7.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for

3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 20 µL total volume containing (a) 50 mM DMT-MM, 100 mM MES at pH 6.0 or 100 mM MOPS at pH 7.0, 1 mM $ZnCl_2$, and 150 mM NaCl; (b) 50 mM EDC/TFP, 100 mM MES at pH 6.0 or 100 mM MOPS at pH 7.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$ and 150 mM NaCl. The sample was incubated at 37 °C for 2 h and separated by 8% PAGE.

Procedure for PCR. In each selection round, two PCR reactions were performed, 10-cycle PCR followed by 30-cycle PCR. First, a 100 µL sample was prepared containing the PAGE-purified selection product, 200 pmol of forward primer, 50 pmol of reverse primer, 20 nmol of each dNTP, and 10 µL of 10X Taq polymerase buffer [1x = 20 mM Tris-HCl, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, and 0.1% Triton X-100]. This sample was cycled 10 times according to the following PCR program: 94 °C for 2 min, 10x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Taq polymerase was removed by phenol/chloroform extraction. Second, a 50 µL sample was prepared containing 1 µL of the 10-cycle PCR product, 100 pmol of forward primer, 25 pmol of reverse primer, 10 nmol of each dNTP, 20 µCi of α-^{32}P-dCTP (800 Ci/mmol), and 5 µL of 10x Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Samples were separated by 8% PAGE.

Cloning and screening of individual DNAzymes. The PCR primers used for cloning were 5'-CGAAGTCGCCATCTCTTC-3' (forward primer; same as in selection) and 5'-TAATTAATTAATTACCCATCAGGATCAGCT-3' (reverse primer). The 10-cycle PCR product from the appropriate selection round was diluted 1000-fold. A 50 µL sample was prepared containing 1 µL of the diluted 10-cycle PCR product from the appropriate selection round, 100 pmol of forward cloning primer, 25 pmol of reverse cloning primer, 10 nmol of each dNTP, and 5 µL of 10x Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30X (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5

min. The sample was separated by 1% agarose gel and extracted using a GeneJET Gel Extraction Kit (Fermentas). The extracted product was quantified by absorbance (A_{260}) and diluted to 5–10 ng/μL. A 4 μL portion of the diluted PCR product was inserted into the pCR2.1-TOPO vector using a TOPO TA cloning kit (Life Technologies). Individual *E. coli* colonies harboring plasmids with inserts were identified by blue-white screening and grown in LB/amp media. Miniprep DNA was prepared using a GeneJET Plasmid Miniprep Kit (Fermentas) and screened for properly sized inserts by EcoRI digestion and agarose gel analysis. Before sequencing, assays of individual DNAzyme clones were performed with PAGE-purified DNA strands prepared by PCR from the miniprep DNA, using the single-turnover assay procedure that will be described in Chapter 4.

3.5 References

(1) Kunishima, M.; Kawachi, C.; Hioki, K.; Terao, K.; Tani, S. Formation of carboxamides by direct condensation of carboxylic acids and amines in alcohols using a new alcohol-and water-soluble condensing agent: DMT-MM. *Tetrahedron* **2001**, *57*, 1551-1558.

(2) Haskali, M. B.; Farnsworth, A. L.; Roselt, P. D.; Hutton, C. A. 4-Nitrophenyl activated esters are superior synthons for indirect radiofluorination of biomolecules. *RSC Med. Chem.* **2020**, *11*, 919-922.

(3) Hui, K. Y.; Holleran, E. M.; Kovacs, J. Protected amino acid tetrafluorophenyl esters for peptide synthesis. *Int. J. Pept. Protein Res.* **1988**, *31*, 205-211.

(4) Lockett, M. R.; Phillips, M. F.; Jarecki, J. L.; Peelen, D.; Lloyd, M. A Tetrafluorophenyl Activated Ester Self-Assembled Monolayer for the Immobilization of Amine-Modified Oligonucleotides. *Langmuir* **2008**, *24*, 69-75.

(5) Pham, G. H.; Ou, W.; Bursulaya, B.; DiDonato, M.; Herath, A.; Jin, Y.; Hao, X.; Loren, J.;

Spraggon, G.; Brock, A. Tuning a Protein-Labeling Reaction to Achieve Highly Site Selective Lysine Conjugation. *ChemBioChem* **2018**, *19*, 799-804.

(6) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. Deoxyribozymes with 2'-5' RNA ligase activity. *J. Am. Chem. Soc.* **2003**, *125*, 2444-2454.

(7) Wang, Y.; Silverman, S. K. Characterization of deoxyribozymes that synthesize branched RNA. *Biochemistry* **2003**, *42*, 15252-15263.

Chapter 4: DNAzymes for Amine and Peptide Lysine Acylation via Intermediate-Reactivity Aryl Ester Acyl Donors*

4.1 Introduction

To this point in our efforts, the observed selection outcomes were divergent. With a thioester acyl donor, no DNAzymes were found due to the insufficiently reactive electrophile. In contrast, with DMT ester or TFP ester as the acyl donors, specific DNA sequences emerged from the selection process, but they lacked rate enhancement beyond a splint because these electrophiles were too reactive. Therefore, we turned our attention to acyl donors with intermediate reactivity.

We introduced two aryl esters in Chapter 3. On one hand, DMT-MM activates a carboxylic acid and forms a multiply substituted aryl ester, which is too reactive to be considered as the acyl ester with intermediate reactivity. On the other hand, TFP forms an aryl ester, the parent compound of which is the simple phenyl ester. There are four fluorine atoms that are strong electron withdrawing groups on the benzene ring of TFP, making TFP ester highly electrophilic. It is logical to assume that the parent compound of TFP ester (the simple phenyl ester) and other phenyl derivative esters with fewer fluorine substitution on the benzene ring have lower electrophilicity, and those esters can be considered as intermediate reactivity electrophiles.

---

*This research has been published:
Yao, T.; Przybyla, J. J.; Yeh, P.; Woodard, A. M.; Nilsson, H. J.; Brandsen, B. M.; Silverman, S. K. DNAzymes for amine and peptide lysine acylation. *Org. Biomol. Chem.* **2021**, *19*, 171-181.
University of Illinois graduate student Peter Yeh synthesized glutaryl-azide small molecule compound.
University of Illinois graduate student Jack Przybyla performed kinetic assays for amine glutarylation with identified DNAzymes.

We therefore investigated the phenyl ester (PE), 4-fluorophenyl ester (4FPE), and 3,5-difluorophenyl ester (35diFPE) acyl donor substrates. For using these aryl esters, we decided to synthesize and purify each new 5'-aryl ester oligonucleotide substrate rather than relying upon in situ activation like we did in Chapter 3. The PE, 4FPE, and 35diFPE oligonucleotide substrates were synthesized from the 5'-CO_2H oligonucleotide, EDC, and the appropriate phenol derivative, followed by HPLC purification.

4.1.1 Phenyl Ester, 4-fluorophenyl Ester, and 3,5-difluorophenyl Ester Formation and Mechanism

The overall mechanism of phenyl ester formation is very similar to TFP ester formation (Figure 4.1A). The negative charge on the 5'-carboxylic acid oligo attacks the carbodiimide group on EDC and forms the intermediate EDC ester that is unstable due to rearrangement into urea byproduct. In the presence of phenol, the phenol attacks the carbonyl group on the unstable EDC ester and replaces the EDC moiety with phenyl moiety, thus forming the phenyl ester. The phenyl ester is a much more stable aryl ester compared to EDC ester, and it is also much less reactive (electrophilic) compared to TFP ester. The mechanism of 4-fluorophenyl ester and 3,5-difluorophenyl ester formation can be described in the same procedures as the phenyl ester. The structure of all three aryl esters are shown in Figure 4.1B.

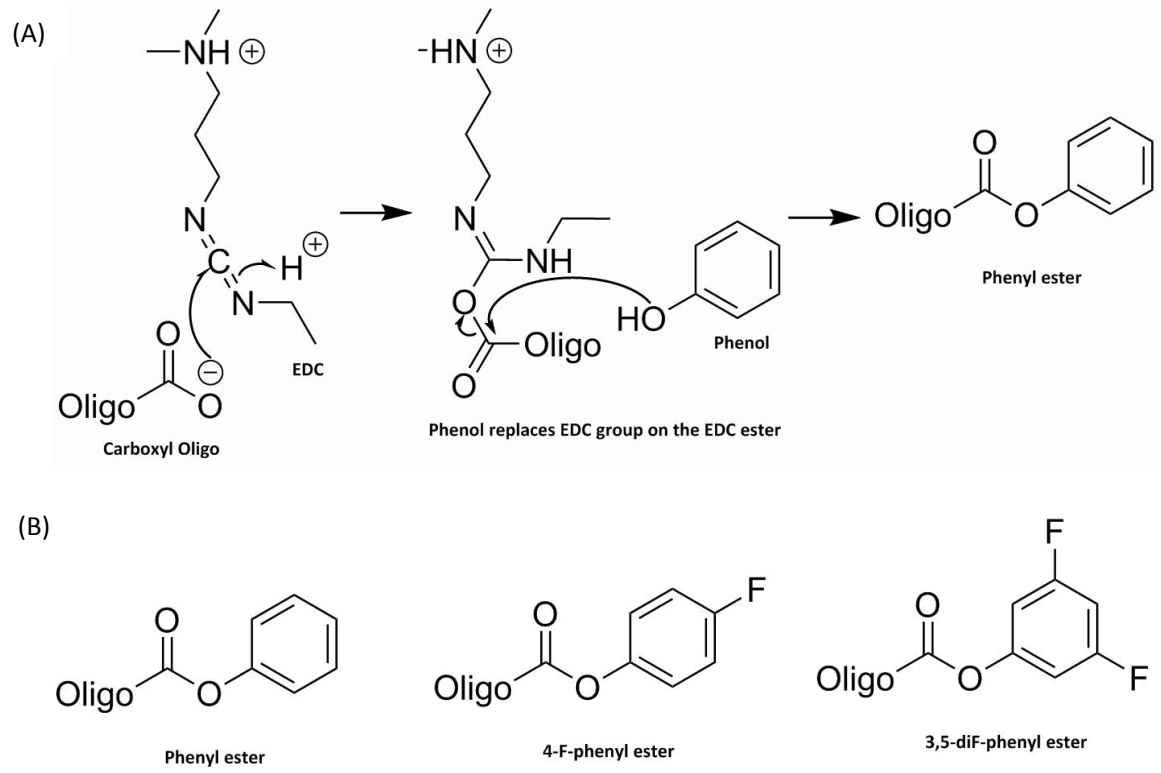


Figure 4.1. (A) Mechanism of phenyl ester formation. (B) Three aryl esters with intermediate reactivity.

4.1.2 In Vitro Selection Design towards DNAzymes for Simple Amine and Peptide Lysine Acylation Using PE, 4FPE and 35diFPE as Acyl Donors

Six selections FL2-FR2 were designed based on the three aryl esters PE, 4FPE, and 35diFPE as electrophiles and 5'-DNA-C_3-NH_2 as nucleophiles. Seven selections HA2-HG2 were designed based on the three aryl esters PE, 4FPE, and 35diFPE as electrophiles and 5'-DNA-HEG-AAAKAA as nucleophiles. These in vitro selection experiments used N_{40} initially random regions with unmodified nucleotides rather than modified nucleotides. The two sets of selections were very similar in every step except they used different amino substrates as nucleophiles and they had different incubation conditions. The key selection steps for two sets in vitro selections FL2-FR2 and HA2-HG2, and their individual incubation conditions are shown in Figure 4.2.

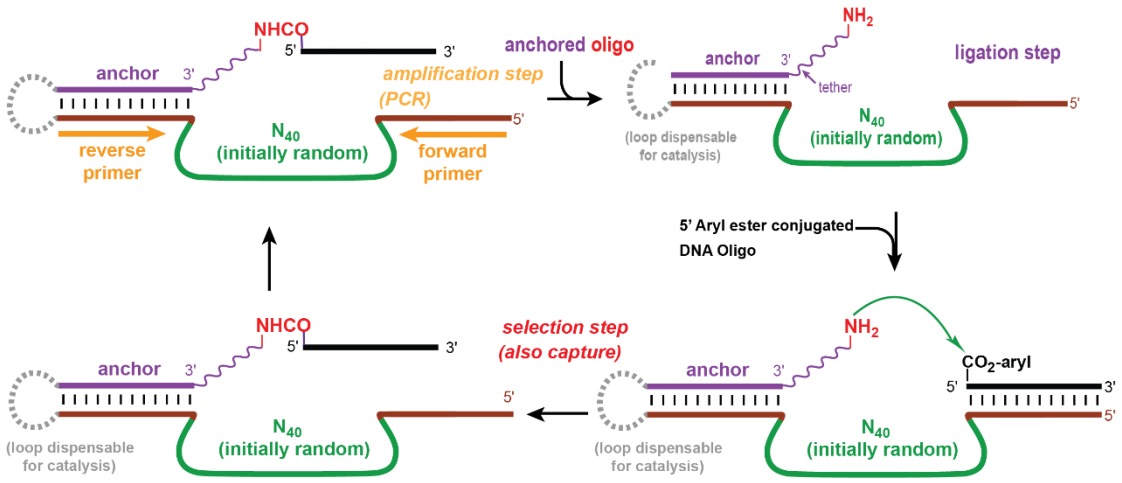
After synthesizing the N_{40} randomized DNA pool with unmodified nucleotides via solid phase oligo synthesizer, the ligation step that covalently links the amine substrate to the DNA pool was carried out. The purified 5'-aryl ester conjugated oligonucleotide was then added into the selection step.

For selections using 5'-DNA-C_3-NH_2 as nucleophile, the key selection step was performed in presence of 50 mM MES pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h (35diFPE); or 50 mM HEPES pH 7.5, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 2 h (35diFPE); or 70 mM HEPES pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h (PE and 4FPE); or 50mM CHES pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h (PE and 4FPE).

For selections using 5'-DNA-HEG-AAAKAA as nucleophile, the key selection step was performed in presence of 50 mM MES pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h (35diFPE); or 70 mM MES pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h (PE and 4FPE); or 70 mM HEPES pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h (PE and 4FPE); or 50mM CHES pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h (PE and 4FPE).

In the selection step, the oligonucleotide part (transferred from the 5'-aryl ester oligonucleotide) was conjugated to the amino group on the pool sequence that provided enough mass difference for PAGE isolation, so no additional capture step was needed. At the end of each selection round, the sequence population was restored by PCR amplification using Taq polymerase, and then subjected to subsequent iteration of rounds.

(A)



(B)

FL2: phenol ester	pH 7.5 (70 mM HEPES, 150 mM NaCl), 40 mM Mg^{2+}/20 mM Mn^{2+}/1 mM Zn^{2+}, 16 h
FM2: phenol ester	pH 9.0 (50 mM CHES, 150 mM NaCl), 40 mM Mg^{2+}, 16 h
FN2: 4-F-phenol ester	pH 7.5 (70 mM HEPES, 150 mM NaCl), 40 mM Mg^{2+}/20 mM Mn^{2+}/1 mM Zn^{2+}, 16 h
FP2: 4-F-phenol ester	pH 9.0 (50 mM CHES, 150 mM NaCl), 40 mM Mg^{2+}, 16 h
FQ2: 3,5-diF-phenol ester	pH 6.0 (50 mM MES, 150 mM NaCl), 40 mM Mg^{2+}, 16 h
FR2: 3,5-diF-phenol ester	pH 7.5 (50 mM HEPES, 150 mM NaCl), 40 mM Mg^{2+}, 2 h

(C)

HA2: phenyl ester	pH 6.0 (70 mM MES, 150 mM NaCl), 40 mM Mg^{2+}/20 mM Mn^{2+}/1 mM Zn^{2+}
HB2: phenyl ester	pH 7.5 (70 mM HEPES, 150 mM NaCl), 40 mM Mg^{2+}/20 mM Mn^{2+}/1 mM Zn^{2+}
HC2: phenyl ester	pH 9.0 (50 mM CHES, 150 mM NaCl), 40 mM Mg^{2+}
HD2: 4-F-phenyl ester	pH 6.0 (70 mM MES, 150 mM NaCl), 40 mM Mg^{2+}/20 mM Mn^{2+}/1 mM Zn^{2+}
HE2: 4-F-phenyl ester	pH 7.5 (70 mM HEPES, 150 mM NaCl), 40 mM Mg^{2+}/20 mM Mn^{2+}/1 mM Zn^{2+}
HF2: 4-F-phenyl ester	pH 9.0 (50 mM CHES, 150 mM NaCl), 40 mM Mg^{2+}
HG2: 3,5-diF-phenyl ester	pH 6.0 (50 mM MES, 150 mM NaCl), 40 mM Mg^{2+}

Figure 4.2. (A) Overview of key steps for selections FL2-FR2 and HA2-HG2. (B) Individual incubation condition for selections FL2-FR2. (C) Individual incubation condition for selections HA2-HG2.

