DEVELOPMENT AND EXECUTION OF SELECTION METHODS FOR LYSINE-ACYLATING DEOXYRIBOZYMES

BY

JOHN J. (JACK) PRZYBYLA

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biochemistry
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Adviser:

Professor Scott K. Silverman

ABSTRACT

DNA catalysts (deoxyribozymes, or DNAzymes) represent an exciting possibility in the world of novel catalysis. Their ability to be amplified via polymerase chain reaction (PCR) allows for *in vitro* selection of DNAzymes that catalyze specific reactions. Through the use of *in vitro* selection, DNAzymes have been identified for a growing number of reactions, including lysine side chain acylation. However, these acylating DNAzymes make use of model substrates that are attached to a complementary DNA anchor that allows for recruitment to a DNAzyme active site. These highly specific substrates do not resemble the large protein or small-molecule substrates that act as valuable targets for novel catalysts. As a result, these DNAzymes do not yet possess the level of utility that makes novel catalysts exciting. This body of work, conducted in the Silverman lab at UIUC, looks to build upon previous discoveries by establishing a selection scheme for lysine acylation in a non-anchored peptide. This includes the development of more diverse and complex substrates, the development and testing of selection conditions, and the development of a necessary "capture" step to enable the selection process. As a precursor to my ultimate project, my contributions towards the identification of amine acylating, and later lysine side chain acylating DNAzymes, will also be included.

At time of writing, the selection methods discussed here are being employed in the Silverman lab to identify DNAzymes that catalyze the acylation of lysine on a non-anchored peptide. If these DNAzymes are successfully identified, they would act as another crucial step towards the identification of "useful" novel DNA catalysts.

ACKNOWLEDGEMENTS

I'd like to thank my mentor for the last 3 years, Scott Silverman, for his incredible insight, unwavering guidance, and compassion. Thank you for making things seem possible when I felt overwhelmed.

My lab mates Prakriti, Shukun, Robert, Morgan, Caroline, and Yves for their kind spirits and good company. You all made the hours of lab work and troubleshooting survivable, and I trust you to teach these new kids coming in how it's done.

John and Suzanne Przybyla, my parents who have shown me nothing but an outpouring of love and support since I began to live by myself and for myself. This road has been complicated and difficult, and it means the world that you've been by my side through it all.

My friends all across the states and the globe, for distracting me with conversation, artwork, projects, and company when it all felt like too much to handle. I have you all for keeping me sane.

Thank you to Prakriti for sharing your successes, your struggles, and your heart.

Thank you to Shukun for sharing your dreams, your work, and your joy.

I'll be seeing you all as we reach ever so closer to the futures we promised ourselves.
Until then!

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Purpose

The purpose of this research has been to identify and characterize single-stranded DNA (ssDNA) that is capable of catalyzing the site-specific addition of an acyl group onto a lysine residue side chain. Deoxyribozymes, or DNAzymes, have been found to catalyze a wide range of reactions[1] including oligonucleotide cleavage/ligation, phosphorylation of oligonucleotides and peptides, and amide cleavage. DNAzymes have also demonstrated sequence-specificity in their catalysis involving peptide substrates.[2] This potentially presents DNAzymes as a tool to be used by researchers investigating the effects of post-translational modifications (PTMs), including the acylation of lysine residues. My project sought to identify DNAzymes that perform sequence-specific lysine acylation of a model peptide, with the ultimate intent of identifying DNAzymes that could perform similar catalysis on more complex protein substrates.

DNA as a Catalyst

The conventional characterization of DNA in biology is as an information carrier, with proteins acting as the primary source of biologically relevant catalysis. RNA stands between the two, both metaphysically and chemically; while mRNA acts as an information carrier between DNA and ribosomes, the ribosome itself represents the ability of RNA to act as a catalyst for amide bond formation. Naturally occurring ribozymes have also been found to be capable of catalyzing self-cleavage[3] and phosphodiester transfer.[4] The chemical composition of ribonucleotides (Fig. 1) allows for both covalent[5] and general acid/base[6] catalysis, hydrogen bonding, and the formation of secondary and tertiary structures. The composition of DNA is notably similar, lacking the 2'-hydroxyl group but otherwise maintaining all other functional groups, and thereby theoretically confers the same mechanisms of catalysis.

DNA represents a remarkable candidate for a catalyst in the specific realm of novel catalysis. At present, arguably the most promising approach to novel enzyme development is on the protein side. Directed evolution of proteins allows for the development of novel enzymes using natural proteins as precursors[7] while computational design allows for the data-directed development of novel active sites,[8] though these active sites must be embedded within a pre-existing protein "body" to function. These approaches are both limited in their necessary reliance on existing proteins, as well as current difficulties in predicting protein folding patterns.[9] The ability to perform an *in vitro* selection experiment on a pool of randomly synthesized proteins could theoretically circumvent these restrictions, though such an approach is not feasible. A majority of randomly synthesized proteins would be unlikely to contain sequences that would properly fold in solution, and the inability to amplify peptide sequences would lead to a prohibitive loss of material after even a single round of selections Both DNAzymes and ribozymes possess the necessary attributes to allow for successful *in vitro* selection experiments, including the ability to be amplified via PCR. This allows for pools to be replenished after selection, thereby allowing for iterated enrichment of catalytically active sequences. Additionally, all conventional nucleobases readily interact in water, thus allowing for the likely formation of some higher order structure in any random sequence. Lastly, random DNA/RNA sequences occupy a smaller sequence space (4^n) than proteins (20^n), thus allowing a greater fraction of any random pool of length "n" to be assayed during selections. Using *in vitro* selection, scientists have been able to identify nucleic acid catalysts that catalyze a number of non-biological reactions.[1,2] The choice between RNA or DNA as a nucleic acid catalyst is largely up to preference,[10] though the use of

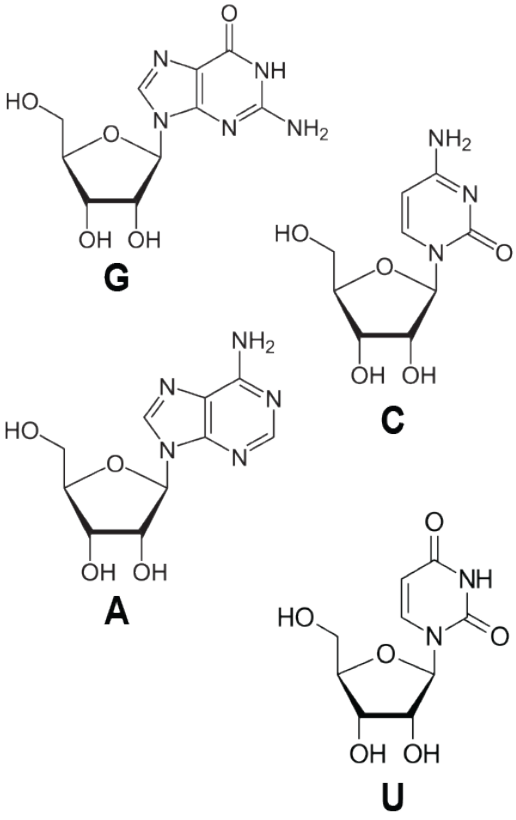

Fig. 1. Structure of ribonucleosides guanosine (G), cytidine (C), adenosine (A), and uridine (U). The abundance of oxygen and nitrogen groups allows for hydrogen bonding, acid/base catalysis, and covalent catalysis.[5]

DNA circumvents the additional transcription and reverse transcription steps of PCR amplification of RNA.

Lysine Acylation

The goal of my project was to identify DNAzymes that catalyze lysine acylation. Lysine acylation describes a large number of reactions associated with cellular signalling,[11] gene expression,[12] enzyme mechanism regulation,[13] and more. Histone acetylation alone has been the target of research inquiries for several decades.[14] Additionally, new acylation pathways, including crotonylation,[15] have been discovered in biological systems and are currently under investigation. The ability to induce site-specific acylations onto protein lysine residues *in vitro* would be an invaluable tool in investigating these reactions and the effects they have. Unfortunately, most *in vitro* methods invite complications. Chemical methods, while efficient, lack specificity. Enzymatic methods rely on the existence of an appropriate enzyme that must be identified, isolated, and made to function in an *in vtiro* environment; needless to say, this process invites several steps that would impede investigations. The *de novo* identification of site-specific lysine-acylating DNAzymes would aid researchers in their investigations of acylation pathways, both old and new, by providing an enzymatic tool that is both specific and readily available.

***In Vitro* Selection**

In vitro selection begins with the synthesis of a random pool. Each DNA molecule in this pool contains a random region of X nucleotides that is flanked on either side by non-random regions (Fig. 2). The purpose of the non-random regions is two-fold. The first is to provide consistent regions for complementary primers to bind during PCR. The second is to provide templates to which DNA-anchored substrates can be recruited to the active site by Watson-Crick base pairing. This anchoring reduces the entropic cost of the reaction, thereby reducing the catalytic load placed on DNAzymes in the pool. The purpose of the random region is to provide an active site. The random region length X introduces another serious consideration – the longer the random region, the more potentially complex the resulting active sites However, the ability of a reasonably sized pool of random sequences to contain each random sequence in its sequence space drops exponentially as the random region length increases. All of my selection schemes made use of a 40 nucleotide (N_{40}) random region, in a pool made to contain ~10^{14} individual

sequences. While this pool size is inherently lacking ($4^{40} \approx 10^{24}$ unique sequences), it has been historically successful in the Silverman lab.[1] Additional selections were conceived using a 20 nucleotide (N_{20}) random region (~10^{12} unique sequences, which is fully contained within the ~10^{14} pool), though my own work never makes use of this pool.

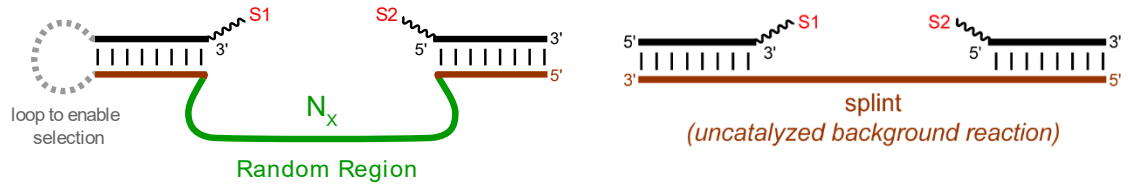

Fig. 2. A generalized structure of the random pool populations used in *in vitro* selection, with random region length X. S1 and S2 denote substrates in the reaction for which DNA sequences are being selected. A ligated loop (~7 nucleotides) is added during the selection process to allow for isolation of reactive sequences by mass shift. Background reactivity is determined using a DNA splint that resembles the non-random regions of pool sequences.

Selections are carried out in a variety of solution conditions which can vary in incubation time, incubation temperature, pH, and metal cofactor availability. At the end of each selection round, sequences for which the reaction occurred are isolated via PAGE mass shift. This population will include both catalytic sequences and sequences for which the reaction occurred uncatalyzed, the latter of which vastly outnumbers the former in early selection rounds. The pool is then amplified using PCR, allowing for additional rounds of selection. Each round of selections enriches the pool with catalytically active sequences while a large number of non-catalytic sequences are discarded. This process is repeated until a substantial yield above background is observed, at which point the sequences are cloned and sequenced; these isolated sequences can then be characterized.

My experiments necessitated a number of additional considerations for the acyl donor, the lysine substrate, and the method by which catalytically active sequences could be isolated. One of the two substrates invariably needed to be tethered in some way to the DNAzyme, as the formation of a DNAzyme-conjugated single turnover product would make reactive sequences distinguishable from non-reactive sequences. Similarly, the substrates had to be capable of providing a substantial mass shift. The approach to these problems varied depending on the selection experiment being performed and will thus be explained in detail during the discussion of each individual experiment.

Preceding Experiments: Amine Acylation

Investigations into lysine-acylating DNAzymes began with selections conducted by senior graduate student Tianjiong (Yves) Yao.[16] This selection made use of a thioester acyl donor (Fig. 3) and a "simple" amine substrate (Fig. 4), both conjugated to DNA oligonucleotide complementary to one of the non-random regions of the pool DNA. The amine group acts as a nucleophile, attacking the thioester carbonyl electrophile. Upon acylation, the thiol-DNA would act as a leaving group, leading to the conjugation of the attached acyl group to the amine. The acyl group itself was modified with a carboxylic acid, allowing for a 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMT-MM) activated "capture" step with a 5'-amino modified oligonucleotide. This capture step would provide a substantial enough mass shift to isolate catalytically active sequences. All selections were conducted using an N_{40} random region length in a variety of pH and metal ion availability conditions. The intent of these earlier selections was to perform an uncomplicated amine acylation using an S-to-N donor, which is conventionally used in a number of synthetic applications including protein synthesis and modification.[17] Additionally, a thioester acyl donor is chemically similar to the coenzyme A acyl donors often used in biological systems, and thus seemed like an enticing candidate electrophile. However, all selections failed to show a significant yield above background after 10 rounds and were thus abandoned.

