In Vitro Selection of Phosphoserine and Phosphothreonine Lyase DNA Enzymes

By

Adam Christopher Wylder

Thesis
for the Bachelor of Science
in
Biochemistry

College of Liberal Arts and Sciences
University of Illinois
Urbana-Champaign, Illinois
2017
ABSTRACT

In nature, the canonical role of DNA is the storing and transferring of genetic information, while proteins and RNA can act as catalysts, or enzymes. Despite this delegation by nature, artificial single-stranded DNA catalysts, or deoxyribozymes, can be identified by in vitro selection. The genotype is an intrinsic property of nucleic acid catalysts; any sequence exhibiting the desired phenotype can be amplified. This is a notable advantage of nucleic acid catalysts over protein catalysts, as protein catalysts, independent from their genotypic information, cannot be amplified. Compared to RNA catalysts, DNA catalysts are more stable, cheaper to synthesize, directly amplifiable by polymerases, and exhibit similar catalytic proficiency. Therefore, the identification of deoxyribozymes through in vitro selection is a tractable endeavor to solve modern chemical and biological problems.

One such biological problem is a lack of chemoselective catalysts for the formation of post-translational modifications on proteins and peptides. This problem can be rectified by the de novo approach of in vitro selection, which allows the identification of novel enzymes to catalyze specific chemical reactions, including the modification of peptide and protein substrates. One post-translational modification of interest is the catalytic β-elimination of phosphate from phosphoserine (pSer) to form dehydroalanine (Dha), a non-proteogenic, electrophilic, α,β-dehydroamino acid. An enzyme that catalyzes this conversion of pSer to Dha is known as a pSer lyase. The actions of pSer lyases are of vital importance to several organisms, as Dha is an intermediate in the synthesis of many classes of cyclic, bioactive peptides, including lantibiotics, thiopeptides, and microcystins. Selections performed in the Silverman lab identified two pSer lyase deoxyribozymes, which expanded the known catalytic scope of DNA. The practical applications of one pSer lyase deoxyribozyme, DhaDz1, were demonstrated through its use in the
chemoenzymatic synthesis of a cyclic cystathionine-containing peptide, a more stable analogue of the biologically active compstatin peptide. While DhaDs1 could act on several model peptide substrates, its activity was not general enough to function on many of the tested pSer-containing peptides, demonstrating the deoxyribozyme’s limited scope of potential substrates. In vitro selection experiments were thus designed to identify pSer lyase deoxyribozymes which could act on a more diverse set of substrates. This selection strategy alternates which chemically distinct peptide is presented to the oligonucleotide sequences, thereby applying pressure for the DNA enzymes to accommodate diverse substrates. Additionally, this alternating selection strategy is being incorporated in selections which aim to identify phosphothreonine (pThr) lyase deoxyribozymes, which convert pThr to dehydrobutyrine (Dhb) through a reaction analogous to that of pSer lyases. Enzymes identified through these selections are expected to have practical applications in both synthesis and phosphoproteomics studies.
ACKNOWLEDGEMENTS

I would like to thank Prof. Scott Silverman for the amazing opportunity and experience he provided me. Working in the Silverman lab gave me extraordinary opportunities to further immerse myself in science and to begin my growth as an independent researcher.

I would also like to thank Dr. Jagadeeswaran Chandrasekar for his training and patience, ultimately teaching me much of what I know today. I am also indebted to the excellent scientists I have had the pleasure of working and speaking with throughout various stages of my development as a scientist.

Finally, I would like to thank my friends and family, who have always provided much-needed support in my life. Specifically, the security offered by my family has been paramount in allowing me to follow my passions and succeed to the best of my ability.
# Table of Contents

Abstract ................................................................................................................................. 2  
Acknowledgments................................................................................................................ 4  
Table of Contents .................................................................................................................. 5  
Introduction............................................................................................................................ 6  
Materials and Methods ......................................................................................................... 10  
Results and Discussion ......................................................................................................... 19  
Conclusions............................................................................................................................. 30  
References .............................................................................................................................. 32
INTRODUCTION

Biological catalysts, or enzymes, are necessary for reactions relevant to life to occur at their necessary rates. Enzymes achieve chemoselective and stereospecific catalysis through several mechanisms, including donating or accepting protons, contributing toward favorable electrostatics, adopting conformations which present the substrate in the catalytically optimal orientation, and preferential transition state binding. Enzymatic activity was long thought exclusive to proteins, polymers of amino acids. Thomas Cech and Sidney Altman challenged this orthodoxy in the 1980s by separately publishing on the fundamental catalytic ability of RNA. While the bases of ribonucleotides have less functional group diversity than the side chains of amino acids, the functional groups found in RNA are sufficient for catalysis. RNA’s ability to adopt a variety of conformations further enhances catalysis.