4.2 Results and Discussion

4.2.1 Splinted Background Assays to Optimize the Selection Condition for Both Simple Amine Substrate 5'-DNA-C_3-NH_2 and Tethered Peptide Substrate 5'-DNA-HEG-AAAKAA

4.2.1.1 Splinted Background Reactions Using 5'-DNA-C_3-NH_2

The same two amine substrates (5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA) as described in Chapter 3 were used as nucleophiles in this stage of thesis work. With the PE, 4FPE and 35diFPE 5'-aryl ester oligonucleotide substrates in hand, we first evaluated their hydrolytic stabilities and uncatalyzed, splinted background reactivities with the 5'-DNA-C_3-NH_2 substrate under likely incubation conditions for in vitro selection. Each of Mg^{2+}, Mn^{2+}, and Zn^{2+} was included at pH 7.5 on the basis of our many prior successful DNAzyme selection efforts using these metal ions at pH 7.5. However, at pH 9.0 only Mg^{2+} can be included, because Mn^{2+} oxidizes and Zn^{2+} precipitates at this higher pH. The result of hydrolytic stability and splinted background reaction yields of PE and 4FPE is shown in Table 4.1.

Entry[a]	Substrate	pH	Metal ions	Time	Intact acyl donor, %[b]	Uncorrected background yield, %[c]
1	PE	7.5	Mg^{2+}/Mn^{2+}/Zn^{2+}	2 h	92.4 ± 1.5	0.1
2	PE	7.5	Mg^{2+}/Mn^{2+}/Zn^{2+}	16 h	81.5 ± 0.2	0.3
3	PE	9.0	Mg^{2+}	2 h	94.2 ± 0.7	0.7
4	PE	9.0	Mg^{2+}	16 h	80.4 ± 5.1	2.4
5	4FPE	7.5	Mg^{2+}/Mn^{2+}/Zn^{2+}	2 h	93.0 ± 0.3	0.02
6	4FPE	7.5	Mg^{2+}/Mn^{2+}/Zn^{2+}	16 h	77.7 ± 0.3	0.6
7	4FPE	9.0	Mg^{2+}	2 h	93.3 ± 0.2	0.5
8	4FPE	9.0	Mg^{2+}	16 h	74.1 ± 4.5	3.5

Table 4.1. Determining the optimal incubation conditions for the PE and 4FPE acyl donors, (cont.)

Table 4.1. (cont.) considering both the stability toward hydrolysis and the uncatalyzed, splinted background reactivity with the DNA-C_3-NH_2 substrate. [a] Incubation conditions: 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C, or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C. In vitro selection experiments were subsequently performed using each of the two acyl donors at each of the two pH values, with incubation time of 16 h (i.e., all four even-numbered entries). [b] Intact acyl donor values were determined by PAGE. Each value is the mean of two or three independent assays, with the error as standard deviation (n = 3; entries 4 and 8) or half of the range (n = 2; all other entries). [c] Uncorrected background yield % values were determined by PAGE using the general assay procedure and a complementary DNA splint rather than a DNAzyme. Each value is the mean of n = 3–6 independent assays (n = 2 for entries 5 and 7), with standard deviation (or half of range) ≤ 0.2%.

The hydrolytic stability and splinted background reaction yield of PE and 4FPE were assayed by PAGE; meanwhile, the hydrolytic stability of 35diFPE was assayed by HPLC and its splinted background reaction yield was assayed by PAGE (Table 4.2).

Entry[a]	Substrate	pH	Metal ions	Time	Intact acyl donor, %[b]	Uncorrected background yield, %[c]
1	35diFPE	6.0	Mg^{2+}	2 h	45	0.5
2	35diFPE	6.0	Mg^{2+}	16 h	30	0.8
3	35diFPE	7.5	Mg^{2+}	2 h	29	1.6
4	35diFPE	9.0	Mg^{2+}	2 h	0	13.3

Table 4.2. Determining the optimal incubation conditions for the 35diFPE acyl donor, considering both the stability toward hydrolysis and the uncatalyzed, splinted background reactivity with the DNA-C_3-NH_2 substrate. [a] Incubation conditions: 50 mM MES, pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 2 h or 16 h; or 50 mM HEPES, pH 7.5, 40 mM $MgCl_2$, and 150 mM (cont.)

Table 4.2. (cont.) NaCl at 37 °C for 2 h; or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 2 h. [b] Intact acyl donor values were determined by HPLC. [c] Uncorrected background yield % values were determined by PAGE using the general assay procedure and a complementary DNA splint rather than a DNAzyme.

Based on the data shown, the optimal selection conditions for using 5'-DNA-C_3-NH_2 substrate as nucleophile and PE or 4FPE or 35FPE as electrophile were determined as follows:

For PE selection, 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h, or 70 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h;

For 4FPE selection, 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h, or 70 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h;

For 35diFPE selection, 50 mM MES, pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h, or 50 mM HEPES, pH 7.5, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 2 h. Based on data in the Table 4.2, there is no selection for 35diFPE at pH 9.0, because 35diFPE is very unstable at pH 9.0 that 100% of 35diFPE hydrolyzed within 2 hours incubation time. There was not enough aryl ester acyl donor electrophile during the complete incubation period, for which will interfere the amine acylation reaction rate. Thus, the pH 9.0 condition for 35diFPE was ruled out.

4.2.1.2 Splinted Background Reactions Using 5'-DNA-HEG-AAAKAA

We then evaluated the hydrolytic stabilities and the uncatalyzed splinted background reactivity of PE, 4FPE and 35diFPE with the 5'-DNA-HEG-AAAKAA substrate under likely incubation conditions for in vitro selection. The hydrolytic stability data of three aryl esters can

also apply to the selection condition using 5'-DNA-HEG-AAAKAA substrate as nucleophile.

There was only one selection condition, which is 50 mM MES, pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h, for using 35diFPE as electrophile. Thus, the splinted background reaction using 35diFPE was assayed directly under the same incubation condition. There were six selection conditions for using PE and 4FPE; however, not all of them needed to be assayed for splinted background reactivity.

The splinted background reaction data showed that the background reactivity is lower at lower pH compared to the reactivity at higher pH. It is logical to assume that if the splinted background reaction of these two aryl esters (PE and 4FPE) with 5'-DNA-HEG-AAAKAA were assayed under the highest possible pH and longest possible incubation time, aka the extreme incubation condition, the resulting background yields would provide us a higher limit to optimize the selection condition. Thus, the splinted background assays for PE, 4FPE and 35diFPE were much simplified and shown in Table 4.3.

Entry[a]	Substrate	pH	Metal ions	Time	Intact acyl donor, %[b]	Uncorrected background yield, %[c]
1	PE	9.0	Mg^{2+}	16 h	80.4 ± 5.1	2.2
2	4FPE	9.0	Mg^{2+}	16 h	74.1 ± 4.5	3.2
3	35diFPE	6.0	Mg^{2+}	16 h	30	0.2

Table 4.3. Determining the optimal incubation conditions for the PE, 4FPE, and 35diFPE acyl donors, considering both the stability toward hydrolysis and the uncatalyzed, splinted background reactivity with the 5'-DNA-HEG-AAAKAA substrate. [a] Incubation conditions: 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h; or 50 mM MES, pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h. [b] Intact acyl donor values were determined by PAGE (PE, 4FPE) or HPLC (35diFPE). [c] Uncorrected background yield % values were determined by PAGE using the general assay procedure and a complementary DNA splint rather than a DNAzyme.

For PE and 4FPE substrates, the listed incubation conditions are the higher limit of possible incubation conditions with the highest possible pH and longest possible incubation time. The yields of splinted background reaction under these listed incubation conditions represent the highest possible splinted background yields for using PE and 4FPE as acyl donor electrophiles. The fact that all splinted background reactivities were within the reasonable range (0.1%-5%) indicates that not only these listed incubation conditions but also milder conditions (e.g. lower pH) can be applied as suitable selection conditions in the actual selection rounds. Thus, all six selection conditions for PE and 4FPE and one selection condition for 35diFPE were validated.

4.2.2 Selections with Simple Amine Substrate (5'-DNA-C_3-NH_2)

4.2.2.1 Selection Progression Data

Based on the data, two particular incubation conditions were chosen for PE and 4FPE acyl donor substrates: lower pH of 7.5 (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h) and higher pH of 9.0 (50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h). Two incubation conditions were chosen for 35diFPE acyl donor substrate as well: lower pH of 6.0 (50 mM MES, pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h) and higher pH of 7.5 (50 mM HEPES, pH 7.5, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h). The 35diFPE was totally hydrolyzed even under pH 6.0 and pH 7.5 condition if more metal ions other than Mg^{2+} was presented, thus only Mg^{2+} was applied to pH 6.0 and pH 7.5 condition in 35diFPE selections.

Six selections FL2-FR2 were designed based on the three aryl esters PE, 4FPE, and 35diFPE as electrophiles and 5'-DNA-C_3-NH_2 as nucleophiles, and the in vitro selections were performed for 7 or 8 rounds (Figure 4.3). The splinted background reaction was employed as the position indicator for acylation product. Unlike the capture yield of capture standard reaction described in Chapter 2 that caps the maximum yield for the actual selection product, however, the yield of

splinted background reaction was totally independent on the actual selection reaction. Since no capture step was involved in selection rounds according to the selection design, the acylation yield in each round was directly correlated to the catalytic activity tendency. Thus, in the selection progression data figures, only the actual selection yields were presented with no comparison to the yields of splinted background reaction.

Figure 4.3. Selection progression data for selections FL2-FR2 using PE, 4FPE, and 35diFPE as electrophiles and 5'-DNA-C_3-NH_2 as nucleophile.

After 4 rounds, FM2 and FP2 selection reached maximum yield plateau about 5% and 7%. The 30-cycle PCR products of 4FM2 and 4FP2 were assayed in trans and their in trans activity was confirmed; thus, 4FM2 and 4FP2 were cloned and sequenced. After 7 rounds, FN2 selection reached maximum yield plateau about 18%; after 8 rounds, FL2 and FQ2 selection reached maximum yield plateau about 19% and 20%. The 30-cycle PCR product of 8FL2, 7FN2 and 8FQ2 were tested in trans, and the in trans activity of 8FL2 and 7FN2 was confirmed. However, there was no in trans activity showing up for 8FQ2 selection. Thus, only 8FL2 and 7FN2 selections were cloned and assayed and 8FQ2 selection was set aside. After 8 rounds, there was no yield increase for FR2 selection; therefore, FR2 selection was stopped after round 8. In summary, 4 out of 6 selections (4FM2, 4FP2, 8FL2 and 7FN2) had the in trans activity and were cloned and sequenced.

One possible explanation for 8FQ2 selection only showing the in cis activity but no in trans activity was that, misfolded bands aberrantly migrated to the "correct" position and being collected and amplified. These aberrantly migrated bands did not have the actual activity, thus no in trans activity. Another possible reason for no in trans activity was that, the in cis activity was loop-dependent. Thus, no activity was observed since no loop was present when assayed in trans.

4.2.2.2 DNAzymes Sequences from Selections with 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA

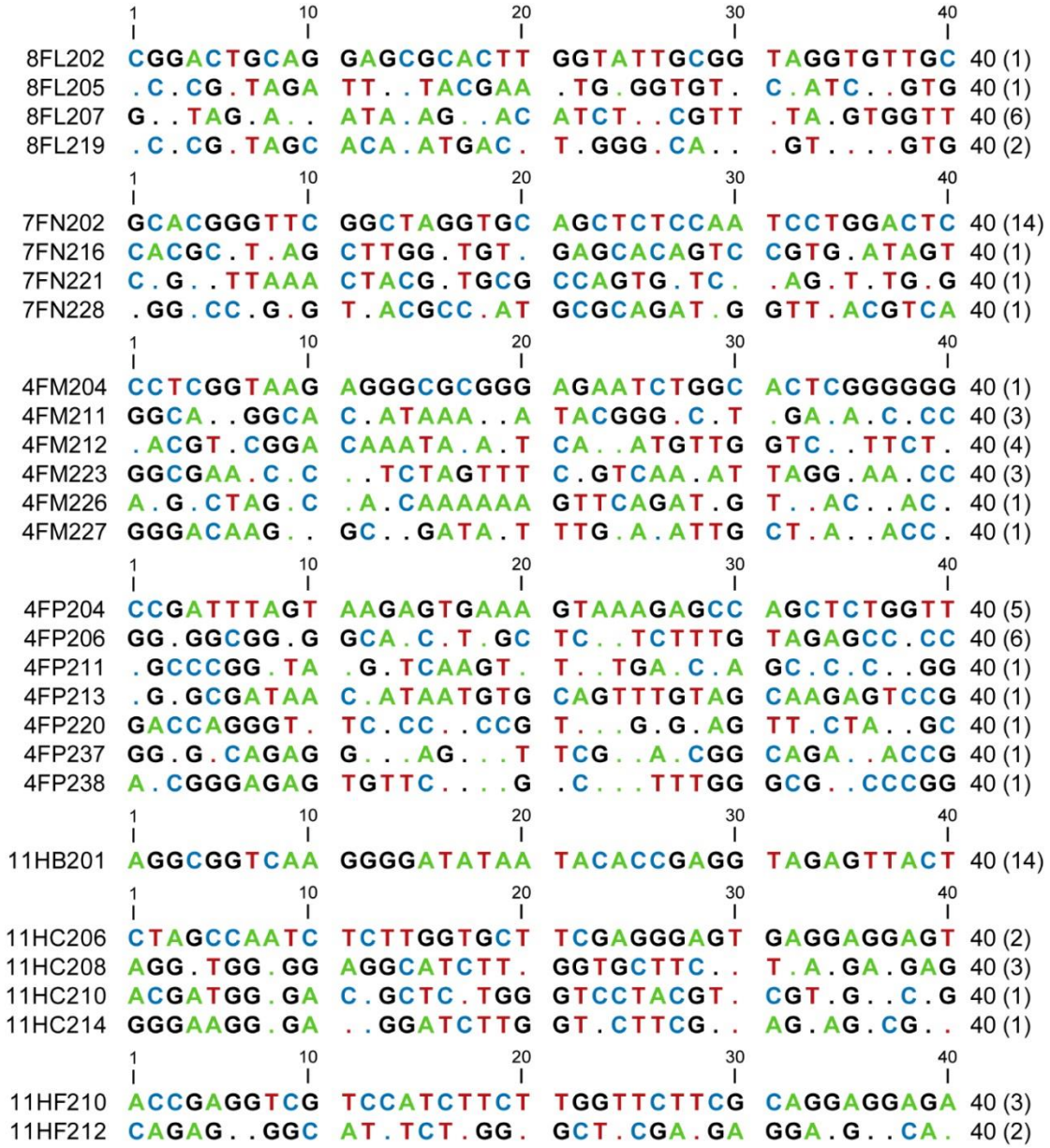


Figure 4.4. Sequences of the DNAzymes identified. Only the initially random (N_{40}) sequences are shown. DNAzymes were used as 5′-CGAAGTCGCCATCTCTTC-N_{40}-ATAGTGAGTCGTATTA-3′. In each alignment, a dot denotes conservation, i.e., the same nucleotide as in the uppermost sequence; a dash denotes a gap. On the far right is shown the sequence length and (in parentheses) the number of times that the particular sequence was found during cloning.