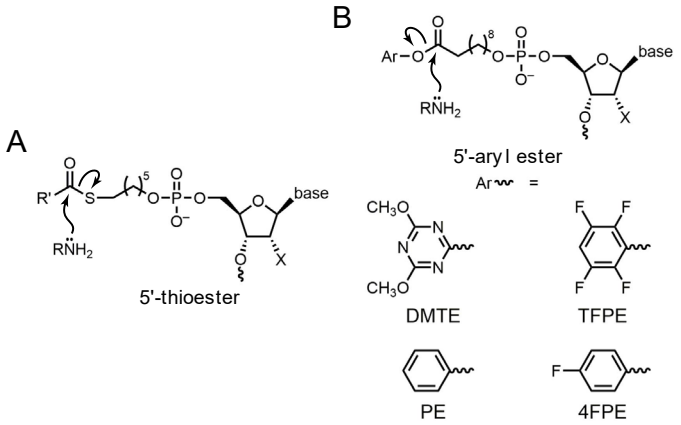

Fig. 3. Structures of reactive ester electrophiles used for *in vitro* selection of amine acylating DNAzymes. (A) Thioester acyl donor oligonucleotide used in the earliest phase of selections. No DNAzymes were identified using this acyl donor. (B) Aryl ester acyl donor oligonucleotides used in later selections. DMTE = 4,6-dimethoxy-1,3,5-triazin-2-yl ester; TFPE = 2,3,5,6-tetrafluorophenyl ester; PE = phenyl ester; 4FPE = 4-fluorophenyl ester.

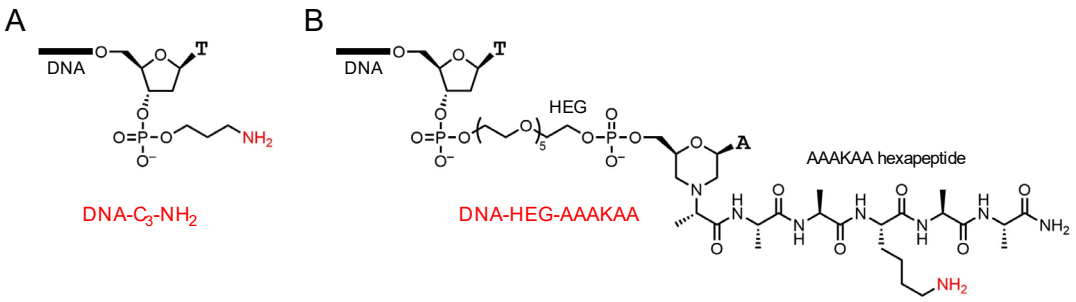

Fig. 4. Structures of amine nucleophiles used for *in vitro* selection of amine acylating DNAzymes. (A) "Simple" amine substrate with a C_3 tether. (B) Model peptide AAAKAA substrate with a HEG tether. Substrate was formed by reductive amination of the free α-NH_2 of the AAAKAA peptide. The HEG tether is 35 atoms longer than the C_3 tether.

A new set of selections was created using a collection of more reactive aryl ester electrophiles (Fig. 3) and the same simple amine nucleophile. The more reactive of these electrophiles, DMTE and TFPE, led to relatively high formation of background product at 34% and 15% in 0.5 min respectively. Strong pool reactivity was observed as early as round 3, though no DNAzyme sequenced from these selections exhibited a catalytic yield substantially above background. It can be argued that these DNAzymes were recapitulating a DNA splint (which was used to determine background reactivity) and were thereby non-catalytic. Selections with the less reactive PE and 4FPE were more successful, exhibiting substantial yield above background (20%~70% observed vs. 3.5% background) by rounds 7-8 in a variety of conditions. DNAzymes from these selections were cloned, sequenced, and assayed for metal ion dependency and general kinetics, with their highest rate enhancements of ~10^3.

When a selection is successful, the next step is to increase the complexity of the substrates in the direction of the ultimate goal. Since the goal was the acylation of proteins, a DNA-tethered model peptide DNA-HEG-AAAKAA (Fig. 4) was made to replace the simple amine substrate. Predictably, no DNAzymes from the above-mentioned selections exhibited a substantial yield above background using this new substrate; they had not been selected for such a function. As such, a new set of selections using the aryl ester acyl donors and the model peptide DNA-HEG-AAAKAA were conducted. Maximal pool reactivity was observed around round 11 in various conditions, once again for the less reactive PE and 4FPE aryl esters. Individual DNAzymes were cloned, sequenced, and assayed as before, exhibiting rate enhancements of ~10^2.

Ideally, DNAzymes would be capable of performing acylations on peptides or proteins in solution without the use of a DNA tether. To this end, the above DNAzymes were assayed for their ability to catalyze the acylation of the untethered AAAKAA peptide, though no catalytic activity was observed. This is likely due to the DNA tether being present throughout the selection process; the identified DNAzymes were not selected to function without it. A new set of selections using an untethered peptide substrate would need to be conducted to identify untethered peptide acylating DNAzymes. Though the lack of a DNA tether significantly increases the entropic cost of the reaction, DNAzymes have been found to be capable of catalyzing tyrosine-RNA conjugation reactions on free peptide substrates.[18,29-31] It is therefore reasonable to believe that acylation of a non-tethered peptide lysine is feasible.

Preceding Experiments: Early Glutarylation Experiments

Though Yves was able to identify DNAzymes that catalyzed acylation using a large oligonucleotide as the acyl group, most if not all biologically relevant protein acylations make use of a small molecule acyl group. Examples include acetylation,[14] succinylation,[19] malonylation,[20] and more.[21] Ideally, DNAzymes would be able to catalyze acylation reactions using small-molecule acyl groups instead of a large acyl-oligonucleotide. To this end, a new acyl donor was developed (Fig. 5), a glutaryl donor specifically. This donor was made via a glutaryl azide compound conjugated to an alkyne-modified DNA tether by CuAAC (copper-catalyzed azide–alkyne cycloaddition). The aryl ester functional group in this new acyl donor is inverted, resulting in the DNA tether acting as the leaving group while the glutaryl group alone becomes conjugated to the amine nucleophile. The glutaryl group itself was chosen due to the presence of an exposed carboxylic acid after acylation which can be targeted for a DMT-MM activated capture using an amino-modified oligonucleotide. The formation of an amide bond between the glutaryl group and the capture oligonucleotide supplements the otherwise negligible mass shift provided by a small molecule and thereby allows for enrichment of catalytically active sequences.

All DNAzymes identified through selections using the previous acyl donor were assayed using the new glutaryl donor oligonucleotide. Three DNAzymes identified for acylation of the simple amine substrate demonstrated substantial glutarylation yield but were not further characterized; no DNAzymes identified using the DNA-HEG-AAAKAA substrate showed any glutarylation yield. Therefore, a new set of selections was conceived for DNAzymes that

explicitly use either the simple amine substrate or the DNA-HEG-AAAKAA substrate with the new glutaryl donor oligo.

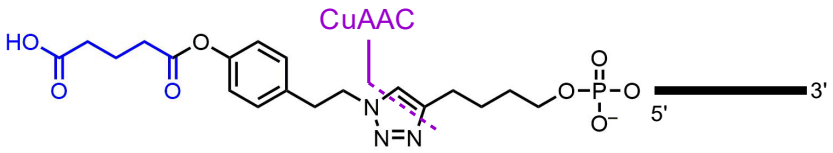

Fig. 5. Structure of the glutaryl donor oligonucleotide used in amine glutarylation assays and selections. Substrate was formed by CuAAC between a glutaryl azide compound and a 5'-alkyne oligonucleotide. A successful glutarylation leads to the loss of the glutaryl group, in blue.

CHAPTER 2: AMINE GLUTARYLATION

Peptide Glutarylation

My work in the Silverman lab began with assisting Yves in his investigations in amine glutarylation. Selections using the glutaryl donor had reached round 9, and I was given two populations to analyze. The first used DNA-HEG-AAAKAA as the nucleophile at pH 7.5 (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl), while the second used DNA-HEG-AAAKAA as the nucleophile at pH 9.0 (50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl). Both populations were incubated for 16 h at 37 °C during selection. Previous selections that used the large oligonucleotide acyl donor found success with both of these conditions, though DNAzymes that made use of DNA-HEG-AAAKAA were overwhelmingly more effective at pH 9.0.[16] This is likely due to the amine nucleophile on the lysine side chain being more strongly deprotonated at pH 9.0, thus reducing the catalytic demand of the DNAzyme to "activate" the nucleophile via deprotonation. Metal ion availability was based on historically successful ion concentrations in the Silverman lab, though only Mg^{2+} was included in pH 9.0 selections since Mn^{2+} becomes oxidized and Zn^{2+} precipitates out of solution at higher pH.

During these selection rounds, the tethered peptide nucleophile was ligated to the pool population to allow for separation via mass shift; the presence of this ligation is referred to as the in cis configuration. Since in cis activity was already confirmed via the selection rounds results, I was first tasked with testing in trans activity, or activity without ligation of substrates to the pool population. To this end, the DNA-HEG-AAAKAA substrate was radiolabeled on its 5' end to act as a traceable marker for yield (whereas the pool population would usually act as the marker, internally radiolabeled during PCR). The capture step using a 5'-amino modified oligonucleotide was also included, since glutarylation alone still did not provide a significant enough mass shift for sufficient separation of products from reactants. Lastly, background reactivity was determined using a DNA splint that contained exactly the two non-random regions of the pool population without a random region between them. This splint would allow for Watson-Crick base pairing and thus recruiting of the substrates but would have no potential for catalysis once substrates were bound.

In trans yields for the pH 7.5 and pH 9.0 population pools were 11.1% and 10.4% respectively when accounting for the efficiency of the capture step (~60%). These yields were not significantly above the background yield at ~5%, but they nonetheless demonstrated some amount of catalysis and were thereby cause for further investigation. Both populations were cloned, and 40 individual colonies expressing successful transformation were isolated for each. The presence of single inserts was confirmed via EcoRI digestion of the cloning plasmid, after which the inserts were amplified via PCR. Ideally, this would present 40 individual DNAzyme candidates from the pool population, each of which can be individually assayed for catalytic activity.

Individual in trans assays of the first 24 candidates of the pH 7.5 population and the first 20 candidates of the pH 9.0 gave yields ranging from ~3.5-8.0%. Overall reactivity in both pools, which acted as a standard, remained at ~5.0%. This was the less exciting outcome, that each sequence in a population pool shares a small but similar level of catalytic activity; more exciting is that a small amount of pool activity would represent a single strongly catalytic sequence being overwhelmed by sequences with low or no activity. Regardless, a total of 21 cloned sequences were chosen for sequencing from which 8 unique DNAzyme sequences were identified, 2 from the pH 7.5 pool and 6 from the pH 9.0 pool. Further in trans assays of these DNAzymes failed to show a yield significantly above background, with the strongest increase in yield being no greater than 2-fold above background (Fig. 6). Additional assays were conducted on controls DNA-HEG-AAAKAA-Glu, a positive control representing a successful glutarylation, and DNA-HEG-rA, a negative control in which the AAAKAA group was omitted. These controls were chosen to confirm that this low reactivity was not due to some error in the capture or selection steps, respectively. As expected, the positive control had a capture yield of ~60%, which was congruent with our previously observed captures, while the negative control had 0% glutarylation yield and 0% capture yield. We concluded that these DNAzymes were unremarkable in their catalysis, and could at best be argued to be recapitulating a splint. Seeing as the pools had already undergone 9 rounds of selection, it was unlikely that further selection rounds would produce stronger catalytic sequences. The selections were abandoned, without assaying the remaining cloned sequences.

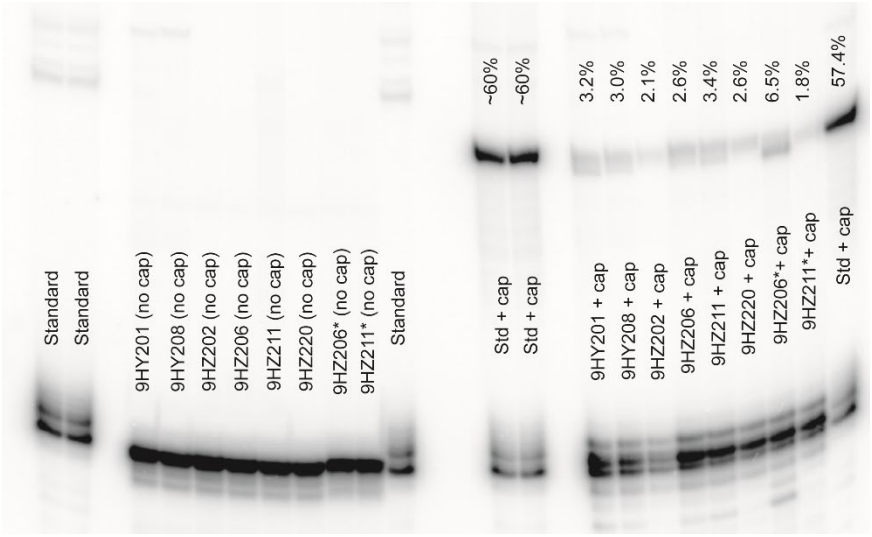

Fig. 6. In trans assay of DNA-HEG-AAAKAA glutarylating DNAzymes on PAGE. 9HY2XX denotes cloned sequence #XX from pH 7.5 selections, 9HZ2XX denotes the same for pH 9.0 selections. 9HZ206 and 9HZ211 had a single deletion mutation in the non-random region; an asterisk (*) denotes sequences that had this deleted nucleotide restored. "Cap" denotes an additional DMT-MM activated capture step. Standard is a pre-glutarylated DNA-HEG-AAAKAA. The highest yield, in consideration of capture efficiency, is 6.5/0.574 = 11.3%, which is about 2-fold higher than the ~5% background yield (not shown).