DNA differs from RNA through both the incorporation of the T nucleobase instead of U, and the omission of a 2′-OH on its sugar. Given this structural similarity, it follows that single-stranded DNA, like single-stranded RNA, can achieve catalytic activity through adopting relevant conformations which present functional groups in a favorable environment. DNA’s fundamental catalytic ability in vitro was confirmed by Breaker and Joyce in their publication on a synthetic, single-stranded, RNA-cleaving DNA enzyme. Inspired by this effort, many DNA enzymes, or deoxyribozymes, have been identified with diverse functionalities, including RNA cleavage, DNA phosphodiester hydrolysis, DNA 5′-phosphorylation, tyrosine phosphorylation, phosphotyrosine and phosphoserine dephosphorylation, porphyrin metalation, as well as many others.

Despite the breadth of known enzymes, their applications are limited to those which mimic their use in nature. Thus, slight deviation in substrate identity or environmental conditions can lead to drastic reduction or elimination of catalytic activity. Additionally, many chemical reactions have
no known biological catalyst. Together, these considerations underscore the importance of developing new enzymes through de novo techniques like in vitro selection, as well as expanding the activity of already-existing enzymes through techniques like directed evolution.

The primary technique used to identify novel nucleic acid catalysts is in vitro selection. This begins with the solid-phase synthesis of fixed-length nucleic acid sequences, which contain a randomized region of length $X$ (N$_X$) flanked on their 5' and 3' ends by fixed regions, which are necessary for PCR amplification and substrate ligation. The population of synthesized single-stranded DNA sequences next is attached to the substrate of interest, producing a population in which every sequence is covalently attached to the substrate. Next, the sequences and their attached substrate are incubated in conditions favorable for the reaction, including metal cation cofactors. During this incubation, sequences catalytically active in the tested conditions modify their attached substrate. Next, a method to separate the sequences with the desired catalytic function from the rest of the population must be performed; typically, denaturing polyacrylamide gel electrophoresis (PAGE) is used. While most sequences in the population migrate within a single band during PAGE, those which adopt a complex conformation while not catalyzing the desired modification may migrate repeatedly to a certain part of the gel, potentially the location where the product of interest migrates. Thus, undesired sequences are purified along with the desired sequences. All sequences retrieved are then amplified by PCR. This process is then repeated until the catalytically active sequences dominate the population, at which point the active sequences are cloned and characterized.

Rational design is another method used to identify enzymes de novo. By knowing the mechanisms and structures of naturally occurring enzymes, researchers can attempt to rationally design a protein taking inspiration from these common motifs. Once the protein is synthesized,
researchers frequently subject it to modification through directed evolution, which differs from the
de novo techniques mentioned above in that it focuses on modulating the activity of already
existing proteins. The technique begins by diversifying via mutagenesis a pool of DNA encoding
proteins of interest. These proteins are expressed and screened for the activity of interest, at which
point the corresponding genes are amplified and subjected to further rounds of mutagenesis.16

Nucleic acid catalysts have several key features which make them ideal candidates for the
identification of catalysts de novo. First, their ability to be amplified (unlike proteins) allows the
use of in vitro selection, and thus the identification of new enzymes without extensive structural
and mechanistic knowledge of associated natural enzymes. Second, a randomly generated polymer
of nucleotides is more likely to adopt a complex conformation than is a randomly generated
polymer of amino acids, as the formation of secondary structural elements in proteins relies heavily
their context within a stabilizing tertiary structure.17 In comparison, nucleic acid polymers can
adopt diverse conformations due to their high conformational freedom. Consider randomly
generated pools of both 40-mer polypeptides and 40-mer nucleic acid sequences; both cases would
explore a sequence space of size $40^n$, where $n$ is the number of potential monomers ($n=4$ for the
former and $n=20$ for the latter). Given a fixed length polymer, there are more unique polypeptides
than unique nucleic acid sequences. Taken collectively, examining $X$ nucleic acid sequences
explores a higher ratio of total sequence space than examining $X$ polypeptides, where each of the
examined nucleic acid sequences are additionally more likely to adopt a structured conformation.