After cloning and sequencing, 6 unique DNAzyme sequences were identified from FM2 selection, from which 4FM212 has a repeat frequency of 4 times, 4FM211 and 4FM223 have a repeat frequency of 3 times, and the rest sequences of FM2 selection have only 1 repeat frequency. 7 unique DNAzyme sequences were identified from FP2 selection, from which 4FP204 and 4FP206 have a repeat frequency of 5 times and 6 times, and the rest sequences of FP2 selection have only 1 repeat frequency. 4 unique DNAzymes sequences were identified from FL2 selection, from which 8FL207 has a repeat frequency of 6 times and the rest sequences have 1 or 2 repeat frequency. 4 unique DNAzymes sequences were identified from FN2 selection, from which 7FN202 has a repeat frequency of 14 times and the rest sequences have 1 repeat frequency. There are no common conserved motifs among individual sequence from the same selection.

4.2.2.3 Kinetic Assays of DNAzymes from 5'-DNA-C_3-NH_2 Selection

Four selections, two acyl donor oligonucleotide substrates (PE and 4FPE) and two incubation conditions (pH 7.5 and 9.0) per substrate, were cloned and sequenced. These four selections can be divided into two groups according to their corresponding selection condition: pH 7.5 group (8FL2 and 7FN2) and pH 9.0 group (4FM2 and 4FP2).

As for pH 7.5 selections, there are overall 8 individual sequences identified. Four sequences (8FL202, 8FL205, 8FL207 and 8FL219) were from the FL2 selection which used PE substrate; four sequences (7FN202, 7FN216, 7FN221 and 7FN228) were from the FN2 selection which used 4FPE substrate. All 8 DNAzymes were assayed under their parent incubation condition as well as the incubation condition including different metal ions combination. Therefore, the metal ion dependence of the eight DNAzymes that use 5'-DNA-C_3-NH_2 and the PE or 4FPE substrates at pH 7.5 were also investigated at the same time (Figure 4.5 and Figure 4.6).

First, substantial amine acylation activity was observed for many of the resulting individual DNAzymes (Figure 4.5 and Figure 4.6). The PE and 4FPE selections each gave four DNAzymes, each with ~50% yield in 24 h. The highest rate enhancements, calculated by taking single-turnover k_{obs} for the DNAzyme and dividing by k_{bkgd} for the uncatalyzed, splinted background reaction using the complementary DNA splint, were 1100 and 760 (each ~10^3) for the 7FN216 and 7FN202 DNAzymes, respectively, which both use the 4FPE substrate at pH 7.5.

Second, by evaluating each DNAzyme with all possible combinations of 40 mM Mg^{2+}, 20 mM Mn^{2+}, and 1 mM Zn^{2+}, which were the concentrations of each ion that were used during the selection process, we found two types of metal ion dependence (Figure 4.5). All eight DNAzymes worked optimally when all three of Mg^{2+}, Mn^{2+}, and Zn^{2+} were included. Five of the eight DNAzymes, such as 7FN216, retained substantial activity with only Mg^{2+} and Zn^{2+} (omitting Mn^{2+}) and had greatly reduced yield with only Mn^{2+} and Zn^{2+} (omitting Mg^{2+}), whereas omitting Zn^{2+} led to no activity. The other three DNAzymes, such as 7FN202, needed all three of Mg^{2+}, Mn^{2+}, and Zn^{2+} for catalysis.

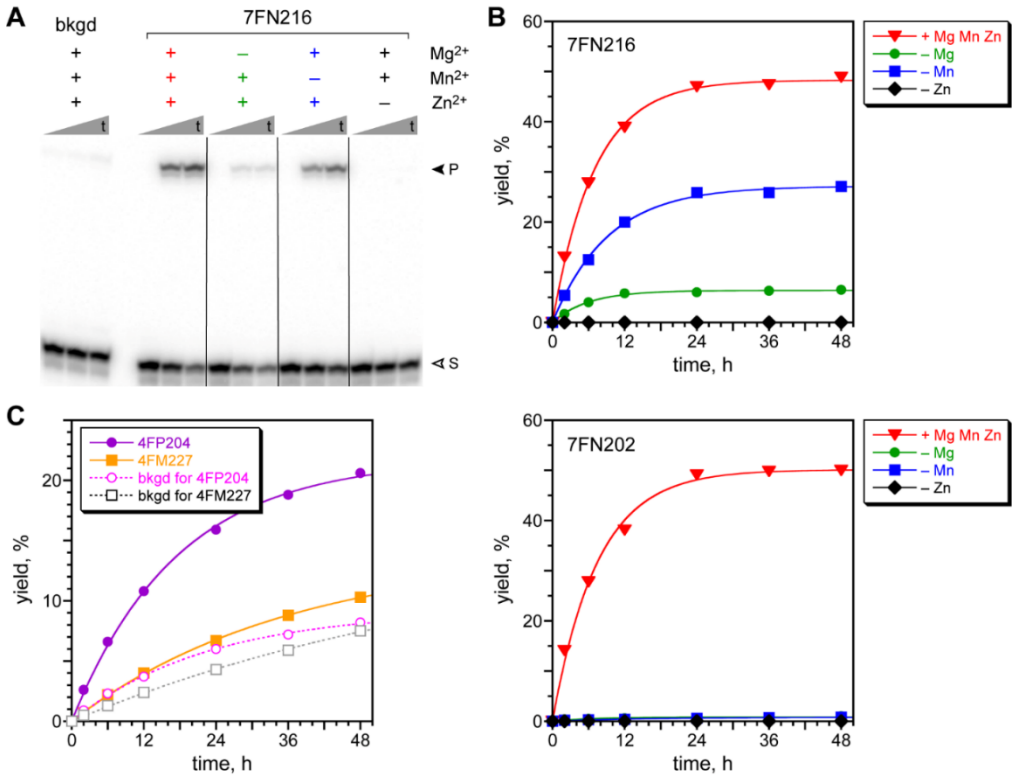


Figure 4.5. Assays of DNAzymes identified by in vitro selection using the DNA-C_3-NH_2 substrate. (A) Representative PAGE image for the 7FN216 DNAzyme with its 4FPE acyl donor substrate. Bkgd = complementary DNA splint in place of DNAzyme, to assess the uncatalyzed background reaction. Incubation conditions: 70 mM HEPES, pH 7.5, combinations of 40 mM $MgCl_2$, 20 mM $MnCl_2$, and 1 mM $ZnCl_2$ as indicated, and 150 mM NaCl at 37 °C. The background reaction was with $Mg^{2+}/Mn^{2+}/Zn^{2+}$. Shown are representative timepoints (t = 0.5 min, 6 h, 48 h; S = substrate, P = product). (B) Kinetic plots for 7FN216 and 7FN202, which have different metal ion dependence. k_{obs} values (h^{-1} ± standard deviation, each n = 3, with $Mg^{2+}/Mn^{2+}/Zn^{2+}$): 7FN216, 0.19 ± 0.06; 7FN202, 0.13 ± 0.02; background (k_{bkgd}), (1.7 ± 0.1) x 10^{-4}. k_{obs} values for the other six DNAzymes are 0.05–0.09 h^{-1}. (C) Kinetic plots for 4FM227 (PE substrate), 4FP204 (4FPE substrate), and a complementary DNA splint as background reaction for each substrate. Incubation conditions: 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C. k_{obs} values (h^{-1}, n = 1): 4FM227, 0.028; 4FP204, 0.057; background (k_{bkgd}): PE, 0.014; 4FPE, 0.044. Data was similar for the other eleven 4FM2 and 4FP2 DNAzymes.

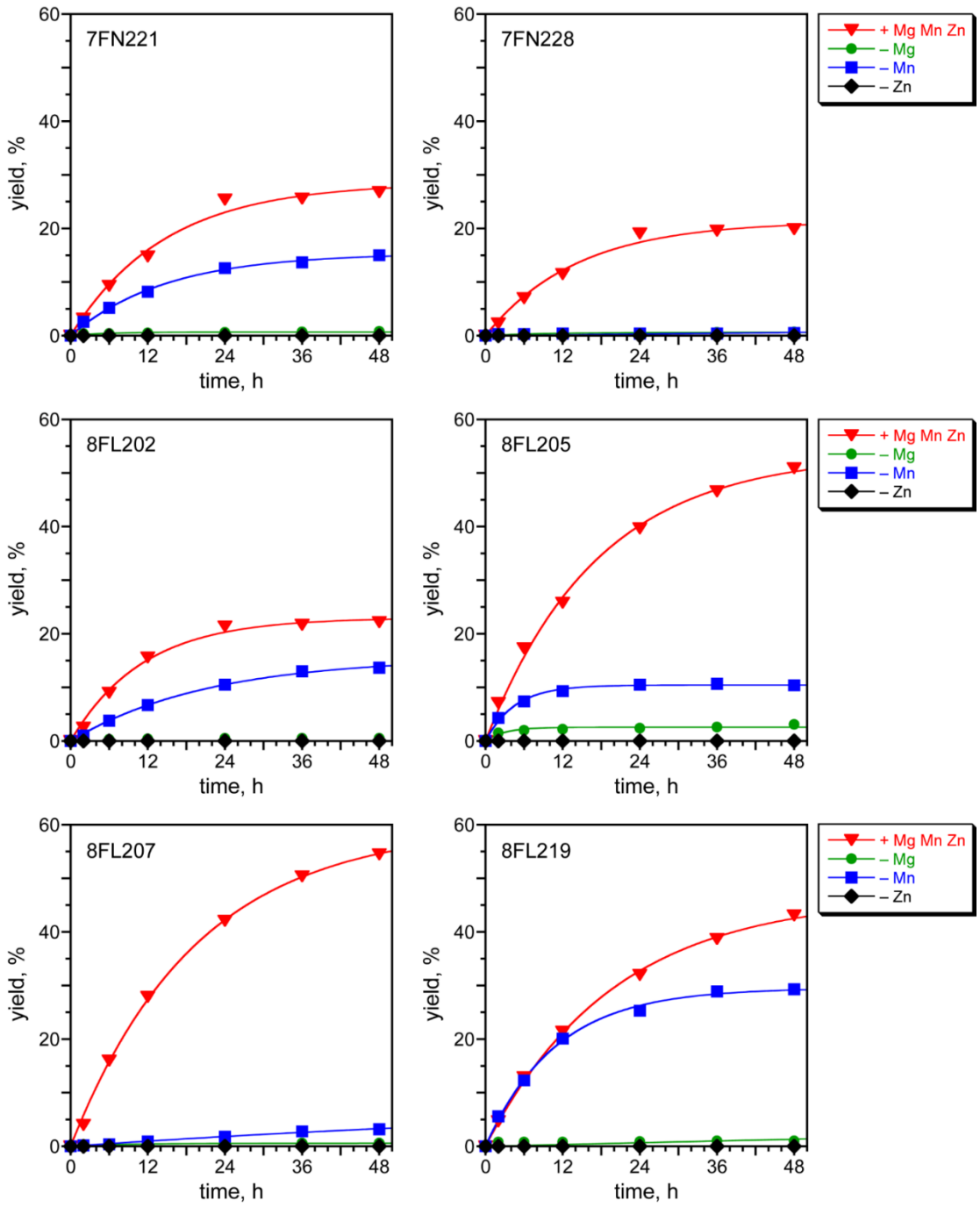


Figure 4.6. Kinetic plots for DNAzymes with 5'-DNA-C_3-NH_2 and PE or 4FPE substrates. k_{obs} values (h^{-1}, $n = 1$, with $Mg^{2+}/Mn^{2+}/Zn^{2+}$): 7FN221, 0.069; 7FN228, 0.071; 8FL202, 0.090; 8FL205, 0.058; 8FL207, 0.051; 8FL219, 0.050; background with complementary DNA splint and PE ester (k_{bkgd} for 8FL2 DNAzymes),1.3 x 10^{-4} h^{-1}. Rate enhancements (k_{obs}/k_{bkgd}), in same DNAzyme order: 410, 420, 690, 450, 390, 380.

As for pH 9.0 selections, there are overall 13 individual sequences identified. Six sequences (4FM204, 4FM211, 4FM212, 4FM223, 4FM226, 4FM227) were from the FM2 selection which used PE substrate; seven sequences (4FP204, 4FP206, 4FP211, 4FP213, 4FP220, 4FP237, 4FP238) were from the FP2 selection which used 4FPE substrate.

However, at pH 9.0 only Mg^{2+} can be included, because Mn^{2+} oxidizes and Zn^{2+} precipitates at this higher pH. Thus, the in trans assays of all 13 DNAzymes were only performed under the parent incubation condition and no metal ions dependence assays were carried out.

In contrast to the pH 7.5 selection, the pH 9.0 selection for each substrate gave only a modest increase in pool yield as the rounds progressed (Figure 4.3). All 13 DNAzymes sequences were assayed in trans and gave only 5–20% yields in 48 hours. Each individual DNAzyme had low rate enhancement of at most 2 (representative data in Figure 4.5C), and these DNAzymes were not studied further. An immediate conclusion is that the two pH 7.5 selections, for which the incubation conditions led to lower background yields (0.3–0.6%), were more successful at providing DNAzymes than the two pH 9.0 selections, which had higher background yields (2.4–3.5%).

4.2.2.4 Acylation Product Confirmation

For one representative DNAzyme from each of the pH 7.5 selection experiments, the acylation product was isolated by PAGE, and its expected mass was confirmed by MALDI mass spectrometry. In addition, all eight of the pH 7.5 DNAzymes were assayed using, as a negative control in place of 5'-DNA-C_3-NH_2, an unmodified DNA oligonucleotide lacking the pendant C_3-NH_2 at its 3'-end. In each case, no product formation was observed (<0.2%), consistent with nucleophilic reactivity of the amino group in the DNAzyme-catalyzed acylation reaction.

4.2.3 Selections with Peptide Lysine Substrate (5'-DNA-HEG-AAAKAA)

With success in identifying DNAzymes that acylate the simple amine nucleophile 5'-DNA-C_3-NH_2, especially at the lower pH of 7.5, our attention was shifted to the substrate that presents a peptide Lys nucleophile, 5'-DNA-HEG-AAAKAA, where our ultimate goal is DNAzyme-catalyzed acylation of peptide and protein Lys residues. Unsurprisingly based on our previous report with DNA-catalyzed Lys phosphoramidate formation,[1] none of the above-described DNAzymes identified using the 5'-DNA-C_3-NH_2 substrate had any detectable activity (<0.2%) when tested with 5'-DNA-HEG-AAAKAA. Therefore, new selection experiments with 5'-DNA-HEG-AAAKAA were designed and performed.

Three incubation conditions were determined to use the same PE and 4FPE acyl donor as electrophiles, and one incubation condition was determined to use 35diFPE acyl donor as electrophile. The conclusion we drawn from last selection was that pH 7.5 was better than pH 9.0 in terms of higher rate enhancement. Therefore, in addition to already established selection conditions at pH 7.5 and pH 9.0 as for 5'-DNA-C_3-NH_2, one extra incubation condition in presence of 70 mM MES pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h was applied for PE and 4FPE selections with 5'-DNA-HEG-AAAKAA. One specific incubation condition in presence of 50mM MES pH 6.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h was used for 35diFPE selection with 5'-DNA-HEG-AAAKAA.

4.2.3.1 Selection Progression Data

Seven selections HA2-HG2 were initially designed based on the three aryl esters PE, 4FPE, and 35diFPE as electrophiles and 5'-DNA-HEG-AAAKAA substrate as nucleophile. However, the lesson was learned that FQ2 selection was inactive in trans when the seven newly designed selections were just initiated. Considering that FQ2 and HG2 selection used the same aryl ester 35diFPE and the same incubation condition, it was unlikely that a catalytic activity would show

up in selection using a less reactive nucleophile (5'-DNA-HEG-AAAKAA) rather than a selection using a more reactive nucleophile (5'-DNA-C_3-NH_2). Thus, HG2 selection was dropped at the beginning immediately after round 1. Other six in vitro selections (HA2-HF2), however, were performed for 7 or 11 rounds (Figure 4.7).

The splinted background reaction was employed as the position indicator for acylation product. Unlike the capture yield of capture standard reaction described in Chapter 2 that caps the maximum yield for the actual selection product, however, the yield of splinted background reaction was totally independent on the actual selection reaction. Since no capture step was involved in selection rounds according to the selection design, the acylation yield in each round was directly correlated to the catalytic activity tendency. Thus, in the selection progression data figures, only the actual selection yields were presented with no comparison to the yields of splinted background reaction.