Simple Amine Glutarylation

Selections using the glutaryl donor oligonucleotide with the simple amine substrate were delayed, as selections with the model peptide seemed feasible and more promising. However, our failure to identify model peptide glutarylation activity led us to further investigate DNAzymes that glutarylate the simple amine nucleophile. We elected to test only the three DNAzymes from previous selections that already demonstrated substantial glutarylation abilities. These would act as confirmatory DNAzyme standards for future selections, while also allowing Yves to graduate at a reasonable time!

All three DNAzymes that catalyzed simple amine glutarylation were identified in Yves's selections using the large oligonucleotide acyl donor in pH 7.5 conditions (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl).[16] Since these DNAzymes were already assayed for in trans activity, the following in trans assays were conducted as a means of characterizing their catalytic activity. Additionally, assays using these simple amine-acylating DNAzymes made use of "decoy" oligonucleotides, which are 60 nucleotide DNA strands directly complementary to the initially random regions (40 nucleotides) of each DNAzyme as well as 10 nucleotides on either side that are complementary to the non-random regions. This oligonucleotide would be added after the selection step of in trans assays to bind to

the DNAzyme, kicking off the substrates and products in the process. This eliminated additional bands on PAGE, which otherwise represented products that remained Watson-Crick base paired to the DNAzyme and which complicated analysis. Remarkably, no capture step was necessary, since the negative charge conferred by the glutaryl group's exposed carboxylic acid provided the otherwise positively charged amine substrate with a substantial enough PAGE shift for visual separation.

My own investigations began with a 48 h assay to find rudimentary enzyme kinetics. Time points were taken at t = 0 h, 16 h, and 48 h. These assays made use of both a hydroxyl-modified oligonucleotide control and non-modified oligonucleotide control, wherein the amine was replaced with a hydroxyl and removed, respectively. A hydroxyl nucleophile is much weaker than a primary amine nucleophile, and should thus have significantly reduced yield. The non-modified oligonucleotide also has an exposed hydroxyl group, thus we expected similarly low reactivity. We observed glutarylation yields for two of the DNAzymes, 8FL205 and 8FL219, of 15-20% at 16 h and 20-30% within 48 h. The remaining DNAzyme, 7FN221, had lower yield of 5-8% at 16 h and ~10% at 48 h (Fig. 7). As was expected, the non-modified oligonucleotide control showed no detectable product formation at any time point. Assays with the hydroxyl nucleophile showed non-negligible yields, though values were inconsistent across time points, and activity was present at t = 0 h – these results do not suggest the formation of a product band, but rather some residual species that is present throughout the entire incubation. The splinted background reaction also showed a yield of ~10% despite having no visible product band, though this is likely due to counts "bleeding over" from strong nearby lanes. Background reaction products observed on later gels showed yields that were consistently <1.0% (Fig. 8).

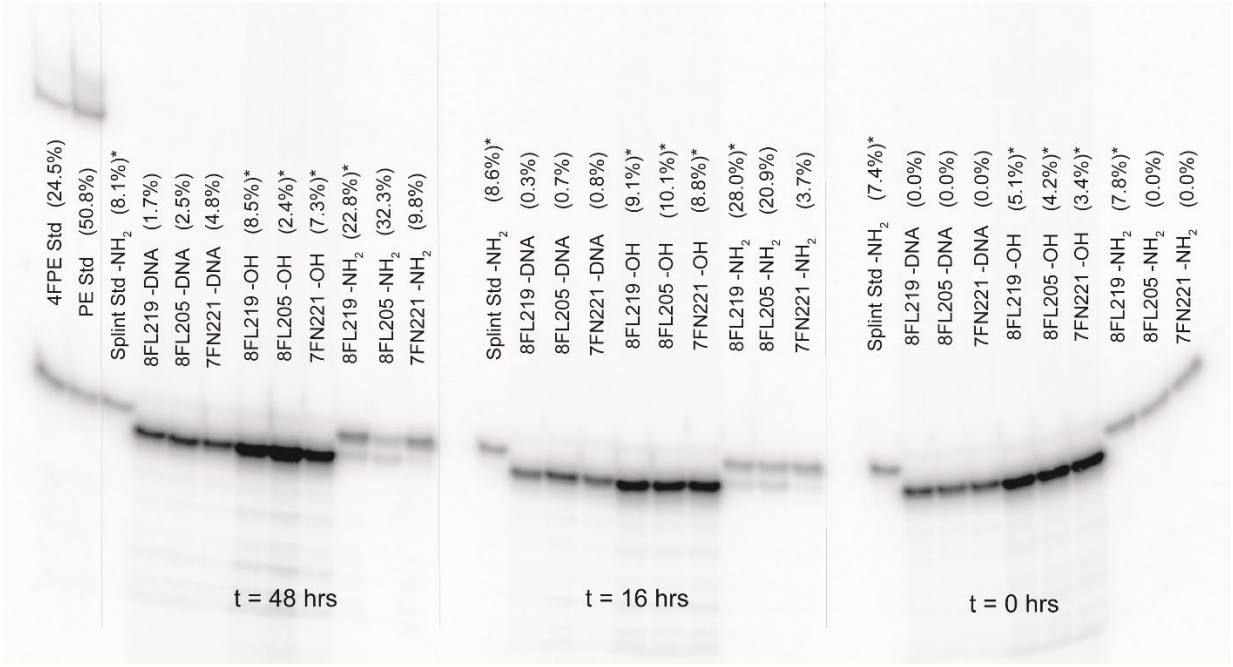

Fig. 7. In trans 48 h kinetic assay for simple amine glutarylating DNAzymes on PAGE. 8FL2XX denotes sequences #XX from pH 7.5 selections (PE acyl donor); 7FN2XX denotes the same from pH 7.5 selections (4FPE acyl donor). All lanes are additionally labelled with the nucleophile substrate, where -DNA denotes the non-modified oligonucleotide control 4FPE/PE standards use identified DNAzymes with the simple amine substrate and the denoted aryl ester acyl donor. Asterisks denote lanes that showed product formation (which should be below the starting bands) at t = 0 h, and thus have artificially high yields at later time points.

Enzymatic products were synthesized, and their identities were confirmed via mass spectrometry. More thorough kinetics and metal ion dependency assays were also conducted, though not by me.

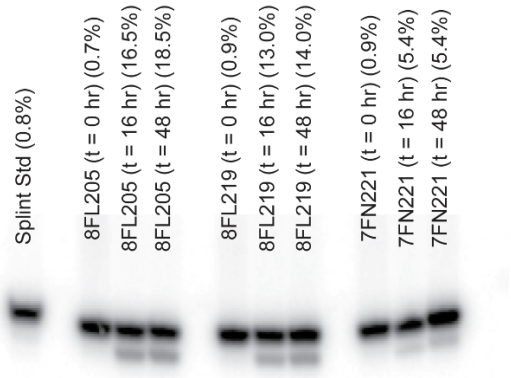

Fig. 8. A later in trans kinetic assay for simple amine-glutarylating DNAzymes on PAGE. Lane distribution leaves more space for each set of samples, reducing the effect of counts "bleeding over" into adjacent lanes and thus reducing the artificially high yields previously seen in the splinted background and t = 0 h time points.

Amine Glutarylation: Outlook

Amine glutarylation, specifically lysine glutarylation, has become increasingly associated with a number of regulatory pathways,[23, 24] including amino acid metabolism and cellular respiration. As a result, it would have been very exciting to identify DNAzymes that were capable of catalyzing a glutarylation reaction on a model peptide. It would be cause to investigate more complex peptide substrates, or perhaps to conceive selections for peptide sequence specificity – and if those went well, we could conceive selections for glutarylation of model protein domains. And though selections using the glutaryl donor with DNA-HEG-AAAKAA failed, the success of both the oligonucleotide-acyl donor with DNA-HEG-AAAKAA and the glutaryl donor with the simple amine substrate strongly suggests that this reaction is achievable with DNAzymes. The exact reason for this failure of the selection process is unknown, though it is unlikely to be the reaction substrates given all other successes. Further, the incubation conditions were reasonable in all other selections including model peptide glutarylation, as evidenced by relatively high (3-5%) splinted background yields. The pool population simply failed to produce a yield substantially above background. To remedy this, we can attempt to re-initiate selections using the same substrates and incubation conditions, but with variable pool lengths. N_{20} and N_{60} have historically seen success in the Silverman lab. Though the N_{60} pool presents an exponentially larger number of sequences that are absent in any synthesized pool, the longer random region length allows for more complex higher order structure, and thus potentially more complex active sites. The N_{20} pool provides a complete sampling of the sequence space by about 10^2 times, though the shorter random region limits the formation of higher order structures. Regardless of which direction is taken, it is clear that further selection experiments are necessary to identify DNAzymes that are capable of catalyzing model peptide glutarylation.

CHAPTER 3: ACYLATION OF FREE PEPTIDE

Experimental Design

The success of amine-acylating DNAzyme selections as well as the identification (though not via selection) of amine-glutarylation DNAzyme gave us a few directions to take the next project. The first was to continue selections for glutarylation, particularly using the DNA-HEG-AAAKAA substrate. The other option was to further increase the complexity of one of the substrates. Since DNAzymes that make use of an acyl donors that confers a small-molecule acyl group had already been identified, we turned our attention to the peptide substrate. Most proteins currently being investigated by the greater scientific community are not conjugated to single-stranded DNA, and have no means of being anchored to a DNAzyme without unwanted modifications. Ideally, DNAzymes would be able to perform acylation reactions on lysine residues without a dedicated DNA component. To this end, we began investigations into "free" peptide substrates.

Selections would ultimately be designed for both the N_{20} and N_{40} pools, with the following pH conditions: pH 6.0 (70 mM MES, pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl), pH 7.5 (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl), and pH 9.0 (50 mM CHES, pH 9.0, 40 mM $MgCl_2$, and 150 mM NaCl). This range of pH values has seen great success in the Silverman lab, including their use in Yves's selections. A wide range of pH values ensures the greatest likelihood of identifying DNAzymes due to our inability to predict exact DNAzyme mechanisms of catalysis and, by extension, the most favorable pH conditions. Incubation time and temperature would be determined via preliminary assays for background reactivity such that background reactivity would not exceed ~5%. Higher values would result in significant carry-over of false positive sequences for which the reaction occurred uncatalyzed, and thus decrease the enrichment factor of catalytic sequences possible in each round of selection. This may lead to a failed selection as seen with earlier amine acylation experiments.

Since the peptide would be necessarily unconjugated to the pool population during selection and lacking in any sort of DNA anchor, the acyl donor would have to be ligated to the pool instead. A successful acylation would lead to the conjugation of the peptide to the entire DNA molecule, after which the peptide could be targeted for capture and thus allow for the

separation of catalytically active sequences from the pool. To this end, a new suite of aryl ester oligonucleotides was conceived that strongly resembled those used in early amine acylation selections (Fig. 3), with the phenol-derivative ring acting as the leaving group. We suspected, however, that the less reactive aryl compounds would provide insufficient reactivity to overcome the decrease in reaction entropy that comes with a free peptide nucleophile. As a result, we focused on a set of more reactive aryl esters derived from more reactive phenol derivatives (Fig. 9) that had a greater number of electron-withdrawing substituents on the ring. Unlike PE and 4FPE, 2,4,6-trifluoro phenol (TriFPE) and TFPE could readily hydrolyze in solution, and thus would need to be synthesized directly before selection. Assays would be performed by one of my lab mates, Prakriti Das, to determine their stability in the proposed selection conditions. If these aryl compounds showed too much background reactivity or degraded too rapidly during early assays, we would focus back on PE and 4FPE.

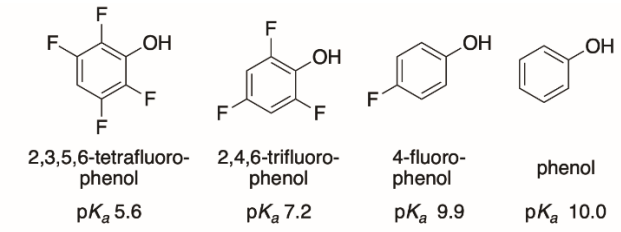

Fig. 9. The new set of reactive aryl compound candidates for acyl donor formation in free peptide acylation selections. The only candidate never previously investigated was TriFPE (2,4,6-trifluoro phenol). pKa values provide a general idea of reactivity.