Identifying DNA catalysts through in vitro selection has many practical advantages over
using RNA. Solid-phase synthesis of RNA is more difficult and resource-intensive than that of
DNA, as the 2’-OH group of ribonucleotides requires extensive protection/deprotection, typically
through usage of silyl ethers.18 Additionally, RNA is less stable than DNA both because the 2’-
OH group of RNA can perform nucleophilic attack on the adjacent phosphodiester linkage, and because RNA is subject to degradation by RNases. Due to these considerations, the Silverman lab focuses on the identification of DNA enzymes to catalyze diverse reactions on various substrates. Two related reactions studied in the Silverman lab are the formation of dehydroalanine (Dha) and dehydrobutyrine (Dhb) from phosphoserine (pSer) and phosphothreonine (pThr), respectively.

Dha and Dhb are interesting as they are two of the only electrophilic amino acids found in nature. Incorporation of these electrophilic sites provides handles for modification by nucleophilic addition. These electrophilic handles are also useful as targets for cyclization in peptides including lantibiotics, thiopeptides, and microcystins as the electrophilic residues are subject to Michael-type addition by internal cysteine (Cys) residues. Natural products containing Dha, Dhb, or other α,β-dehydroamino acids exhibit several therapeutic properties, including antibacterial, antifungal, and anticancer activities.

In nature, Dhb has been shown to be formed through the phosphorylation of threonine followed by β-elimination. Enzymes which catalyze this β-elimination step are known as pThr lyases. By extension, an enzyme which catalyzes the β-elimination of a phosphorylated serine are known as pSer lyases, although no pSer lyases have been found in nature. Harsh, basic conditions can also catalyze the β-elimination of a phosphate from pSer/pThr. One potential application of this method is as a phosphopeptide enrichment tool for usage in phosphoproteomics studies. Following base treatment (typically ~1 M hydroxide) of peptidic fragments or full-length proteins containing pSer or pThr residues, the resulting electrophilic sites can undergo Michael-type addition by ethanedithiol, which can then react with biotin maleimide, forming a complex captured by resin-linked streptavidin. However, the applicability of this methodology is restricted, as the harsh conditions in the elimination step have the potential to cause numerous, undesired side
reactions. To help overcome this issue, Davis et. al published the use of bis-alkylation-elimination to convert cysteine (Cys) into Dha. Unfortunately, any additional Cys residues will undergo conversion to Dha as well. This is problematic during total synthesis of lantibiotics, which would need to be synthesized from a precursor with two Cys residues, as there needs to be both a Dha and Cys residue for cyclization.

The Silverman lab’s initial work on phosphoserine lyase deoxyribozymes expanded the known catalytic ability of DNA and demonstrated practical applications of pSer lyase deoxyribozymes through the chemoenzymatic synthesis of a more stable compstatin analogue. The work described in this thesis seeks to overcome current synthetic limits in the installation of pSer and pThr residues through the identification of both pSer lyases which can work on a diverse selection of substrates, as well as the first pThr lyase deoxyribozymes.

MATERIALS AND METHODS

Collaborations

The selections to identify DhaDz1 and DhaDz2, as well as their cloning and characterizations, were performed by former Silverman lab graduate student, Dr. Jagadeeswaran Chandrasekar. I assisted Dr. Chandrasekar in performing several DhaDz1/DhaDz2-catalyzed reactions, as well as in the solid-phase synthesis and enzymatic preparation of compstatin. Dr. Chandrasekar largely designed the initial selection scheme described in this thesis, synthesized the initial batch of substrates used, and acted as a guide for the rest of the work.