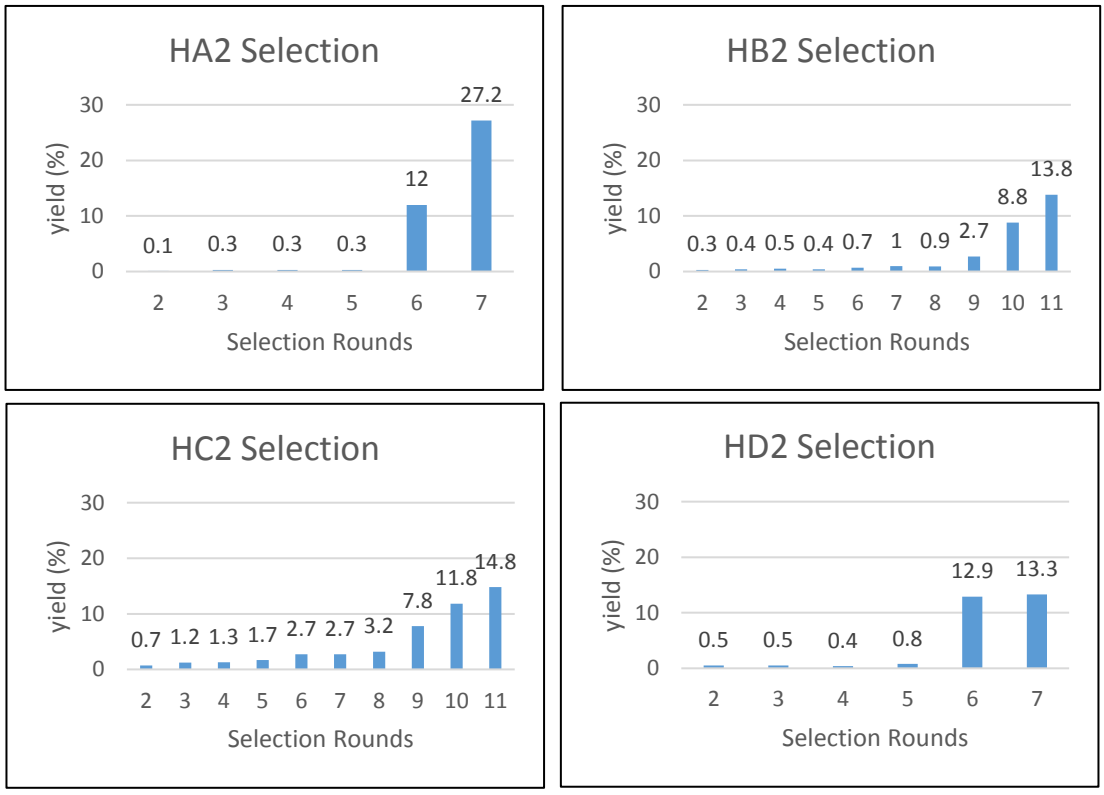


Figure 4.7. (cont.)

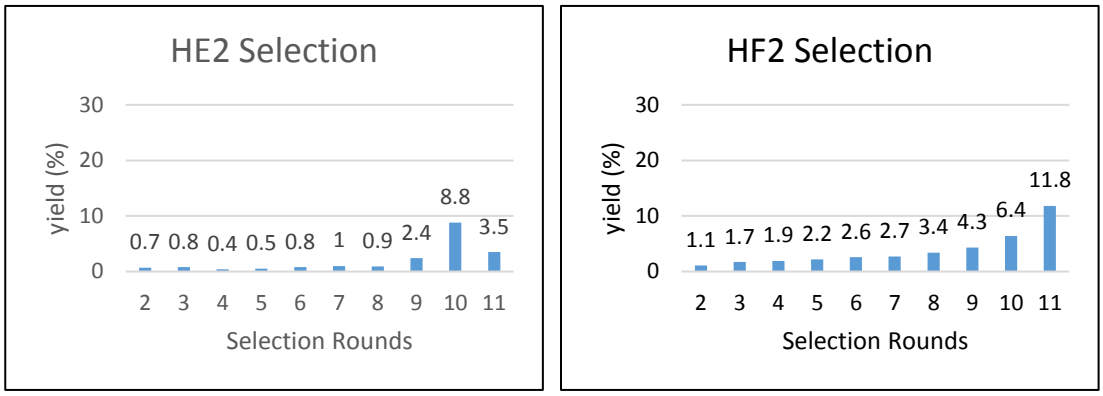


Figure 4.7. Selection progression data for selections HA2-HF2 using PE, 4FPE, and 35diFPE as electrophiles and 5'-DNA-HEG-AAAKAA as nucleophile.

After 7 rounds, HA2 and HD2 selection reached maximum yield plateau about 27% and 13%. The 30-cycle PCR products of 7HA2 and 7HD2 were assayed in trans; however, they were inactive in trans, and 7HA2 and 7HD2 were dropped after round 7.

After 10 rounds, HE2 selection reached maximum yield plateau about 9%. However, multiple bands were observed in the product region of the gel, and most of these were assigned to noncatalytic DNA sequences that misfold and therefore migrate aberrantly. Consistent with this observation, the 10HE2 pool was inactive in trans. Thus, HE2 selection was dropped after round 10.

After 11 rounds, HB2, HC2 and HF2 selection reached maximum yield plateau about 14%, 15% and 12%. The 11HB2, 11HC2 and 11HF2 sequence pools were tested in trans, and their in trans activity was confirmed. In summary, among seven initiated selections, only three selections (11HB2, 11HC2 and 11HF2) were cloned and sequenced.

4.2.3.2 DNAzymes Sequences from Selections with 5'-DNA-HEG-AAAKAA

After cloning and sequencing, only one unique DNAzyme sequence, 11HB201, was

identified from the HB2 selection. However, this sequence has a dominant repeat frequency of 14 times (Figure 4.4). 4 unique DNAzyme sequences were identified from the HC2 selection, from which 11HC206 has a repeat frequency of 2 times, and 11HC208 has a repeat frequency of 3 times, and the rest sequences of the HC2 selection have only 1 repeat frequency. 2 unique DNAzymes sequences were identified from the HF2 selection, from which 11HC210 has a repeat frequency of 3 times and 11HF212 has a repeat frequency of 2 times. There is no common conserved motifs among individual sequence from the same selection.

4.2.3.3 Kinetic Assays of DNAzymes from 5'-DNA-HEG-AAAKAA Selection

Lys acylation activity was observed for many of the resulting individual DNAzymes (Figure 4.8). The selection experiment at pH 7.5 with the PE substrate gave one single DNAzyme sequence, 11HB201, with modest 15% yield consistent with that of the uncloned round 11 pool as a whole. Since the only DNAzyme from pH 7.5 selection has a modest 15% yield in 48 h, it is not of great interest to investigate the metal ion dependence of such DNAzyme. Therefore, no metal ion dependence assay was performed for 11HB201.

In contrast, the two pH 9.0 selections both led to several distinct DNAzymes (Figure 4.8), most of which have high yields (>75%). The rate enhancements at pH 9.0 with Mg^{2+} for 11HC206 (PE substrate) and 11HF210 (4FPE substrate) were 86 and 60, respectively. These values are about an order of magnitude lower than the rate enhancements for the best DNAzymes identified for acylation of the 5'-DNA-C_3-NH_2 substrate, such as 7FN216 and 7FN202. This is due to the different pH values, 9.0 for the 5'-DNA-HEG-AAAKAA DNAzymes versus 7.5 for the 5'-DNA-C_3-NH_2 DNAzymes, where k_{bkgd} is substantially greater at the higher pH, and a greater k_{bkgd} leads to a lower rate enhancement.

Control experiments were performed using 5'-DNA-C_3-NH_2 in place of 5'-DNA-HEG-AAAKAA for all seven of the new DNAzymes. The results with the three DNAzymes of Figure

4.8 are representative and also consistent with our previous report with DNA-catalyzed Lys phosphoramidate formation.[1] In 24 h, the yields with DNA-C_3-NH_2 were 11HB201, <0.1%; 11HC206, 0.8%; and 11HF210, 1.4%. The corresponding k_{obs} values are calculated to be >6,400-fold, 360-fold, and 210-fold lower, respectively than k_{obs} of the same DNAzymes with DNA-HEG-AAAKAA (Figure 4.8), which supports the conclusion of selective DNAzyme-catalyzed nucleophilic reactivity of the Lys amino group of the AAAKAA hexapeptide. Because primary amino groups are not indiscriminately acylated by these DNAzymes, productive catalytic interactions are likely between each DNAzyme and its tethered peptide substrate.

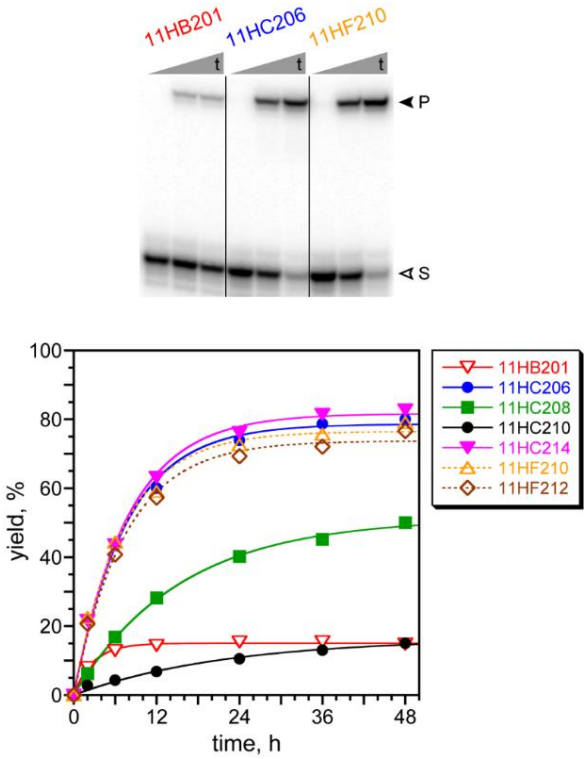


Figure 4.8. Assays of DNAzymes identified by in vitro selection using the DNA-HEG-AAAKAA substrate with the PE and 4FPE acyl donors. Incubation conditions for 11HB201 (PE substrate): 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C. Incubation conditions for all 11HC2 (PE substrate) and 11HF2 (4FPE substrate) DNAzymes: 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C. The PAGE assays for three DNAzymes are shown with representative timepoints (t = 0.5 min, 6 h, 48 h; S = substrate, (cont.)

Figure 4.8. (cont.) P = product). Representative kinetic plots are shown for several DNAzymes. k_{obs} values (h^{-1} ± standard deviation, n = 3): 11HB201, 0.27 ± 0.08; 11HC206, 0.12 ± 0.02; 11HF210, 0.12 ± 0.02. Additional k_{obs} values (h^{-1} with % yield at 48 h, n = 1): 11HC208, 0.065 (50%); 11HC210, 0.046 (15%); 11HC214, 0.13 (83%); 11HF212, 0.13 (77%). Background assays used the DNA-HEG-AAAKAA substrate and an exactly complementary splint in place of a DNAzyme, with k_{bkgd} values (h^{-1}) and % yield at 48 h as follows: pH 7.5 PE, 0.00013 ± 0.00004 (0.6%; n = 3); pH 9.0 PE, 0.0014 ± 0.0001 (5.8%; n = 3); pH 9.0 4FPE, 0.0020 ± 0.0001 (8.0%; n = 4).

4.2.3.4 Acylation Product Confirmation

For the only DNAzyme from the pH 7.5 selection experiment and one representative DNAzyme from each of the pH 9.0 selection experiments, the acylation product was isolated by PAGE, and its expected mass was confirmed by MALDI mass spectrometry. Interestingly, in an outcome opposite to that of the above-described selections with the simpler DNA-C_3-NH_2 substrate, here the selections with 5'-DNA-HEG-AAAKAA at the higher pH of 9.0 were more successful in leading to active DNAzymes.

4.2.4 Assays with Free Peptide Substrates

In the longer term, peptide-modifying DNAzymes will have their greatest utility when they can function with free (untethered, not DNA-anchored) peptide substrates. We assayed the DNAzymes that were identified with the DNA-HEG-AAAKAA substrate for their ability to function with 2 mM of free AAAKAA hexapeptide that is not tethered to the DNA anchor oligonucleotide. Unfortunately, in all cases, no activity was observed (<1.5% by PAGE-shift analysis, using 3'-^{32}P-radiolabeled 5'-aryl ester oligonucleotide; the untethered synthetic precursor oligonucleotide that was formerly connected to AAAKAA was included in these experiments). This result is unsurprising, given that the DNA-anchored AAAKAA was presented to the

DNAzyme population in every selection round, and an analogous tether requirement by emergent DNAzymes has been encountered in many of our prior selections. In the future, we intend to perform lysine acylation selection experiments in which an azide-modified free peptide is used in the selection step, thereby enforcing a strict pressure for the resulting DNAzymes to function with the free peptide.[2]

4.3 Summary

We learned and were inspired from previous lessons that thioesters were too unreactive to support DNAzyme catalysis, whereas the more electrophilic DMTE and TFPE were too reactive. In this chapter, we sought intermediate-reactivity acyl donor electrophile and successfully established that DNAzymes can catalyze amine acylation, including acylation of a Lys residue in a short DNA-anchored peptide.[3] Key to this success was identifying suitably substituted aryl esters (PE and 4FPE) as the electrophilic acyl donors, along with appropriate incubation conditions that balance electrophile stability and reactivity. The observation that different pH values (and therefore different degrees of uncatalyzed background reactivity) supported emergence of the best DNAzymes for the two different substrates, 5'-DNA-C_3-NH_2 and 5'-DNA-HEG-AAAKAA, suggests that pH is an important experimental variable to explore in our future studies.

We also observed that several DNAzymes can function for amine glutarylation, even though these DNAzymes were not directly identified by selection for amine acylation using the glutaryl donor substrate. These experiments and DNAzymes are described in Chapter 5.

4.4 Materials and methods

4.4.1 Preparation of Oligonucleotides

DNA oligonucleotides and oligonucleotide conjugates were obtained from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an ABI 394 instrument using reagents from Glen Research. All oligonucleotides and conjugates were purified by 7 M urea denaturing PAGE with running buffer 1X TBE (89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3), extracted from the polyacrylamide with TEN buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaCl), and precipitated with ethanol as described previously.[4,5] Hexapeptide AAAK(tfa)AA was prepared by solid-phase synthesis using Fmoc Rink amide MBHA resin as described.[6] The trifluoroacetyl protecting group was removed after conjugation to the DNA anchor oligonucleotide. The coupling reagents DMT-MM and EDC/TFP were commercially available (Fisher Scientific).

Conjugation of AAAKAA to DNA anchor oligonucleotide. The DNA anchor oligonucleotide was 5'-GGATAATACGACTCACTAT-HEG-rA-3', where the 3'-terminal ribonucleotide was oxidized by $NaIO_4$ and used for conjugation to the peptide N-terminus. A 100 µL sample containing 1.0 nmol of DNA anchor oligonucleotide in 100 mM HEPES, pH 7.5, and 10 mM $NaIO_4$ was incubated at room temperature for 1 h. The oxidized product was precipitated to remove excess $NaIO_4$ by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The precipitated product was dissolved in 65 µL of water and used directly in the next step. A 100 µL sample containing the $NaIO_4$-oxidized DNA anchor oligonucleotide and 100 nmol (100 equiv) of AAAK(tfa)AA hexapeptide in 100 mM NaOAc, pH 5.2, 50 mM $NiCl_2$, and 10 mM $NaCNBH_3$ was incubated at 37 °C for 14 h. The DNA-anchored hexapeptide was precipitated by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The trifluoroacetyl protecting group was removed by incubation in 100 µL of 30% NH_4OH for 1 h at room temperature. NH_4OH was removed in a speedvac and the

conjugation product was purified by 20% PAGE.