The peptide itself presented some interesting opportunities. Previous selections used an unambiguous model peptide with a single lysine and no other significant side chains. While this was ideal for our first amine acylation selections, it does not reflect the level of complexity seen in biologically significant proteins. A more complex peptide substrate not only more greatly resembles these desirable protein targets, but can also allow for the identification of DNAzymes with sequence-specific catalysis![2] As such, we elected to conceive and synthesize a number of hexapeptide substrates, which will be elaborated upon later in this chapter. All candidates contained a single lysine residue at the C-terminus (XXXXXK) to act as a nucleophile and omitted other nucleophiles with nearby pKa values (cysteine and tyrosine, for example). The position of lysine was changed compared to previous selections to act as a precedent for future divergent selections, which would focus on reactions involving small molecules. Since the lysine residue in previous experiments was relatively unobstructed by other residue side chains, we did

not suspect that this change would significantly change overall yields. All sequences also contained an adjacent tryptophan residue (XXXXWK), as we hypothesized that a nearby ring structure would have favorable interactions with DNA nucleobases and aid in recruiting the substrate to an active site.

The final consideration was in regard to mass shift. A hexapeptide is substantially less massive than any sequence in either pool population. As such, a successful acylation would not result in a distinguishable product band. To circumvent this, we coupled 6-azidohexanoic acid onto the N-terminus of all peptide candidates. This would not only remove the α-NH_2 as a potentially competing nucleophile, but would also allow for a capture step using CuAAC. To this end, we synthesized a 3'-alkyne oligonucleotide for CuAAC with this terminal azido group. We also needed to determine ideal CuAAC conditions, however, since high Cu^+ concentrations can be damaging to DNA[25] and could thereby artificially reduce final capture yield values.

Acyl Donor Stability Assays

Before each selection step, the acyl donor would need to be "pre-activated" with the appropriate reactive aryl group. This would be accomplished through incubation of a carboxyl-modified oligonucleotide with the phenol-derivative compound and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). Carboxylic acid groups attack EDC, forming a strong EDC leaving group for the attack via the phenol compound. All pre-activations would be conducted in one of the three pH conditions (6.0, 7.5, or 9.0 as detailed above). Prakriti determined that the best incubation conditions were 2 h at room temperature (50 mM buffer, 50 mM aryl compound, 50 mM EDC, 150 mM NaCl), which would regularly lead to >90% aryl-ester formation (determined via mass spec) in all pH conditions.

Stability assays were performed by incubating the TFPE pre-activated aryl ester in selections conditions as detailed above in the absence of other substrates or pools. Time points were taken at 10 min, 2 h, and 16 h. TFPE under all selection conditions was >90% intact at 10 min and 2 h. In both pH 6.0 and pH 9.0 conditions, TFPE was 74-77% intact at 16 h; it remained ~50% intact after 16 h in pH 7.5 conditions. We deemed this stability sufficient given the following considerations: first, the presence of EDC in solution meant that hydrolyzed aryl esters could readily reform during selection; second, sequences may be able to perform significant

catalysis early in the aryl ester's half-life. Since TriFPE is inherently less reactive than TFPE, we did not deem it necessary to conduct TriFPE aryl ester stability assays before starting selections.

Ligation Step Confirmation

After acyl donor formation, the first step of selection would be to ligate the acyl donor to the pool population. However, due to the general reactivity of the proposed acyl donors, the relatively lengthy ligation step would be conducted first, followed by pre-activation. I was tasked with identifying optimal conditions for ligation. Our lab had seen relative success with 16 h incubations at 37 °C using T4 DNA ligase, so my relatively brief search began with these conditions. A set of three separate DNA splints were assayed, each complementary to the 5' non-random region of the DNA pool and the acyl-precursor carboxyl oligonucleotide but each varying in length by 4 nucleotides (33, 29, and 25 nucleotides). The shortest of these splints binds to 8 nucleotides in the non-random region while the remaining nucleotides bind to the carboxyl oligonucleotide; the longer splints bind to an increasing length of the non-random region. The longer sequences thus have a greater affinity for the random pool, though this increased affinity may not be necessary for a successful ligation.

Incubation with the two longer splints resulted in a ligation yield of ~90%, while the shortest splint resulted in ~76-77% ligation yield. These yields, though high, could result in the loss of reactive sequences before the selection step ever occurs. Historically, however, a loss of more than 10% of the pool during ligation has still led to successful selections, and so we thought it safe to proceed with the 33 nucleotide splint.

Peptide Design and Synthesis

When designing hexapeptide candidates, our primary considerations (beyond those mentioned in Experimental Design) were water solubility and semi-arbitrary amino acid sequence composition. High water solubility would enable us to conduct selections without complicating solution conditions with organic solvent, while semi-arbitrary amino acid sequences would lead to sequence motifs that could be changed during future sequence specificity assays. Our first candidate was N_3-GQGQWK, based on the tetrapeptide N_3-GQWK that Prakriti had been attempting to synthesize for her investigations into reactions involving small molecules. This hexapeptide was theoretically sufficiently water soluble, though synthesis

proved challenging. Yields were unusually low (<1% yield) though still usable, and purification by HPLC resulted in a number of peaks that all had a strong absorbance at 280 nm, implying a number of incomplete couplings after the addition of tryptophan. The product closest to the target mass of 841 Da was a peptide with mass 845-847 Da, and though this mass difference is small, it was too substantial to warrant identification as the desired peptide. This discrepancy could be due to a number of reasons, including complications in coupling and/or insufficient removal of side chain protecting groups. Regardless, isolation of the correct peptide product seemed impossible. Additionally, none of our hypotheses accounted for such a small mass shift, and identification of the higher mass product seemed impossible with the amino acids, protecting groups, and reagents we were using. Synthesis of the N_3-GQWK peptide was also unsuccessful for unclear reasons. As a result, the N_3-GQGQWK hexapeptide candidate was abandoned, and future peptide candidates would avoid the GQ sequence motif entirely.

Our next candidate was N_3-AEQSWK. Our only concern with this peptide was the presence of a serine residue, which could technically act as a competing nucleophile in place of the lysine. However, seeing as a lysine side chain nucleophile is much stronger than a serine side chain nucleophile, we did not suspect that any reactive sequences would preferentially make use of a serine nucleophile. Even if they did, acylation would result in the formation of an ester bond instead of an amide bond; we could confirm the presence of this ester through cleavage via a strong base. Early synthesis attempts failed often due to the poor coupling of 6-azidohexanoic acid, though the rest of the peptide seemed to have far fewer failed couplings than N_3-GQGQWK. We ultimately found a successful 6-azidohexanoic acid coupling procedure conducted by previous graduate student Chih-Chi (Jimmy) Chu, which involved a different set of coupling reagents and an overnight incubation as opposed to our standard ~1-2 h incubations. With this, the correct peptide substrate was both easily synthesized and isolated, though yields were still low (~10%).

Assays conducted using N_3-AEQSWK were complicated by the presence of an unexpected band near both the reactant bands and the product bands (Fig. 10). We ultimately determined that these bands were as a result of the remaining EDC from acyl donor pre-activation complexing with the carboxylic acid on the glutamic acid side chain. Both the loss of a negatively charged group and the additional mass provided by the EDC would result in the observed upwards PAGE shift. To avoid potential complications surrounding multiple product bands, and without risking acyl donor stability through a purification process, this hexapeptide candidate was also abandoned.

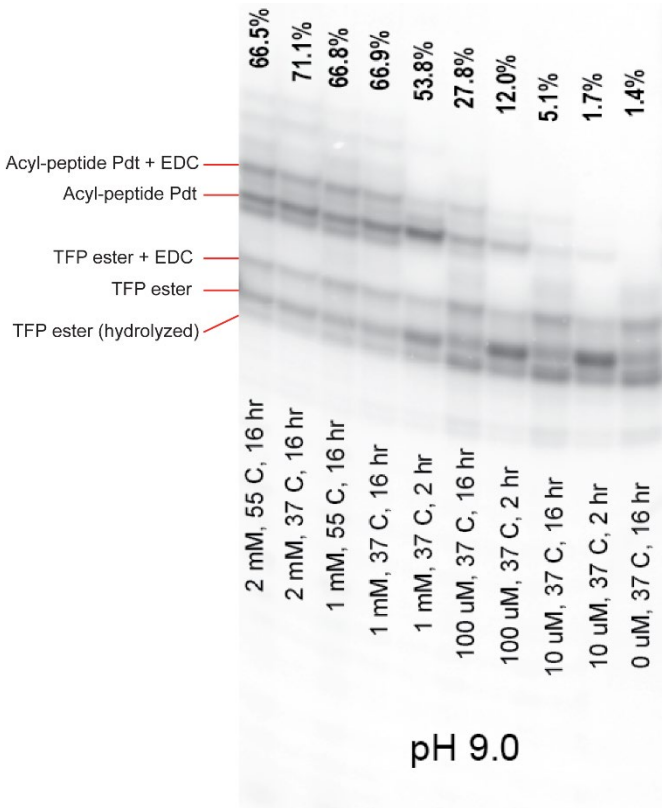

Fig. 10. An excerpt from a background reactivity assay using TFPE "pre-activated" ester and the N_3-AEQSWK peptide. Though the reaction was successful, the presence of EDC in solution led to a number of undesirable EDC-adduct bands.

Wanting to capitalize on the relative success of N_3-AEQSWK, we chose our final hexapeptide candidate to be N_3-ASQSWK. Synthesis was once again uncomplicated, with yields that would gradually approach ~10% in later syntheses. Due to the relatively high background yields seen with high concentrations of N_3-AEQSWK, these synthesis yields were deemed sufficient pending further investigation. We elected not to perform a new set of background reactivity assays using N_3-ASQSWK and instead continued working towards completing the selection scheme – results with N_3-AEQSWK were sufficiently clear even with the additional product bands, and we assumed that N_3-ASQSWK would not vary substantially in reactivity with the aryl ester acyl donors.

Initial Background Reactivity and Selection Conditions

Background reactivity was determined using TFPE and TriFPE activated acyl donors along with the N_3-AEQSWK peptide. Since the peptide substrate is inherently free of a DNA tether, the DNA splint usually associated with background reactivity was omitted. A number of variables were tested including peptide concentration (10 μM, 100 μM, 1 mM, and 2 mM N_3-AEQSWK), incubation time (2 h and 16 h), and incubation temperature (37 °C and 55 °C) in all proposed pH conditions (pH 6.0, pH 7.5, and pH 9.0, as detailed above). Ideal selection conditions would produce product yields high enough to be visible on PAGE, but low enough to allow for significant rate enhancement by DNAzymes. Lastly, a set of conditions needed to be identified for the high-yield formation of a "forced" background product that could act as a standard as well as a marker on PAGE for the desired product.

With TFPE (Fig. 11), background reactivity increased with higher peptide concentrations and longer incubation times, though the difference in yield resulting from variable incubation temperatures was negligible. As expected, pH 9.0 conditions produced a far higher range of yields (~1-70%), likely because the lysine nucleophile would be most strongly deprotonated at pH 9.0. pH 7.5 and pH 6.0 conditions had much lower yields overall (~1-20%). We selected pH 9.0, 1 mM peptide, 16 h, and 37 °C as our forced background conditions (66.9% yield). For our selection conditions, we chose the following: pH 9.0, 10 μM peptide, 16 h, and 37 °C (5.1% yield); pH 7.5, 10 μM peptide, 16 h, and 37 °C (2.7% yield); and pH 6.0, 100 μM, 16 h, and 37 °C (1.5% yield). These conditions satisfied our desired range of pH values while also maintaining consistent time points for convenience.

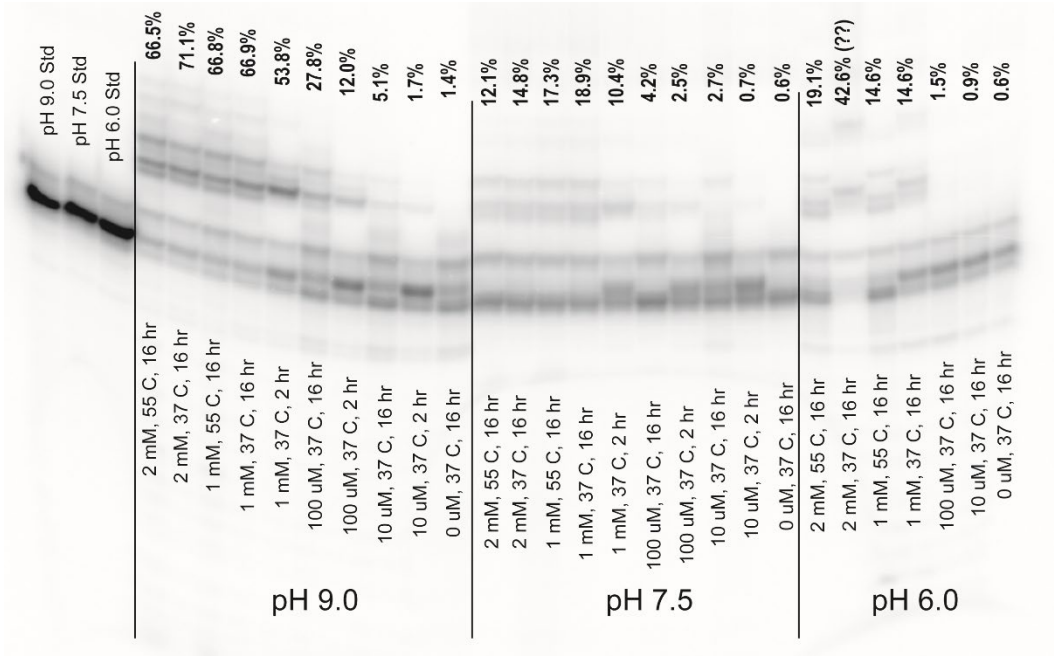

Fig. 11. The complete TFPE background reactivity assay using N_3-AEQSWK. All lanes are labelled with peptide substrate concentration, incubation temperature, and incubation time. Yields were calculated by including the speculative EDC adduct bands for both the reactants and products. As we determined later, the pH values as listed do not reflect the actual pH conditions in solution; for the sake of consistency, these values refer only to the selection conditions as described above. Standards include only the aryl ester pre-activated in the described conditions.