Synthesis of Oligonucleotides

DNA oligonucleotides were either obtained from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an ABI 394 instrument using reagents
from Glen Research. All oligonucleotides were purified by 7 M urea denaturing PAGE with running buffer 1× TBE (89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3), extracted from the polyacrylamide with TEN buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaCl), and precipitated as described previously.\textsuperscript{32,33}

**Synthesis of Peptides**

*Synthesis of peptides for use in selection.* All amino acid monomers, including those for phosphoserine (pSer, $S^p$), homocysteine (Hcy), dehydrobutyrine (Dhb), trifluoroacetylysine (KTfa), and $S$-methylcysteine (CysMe, CMe) were obtained from Chem-Impex (Wood Dale, IL). All peptides were prepared by solid-phase synthesis, using Fmoc Rink amide MBHA resin from Chem-Impex, with $N,N,N,N'$-tetramethyl-$O$-(7-azabenzotriazol-1-yl)uronium hexafluorophosphate (HATU) as the coupling agent.\textsuperscript{34} Synthesis of the peptides AAAS$^p$AA, AAAT$^p$AA, AAAT$^p$AAA, AFFS$^p$FFA, AFFT$^p$FFA, AEES$^p$EEA, AEET$^p$EEA, AKKS$^p$KKA, AKKT$^p$KKA, and GIT$^p$PVAF was performed at 0.2 mmol scale, initiated using 260 mg of Rink amide resin with a loading capacity of 0.77 mmol/g. All steps were monitored by ninhydrin test. For each coupling, 5 equivalents (1.0 mmol) of Fmoc-amino acid, 4.9 equivalents (373 mg, 0.98 mmol) of HATU, and 10 equivalents (350 μL, 2.0 mmol) of $N,N$-diisopropylethylamine (DIPEA) were mixed in 5 mL of anhydrous DMF. The coupling reaction was initiated by adding this mixture to the resin and agitating by bubbling of nitrogen for 60 min, followed by washing with DMF (3 × 10 mL). The N-terminus of the newly coupled residue was deprotected by three repetitions of a cycle starting with agitating the resin in 20% piperidine in DMF (5 mL) for 5 min, and ending by washing with DMF (3 × 5 mL). The peptide was cleaved from the solid support by stirring the resin in a separate vial for 90 min with 5 mL of trifluoroacetic acid (TFA), 125 μL of water, and 50 μL of triisopropylsilane. The resin was removed by filtration, and the resulting solution was dried via
rotary evaporator. To the resulting material 20 mL of cold diethyl ether was added, and after filtration and HPLC purification, a white solid was obtained. The HPLC gradient was 2-12% CH₃CN over 60 minutes (with the sample expected at ~30 minutes), except for purification of AFFTPFFA, which used a gradient of 30-50% ACN over 60 minutes. The mass of the dried purified product was measured to quantify the moles of peptide.

Synthesis of the peptides AAA(Dhb)AA and AFF(Dhb)FFA was performed as described above, with the following procedural modifications. Synthesis was performed at 0.1 mmol scale, initiated using 130 mg of 0.77 mmol/g Rink amide resin. For each coupling of a standard amino acid, 5 equivalents (0.5 mmol) of Fmoc-amino acid, 4.9 equivalents (187 mg, 0.49 mmol) of HATU, and 10 equivalents (175 μL, 1.0 mmol) of DIPEA were mixed in 3 mL of anhydrous DMF and agitated with the resin for 60 min. For Fmoc-α,β-dehydro-2-aminobutyric acid, 5 equivalents (0.2 mmol) of Fmoc-amino acid, 1.96 equivalents (75 mg, 0.20 mmol) of HATU, and 4 equivalents (70 μL, 0.4 mmol) of DIPEA were mixed in 3 mL of anhydrous DMF and agitated with the resin for 120 min. For the step involving deprotection of Fmoc-α,β-dehydro-2-aminobutyric acid, extra care was taken to utilize dry solvent to avoid hydrolysis of the enamine. The peptide was cleaved from the solid support by stirring the resin in a separate vial containing 3 mL of TFA, 100 μL of water, 100 μL of ethanedithiol, and 40 μL of triisopropylsilane for 90 min. These peptides were purified and quantified as above.