Synthesis of 5'-CO_2H-C_9-DNA oligonucleotide. The carboxylic acid DNA oligonucleotide sequence was 5'-CO_2H-C_9-GAAGAGATGGCGACTTCG-3'. The 5'-GAAGAGATGGCGACTTCG-3' was the precursor oligonucleotide, which was prepared by ABI 394 solid-phase synthesizer, for the eventual 5'-CO_2H-C_9-DNA oligonucleotide. The precursor column was dried and stored in -20 °C condition before usage. 5'-Carboxyl-C_{10}-Modifier (Glen Research, 100 μmol) was dissolved in 1 mL anhydrous acetonitrile under the protection of argon. 500 μL of dissolved 5'-Carboxyl-C_{10}-Modifier was mixed with 500 μL activator, and this mixture was applied to the precursor column for 12 min coupling reaction under argon. The mixture was removed, and the precursor column was washed with 1 mL anhydrous acetonitrile 3 times. 1 mL oxidizing reagent was applied to the precursor column for 15 min oxidization reaction under argon. The oxidizing reagent was removed, and the precursor column was washed with 1 mL anhydrous acetonitrile 3 times. 500 μL Cap A reagent was mixed with 500 μL Cap B reagent, and this mixture was applied to the precursor column for 5 min capping reaction under argon. The mixture was removed, and the precursor column was washed with 1 mL anhydrous acetonitrile 3 times, and the column was dried with argon for 5 min. The CPG from precursor column was transferred to a back ring tube, and then 400 μL 0.4 M NaOH in H_2O/MeOH (1:4, v/v) was added into the tube and incubated for 48 hours at room temperature. The black ring tube (with CPG and NaOH mixture inside) was sonicated for 5 min, and the CPG was spun down. The supernatant was transferred to a new clean tube, the CPG was rinsed with 200 μL water, and the rinsed water was combined with the supernatant. The supernatant was ethanol precipitated and finally dissolved in 200 μL water. 50 μL of unpurified 5'-CO_2H-C_9-DNA oligonucleotide solution was taken for HPLC purification each time. The HPLC is using a Shimadzu Prominence instrument with a Phenomenex Gemini-NX C_{18} column (5 μm, 10 × 250 mm), solvent A (20 mM triethylammonium acetate in 50% acetonitrile/50% water, pH 7.0), solvent B (20 mM triethylammonium acetate in water, pH 7.0), and flow rate of 3.5 mL/min. For 5'-CO_2H-C_9-DNA oligonucleotide, the gradient was 15% solvent

A/85% solvent B at 0 min to 30% solvent A/70% solvent B at 45 min, and the elution time was between 24-25 min.

4.4.2 Mass Spectrometry of Oligonucleotides

Data were acquired on a Bruker UltrafleXtreme MALDI-TOF mass spectrometer with matrix 3-hydroxypicolinic acid in positive ion mode at the UIUC School of Chemical Sciences Mass Spectrometry Laboratory. All *m*/*z* values are for $[M+H]^+$. Samples were desalted by Millipore C_{18} ZipTip before analysis.

Data for oligonucleotides and conjugates were as follows. For those that bind to the left-hand DNAzyme binding arm, the DNA sequence was 5'-GGATAATACGACTCACTAT-3'. For those that bind to the right-hand DNA binding arm, the DNA sequence was 5'-GAAGAGATGGCGACTTCG-3'.

Oligonucleotides and conjugates that bind to left-hand DNAzyme binding arm

DNA-C_3-NH_2 *m*/*z* calcd. 5934.9, found 5931.1, Δ= –0.06%

DNA-HEG-AAAKAA *m*/*z* calcd. 6934.6, found 6938.9, Δ= +0.06%

Oligonucleotides and conjugates that bind to right-hand DNAzyme binding arm

5'-CO_2H-C_9-DNA *m*/*z* calcd. 5839.9, found 5840.3, Δ= +0.007%

PE substrate *m*/*z* calcd. 5916.0, found 5916.8, Δ= +0.01%

4FPE substrate *m*/*z* calcd. 5934.0, found 5932.5, Δ= –0.03%

5'-alkyne-C_6-DNA *m*/*z* calcd. 5750.8, found 5747.6, Δ= –0.04%

The products of several representative individual DNAzymes were analyzed by MALDI mass spectrometry. Each product was prepared from a 21 µL sample containing 200 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate, 220 pmol of DNAzyme, and 240 pmol of acyl donor substrate. The sample was annealed in 5 mM HEPES, pH 7.5 or CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The DNAzyme-catalyzed reaction was initiated by bringing the sample to 30 µL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C for 48 h, precipitated with ethanol, and purified by 20% PAGE.

DNAzyme products

8FL205 with DNA-C_3-NH_2, PE, pH 7.5 *m*/*z* calcd. 11756.8, found 11751.8, Δ= –0.04%

7FN216 with DNA-C_3-NH_2, 4FPE, pH 7.5 *m*/*z* calcd. 11756.8, found 11753.1, Δ= –0.03%

11HB201 with DNA-HEG-AAAKAA, PE, pH 7.5 *m*/*z* calcd. 12756.5, found 12758.5, Δ= +0.02%

11HC206 with DNA-HEG-AAAKAA, PE, pH 9.0 *m*/*z* calcd. 12756.5, found 12761.5, Δ= +0.04%

11HF210 with DNA-HEG-AAAKAA, 4FPE, pH 9.0 *m*/*z* calcd. 12756.5, found 12757.7, Δ= +0.01%

4.4.3 In Vitro Selection Procedure Steps

The key selection step of each round using a 5'-aryl ester oligonucleotide substrate is shown in Figure 4.9. The DNA anchor oligonucleotide sequence was 5'-GGATAATACGACTCACTAT-3'. The 5'-aryl ester oligonucleotide sequence was 5'-GAAGAGATGGCGACTTCG-3'. The random DNAzyme pool was 5'-CGAAGTCGCCATCTCTTC-N_{40}-ATAGTGAGTCGTATTAAGCTGATCCTGATGG -3'. PCR primers were 5'-CGAAGT-CGCCATCTCTTC-3' (forward primer) and 5'-$(AAC)_{18}$XCCATCAGGATCAGCT-3', where X is the HEG spacer to stop Taq polymerase (reverse primer). In each round, the ligation step to attach the DNAzyme pool at its 3'-end with the 5'-end of the amine-containing substrate was performed using a DNA splint and T4 DNA ligase. The splint sequence was 5'-ATAGTGAGTCGTATTATCCCCATCAGGATCAGCTTAATACGACTCACTAT-3'.

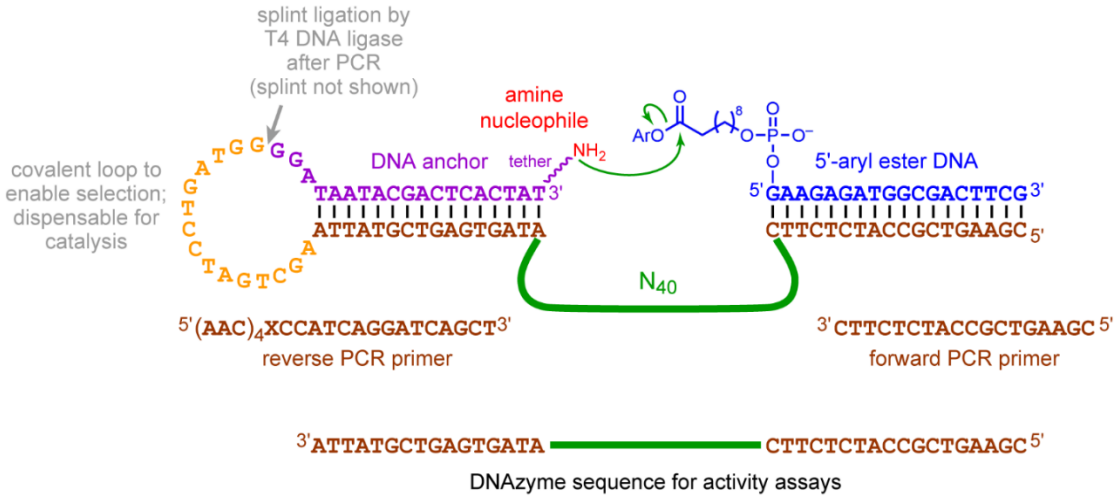


Figure 4.9. Nucleotide details of the in vitro selection experiments with a 5'-aryl ester substrate.

Procedure for ligation step in round 1. A 25 µL sample containing 600 pmol of DNA pool, 750 pmol of DNA splint, and 900 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 3 µL of 10x T4 DNA ligase buffer (400

mM Tris, pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, and 5 mM ATP) and 2 µL of 5 U/µL T4 DNA ligase (Thermo Fisher). The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for ligation step in subsequent rounds. A 17 µL sample containing the PCR-amplified DNA pool (~5–10 pmol), 30 pmol of DNA splint, and 50 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 2 µL of 10x T4 DNA ligase buffer and 1 µL of 1 U/µL T4 DNA ligase. The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for selection step in round 1. Each selection experiment was initiated with 200 pmol of the ligated pool. A 20 µL sample containing 200 pmol of ligated pool and 300 pmol of 5'-aryl ester DNA was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA or 5 mM CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 40 µL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The Mn^{2+} was added from a 10x stock solution containing 200 mM $MnCl_2$. The Zn^{2+} was added from a 10x stock solution containing 10 mM $ZnCl_2$, 20 mM HNO_3, and 200 mM HEPES at pH 7.5; this stock solution was freshly prepared from a 100x stock of 100 mM $ZnCl_2$ in 200 mM HNO_3. The metal ion stocks were added last to the final sample. The sample was incubated at 37 °C for 16 h and separated by 8% PAGE.

Procedure for selection step in subsequent rounds. A 14 µL sample containing the ligated pool and 30 pmol of 5'-aryl ester DNA was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA or 5 mM CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to

20 µL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C for 16 h and separated by 8% PAGE.

Procedure for PCR. In each selection round, two PCR reactions were performed, 10-cycle PCR followed by 30-cycle PCR. First, a 100 µL sample was prepared containing the PAGE-purified selection product, 200 pmol of forward primer, 50 pmol of reverse primer, 20 nmol of each dNTP, and 10 µL of 10x Taq polymerase buffer [1x = 20 mM Tris-HCl, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, and 0.1% Triton X-100]. This sample was cycled 10 times according to the following PCR program: 94 °C for 2 min, 10x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Taq polymerase was removed by phenol/chloroform extraction. Second, a 50 µL sample was prepared containing 1 µL of the 10-cycle PCR product, 100 pmol of forward primer, 25 pmol of reverse primer, 10 nmol of each dNTP, 20 µCi of α-^{32}P-dCTP (800 Ci/mmol), and 5 µL of 10x Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Samples were separated by 8% PAGE.

Cloning and screening of individual DNAzymes. The PCR primers used for cloning were 5'-CGAAGTCGCCATCTCTTC-3' (forward primer; same as in selection) and 5'-TAATTAATTAATTACCCATCAGGATCAGCT-3' (reverse primer). The 10-cycle PCR product from the appropriate selection round was diluted 1000-fold. A 50 µL sample was prepared containing 1 µL of the diluted 10-cycle PCR product from the appropriate selection round, 100 pmol of forward cloning primer, 25 pmol of reverse cloning primer, 10 nmol of each dNTP, and 5 µL of 10x Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. The sample was separated by 1% agarose gel and extracted using a GeneJET Gel Extraction Kit (Fermentas). The extracted product was quantified by absorbance (A_{260}) and diluted to 5–10

ng/μL. A 4 μL portion of the diluted PCR product was inserted into the pCR2.1-TOPO vector using a TOPO TA cloning kit (Life Technologies). Individual *E. coli* colonies harboring plasmids with inserts were identified by blue-white screening and grown in LB/amp media. Miniprep DNA was prepared using a GeneJET Plasmid Miniprep Kit (Fermentas) and screened for properly sized inserts by EcoRI digestion and agarose gel analysis. Before sequencing, assays of individual DNAzyme clones were performed with PAGE-purified DNA strands prepared by PCR from the miniprep DNA, using the single-turnover assay procedure that is described in this chapter.

4.4.4 Preparative Procedures for Oligonucleotides and Conjugates

The 5'-aryl ester oligonucleotide (PE and 4FPE) substrates were each prepared by EDC activation and reaction with phenol and 4-fluorophenol. A 40 μL sample containing 3 nmol of 5'-CO_2H oligonucleotide, 50 mM EDC, and 50 mM phenol or 4-fluorophenol was incubated in 100 mM MES, pH 6.0 and 150 mM NaCl at RT for 12 h. The PE or 4FPE substrate was purified by HPLC, using a Shimadzu Prominence instrument with a Phenomenex Gemini-NX C_{18} column (5 μm, 10 × 250 mm), solvent A (20 mM triethylammonium acetate in 50% acetonitrile/50% water, pH 7.0), solvent B (20 mM triethylammonium acetate in water, pH 7.0), and flow rate of 3.5 mL/min. For PE, the gradient was 20% solvent A/80% solvent B at 0 min to 60% solvent A/40% solvent B at 60 min. For 4FPE, the gradient was 25% solvent A/75% solvent B at 0 min to 55% solvent A/45% solvent B at 60 min. The separated sample was lyophilized, precipitated with ethanol, and quantified by UV absorbance (A_{260}). The typical isolated yield of PE/4FPE oligonucleotide substrate was 500 pmol (17%).

4.4.5 HPLC Chromatograms for Purification of the PE and 4FPE Oligonucleotide Substrates

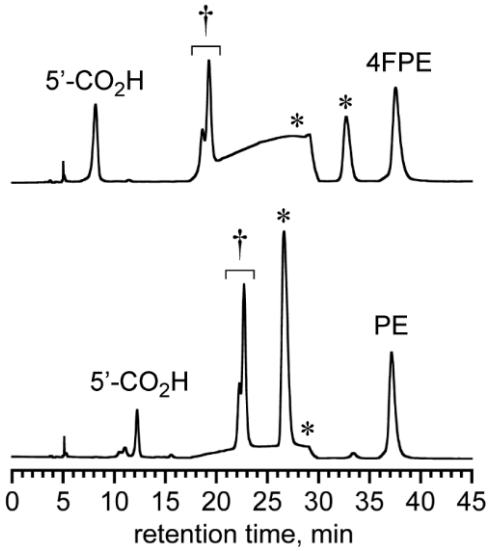


Figure 4.10. HPLC chromatograms (260 nm detection) for purification of the PE and 4FPE oligonucleotide substrates. The HPLC procedures are provided as above. Note that the gradients were different for the two purifications. The peaks assigned as the PE and 4FPE substrates were validated by MALDI mass spectrometry, as were the peaks assigned to the 5'-CO_2H DNA oligonucleotide (precursor to PE or 4FPE substrate). The peaks marked with daggers (†) are oligonucleotides, as assessed from their full UV spectra (PDA detector), and are likely the *N*-acylurea dead-end products formed by rearrangement after EDC activation to form the transient *O*-acylisourea products. The peaks marked with asterisks (*) are small-molecule compounds rather than oligonucleotides, as assessed from their full UV spectra (PDA detector).

4.4.6 Assays for PE and 4FPE Acyl Donor Stability and Background Reactivity

We sought to identify incubation conditions in which each acyl donor substrate satisfies two key criteria: (a) stability to the extent of at least 25% acyl donor remaining intact (i.e., unhydrolyzed) at the end of the incubation period in the absence of amine nucleophile, and ideally higher than 25% remaining intact; and (b) observable but modest background amine acylation reactivity (amide formation) in the presence of the amine nucleophile. We chose criterion (a)

because this extent of stability, under the inherently single-turnover arrangement of in vitro selection (i.e., each candidate DNAzyme molecule is persistently Watson-Crick bound to a single acyl donor oligonucleotide molecule for the duration of the reaction), means that a substantial fraction of the population's DNA sequences have the opportunity to catalyze amine acylation at some point during the incubation and therefore survive the selection round. We chose criterion (b) because the negative outcome with the thioester selections implies that beginning with absolutely no observable background reactivity could result in the inability of any DNAzyme sequences to emerge from the selection process. That is, even appreciable rate enhancement relative to a very low background reactivity could still result in undetectable DNAzyme-catalyzed reactivity of the population after enrichment of catalytically active sequences through many selection rounds.