TriFPE background assays omitted both the increased incubation temperature variable as well as the highest peptide concentration (2 mM peptide), as neither had an observable impact on yields in the TFPE assay. Additionally, since forced background conditions had already been determined, only selection conditions were considered. For TriFPE, we chose the following conditions: pH 9.0, 100 µM peptide, 16 h, and 37 °C (2.9% yield); pH 7.5, 1 mM peptide, 16 h, and 37 °C (5.0% yield); and pH 6.0, 1 mM peptide, 16 h, and 37 °C (1.9% yield).

An additional set of background reactivity assays were performed using a 3,5-difluoro-phenol (DFPE) activated acyl donor. The intent of these assays was to produce a third set of selection conditions (along with TFPE and TriFPE conditions) which would ultimately contribute to a more complex selection scheme and a potentially broader range of DNAzymes. While reasonable background reactivity was observed (~0.5-1.5%), DFPE pre-activation was significantly less efficient at pH 6.0 and pH 9.0 than all other pre-activated aryl esters (>50%). Thus, all DFPE pre-activations would need to be conducted at pH 7.5 after which the buffer would need to be overwhelmed. Confirmatory assays for DFPE buffering conditions were conducted by Prakriti. However, she determined in her investigations that all of our previous

conditions were insufficiently buffered, with the expected pH 6.0-9.0 range resembling something closer to pH 5.0-6.5. It was determined that the significant quantities of phenol compound and EDC left over from pre-activation were shifting the pH values lower than expected. This introduced a number of problems, the first being that both aryl ester stability and background reactivity would need to be determined again in correctly buffered conditions. While the Silverman lab has seen success with selection around pH 5.0, we deemed the resulting range of pH values too narrow to result in a successful selection. While it would also be possible to isolate the acyl donor after pre-activation, the lengthy process would risk serious degradation of the aryl ester. Additionally, removal of EDC would completely remove the possibility of degraded aryl ester reforming during the selection step itself. We determined that the best course of action was to find a new set of pre-activation and buffering conditions that would allow for our initially desired range of pH 6.0-9.0 without sacrificing the integrity of the already promising TFPE and TriFPE. Once those conditions were set, a new background reactivity assay would be conducted and new selection conditions would be determined. If the higher pH range proved too reactive with our chosen aryl esters, we could fall back onto the less reactive aryl esters used by Yves – thought we still strongly suspected that they would be insufficiently reactive with a non-tethered peptide – and if all else failed, we could conduct selections at pH 5.0-6.5.

Finalized Background Reactivity and Selection Conditions

Experiments and assays conducted to determine properly buffered pre-activation and selection conditions, as well as aryl ester stability assays, were conducted by Prakriti Das. She determined that a pH range of 6.5-8.5 was feasible for both TFPE and TriFPE. For TFPE, these conditions were as follows: pH 6.5 (105 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl); pH 7.4 (220 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl); and pH 8.5 (240 mM CHES, pH 9.0, 40 mM $MgCl_2$, 150 mM NaCl). For the less acidic TriFPE, the conditions were as follows: pH 6.5 (70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl); pH 7.4 (90 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl); and pH 8.5 (140 mM CHES, pH 9.0, 40 mM $MgCl_2$, 150 mM NaCl). While the use of DFPE was still feasible, we deemed its constrictive pre-activation conditions too prohibitive compared to the already complex pre-activations for TFPE and TriFPE. If background reactivity for either of the

more reactive aryl esters was too high, however, further potential conditions for DFPE would be determined.

In parallel to Prakriti's investigations, I began investigations into the capture step.

Early Capture Step Assays

The intention of a capture step is to join a large oligonucleotide to the product of reactions that result in negligible mass shifts. This supplemental mass shift allows for the separation of catalytically active sequences from the pool during selection, though it can also be employed during in trans assays to increase band separation if necessary. In our selections, a small peptide (~841 Da) would become joined to a member of the pool population that was ligated to the acyl donor (>10^5 Da) – a capture step would be necessary to differentiate active sequences from the pool during selection.

All peptide candidates were designed with a 6-azidohexanoic acid group coupled onto the α-NH_2 explicitly to allow for capture via CuAAC. As such, two capture oligonucleotides, one 18 nucleotides and the other 49 nucleotides in length, were synthesized. Each capture oligonucleotide was made to be completely distinct from both non-random regions of the random DNA pool to prevent significant Watson-Crick base pairing, and each was modified with a 3'-alkyne group that could be targeted for CuAAC. The purpose of having two differently sized capture oligos was to allow for alternating capture during selection rounds. It may be possible that certain non-catalytic DNA sequences possess higher order structure that would obstruct movement through PAGE and thus lead to a gel position similar to that of a captured catalytically active sequence, despite these sequences not performing catalysis. Alternating the length of the capture oligonucleotide between rounds, and thus alternating the size of the mass shift between rounds, ensures that these anomalously migrating sequences are not selected for.

Early capture assays were as minimal as possible, including only the pre-formed acylation product (which was formed using forced background conditions) and the short alkyne capture oligonucleotide; the pool was omitted. CuAAC reagent concentrations were based on those that were previously successful in selections: 0.4 mM $CuCl_2$, 0.8 mM Sodium ascorbate, 2.8 mM tris-hydroxypropyltriazolylmethylamine (THPTA), 50 mM HEPES, pH 7.5, and 150 mM NaCl. Samples were incubated at 37 °C for 1 h. Under these conditions, the yield of CuAAC capture was ~30%. While selections have seen success with capture yields this low, we

suspected that the loss of ~70% of our catalytically active pool population in each round would be detrimental to the success of our selections. Previously successful capture step schemes in the lab made use of a DNA splint that performed Watson-Crick base pairing with both the capture oligo and the acyl donor oligonucleotide. This reduces the entropy of reaction and thus makes capture more efficient. As a result, all future capture step assays included an appropriate DNA splint, as well as an annealing step to allow both substrates to bind.

A comprehensive capture assay was performed to determine the effect of splinting on capture efficiency. However, a number of other variables were investigated in parallel to ascertain ideal capture conditions. The first consideration was CuAAC reagent concentration; though previous selections were successful with the above conditions, CuAAC formation of the glutaryl donor used in Yves's later experiments made use of scaled-up concentrations (5 mM $CuCl_2$, 10 mM sodium ascorbate, and 35 mM THPTA – 12.5× our original concentrations). We elected to test 1× (0.4 mM $CuCl_2$, 0.8 mM sodium ascorbate, and 2.8 mM THPTA), 2.5×, 5×, and 12.5× CuAAC reagent concentrations, though we suspected that higher $CuCl_2$ concentrations would be potentially damaging to the DNA[25] and thus lead to lower yields. We also elected to test incubation time, with time points at 1 h and 4 h. The 1 h time point was conventional for most CuAAC reactions conducted in lab, while the 4 h time point was to act as a same-day investigation into longer capture incubations. If a 4 h incubation resulted in a much higher capture yield, we would extend the capture step to an overnight incubation (12-16 h) in order to maximize yield; if it produced similar capture yields to the 1 h incubation, we would continue to incubate for 1 h.

The results of this capture assay (Fig. 12) clearly demonstrated the importance of a splint, as well as the general unimportance of the other variables. All splinted capture samples but one showed a capture yield of >90% at both 1 h and 4 h, regardless of CuAAC reagent concentrations. The only exception was the sample with the highest CuAAC reagent concentrations (12.5×) at the 4 h time point, while showed a yield of ~50%. All samples without a splint showed a capture yield of ~60%, once again with the exception of those with the highest CuAAC reagent concentrations (12.5×), which had yields of ~20% at both 1 h and 4 h. These results confirmed our suspicions that higher $CuCl_2$ concentrations would be detrimental to capture yield, though they also confirmed that higher CuAAC reagent concentrations were entirely unnecessary. Going forward, all capture experiments would use the lowest CuAAC

reagent concentrations (1×) as well as a splint, and all capture reactions would be incubated for 1 h.

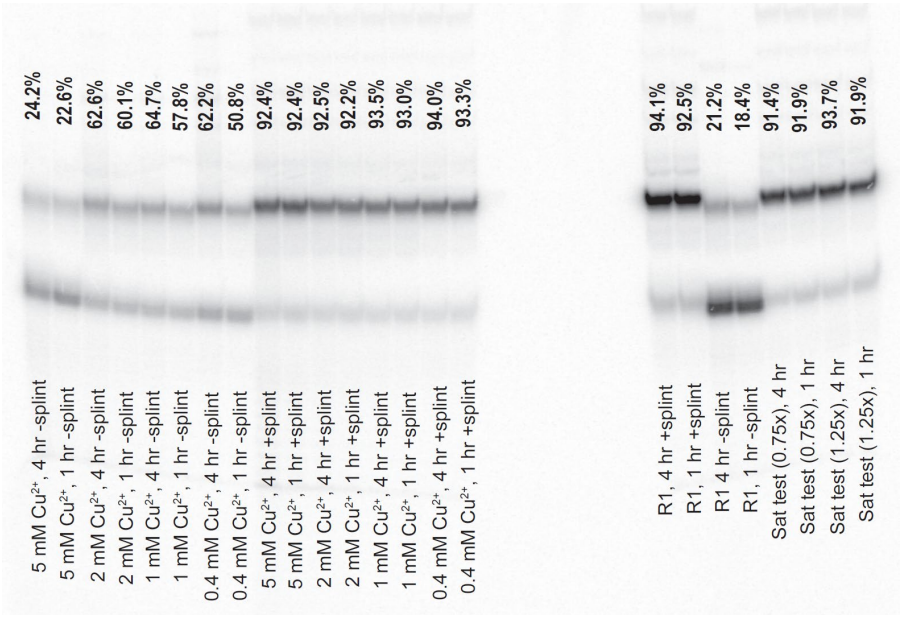

Fig. 12. Initial capture step assay (no pool) including satellite experiments for scaled-up samples and saturation testing. Lanes are labelled with CuAAC reagent concentrations (all reagents are scaled linearly with Cu^{2+}), incubation time, and the presence/absence of splint. "R1," or Round 1, refers to the scaled-up sample concentrations used during the first round of selections, and all were conducted with 0.4 mM $CuCl_2$ (with all other CuAAC reagents scaled appropriately). All other lanes use smaller sample concentrations similar to that of subsequent selection rounds.

In conjunction with the above assay, a satellite experiment was performed to determine capture efficiency in more niche circumstances. First, a capture assay with heavily scaled up amounts of forced background product was conducted using 1× CuAAC reagents in the presence and absence of a DNA splint. This assay was meant to resemble conditions in the first round of selection, in which the ~200 pmol of the random DNA pool undergoes selection and capture; in later rounds, selection is performed on the PCR product of the previous round (~5-10 pmol). Confirmation of capture at this scale would be crucial to avoid losing significant material in the first round of selection. Second, a "saturation" capture assay was conducted with CuAAC reagent concentrations slightly above (1.25×) and below (0.75×) our lowest CuAAC reagent concentration. All samples would make use of a DNA splint. The purpose of this experiment was to determine whether or not our lowest concentration of reagents was sufficiently saturating; if capture yield changed significantly in either case, we would need to modulate our reagent concentrations appropriately.

In hindsight, these additional assays may have been excessive considering how similar their results were to the comprehensive capture assay. The scaled-up capture samples showed >90% yield at 1 h and 4 h when splinted and ~20% yield without a splint. All samples in the saturation assay showed >90% yield at 1 h and 4 h, thus confirming that our 1× conditions were sufficiently saturating.

Advanced Capture Step Assays

Thus far, no capture assays had incorporated the DNA pool at any point. Inevitably, we would need to confirm that the capture conditions identified above would work in the presence of the large oligonucleotides that comprised the random pool. As such, an assay was conducted to ascertain the effect of the presence of the ligated random pool on capture efficiency. Though the selection would ultimately make use of both the N_{40} and N_{20} pools, this assay was conducted using only the N_{40} pool. The random pool was first ligated to either the acyl donor oligo or the previously synthesized acyl donor/peptide conjugate. The former would act as a mock selection, undergoing pre-activation, selection in forced background conditions, and capture; this would not only provide values for capture efficiency in the presence of pool, but would also confirm that the three-step selection process was functional as a whole. The latter would be a positive control for selection that would exclusively undergo a capture step and provide capture efficiency directly. Both groups of samples would undergo capture with both the short and long capture oligonucleotides to determine the effect of capture oligonucleotide length on capture efficiency as well. Lastly, standards would be included for pre-activation and forced selection to ensure the integrity of each step. All capture conditions would use 1× CuAAC reagent concentrations (0.4 mM $CuCl_2$, 0.8 mM sodium ascorbate, and 2.8 mM THPTA) and would be incubated at 37 °C for 1 h.

Pool DNA ligated to the forced background product had little effect on capture efficiency, with capture yields for the short and long capture oligonucleotides around 84% and 87% respectively. However, pool DNA that went through the full selection process had significantly reduced capture yields of 5-7%. The success of the capture step with the forced background product would imply that the capture step was functioning normally. Thus, the poor yield was likely due to failures in either the pre-activation of the acyl donor or the selection step. The yields for neither of these steps were easily determined, since neither step provided a

significant enough mass shift for a product band to be distinguishable from the rest of the pool. Fortunately, we elected to run an assay in parallel that omitted pool ligation to act as a control.

Results in the absence of the random pool were very similar, with capture of the forced background product showing capture yields of >90% as expected, though the acyl donor that underwent the full selection process had yields of <10%. The pre-activation standard, fully quantifiable in the absence of the massive random pool, showed >90% formation of the aryl ester, which was expected. The selection step standard showed ~7% yield, which was an order of magnitude lower than what was expected. These results implied that the selection step failed in both experiments, though the cause of the failure was unlikely to be as a result of the pool. However, the capture step with both capture oligonucleotides demonstrated sufficient capture yield in both the presence and absence of pool, and we thus deemed it appropriate for selections. From this point forward, however, we decided to include two standards in all selections rounds in order to directly determine the failure point in any round of selection. The first would use the forced background product ligated to pool, which would undergo CuAAC capture and act as a standard for the capture step alone. The second would use a pre-activated TFPE acyl donor that was ligated to the pool and incubated in forced background conditions, which then underwent CuAAC capture. This would act as a standard for the selection step, since the forced background yield could be easily calculated using the capture yield observed in the first standard.

A single scaled-up sample of the random pool, meant to resemble the first round of selection, was also ligated to the forced background product and underwent capture. We deemed that the capture yield, while lower (64%) than that of the less concentrated samples above, was still sufficient for initiation.

Wary of other potential failure points in the capture step, we elected to investigate some theoretical problems that could artificially lower capture yield. Our primary concern was that, while there was an ethanol precipitation step between selection and capture, the selection step still involved adding peptide substrate that was several orders of magnitude more concentrated than that of the actual pool sequences and by extension the 3'-alkyne modified capture oligonucleotides, which are added in a concentration relative to the pool concentration. If a substantial fraction of the peptide substrate was carried over through the ethanol precipitation, it is reasonable to assume that a non-negligible portion of the capture oligonucleotide would react with the unreacted peptide substrate, rather than the product formed by catalytically active

sequences. This would artificially decrease the concentration of the capture oligonucleotide and could drive down capture yields. As such, we performed an assay wherein both scaled-up and non-scaled up pool samples would undergo a mock selection using forced background conditions. These conditions already made use of the highest peptide concentration we intended to use (1 mM) and would thus contain the highest level of potential peptide carried over through the purification step. However, an additional phenol:chloroform:isoamyl alcohol (PCI) (25:24:1) extraction step would be conducted in place of an ethanol precipitation for half the samples. We expected that this step would be more efficient at removing all peptide substrate at the potential expense of losing part of the pool during the process.

Results showed that ethanol precipitation and PCI extractions were more or less equivalent when it came to capture yields (Fig. 13). All samples, scaled up or otherwise, had similar capture yields of 65-70% after undergoing a full mock selection. This was encouraging, since these values represented the multiplicative product of pre-activation, selection, and capture yields using pool-ligated acyl donors. In consideration of forced background yields (~70%), this meant that capture yields were >90% regardless of how concentrated the selection sample was. Additionally, it indicated that peptide substrate carrying over from the selection step after purification was not a significant issue, at least within the range of peptide concentrations we intended to use.

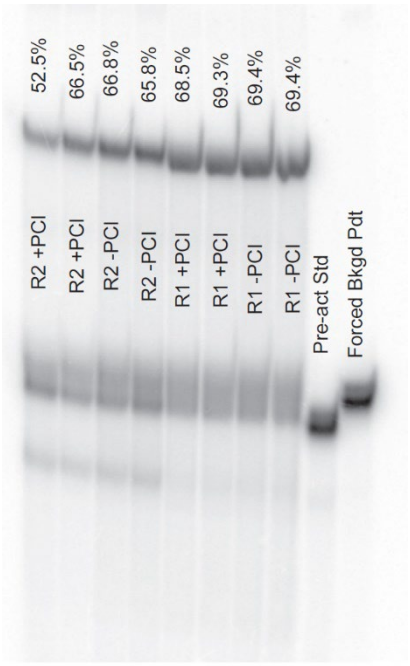

Fig. 13. Capture step assay (with pool) to determine the effect of an additional PCI extraction step on capture yield. R1 refers to scaled-up samples that resembles those used in the first round of selection. R2 refers to smaller samples that resemble all subsequent rounds.

A secondary assay was conducted to determine whether an artificial peptide substrate spike would negatively affect capture yields, and in turn if a PCI extraction could prevent these negative effects. To conserve the random pool, several non-ligated acyl donor samples were pre-activated and placed in forced background conditions. These samples each underwent a PCI extraction to completely remove all peptide substrate. Samples were then spiked with 10 μM, 30 μM, 100 μM, and 0 μM (positive control) peptide, which represented 10-30% carryover from 100 μM peptide and 10%

carryover from 1 mM peptide selection conditions. Half of these samples underwent an additional PCI extraction, after which all samples underwent CuAAC capture.

Results indicated that peptide carryover had an observable, though not substantial deleterious effect on capture yields (Fig. 14). Samples with 0 µM additional peptide as well as all samples that underwent an additional PCI extraction step showed capture yields of >80%, while capture yields for samples with artificial peptide spikes showed yields as low as ~60%. While this is not a substantial decrease in capture efficiency, these results do imply that excess peptide conjugates to the capture oligonucleotide as we hypothesized. However, given our previous assay testing the efficacy of PCI extraction and ethanol precipitation as purification methods, it is unlikely that enough peptide is carried over with either method to significantly hinder capture yields. We ultimately elected to use ethanol precipitation, as it minimizes the amount of DNA lost through purification while being equally as effective for our purposes as a PCI extraction.

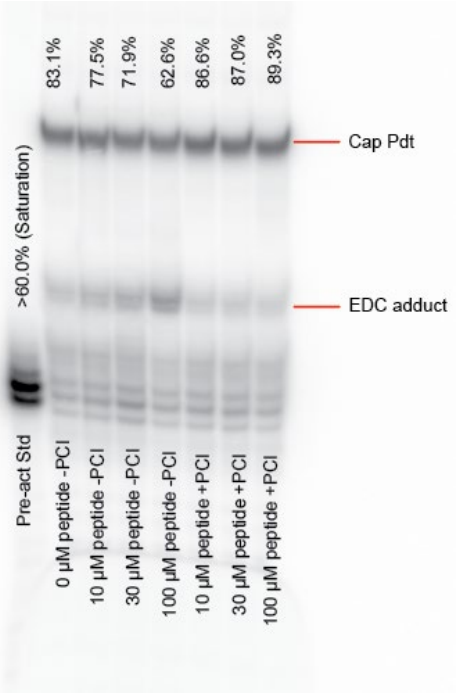

Fig. 14 Capture step assay (no pool) to determine the effect of theoretical azido-peptide carryover from purification steps. All lanes are labelled with the final concentration of peptide added after a preliminary PCI extraction, after which an additional PCI extraction was either performed or omitted. All samples show a clear EDC adduct band, which was included in yield calculations. The pre-activation standard band reached saturation limit on the plate reader and thus could not be accurately calculated.

Selection Scheme Overview and Early Selection Results

Our finalized selection scheme includes a total of nine selections, six of which make use of the N_{40} pool. As mentioned previously, all selection conditions were determined and tested by Prakriti Das. Three of these selections make use of a pre-activated TFPE acyl donor, while the other three make use of a pre-activated TriFPE acyl donor. The final three selections on our scheme make use of a TFPE acyl donor but use the N_{20} pool instead of N_{40}. Each of these sets of three uses pH 6.5 (105 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl), pH 7.4 (220 mM HEPED, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150mM NaCl), and pH 8.5 (240 mM CHES, pH 9.0, 40 mM $MgCl_2$, 150 mM NaCl) incubation conditions, with varying peptide substrate concentrations that allow for reasonable background reactivity. All selection steps are incubated at 37 °C for 16 h, after which the selection product undergoes a CuAAC mediated capture reaction with an appropriate capture oligonucleotide (varies depending on selection round). The products of the capture step are then isolated and amplified via PCR to allow for the next round of selections. As always, the first round of selections makes use of ~200 pmol of random pool, while all following rounds make use of a PCR product (5-10 pmol). All other substrates and reactants are scaled appropriately (see Appendix).

At time of writing, selections are being conducted by Prakriti Das. We are at round 5 of selections, and while pool reactivity is visible, it is not sufficiently above background. As such, it is unknown whether any sequence in the pool is doing anything more than recapitulating a splint. More rounds of selection will be necessary to determine if significant catalytic activity exists within the pools. These results are not uncommon at round 5; our previous selections for amine acylation showed sufficient activity above background at rounds 7-8, for example.[16]

CHAPTER 4: CONCLUSION AND FUTURE DIRECTIONS

Though our selections for acylation of non-tethered lysine side chains have yet to reveal any DNAzyme candidates, our lab's successes with tethered lysine acylation[16] and non-tethered peptide reactivity[18,29-31] strongly suggest that these selections will inevitably be fruitful. If they were to fail, it would suggest that the reaction as presented was too complex for the active sites present in the pool population. This is unlikely to be due to the complexity of the peptide, since a vast majority of our DNAzymes that act on non-tethered peptide use long and relatively complex peptide substrates. It is also unlikely to be as a result of acyl donor reactivity. Thus, it may be worthwhile to reinitiate selections with an N_{60} pool using the same substrates. We could additionally investigate more acidic (~pH 5) conditions than those used in our current selection scheme, though more basic conditions (>pH 9) would certainly lead to denaturation of the pool population. If the selections succeed, however, an exciting collection of opportunities arise. Beyond characterization of DNAzymes, which includes metal cofactor dependency testing and kinetic assays, our substrate design allows for a number of additional assays to determine the specificity and flexibility of our DNAzymes.

Sequence Specificity Assays

The peptide substrate was purposefully designed with the XXXXWK motif in mind, and it is likely that a number of our DNAzyme candidates will have an active site with high affinity for some sequence motif in the N_3-ASQSWK peptide. This allows us to assay each DNAzyme for sequence specificity by presenting each candidate with a suite of mutated peptide substrates (ASQSAK, for instance). This would allow us to determine for each candidate which residues are necessary for substrate detection and catalysis. Sequence specificity provides us with a proxy for site specificity, which is a crucial feature of most proteins associated with PTMs[26] and thus a feature we ultimately want to incorporate into any DNAzyme toolkit. So long as the desired lysine acylation site has a specific sequence around it, we would ideally be able to provide a DNAzyme selected to acylate that specific sequence. Our lab has seen success with these assays before,[2] and we are thus optimistic at the prospect of sequence-specific DNAzymes.

Glutarylation Assays/Selections

As seen with Yves's work, we could also assay our DNAzymes with different acyl donors, including small-molecule acyl donors. This can entail any number of acyl groups, though we would be likely to choose a glutaryl group or any similar group with multiple carboxylic acids as to allow for a DMT-MM activated amino capture. The glutaryl donor we used is much less reactive than the TFPE and TriFPE donors we are currently using, so we may need to synthesize, and perhaps even look into pre activating, a suitable glutaryl donor candidate for these assays. As seen with Yves's work, it is fully possible that some DNAzyme candidates will be able to make use of the new glutaryl substrate, while other DNAzymes will be completely incapable of catalyzing a glutarylation. In the former case, if glutarylation yields are low, we could perform reselections in which certain key motifs in the active site are maintained in the random region in hopes of identifying highly similar but distinct DNAzymes with greater glutarylation activity. In the latter case, it is still feasible to prepare and conduct an entirely new set of selections that use the glutaryl donor throughout, despite the failure of our selections for tethered peptide glutarylation.