For each of the PE and 4FPE acyl donor substrates, we identified suitable incubation conditions that satisfy both criteria (a) and (b), as shown in Table 4.1. We did not yet know whether the most appropriate incubation conditions for each substrate would involve relatively high or low background reactivity within the explored range. Instead, the optimum background reactivity for identifying DNAzymes, as dictated by choice of incubation conditions (pH, metal ions, and time), would have to be determined empirically through the in vitro selection process itself. Based on the tabulated data, we chose two particular incubation conditions for each acyl donor substrate: lower pH of 7.5 (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h) and higher pH of 9.0 (50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 16 h). Each of Mg^{2+}, Mn^{2+}, and Zn^{2+} was included at pH 7.5 on the basis of our many prior successful DNAzyme selection efforts using these metal ions at pH 7.5. However, at pH 9.0 only Mg^{2+} can be included, because Mn^{2+} oxidizes and Zn^{2+} precipitates at this higher pH.

Per Table 4.1 the evaluated incubation conditions lead to background yields up to 3.5%, where these values are uncorrected for partial hydrolysis of the acyl donor during the incubation period. For example, considering entry 8, in the limiting case that only 74% of intact acyl donor is present

for the entire duration of the incubation period (i.e., all of the 26% of hydrolysis occurs immediately at the start of the incubation time), the observed background yield of 3.5% would be relative to a maximum possible background yield of 74%, for a corrected background yield of 4.7%. However, it is unlikely that most of the loss of acyl donor occurs immediately upon starting the incubation period. Instead, as indicated by comparing entries 7 and 8, the loss of acyl donor is more gradual, and the properly corrected background yield would lie somewhere between 3.5% and 4.7%. Because, for example, a background yield of 5% (meaning that 5% of all sequences survive the selection round regardless of their catalytic ability) limits the per-round enrichment to 20-fold, which is still substantial, we decided that the range of conditions represented by the four even-numbered entries in Table 4.1 (all with incubation time of 16 h rather than 2 h) were suitable for performing in vitro selection.

For the intact acyl donor assays of Table 4.1, a 14 µL sample containing 1 pmol of 3'-^{32}P-radiolabeled acyl donor oligonucleotide (PE or 4FPE substrate) was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA or 5 mM CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The assay was initiated by bringing the sample to 20 µL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C for 2 h or 16 h and separated by 20% PAGE. The extent of PE/4FPE hydrolysis was calculated from the the bands corresponding to PE/4FPE and 5'-CO_2H oligonucleotides.

For the background yield assays of Table 4.1, a 14 µL sample containing 0.5 pmol of 5'-^{32}P-radiolabeled DNA-C_3-NH_2, 10 pmol of splint oligonucleotide (5'-CGAAGTCGCCATCTCTTCATAGTGAGTCGTATTATCC-3'), and 20 pmol of nonradiolabeled PE or 4FPE substrate was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA or 5 mM CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling

on ice for 5 min. The background reaction was initiated by bringing the sample to 20 µL total volume containing 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C for 2 h or 16 h and separated by 20% PAGE. The background yield was calculated from the bands corresponding to DNA-C_3-NH_2 and acylation product.

4.4.7 Single-Turnover Deoxyribozyme Assay Procedure

The amino substrate (DNA-C_3-NH_2 or DNA-HEG-AAAKAA) was 5'-^{32}P-radiolabeled using γ-^{32}P-ATP and polynucleotide kinase. A 14 µL sample containing 0.2 pmol of 5'-^{32}P-radiolabeled amino substrate, 10 pmol of deoxyribozyme and 30 pmol of 5'-aryl ester DNA was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The DNA-catalyzed amine acylation reaction was initiated by bringing the sample to 30 µL total volume containing 70 mM HEPES, pH 7.5, 1 mM $ZnCl_2$, 20 mM $MnCl_2$, 40 mM $MgCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C. At appropriate time points, 2 µL aliquots were quenched with 5 µL stop solution (80% formamide, 1x TBE [89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3], 50 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol). Samples were separated by 20% PAGE and quantified with a PhosphorImager. Values of k_{obs} were obtained by fitting the yield versus time data directly to first-order kinetics; i.e., yield = $Y \bullet (1 - e^{-kt})$, where $k = k_{obs}$ and Y is the final yield. Each k_{obs} value is reported with error calculated as the standard deviation from the indicated number of independent determinations.

4.5 References

(1) Benjamin M. Brandsen, T. E. V., Amit Sachdeva, Nora A. Ibrahim, and Scott K. Silverman. DNA-Catalyzed Lysine Side Chain Modification. *Angew. Chem. Int. Ed.* **2014**, *53*, 9045-9050.

(2) Chu, C.; Wong, O.; Silverman, S. K. A Generalizable DNA-Catalyzed Approach to Peptide-Nucleric Acid Conjugation. *ChemBioChem* **2014**, *15*, 1905-1910.

(3) Yao, T.; Przybyla, J. J.; Yeh, P.; Woodard, A. M.; Nilsson, H. J.; Brandsen, B. M.; Silverman, S. K. DNAzymes for amine and peptide lysine acylation. *Organic & Biomolecular Chemistry* **2021**, *19*, 171-181.

(4) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. Deoxyribozymes with 2'− 5' RNA ligase activity. *J. Am. Chem. Soc.* **2003**, *125*, 2444-2454.

(5) Wang, Y.; Silverman, S. K. Characterization of deoxyribozymes that synthesize branched RNA. *Biochemistry* **2003**, *42*, 15252-15263.

(6) Story, S. C.; Aldrich, J. V. Preparation of protected peptide amides using the Fmoc chemical protocol: Comparison of resins for solid phase synthesis 1. *Int. J. Pept. Protein Res.* **1992**, *39*, 87-92.

Chapter 5: Efforts towards DNAzymes for Simple Amine and Peptide Lysine Glutarylation

5.1 Introduction

5.1.1 Glutaryl Aryl Ester as Acyl Donor

According to the amine acylation reaction described in Chapter 4, a successful DNAzyme-catalyzed acylation reaction joins the acyl donor oligonucleotide to the amine acceptor, and the small-molecule phenol derivative of the 5'-aryl ester oligonucleotide is the leaving group. Ideally, amine-acylating DNAzymes will instead be able to use an acyl donor that transfers a small-molecule acyl group rather than a large acyl-oligonucleotide. To explore this possibility, we synthesized a glutaryl donor oligonucleotide in which the orientation of the aryl ester functional group was inverted, such that acylation results in glutarylation of the amine nucleophile. This was achieved by first preparing a glutaryl-azide small-molecule compound, which was then used in a CuAAC (copper-catalyzed azide-alkyne cycloaddition) reaction with a 5'-alkyne-modified oligonucleotide, to form the glutaryl donor oligonucleotide (Figure 5.1). The synthesis of the glutaryl-azide small-molecule compound was performed by previous Silverman lab graduate student Peter Yeh.

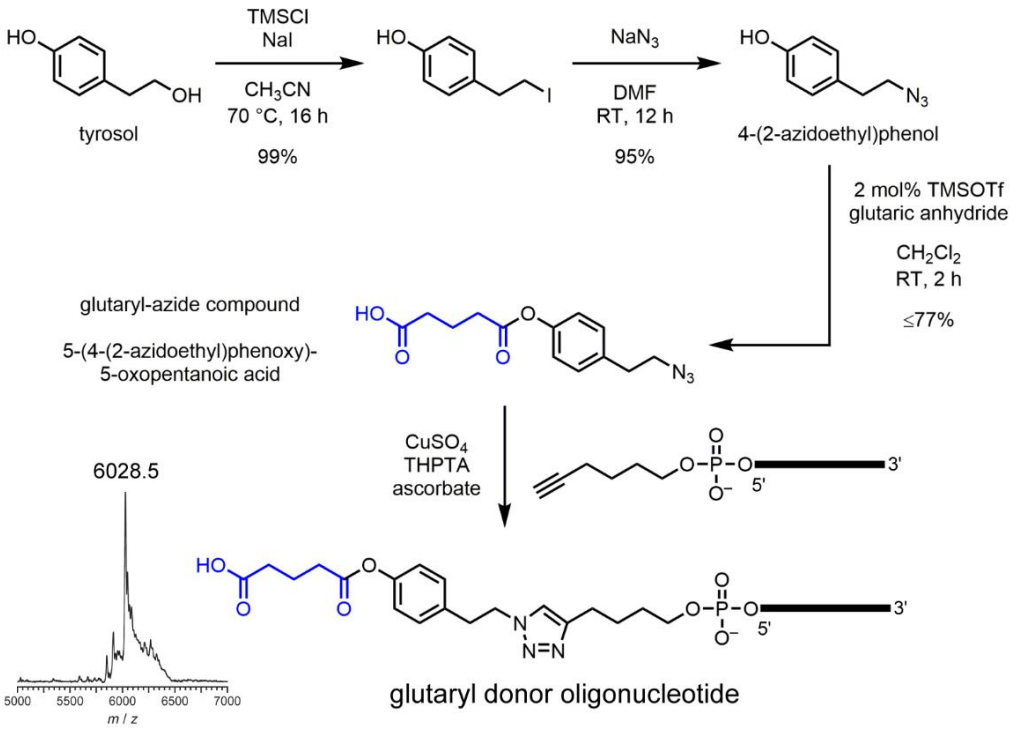


Figure 5.1. The glutaryl donor oligonucleotide was prepared by synthesis of the illustrated glutaryl-azide compound and subsequent CuAAC reaction with 5'-alkyne-C_4-GAAGAGATGGCGACTTCG-3'.

The glutaryl donor oligonucleotide is a *p*-alkyl-substituted aryl ester of glutaric acid, where the *p*-alkyl group is inherently electron-donating. Therefore, we expect the glutaryl donor oligonucleotide to be somewhat less reactive than the PE substrate and perhaps substantially less reactive than the fluoro-substituted esters of such as 4FPE. Also, the spatial presentations of the acyl donors are different (e.g., the linkers to the oligonucleotide are not the same), which may suppress the ability of any DNAzymes identified by in vitro selection with PE or 4FPE acyl donors to function with the glutaryl donor oligonucleotide.

5.1.2 In Vitro Selection Design towards DNAzymes for Amine Glutarylation

Three in vitro selections HX2-HZ2 were designed based on using glutaryl oligonucleotide acyl donor as electrophile and 5'-DNA-HEG-AAAKAA as nucleophile. These in vitro selection experiments used N_{40} initially random regions with unmodified nucleotides rather than modified nucleotides. The key steps of in vitro selections HX2-HZ2 are shown in Figure 5.2. The HX2-HZ2 selection was performed by previous graduate Peter Yeh, summer rotation student Austin Woodard, and graduate Yves Yao (myself).

After synthesizing the N_{40} randomized DNA pool with regular nucleotides via solid-phase oligo synthesizer, the ligation step that covalently links the amine substrate to the DNA pool was carried out. The purified 5'-glutaryl donor oligonucleotide aryl ester was then added into the selection step.

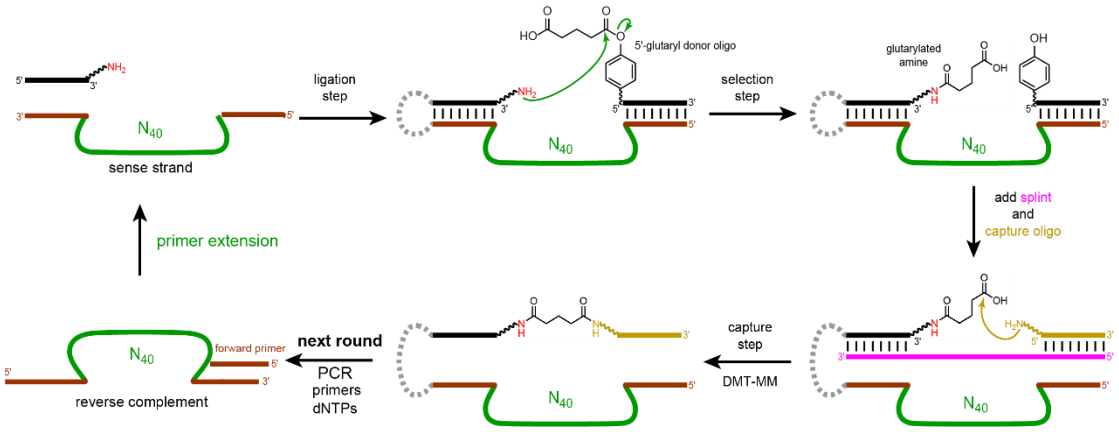


Figure 5.2. Overview of in vitro selection using a glutaryl oligonucleotide acyl donor as electrophile.

The incubation conditions and capture steps of in vitro selections HX2-HZ2 using glutaryl oligonucleotide acyl donor as electrophile are shown in Figure 5.3. After synthesizing the N_{40} DNA pool with unmodified nucleotides, the ligation step that covalently links the amine substrate to the DNA pool was carried out. The subsequent selection step for amine acylation was performed in presence of 5'-glutaryl donor oligonucleotide in 70 mM MES, pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 12-14 h; or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 12-14 h; or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl at 37 °C for 12-14 h, where Zn^{2+}, Mn^{2+}, and Mg^{2+} are metal ion cofactors required by most deoxyribozymes' reactions.

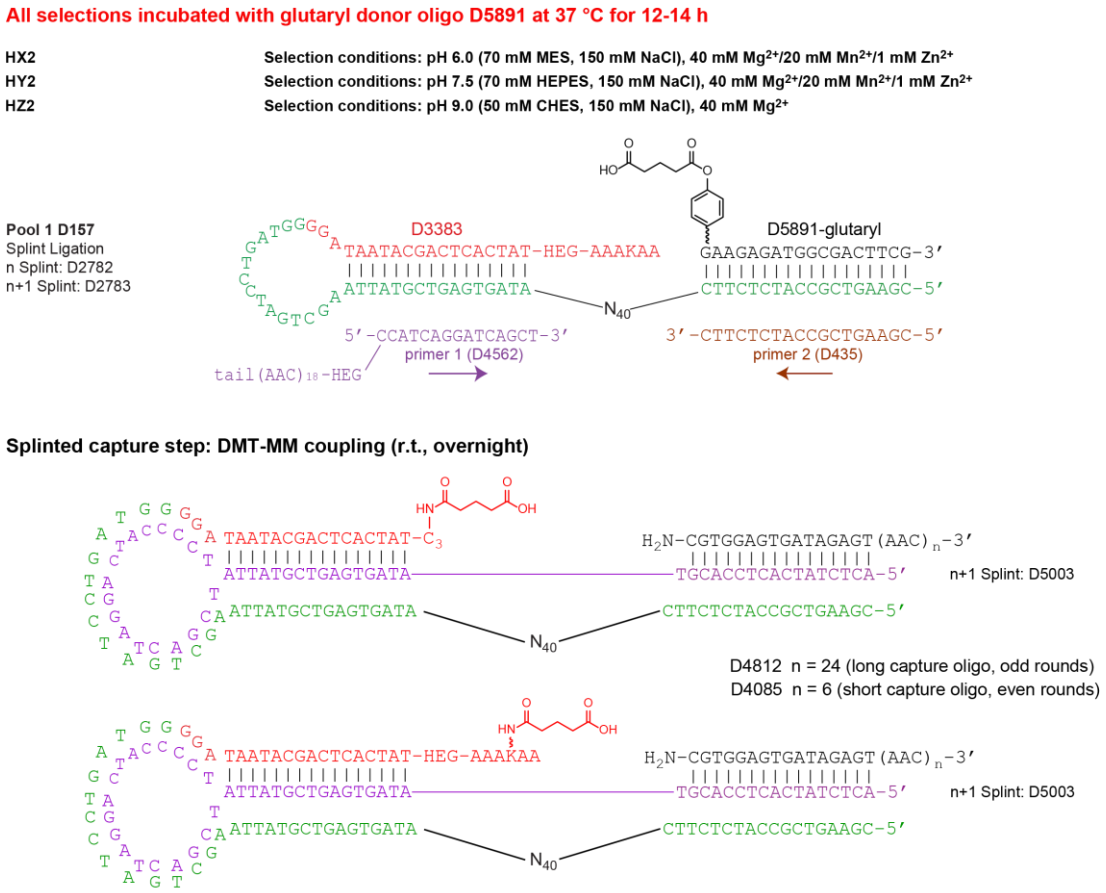


Figure 5.3. Incubation conditions of in vitro selections HX2-HZ2 and capture reactions.