These selections for non-anchored peptide glutarylation would necessitate a different means of separating catalytically active sequences from the pool, since the glutarylation product of a non-anchored peptide would be entirely dissociated from the pool. A successful glutarylation would instead leave only the exposed phenol of the glutaryl donor attached to the DNA pool. This may already introduce a slight PAGE shift, since the phenol, which is neutrally charged in all of our usual conditions, would migrate more slowly on PAGE than the negatively charged intact glutaryl donor. However, if this shift is not substantial enough to allow for sufficient separation of catalytically active sequences, we would need to find some additional way to further separate these sequences. The most tempting possibility would be to use the reactions already employed in our previous selections, specifically targeting the phenol via "activation" of a carboxyl-modified oligonucleotide. This would act as a capture step, with the exposed phenol attacking an EDC activated carboxyl-modified capture oligonucleotide to provide a mass shift. Unfortunately, this option is not feasible with the glutaryl donor due to the presence of a carboxylic acid on the glutaryl group, which would also attack EDC and compete with the capture oligonucleotide. However, our intention behind the use of a glutaryl group was to have this carboxylic acid as a target for capture with an amino-modified oligonucleotide.

Since the glutaryl group is no longer the capture target, we could replace it with an acyl group without a terminal carboxylic acid and attempt selections.

If we chose to maintain the glutaryl donor, we could instead target the glutaryl group of the intact glutaryl donor itself, thus isolating catalytically active sequences by capturing non-catalytic sequences. This would involve a DMT-MM activated capture using an amino-modified oligonucleotide, though capture yield would need to be close to ~100% as to avoid carry over of non-catalytically active sequences. We have yet to establish a capture procedure this effective, though it may be worth investigating more aggressive conditions assuming they do not deteriorate the glutaryl donor.

Alternating Tether Selections

An alternative means of identifying DNAzymes that perform small-molecule acylation of a non-tethered peptide lysine substrate entails the use of alternating tethers (Fig. 15). During any round of the selection process, either the peptide substrate or the acyl donor will be given a DNA tether that allows for Watson-Crick base pairing with and ligation to one of the non-random regions of the pool population. The other substrate will remain non-tethered. In the immediately following round, the tethered substrate would have its tether removed, and the non-tethered substrate would be given an appropriate tether instead. In either case, a conventional capture step that targets the reaction product will allow for separation of catalytically active sequences from the pool.

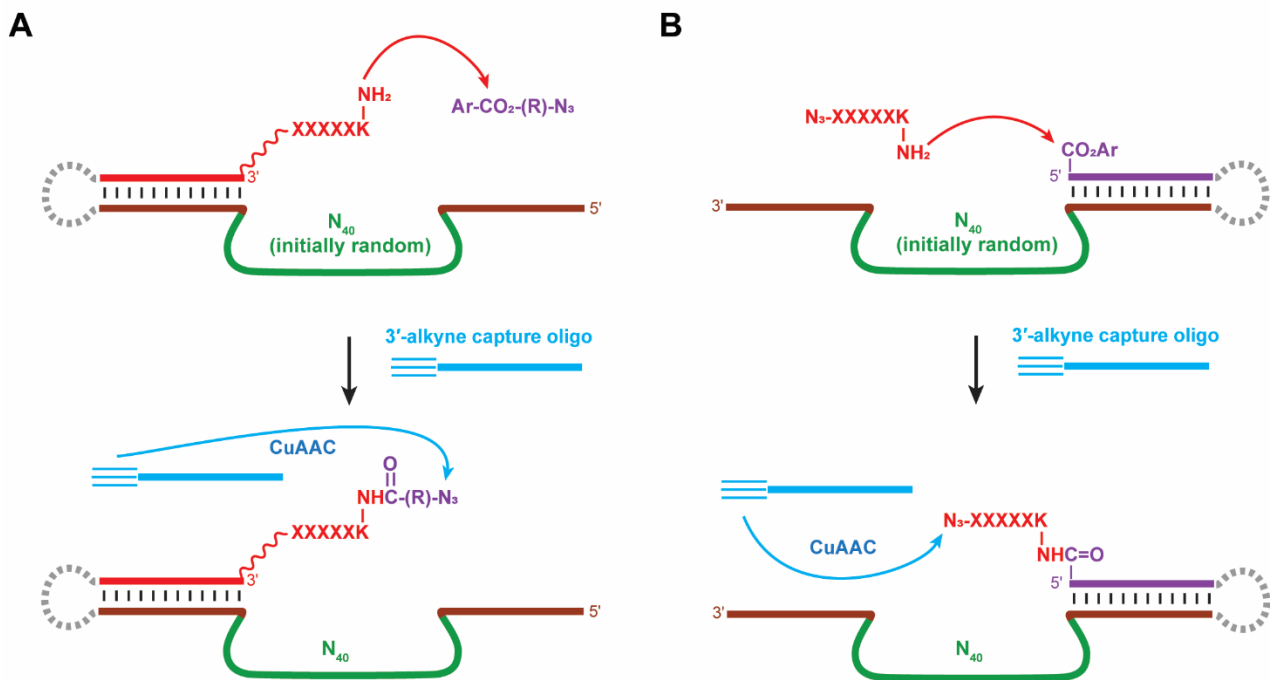

Fig. 15. An example selection scheme for alternating tether selections. **A.** Selections using azido-modified acyl donor and an anchored hexapeptide with a lysine nucleophile where the 3'-alkyne capture oligo is depicted in blue. This R group can a non-complementary (to the non-random region) oligonucleotide or a small molecule group. **B.** Selections using an azido-modified hexapeptide and an anchored acyl donor oligonucleotide.

This selection scheme, if successful, can have one of two results. The first and less exciting result is a DNAzyme that only performs catalysis when a DNA tether is present, though this tether can be on either substrate. The second and more exciting result is a DNAzyme that performs catalysis without the need for any DNA tether. The ability to catalyze something like a lysine acylation with entirely non-tethered substrates would be very valuable in the development of DNAzymes as a tool for inducing PTMs.

Multiple Turnover DNAzymes

If our selections were successful, they would result in single turnover DNAzymes that catalyze the formation of a single amide bond. This is by design as to allow for separation of active sequences from the pool. It is also a direct result of the orientation of the phenol derivative in the acyl donor. However, the ability of TFPE and TriFPE to spontaneously reform after hydrolysis suggests the feasibility of assays for multiple-turnover DNAzymes. Once selections are complete and DNAzymes have been identified, the ability of these DNAzymes to catalyze multiple turnover using a non-ligated acyl donor oligonucleotide could be determined via in trans assay. We could make use of a ligated acyl donor with an inverted phenol, like the glutaryl

donor, such that acylation results in the conjugation of the peptide to a small molecule. This would allow for continuous EDC-mediated activation with a carboxyl-modified small molecule, such that a new acyl group could be recruited after each successful acylation. However, DNAzymes would need to first be assayed for their ability to perform catalysis using the inverted-phenol acyl donor.

These assays could only be performed on already identified DNAzymes, since performing a full selection under multiple turnover conditions would be impossible. Multiple turnover necessitates that the product dissociates from the DNAzyme while any ligated reactant reforms, giving us no distinct chemical groups to target for capture and thus separation from the pool. However, multiple turnover is not a crucial feature of DNAzyme catalysis, and arguably does not warrant a dedicated selection.

Conclusion

In this body of work, I, with the assistance of my fellow researchers in the Silverman lab, developed a selection scheme for the identification of DNAzymes that perform an acylation on a non-tethered lysine in a hexapeptide. This began with my work assisting Yves in assaying and characterizing his DNAzymes selected for simple amine and model peptide acylation for their ability to perform glutarylations. These successes laid the groundwork for our investigations into non-tethered peptide nucleophiles, specifically more complex peptide sequences that would allow for sequence specificity assays at a later stage of selections. We developed and optimized ligation and capture conditions, as well as a suite of appropriate selection incubation conditions using a range of buffered conditions and aryl ester acyl donors. We've thus enabled currently ongoing selections which, at time of writing, have yet to present catalytically active sequences. However, given the efficacy of Yves's selections, as well as historical selections in the Silverman lab, we expect these selections to ultimately be successful.

The development of these selections and the results that are soon to follow represent yet another step towards our ultimate goal of a practical DNAzyme toolkit. Ideally, our DNAzymes would be able to conduct sequence-specific lysine acylation on large proteins using a broad range of potential acyl donors, without the use of a DNA tether on the protein substrate. Additional features such as multiple turnover could be valuable, though a single-turnover DNAzyme that performs a specific and highly valuable modification is still valuable. We have a number of steps to take to achieve this goal, all of which fall beyond the purview of my work. These include

selections that make use of model protein subdomains and selections for DNAzymes that perform lysine acylations with small-molecule acyl donors. However, our current work will likely provide us the opportunity to investigate more complex peptide substrates and new acyl donors, while also providing us with the necessary precedent to warrant future selections using more biologically relevant protein/small-molecule acyl substrates.

CHAPTER 5: EXPERIMENTAL METHODS

General procedure for preparation of 5'-radiolabeled oligonucleotides. A 25 µL solution containing 10 pmol of an appropriate oligonucleotide (no 5' modifications), 10 µCi of α-^{32}P-ATP, 1 µL of 10 U/µL stock of T4 polynucleotide kinase (PNK) (Thermo Scientific), and 1× PNK buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.5) (Thermo Scientific) was incubated at 37 °C for 1 h. This solution was incubated at 95 °C for 3 min to inactivate the PNK. Samples were purified on PAGE (8-20%, depending on the size of the oligonucleotide).

General procedure for preparation of 3'-radiolabeled oligonucleotides. A 25 µL solution containing 10 pmol of an appropriate oligonucleotide (no 3' modifications), 10 µCi of α-^{32}P-CTP, 1 µL of 20 U/µL stock of terminal deoxynucleotidyl transferase (TdT) (Thermo Scientific), and 1× TdT buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, pH 7.9) (Scientific) was incubated at 37 °C for 1 h. Samples were purified on PAGE (8-20%, depending on the size of the oligonucleotide).

TEN extraction for urea PAGE purification. Excised bands from urea PAGE gels were incubated in 350 µL TEN buffer (10 mM Tris-HCl, 1 mM EDTA, 300 mM NaCl, pH 8.0) at either 37 °C for 2 h or room temperature for 12-16 h. This incubation was conducted twice, with fresh TEN buffer for each incubation. These 350 µL portions were then combined and precipitated with ethanol.

Preparation of PE or 4FPE aryl ester oligonucleotides. The 5'-aryl ester oligonucleotide (PE or 4FPE) substrates were each prepared by EDC activation and reaction with phenol or 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 room temperature for 12 h. The PE or 4FPE substrate was purified by HPLC, using a Shimadzu Prominence instrument with a Phenomenex Gemini-NX C18 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 or 4FPE oligonucleotide substrate was 500 pmol (17%).

Preparation of the glutaryl donor oligonucleotide. The glutaryl donor oligonucleotide was prepared by synthesis of the illustrated glutaryl-azide compound and subsequent CuAAC reaction with 5′-alkyne-C6-GAAGAGATGGCGACTTCG-3′ (5′-alkyne modifier from Glen Research). Synthesis of the glutaryl azide was performed by graduate student Peter Yeh as described.[16] The CuAAC conjugation was performed as follows. 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%).

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 10× Taq polymerase buffer [1× = 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, 10× (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 10× Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30× (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′-TAATTAATTAATTA-CCCATCAGGATCAGCT-3′ (reverse primer). The 10-cycle PCR product from the appropriate selection round was diluted 10^3-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 10× Taq polymerase buffer. This sample was cycled 30 times according to the following PCR program: 94 °C for 2 min, 30× (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 described in the main text.

General in trans assay procedure for DNA-HEG-AAAKAA glutarylating DNAzymes. A 14 μL sample containing a 30-cycle PCR product of a DNAzyme clone (5-10 pmol), 0.75 pmol of 5'-radiolabeled DNA-HEG-AAAKAA, and 20 pmol of the glutaryl donor oligonucleotide 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 20% PAGE.

DMT-MM-activated capture of glutarylation products. An 18 μL solution containing the precipitated in trans assay product, 50 pmol of 5 '-amino modified capture oligonucleotide (5′-amino modifier from Glen Research), and 30 pmol of a splint complementary to both DNA-HEG-AAAKAA and the 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 step was initiated by bringing the sample to 20 μL containing 50 mM DMT-MM. Sample was incubated at room temperature for 12-16 h and separated by 20% PAGE.

In trans glutarylation assay for 8FL205, 8FL219, and 7FN221. Individual DNAzyme stocks were synthesized by Integrated DNA Technologies. A 14 μL sample containing 10 pmol DNAzyme, 1 pmol of 5'-radiolabeled 3'-amino modified DNA-C_3-NH_2 (3 '-amino CPG from Glen Research), and 30 pmol of the glutaryl donor oligonucleotide 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 DNAzyme-catalyzed 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 sample was incubated at 37 °C for 48 h, and 2 μL portions were taken at t = 0 h, 16 h, and 48 h. 100 pmol of a corresponding decoy oligonucleotide was added to these portions, which would displace the DNAzyme from products and reactants. Samples were separated by 20% PAGE.

Synthesis of simple amine glutarylation products. The products of several representative individual DNAzymes were analyzed by MALDI mass spectrometry. 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. Each product was prepared from a 21 μL sample containing 200 pmol of DNA-C_3-NH_2 substrate, 220 pmol of DNAzyme, and 240 pmol of acyl donor substrate. The sample 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 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. 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%

Procedure for ligation step assay. A 17 μL sample containing the PCR-amplified DNA pool (5–10 pmol), 30 pmol of DNA splint, and 50 pmol of 5'-carboxyl oligonucleotide (5'-carboxy modifier from Glen Research) 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 10× T4 DNA ligase buffer and 1 μL of 1 U/μL T4 DNA ligase. The sample was incubated at 37 °C for 12-16 h and separated by 8% PAGE. The ligation step assay was performed as above, though the DNA splint was varied between samples as described in Chapter 2.

General procedure for pre-activation of TriFPE and TFPE. TriFPE and TFPE acyl donor oligonucleotides were synthesized via a 5'-carboxyl modified oligonucleotide directly before the selection reaction due to their relative instability in solution. A 10 μL sample containing 10 pmol of 5'-carboxyl oligonucleotide (5'-carboxyl modifier from Glen Research) and 0.5 pmol of 3'-^{32}P-radiolebeled 5'-carboxyl oligonucleotide was incubated in 50 mM EDC, 50 mM TFP or TriFP, 150 mM NaCl, and one of 50 mM MES, pH 6.0, HEPES, pH 7.5, or CHES, pH 9.0 room temperature for 2 h. This sample was used directly in selection and background reactivity assays as the pre-activated acyl donor compound.

Solid-phase peptide synthesis (SPPS) procedure. For every synthesis, 150 μmol (~270 mg at 0.57 mmol/g) of Rink amide MBHA resin (100-200 mesh) was left to swell in 10 mL of *N*,*N*-dimethylmethylformamide (DMF) for 10 min. The following steps were then repeated in order until the full peptide was synthesized.

Deprotection Step. Resin was incubated in ~5 mL of 20% piperidine/80% DMF (v/v) solution under N_2 for 5 min, after which it was washed with 3 × 5 mL portions of DMF. This deprotection step was performed twice consecutively to ensure complete deprotection of the amine group on resin/elongating peptide.

Coupling Step. A 5 mL DMF solution containing 200 mM of *O*-(7-Azabenzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate (HATU) and 200 mM of an appropriate amino acid (Chem-Impex Int'l Inc.) was activated via the addition of 440 mM *N*,*N*-diisopropylethylamine (DIPEA). The deprotected resin was incubated in this solution under N_2 for 2 h, after which it was washed with 3 × 5 mL portions of DMF. A Kaiser test was conducted on a small portion of the resin after every coupling step to confirm a successful coupling via the absence of a deprotected amine. If this test indicated that coupling failed, a second coupling with 100 mM of HATU and 100 mM of the amino acid was conducted, followed by an additional Kaiser test.

6-Azidohexanoic acid coupling step. A 5 mL DMF solution containing 200 mM of 6-azidohexanoic acid, 200 mM of hydroxybenzotriazole (HOBt), and 200 mM of *N*,*N*'-diisopropylcarbodiimide (DIC) was prepared. The deprotected resin was incubated in this solution under N_2 for 16 h, after which it was washed with 3 × 5 mL portions of DMF.

Capping Step. This step was included to prevent the elongation of peptides that failed to couple in an otherwise successful coupling as indicated by Kaiser test. The resin was incubated in 7 mL of 1.4 M acetic anhydride and 2.2 M DIPEA in DMF under N_2 for 15 min, after which it was washed with 3 × 5 mL portions of DMF.

Cleavage of SPPS product from resin. The dried resin was incubated in 5 mL solution containing 95% trifluoroacetic acid (TFA), 2.5% H_2O, and 2.5% triisopropylsilane (TIPS) while stirring for 2 h. This solution was filtered through a 15 mL course porosity filter into ~60 mL of cold diethyl ether and left to cool at -80 °C for 12-16 h. The precipitate was centrifuged at 1000 rpm, washed with additional cold ether, and dried under N_2. Dried samples were purified via HPLC using a Shimadzu Prominence instrument with a Phenomenex Gemini-NX C18 column (5 μm, 10 × 250 mm), solvent A (0.1% TFA/99.9% H_2O v/v, pH 7.0), solvent B (100% acetonitrile), and flow rate of 2.0 mL/min. Purified samples were lyophilized and quantified by absorbance (A_{205}). The identity of each peptide was

confirmed via mass spectrometry using the Waters Q-TOF Ultima ESI mass spectrometer in positive ion mode at the UIUC School of Chemical Sciences Mass Spectrometry Laboratory.

Early background reactivity assays for TriFPE and TFPE. A 10 μL pre-activation sample was brought to 20 μL such that it contained one of 70 mM MES, pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, and 1 mM $ZnCl_2$; 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, and 1 mM $ZnCl_2$; or 50 mM CHES, pH 9.0, and 40 mM $MgCl_2$. Each sample also contained some concentration of N_3-AEQSWK peptide (0 μM, 10 μM, 100 μM, 1 mM, or 2 mM). Samples were incubated at either 37 °C or 55 °C for 2 h or 16 h, after which they were separated by 20% PAGE.

Stability assays for TriFPE and TFPE. The setup for these assays is exactly similar to the above described background reactivity assays, though peptide was omitted. 2 μL portions were taken at t = 10 min, 2 h, and 16 h. Degradation of the pre-activated aryl ester compound was determined via mass spectrometry. 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.

CuAAC capture of azido-peptide acylation products for scaled-up samples. A 7.5 μL solution containing 2.8 mM THPTA, 0.8 mM sodium ascorbate, and 0.4 mM $CuCl_2$ was prepared in advance to allow for Cu^{2+}/THPTA complex formation. To initiate capture, this was added 17.5 μL solution containing the selection product (~200 pmol), 250 pmol of splint, 300 pmol of 3'-alkyne modified oligonucleotide (3'-alkyne CPG from Glen Research), and 150 mM NaCl in 50 mM HEPES, pH 7.5. Sample was incubated at RT for 1 h, and capture products were separated by 8% PAGE.

CuAAC capture of azido-peptide acylation products in small scale capture assays. A 7.5 μL solution containing 2.8 mM THPTA, 0.8 mM sodium ascorbate, and 0.4 mM $CuCl_2$ was prepared in advance to allow for Cu^{2+}/THPTA complex formation. To initiate capture, this was added to a 17.5 μL solution containing the selection product (~5-10 pmol), 50 pmol of splint, 100 pmol of 3'-alkyne oligonucleotide (3'-alkyne CPG from Glen Research), and 150 mM NaCl in 50 mM HEPES, pH 7.5. The sample was incubated at room temperature for 1 h, and capture products were separated by 8% PAGE (if ligated to pool) or 20% PAGE (if not ligated to pool). This procedure was used for all capture assays with variations as described in Chapter 3.

Mass spectrometry of oligonucleotides and conjugates. 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.

Mass values for oligonucleotides and conjugates were as follows. At time of writing, comprehensive mass spectrometry of TFPE and TriFPE has not been conducted.

<u>*Oligonucleotides and conjugates that bind to the left-hand DNAzyme binding arm*</u>

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%

<u>*Oligonucleotides and conjugates that bind to the right-hand DNAzyme binding arm*</u>

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%

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

REFERENCES

1. "Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes", S. K. Silverman, *Trends Biochem. Sci.* **2016**, *41*, 595-609.
2. "Identification of Sequence-Selective Tyrosine Kinase Deoxyribozymes", S. M. Walsh, S. N. Konecki, S. K. Silverman, *J. Mol. Evol.* **2015**, *81*, 218-224.
3. "The catalytic diversity of RNAs", M. J. Fedor, J. R. Williamson, *Nat. Rev. Mol. Cell Biol.* **2005,** *6,* 399-412.
4. "The chemical repertoire of natural ribozymes", J. A. Doudna, T. R. Cech, *Nature* **2002**, *418*, 222-228.
5. "Selective Inactivation of Functional RNAs by Ribozyme-Catalyzed Covalent Modification", D. H. Burke et al., *ACS Synth. Biol.* **2017**, *6*, 3, 528–534.
6. "Mechanism of peptide bond synthesis on the ribosome", S. Trobro, J. Aqvist, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 35.
7. "Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution," H. Renata et al. *Angew. Chem. Int. Ed.* **2015**, *54*, 3351
8. "Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico," A. Broom et al. *Nat. Commun.* **2020**, *11*, 4808.
9. "Challenges in the computational design of proteins", M. Suárez, A. Jaramillo, *J. Roy. Soc., Interface* **2009**, *6*, 477-491.
10. "DNA and RNA Can Be Equally Efficient Catalysts for Carbon-Carbon Bond Formation", M. Chandra, S. K. Silverman, *J. Am. Chem. Soc.* **2008**, *130*, 2936-2937.
11. "The growing landscape of lysine acetylation links metabolism and cell signaling", C. Choudhary, B. T. Weinert, Y. Nishida, E. Verdin, M. Mann, Nat. *Rev. Mol. Cell Biol.* **2014**, *15*, 536-550.
12. "Histone acylation marks respond to metabolic perturbations and enable cellular adaptation", E. Cho et al., *Exp. Mol. Med.* **2020**, *52*, 2005-2019.
13. "Ancient Regulatory Role of Lysine Acetylation in Central Metabolism", S. H. Payne et al., *mBio* **2017**, *8*, 1894-1917.
14. "Histone acetylation: molecular mnemonics on the chromatin", J. Graff, L. Tsai, *Nat. Rev. Neurosci.* **2013**, *14*, 97-111.
15. "Protein lysine crotonylation: past, present, perspective", G. Jiang, C. Li, M. Lu, K. Lu, H. Li, *Cell Death Dis.* **2021**, *12*, 703.
16. "DNAzymes for Amine and Peptide Lysine Acylation", T. Yao, J. J. Przybyla, P. Yeh, A. M. Woodard, H. J. Nilsson, B. M. Brandsen, S. K. Silverman, *Org. Biomol. Chem.* **2021**, *19*, 171-181.
17. "Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology", H. M. Burke, L. McSweeney, E. M. Scanlan, *Nat. Commun.* **2017**, *8*, 15655.
18. "A Generalizable DNA-Catalyzed Approach to Peptide-Nucleic Acid Conjugation", C. Chu, O. Wong, S. K. Silverman, *ChemBioChem* **2014**, *15*, 1905-1910.
19. "Comprehensive succinylome profiling reveals the pivotal role of lysine succinylation in energy metabolism and quorum sensing of *staphylococcus epidermidis*", X. Gao et al., *Front. Microbiol.* **2021**, *11*, 632367.
20. "The first identification of lysine malonylation substrates and its regulatory enzyme", Y. Zhao et al., *Mol. Cell. Proteomics*. **2011**, *10*, 110.
21. "Post-Translational Modification of Proteins in Toxicological Research: Focus on Lysine Acylation", S. Lee, *Toxicol. Res.* **2013**, *29*, 81-86.
22. "Nucleic Acid Catalysis: Metals, Nucleobases, and Other Cofactors", W. L. Ward, K. Plakos, V. J. DeRose, *Chem. Rev.* **2014**, *114*, 4318-4342.
23. "Lysine Glutarylation Is a Protein Post-Translational Modification Regulated by SIRT5", Y. Zhao et al., *Cell Metab*. **2014**, *19*, 605-617.

24. “Characterization and identification of lysine glutarylation based on intrinsic interdependence between positions in the substrate sites”, K. Huang, H. Kao, J. B. Hsu, S. Weng, T. Lee, *BMC Bioinform.* **2019**, *19*, 384.
25. “The relationship of copper to DNA damage and damage prevention in humans”, M. C. Linder, *Mutat. Res.* **2012**, *733*, 83-91.
26. “Site-specific functionalization of proteins and their applications to therapeutics antibodies,” R. Vught, R. J. Pieters, E. Breukink, *Comput. Struct. Biotechnol. J.* **2014**, *9*, 201402001.
27. “Metallodendritic grafted core-shell γ-Fe2O3 nanoparticles used as recoverable catalysts in Suzuki C-C coupling reactions,” D. Rosario-Amorin, M. Gaboyard, R. Clérac, L. Vellutini, S. Nlate, K. Heuzé, *Chem. Eur. J.* **2012**, *18*, 3305-3315.
28. “An Extremely Powerful Acylation Reaction of Alcohols with Acid Anhydrides Catalyzed by Trimethylsilyl Trifluoromethanesulfonate,” P. A. Procopiou, S. P. D. Baugh, S. S. Flack, G. G. Inglis, *J. Org. Chem*. **1998**, *63*, 2342-2347.
29. "Catalytic DNA with Phosphatase Activity", J. Chandrasekar, S. K. Silverman, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5315-5320.
30. "Phosphoserine Lyase Deoxyribozymes: DNA-Catalyzed Formation of Dehydroalanine Residues in Peptides", J. Chandrasekar, A. C. Wylder, S. K. Silverman, *J. Am. Chem. Soc.* **2015**, *137*, 9575-9578.
31. "DNA-Catalyzed Introduction of Azide at Tyrosine for Peptide Modification", P. Wang, S. K. Silverman, *Angew. Chem. Int. Ed.* **2016**, *55*, 10052-10056.