These selection experiments used a capture reaction in which the carboxyl group of the glutaryl moiety transferred to the amine in the selection step was subsequently captured using a 5'-amino oligonucleotide, DMT-MM, and a DNA splint. At the end of each selection round, the sequence population was restored by PCR amplification using Taq polymerase, and subjected to subsequent iteration of rounds.

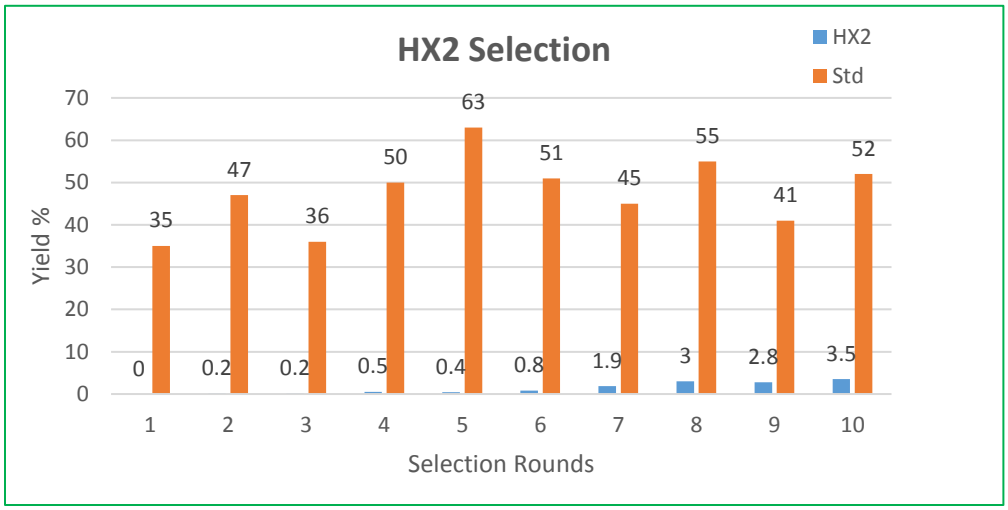
Distinguished from prior selections described in Chapter 2, the capture steps in selection HX2-HZ2 used two capture oligonucleotides with different length. One longer capture oligonucleotide was used in odd selection rounds, meanwhile a short capture oligonucleotide was applied in even rounds. The purpose of using two capture oligos with different size was to avoid false positive capture products due to aberrant migration of misfolded product sequences.

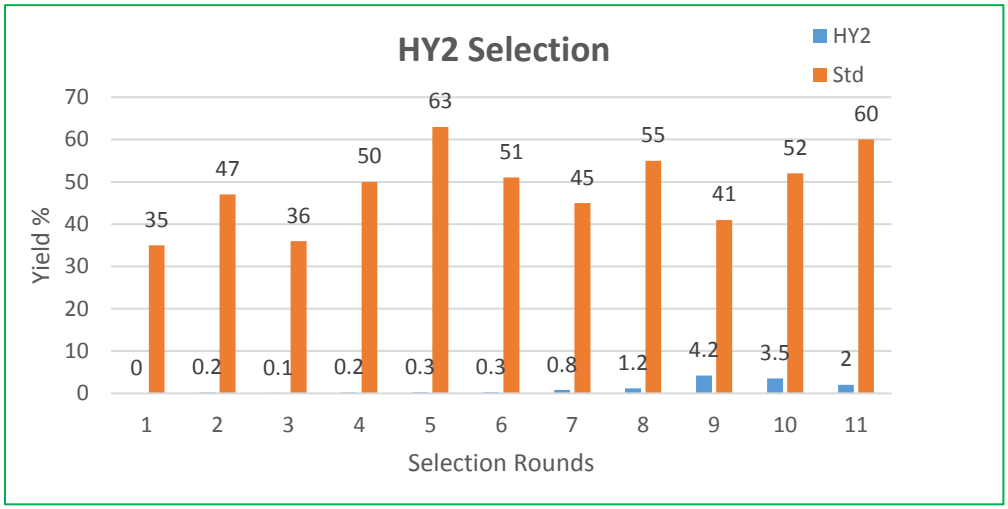
5.2 Results and Discussion

5.2.1 In Vitro Selection Progression Data of HX2-HZ2

Three in vitro selections HX2-HZ2 were performed for 10 or 11 rounds (Figure 5.4). Each selection used N_{40} random region DNA pool with canonical nucleotides (dNTP). All three selections used 5'-DNA-HEG-AAAKAA as their nucleophile and 5'-glutaryl donor oligonucleotide as acyl donor electrophile.

After the selection step, an extra 5'-amino oligonucleotide that serves as mass addition to the acylation product, for which enables a visible gel band shift, was introduced in the following capture reaction. The capture reaction was performed to capture the carboxyl handle within the transferred glutaryl moiety; meanwhile, glutarylated 5'-DNA-HEG-AAAKAA was assayed as capture standard. Their yields were shown in Figure 5.4.





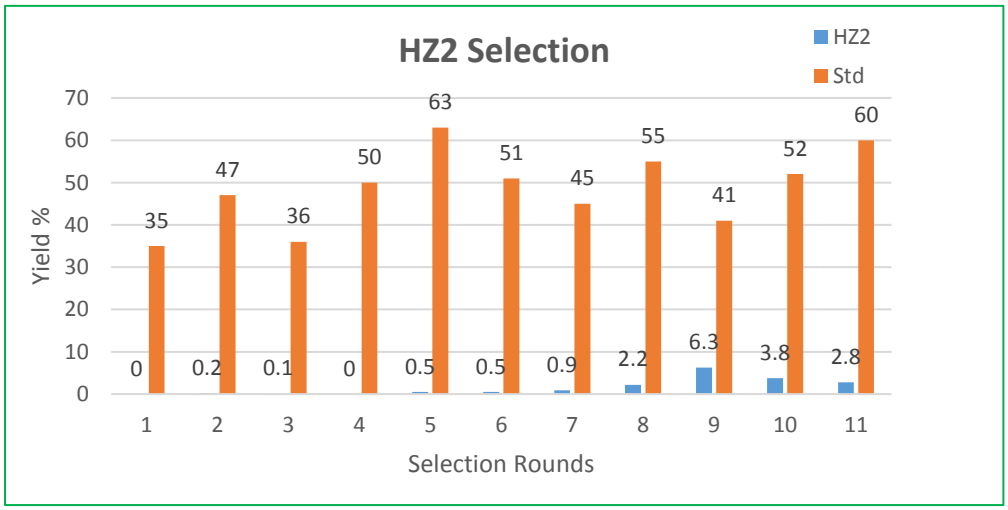


Figure 5.4. Selection progression data for selections HX2-HZ2 using 5'-DNA-HEG-AAAKAA as nucleophile and glutaryl donor oligonucleotide as electrophile.

The capture yield of selection HX2 was 0.2%-3.5% in 10 rounds, and the highest yield was 3.5% at round 10. The capture yield of selection HY2 was between 0.2%-4.2% in 11 rounds, and the highest yield was about 4.2% at round 9. The capture yield of selection HZ2 was between 0.1%-6.3% in 11 rounds, and the highest yield was about 6.3% at round 9. Thus, the corresponding round DNA pools having the highest yield from these three selections (10HX2, 9HY2, 9HZ2) were subject to in trans assays, without covalently ligating 5'-DNA-HEG-AAAKAA to the DNA sequence pool. Provided that the glutaryl moiety was successfully transferred onto DNA sequence pool, the additional glutaryl moiety changed the DNA sequence charge from +1 (NH_3^+) to -1 (COO^-). Since DNA runs from positive end to negative end, despite the trivial mass addition, a visible downward band shift (towards negative end) can be seen directly in the gel exposure assay using 20% gel and radiolabeled amine substrates. Thus, no capture reaction were performed for the in trans assays. However, in each instance, no in trans activity (<0.5%) was observed.

5.2.2 Glutarylation Activity of Existing DNAzymes

The eight pH 7.5 DNAzymes identified from selections FL2 and FN2 were assayed with their parent amine substrate 5'-DNA-C_3-NH_2 as nucleophile and glutaryl donor oligonucleotide as electrophile. Similarly, the seven DNAzymes identified from selections HB2, HC2 and HF2 were assayed with their parent amine substrate 5'-DNA-HEG-AAAKAA as nucleophile and glutaryl donor oligonucleotide as electrophile.

Three out of the eight pH 7.5 DNAzymes identified from selections FL2 and FN2 showed the capability of catalyzing substantial glutarylation of the 5'-DNA-C_3-NH_2 substrate. In contrast, none of the seven DNAzymes identified from selections HB2, HC2 and HF2 catalyzed observable glutarylation of 5'-DNA-HEG-AAAKAA above the splinted background. Thus, the three DNAzymes capable of catalyzing glutarylation, 8FL205, 8FL219, and 7FN221, were further assayed in kinetic experiments.

5.2.3 Kinetic Assays of Identified DNAzymes from Prior Selections

Since the originally designed in vitro selections HX2-HZ2 did not provide any DNAzyme with in trans catalytic activity for amine glutarylation, only the three already identified DNAzymes (8FL205, 8FL219, and 7FN221) from prior selections were tested in the in trans kinetic assays. These three DNAzymes were assayed under their parent selection condition (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C for 16 h), while in presence of 5'-glutaryl oligonucleotide donor rather than their parent acyl donor substrates (PE or 4FPE). The glutarylation yield in 48 h was as high as 19% for 8FL205, 14% for 8FL219 and 5.4% for 7FN221, as compared to the splinted background yield of <0.8%. These findings establish the feasibility of DNAzyme-catalyzed amine glutarylation.

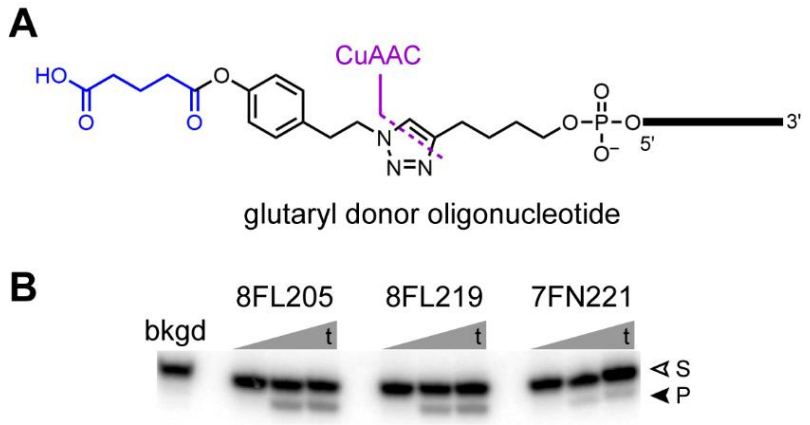


Figure 5.5. DNAzyme assays with the glutaryl donor. (A) Glutaryl donor oligonucleotide structure, as prepared by CuAAC between a synthesized glutaryl-azide small-molecule compound and a 5'-alkyne-modified oligonucleotide. The glutaryl fragment (blue) is transferred during an amine acylation reaction. (B) Assay results using the glutaryl donor with three DNAzymes identified by in vitro selection for acylation of DNA-C_3-NH_2 with the PE (8FL205, 8FL219) or 4FPE (7FN221) acyl donor substrate. Bkgd = complementary DNA splint in place of DNAzyme, to assess the uncatalyzed background glutarylation reaction. Incubation conditions: 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl at 37 °C. Shown are (cont.)

Figure 5.5. (cont.) representative timepoints (t = 0.5 min, 16 h, 48 h; S = substrate, P = product). The respective yields at 48 h were 19%, 14%, and 5.4%, versus <0.8% for splinted background (no product band detectable). For each DNAzyme, no product was detectable with unmodified DNA in place of DNA-C_3-NH_2. Experiment performed by current Silverman lab graduate student Jack J. Przybyla

5.2.4 Structural and Mechanistic Considerations

Using mfold,[1] we systematically predicted the secondary structures of all 28 of the new DNAzymes reported in this study. Each DNAzyme is predicted to have 1–7 plausible secondary structures, each with typically modest folding free energy in the range of –4 to –1 kcal/mol, although with three examples of folding free energies in the –9 to –5 kcal/mol range. In many cases, the various mfold-predicted secondary structures for a single DNAzyme are incompatible with one another. We have not endeavored to synthesize and study the large number of DNAzyme mutants (including covariations) that would be required to assess experimentally the validity of the predicted secondary structures. Importantly, doing so would not provide much if any useful, actionable information for our future studies of amine acylation DNAzymes. In parallel, the three-dimensional structures of these new DNAzymes are unknown, in the context that only two DNAzyme structures of any kind have been reported.[2,3] Future work will include comprehensive analysis of the secondary and tertiary structures and mechanisms of amine acylation DNAzymes.

5.3 Summary of thesis research

In Chapter 2, we learned that thioesters were too unreactive to support the emergence of DNAzymes. In Chapter 3, we learned that more electrophilic DMTE and TFPE were too reactive to allow the DNAzymes showing significant rate enhancement beyond splinted background reaction. In Chapter 4, we established experimentally that DNAzymes can catalyze amine acylation, including acylation of a Lys residue in a short DNA-anchored peptide.[4] Key to this success was identifying suitably substituted aryl esters (PE and 4FPE) as the electrophilic acyl donors, along with appropriate incubation conditions that balance electrophile stability and reactivity. In Chapter 5, we observed that several DNAzymes identified from other selections using aryl ester acyl donors are capable of catalyzing the conjugation of small molecule to amine group. The observation that several DNAzymes can function for amine glutarylation, even though these DNAzymes were not directly identified by selection for amine acylation using the glutaryl donor substrate, bodes well for the longer-term prospects of DNAzymes for amine and Lys acylation with biologically relevant small acyl groups.

For several other DNAzyme-catalyzed activities, we previously found that performing in vitro selection using a peptide substrate with mixed amino acid composition led to DNAzymes that require those specific peptide sequences in their substrates.[5-7] By analogy, we anticipate that for DNAzyme-catalyzed peptide Lys acylation, future selection experiments using mixed-composition Lys-containing peptides will provide sequence-selective Lys-acylating DNAzymes, including those that function with free peptide substrates when an appropriate selection pressure is imposed.[5] Expanding the substrate tolerance of such DNAzymes from peptides to proteins is a further challenge, but worth undertaking considering the difficulty inherent to achieving nonenzymatic site-selective Lys modification of native proteins.[8-14]

5.4 Materials and Methods

5.4.1 Preparation of Oligonucleotides

DNA oligonucleotides and oligonucleotide conjugates were obtained from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an ABI 394 instrument using reagents from Glen Research. All oligonucleotides and conjugates were purified by 7 M urea denaturing PAGE with running buffer 1x TBE (89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3), extracted from the polyacrylamide with TEN buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaCl), and precipitated with ethanol as described previously.[15,16] Hexapeptide AAAK(tfa)AA was prepared by solid-phase synthesis using Fmoc Rink amide MBHA resin as described.[17] The trifluoroacetyl protecting group was removed after conjugation to the DNA anchor oligonucleotide. The coupling reagents DMT-MM and EDC/TFP were commercially available (Fisher Scientific).

Conjugation of AAAKAA to DNA anchor oligonucleotide. The DNA anchor oligonucleotide was 5'-GGATAATACGACTCACTAT-HEG-rA-3', where the 3'-terminal ribonucleotide was oxidized by $NaIO_4$ and used for conjugation to the peptide N-terminus. A 100 µL sample containing 1.0 nmol of DNA anchor oligonucleotide in 100 mM HEPES, pH 7.5, and 10 mM $NaIO_4$ was incubated at room temperature for 1 h. The oxidized product was precipitated to remove excess $NaIO_4$ by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The precipitated product was dissolved in 65 µL of water and used directly in the next step. A 100 µL sample containing the $NaIO_4$-oxidized DNA anchor oligonucleotide and 100 nmol (100 equiv) of AAAK(tfa)AA hexapeptide in 100 mM NaOAc, pH 5.2, 50 mM $NiCl_2$, and 10 mM $NaCNBH_3$ was incubated at 37 °C for 14 h. The DNA-anchored hexapeptide was precipitated by addition of 10 µL of 3 M NaCl and 330 µL of ethanol. The trifluoroacetyl protecting group was removed by incubation in 100 µL of 30% NH_4OH for 1 h at room temperature. NH_4OH was removed in a speedvac and the

conjugation product was purified by 20% PAGE.

Synthesis of glutaryl-azide compound. 4-(2-Azidoethyl)phenol was prepared from tyrosol in two steps as described.[18] Then, 4-(2-azidoethyl)phenol was esterified using glutaric anhydride and TMSOTf by adaptation of a reported method,[19] forming the glutaryl-azide compound, formally, 5-(4-(2-azidoethyl)phenoxy)-5-oxopentanoic acid. A mixture of 4-(2-azidoethyl)phenol (166 mg, 1.02 mmol) and glutaric anhydride (125 mg, 1.10 mmol) was dissolved in CH_2Cl_2 (2 mL). Neat TMSOTf (3.6 μL, 0.02 mmol) was added, and the sample was stirred at room temperature for 2 h. The reaction was monitored by TLC (CH_2Cl_2:MeOH 6:1); the R_f of 4-(2-azidoethyl)phenol is 0.65, and the R_f of the glutaryl-azide compound is 0.56. Additional CH_2Cl_2 (3 mL) and saturated aqueous $NaHCO_3$ (5 mL) were added. The organic layer was separated, extracted twice with saturated aqueous NaCl (5 mL), and dried over $MgSO_4$. The solvent was evaporated to give the crude glutaryl-azide compound as a yellow oil (216 mg, ≤77% yield; ESI HRMS *m*/*z* calcd. for $[M+Na]^+$ 300.0960, found 300.0957), which was used directly in the CuAAC conjugation reaction with 5'-alkyne DNA to form the glutaryl donor oligonucleotide. This synthesis was performed by previous Silverman Lab graduate student Peter Yeh.

Preparation of glutaryl donor oligonucleotide. The glutaryl donor oligonucleotide (Figure 5.1) was prepared by glutaryl-azide compound and subsequent CuAAC reaction with 5'-alkyne-C_4-GAAGAGATGGCGACTTCG-3' (5'-alkyne modifier from Glen Research). For this purpose, 2.2 mg of the glutaryl-azide compound was dissolved in 30 μL of DMSO to make a 250 mM stock solution. The CuAAC conjugation was performed as followed. A 20 μL sample containing 1.37 nmol of 5'-alkyne DNA oligonucleotide, 250 nmol of glutaryl-azide compound (12.5 mM), 40 mM THPTA, 5 mM $CuCl_2$, and 10 mM sodium ascorbate was incubated in 100 mM HEPES, pH 7.0 at room temperature for 30 min. The sample was extracted with phenol/chloroform, precipitated with ethanol, and quantified by UV absorbance (A_{260}). The isolated yield of glutaryl donor oligonucleotide was 1.15 nmol (84%).

5.4.2 Mass Spectrometry of Oligonucleotides

Data were acquired on a Bruker UltrafleXtreme MALDI-TOF mass spectrometer with matrix 3-hydroxypicolinic acid in positive ion mode at the UIUC School of Chemical Sciences Mass Spectrometry Laboratory. All *m*/*z* values are for $[M+H]^+$. Samples were desalted by Millipore C_{18} ZipTip before analysis.

Data for oligonucleotides and conjugates were as follows. For those that bind to the left-hand DNAzyme binding arm, the DNA sequence was 5'-GGATAATACGACTCACTAT-3'. For those that bind to the right-hand DNA binding arm, the DNA sequence was 5'-GAAGAGATGGCGACTTCG-3'.

Oligonucleotides and conjugates that bind to left-hand DNAzyme binding arm

DNA-C_3-NH_2 *m*/*z* calcd. 5934.9, found 5931.1, Δ= –0.06%

DNA-HEG-AAAKAA *m*/*z* calcd. 6934.6, found 6938.9, Δ= +0.06%

Oligonucleotides and conjugates that bind to right-hand DNAzyme binding arm

5'-alkyne-C_6-DNA *m*/*z* calcd. 5750.8, found 5747.6, Δ= –0.04%

glutaryl donor oligo *m*/*z* calcd. 6028.1, found 6028.5, Δ= +0.007%

The products of several representative individual DNAzymes were analyzed by MALDI mass spectrometry. Each product was prepared from a 21 μL sample containing 200 pmol of DNA-C_3-NH_2 or DNA-HEG-AAAKAA substrate, 220 pmol of DNAzyme, and 240 pmol of acyl donor substrate. The sample was annealed in 5 mM HEPES, pH 7.5 or CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The DNAzyme-catalyzed reaction was initiated by bringing the sample to 30 μL total volume containing 70 mM

HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C for 48 h, precipitated with ethanol, and purified by 20% PAGE.

DNAzyme products

8FL205 with DNA-C_3-NH_2, glutaryl donor, pH 7.5 *m*/*z* calcd. 6049.0, found 6047.2, Δ= –0.03%

8FL219 with DNA-C_3-NH_2, glutaryl donor, pH 7.5 *m*/*z* calcd. 6049.0, found 6048.5, Δ= –0.008%

7FN221 with DNA-C_3-NH_2, glutaryl donor, pH 7.5 *m*/*z* calcd. 6049.0, found 6047.9, Δ= –0.02%

5.4.3 In Vitro Selection Procedure Steps

The key selection step of each round using a 5'-glutaryl oligonucleotide donor substrate is shown in Figure 5.2. The DNA anchor oligonucleotide sequence was 5'-GGATAATACGACTCACTAT-3'. The 5'- glutaryl oligonucleotide donor sequence was 5'-GAAGAGATGGCGACTTCG-3'. The random DNAzyme pool was 5'-CGAAGTCGCCATCTCTTC-N_{40}-ATAGTGAGTCGTATTAAGCTGATCCTGATGG -3'. PCR primers were 5'-CGAAGTCGCCATCTCTTC-3' (forward primer) and 5'-$(AAC)_{18}$XCCATCAGGATCAGCT-3', where X is the HEG spacer to stop Taq polymerase (reverse primer). In each round, the ligation step to attach the DNAzyme pool at its 3'-end with the 5'-end of the amine-containing substrate was performed using a DNA splint and T4 DNA ligase. The splint sequence was 5'-ATAGTGAGTCGTATTATCCCCATCAGGATCAGCTTAATACGACTCACTAT-3'.

Procedure for ligation step in round 1. A 25 µL sample containing 600 pmol of DNA pool,

750 pmol of DNA splint, and 900 pmol of DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 3 µL of 10x T4 DNA ligase buffer (400 mM Tris, pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, and 5 mM ATP) and 2 µL of 5 U/µL T4 DNA ligase (Thermo Fisher). The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for ligation step in subsequent rounds. A 17 µL sample containing the PCR-amplified DNA pool (~5–10 pmol), 30 pmol of DNA splint, and 50 pmol of DNA-HEG-AAAKAA substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. To this solution was added 2 µL of 10x T4 DNA ligase buffer and 1 µL of 1 U/µL T4 DNA ligase. The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for selection step in round 1. Each selection experiment was initiated with 200 pmol of the ligated pool. A 20 µL sample containing 200 pmol of ligated pool and 300 pmol of 5'-glutaryl oligonucleotide donor was annealed in 5 mM MES, pH 6.0, 15 mM NaCl, and 0.1 mM EDTA or 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA or 5 mM CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 40 µL total volume containing 70 mM MES, pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The Mn^{2+} was added from a 10X stock solution containing 200 mM $MnCl_2$. The Zn^{2+} was added from a 10x stock solution containing 10 mM $ZnCl_2$, 20 mM HNO_3, and 200 mM HEPES at pH 7.5; this stock solution was freshly prepared from a 100x stock of 100 mM $ZnCl_2$ in 200 mM HNO_3. The metal ion stocks were added last to the final sample. The sample was incubated at 37 °C for 16 h and separated by 8% PAGE.

Procedure for selection step in subsequent rounds. A 14 μL sample containing the ligated pool and 30 pmol of 5'-glutaryl oligonucleotide donor was annealed in 5 mM MES, pH 6.0, 15 mM NaCl, and 0.1 mM EDTA or 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA or 5 mM CHES, pH 9.0, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 20 μL total volume containing 70 mM MES, pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl or 50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl. The sample was incubated at 37 °C for 16 h and separated by 8% PAGE.

Procedure for capture step in round 1. A 90 μL sample containing the selection product, 300 pmol of DNA splint, and 400 pmol of 5'-amino capture oligonucleotide was annealed in 100 mM MOPS, pH 7.0, and 1 M NaCl by heating at 95 °C for 3 min and cooling on ice for 5 min. The capture reaction was initiated by bringing the sample to 100 μL total volume containing 100 mM MOPS, pH 7.0, 1 M NaCl, and 50 mM DMT-MM. The sample was incubated at 37 °C for 12 h.

Procedure for capture step in subsequent rounds. A 20 μL sample containing the selection product, 50 pmol of DNA splint, and 100 pmol of 5'-amino capture oligonucleotide was annealed in 100 mM MOPS, pH 7.0, and 1 M NaCl by heating at 95 °C for 3 min and cooling on ice for 5 min. The capture reaction was initiated by bringing the sample to 25 μL total volume containing 100 mM MOPS, pH 7.0, 1 M NaCl, and 50 mM DMT-MM. The sample was incubated at 37 °C for 12 h.

Procedure for PCR. In each selection round, two PCR reactions were performed, 10-cycle PCR followed by 30-cycle PCR. First, a 100 μL sample was prepared containing the PAGE-purified selection product, 200 pmol of forward primer, 50 pmol of reverse primer, 20 nmol of each dNTP, and 10 μL of 10X Taq polymerase buffer [1X = 20 mM Tris-HCl, pH 8.8, 10 mM

$(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, and 0.1% Triton X-100]. This sample was cycled 10 times according to the following PCR program: 94 °C for 2 min, 10x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Taq polymerase was removed by phenol/chloroform extraction. Second, a 50 µL sample was prepared containing 1 µL of the 10-cycle PCR product, 100 pmol of forward primer, 25 pmol of reverse primer, 10 nmol of each dNTP, 20 µCi of α-^{32}P-dCTP (800 Ci/mmol), and 5 µL of 10x Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. Samples were separated by 8% PAGE.

Cloning and screening of individual DNAzymes. The PCR primers used for cloning were 5'-CGAAGTCGCCATCTCTTC-3' (forward primer; same as in selection) and 5'-TAATTAATTAATTACCCATCAGGATCAGCT-3' (reverse primer). The 10-cycle PCR product from the appropriate selection round was diluted 1000-fold. A 50 µL sample was prepared containing 1 µL of the diluted 10-cycle PCR product from the appropriate selection round, 100 pmol of forward cloning primer, 25 pmol of reverse cloning primer, 10 nmol of each dNTP, and 5 µL of 10X Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30x (94 °C for 30 s, 47 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min. The sample was separated by 1% agarose gel and extracted using a GeneJET Gel Extraction Kit (Fermentas). The extracted product was quantified by absorbance (A_{260}) and diluted to 5–10 ng/µL. A 4 µL portion of the diluted PCR product was inserted into the pCR2.1-TOPO vector using a TOPO TA cloning kit (Life Technologies). Individual *E. coli* colonies harboring plasmids with inserts were identified by blue-white screening and grown in LB/amp media. Miniprep DNA was prepared using a GeneJET Plasmid Miniprep Kit (Fermentas) and screened for properly sized inserts by EcoRI digestion and agarose gel analysis. Before sequencing, assays of individual DNAzyme clones were performed with PAGE-purified DNA strands prepared by PCR from the miniprep DNA, using the single-turnover assay procedure that is described in Chapter 4.

5.5 References

(1) Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406-3415.

(2) Ponce-Salvatierra, A.; Wawrzyniak-Turek, K.; Steuerwald, U.; Höbartner, C.; Pena, V. Crystal structure of a DNA catalyst. *Nature* **2016**, *529*, 231-234.

(3) Liu, H.; Yu, X.; Chen, Y.; Zhang, J.; Wu, B.; Zheng, L.; Haruehanroengra, P.; Wang, R.; Li, S.; Lin, J.; Li, J.; Sheng, J.; Huang, Z.; Ma, J.; Gan, J. Crystal structure of an RNA-cleaving DNAzyme. *Nat. Commun.* **2017**, *8*, 2006.

(4) Yao, T.; Przybyla, J. J.; Yeh, P.; Woodard, A. M.; Nilsson, H. J.; Brandsen, B. M.; Silverman, S. K. DNAzymes for amine and peptide lysine acylation. *Org. Biomol. Chem.* **2021**, *19*, 171-181.

(5) Chu, C.; Wong, O.; Silverman, S. K. A Generalizable DNA-Catalyzed Approach to Peptide-Nucleic Acid Conjugation. *ChemBioChem* **2014**, *15*, 1905-1910.

(6) Walsh, S. M.; Konecki, S. N.; Silverman, S. K. Identification of Sequence-Selective Tyrosine Kinase Deoxyribozymes. *J. Mol. Evol.* **2015**, *81*, 218-224.

(7) Wang, P.; Silverman, S. k. DNA-Catalyzed Introduction of Azide at Tyrosine for Peptide Modification. *Angew. Chem. Int. Ed.* **2016**, *55*, 10052-10056.

(8) Takaoka, Y.; Nishikawa, Y.; Hashimoto, Y.; Sasaki, K.; Hamachi, I. Ligand-directed dibromophenyl benzoate chemistry for rapid and selective acylation of intracellular natural proteins. *Chem. Sci.* **2015**, *6*, 3217-3224.

(9) Chilamari, M.; Kalra, N.; Shukla, S.; Rai, V. Single-site labeling of lysine in proteins through a metal-free multicomponent approach. *Chem. Commun.* **2018**, *54*, 7302-7305.

(10) Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V. Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **2018**, *140*, 15114-15123.

(11) Forte, N.; Benni, I.; Karu, K.; Chudasama, V.; Baker, J. R. Cysteine-to-lysine transfer antibody fragment conjugation. *Chem. Sci.* **2019**, *10*, 10919-10924.

(12) Mortensen, M. R.; Skovsgaard, M. B.; Gothelf, K. V. Considerations on probe design for affinity guided protein conjugation. *ChemBioChem* **2019**, *20*, 2711-2728.

(13) Yamada, K.; Ito, Y. Recent Chemical Approaches for Site-Specific Conjugation of Native Antibodies: Technologies toward Next-Generation Antibody-Drug Conjugates. *ChemBioChem* **2019**, *20*, 2729-2737.

(14) Nielsen, T.; Märcher, A.; Drobňáková, Z.; Hučko, M.; Štengl, M.; Balšánek, V.; Wiberg, C.; Nielsen, P. F.; Nielsen, T. E.; Gothelf, K. V.; Cló, E. Disulphide-mediated site-directed modification of proteins. *Org. Biomol. Chem.* **2020**, *18*, 4717-4722.

(15) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. Deoxyribozymes with 2'− 5' RNA ligase activity. *J. Am. Chem. Soc.* **2003**, *125*, 2444-2454.

(16) Wang, Y.; Silverman, S. K. Characterization of deoxyribozymes that synthesize branched RNA. *Biochemistry* **2003**, *42*, 15252-15263.

(17) Story, S. C.; Aldrich, J. V. Preparation of protected peptide amides using the Fmoc chemical

protocol: Comparison of resins for solid phase synthesis 1. *Int. J. Pept. Protein Res.* **1992**, *39*, 87-92.

(18) Rosario-Amorin, D.; Gaboyard, M.; Clérac, R.; Vellutini, L.; Nlate, S.; Heuzé, K. Metallodendritic grafted core-shell g-Fe_2O_3 nanoparticles used as recoverable catalysts in Suzuki C-C coupling reactions. *Chem. Eur. J.* **2012**, *18*, 3305-3315.

(19) Procopiou, P. A.; Baugh, S. P. D.; Flack, S. S.; Inglis, G. G. A. An Extremely Powerful Acylation Reaction of Alcohols with Acid Anhydrides Catalyzed by Trimethylsilyl Trifluoromethanesulfonate. *J. Org. Chem.* **1998**, *63*, 2342-2347.