**Synthesis of DNA-Anchored Peptides**

The DNA-anchored pSer-, pThr-, and Dhb-containing peptide was synthesized by reductive amination, yielding a linkage between the N-terminal α-amino group of the peptide and the 3’-end of the oligonucleotide. To begin, 1.0 nmol of the starting 3’-rA oligonucleotide was oxidized in a 100 μL solution of 100 mM HEPES, pH 7.5, and 10 mM NaIO₄ for 1 h. Excess
NaIO$_4$ was removed by ethanol precipitation, and the precipitated product was dissolved in 65 μL of water. The product, along with 100 equivalents (100 nmol) of the peptide, was incubated in 100-μL of 100 mM NaOAc, pH 5.2, 50 mM, 50 mM NiCl$_2$, and 10 mM NaCNBH$_3$ at 37°C for 14 h. This product was precipitated by addition of 220 μL of water, 30 μL of 3 M NaCl, and 900 μL of ethanol and purified by denaturing 20% PAGE.

DNA-anchored AAA(Dha)AA peptide was prepared through usage of a S-methylcysteine, CysMe (CMe), containing precursor hexapeptide. To convert this precursor into the desired AAAAS$^P$AA peptide, a 50 μL sample containing 500 pmol of DNA-anchored AAA(CMe)AA, 100 mM HEPES, pH 7.5, and 10 mM NaIO$_4$ was incubated at room temperature for 1 h to oxidize the thioether to a sulfone. This product was precipitated by addition of 220 μL of water, 30 μL of 3 M NaCl, and 900 μL of ethanol. The sulfone product was dissolved in 20 μL of 10 mM NaOH, pH 12.0, and incubated at 37 °C for 4 h to form the final elimination product, DNA-anchored AAA(Dha)AA, which was precipitated by addition of 250 μL of water, 30 μL of 3 M NaCl, and 900 μL of ethanol.

Each DNA-anchored oligonucleotide contains a hexa(ethylene glycol) [HEG] spacer, whose conformational freedom and longer tether length relative to the commonly employed C3 tether may allow the tethered substrate to instead appear free in solution, potentially allowing identification of enzymes which can function on free substrates.

**Procedure for Synthesis of Long 5′-thiol Capture Oligonucleotide**

A 34 μL sample containing 2 nmol of short 5′-thiol capture oligonucleotide, 2.5 nmol of the splint oligonucleotide 5′-TATCAGTATATGCGATGCTTCCGTCCATCTCTTCC-3′, and 3 nmol of 5′-phosphorylated 5′-AGCATCGCATATACTGATAAACAAACAACAACAAACAACACAC-AC-3′ was annealed by heating at 95°C for 3 min and cooling on ice for 5 min in 5 mM Tris, pH
7.5, 15 mM NaCl, and 0.1 mM EDTA. Following annealing, 4 μL of 10× T4 DNA ligase buffer (Fermentas) and 2 μL of 5 U/μL T4 DNA ligase (Fermentas) was added. The sample was incubated at 37 °C for 12 h, and purified by 12% PAGE to obtain the long 5′-thiol capture oligonucleotide used in selections MA1, MB1, and QM1-LX1. The sample was freshly treated with DTT prior to each use, as with the short 5′-thiol capture oligonucleotide (see Procedure for preparation of 5′-thiol capture oligonucleotide).

**Procedure for Capture Validation Assays**

To measure the capture yield of the standards in cis, the 5′-terminus of a DNA-anchored Dhb-containing peptide [AAA(Dhb)AA, AEE(Dhb)EEA, AKK(Dhb)KKA, AFF(Dhb)FFA, or GI(Dhb)PVAF] was first ligated to the 3′-terminus of the random oligonucleotide pool using a DNA splint and T4 DNA ligase (see Procedure for ligation step). The DNA-anchored peptide was 5′-radiolabeled using γ-32P-ATP and Optikinase, which was used because it lacks the 3′-phosphatase activity of the typically used T4-PNK. A 10 μL sample containing 0.1 pmol of 5′-radiolabeled DNA-anchored peptide linked to the pool, 20 pmol of capture splint, and 100 pmol of the short 5′-thiol capture oligonucleotide was incubated 37 °C for 12 h in various buffer conditions following incubation in selection conditions (see Procedure for selection). Capture was also validated using the DNA-anchored standards prior to ligation with the random oligonucleotide pool using the same method described above. The “background rate” of the capture reaction was determined using the procedures mentioned above, except by using DNA-anchored pSer/pThr-containing peptides.
In Vitro Selection Procedure

An overview of the key selection and capture steps of each round is shown in Figure 1. The random deoxyribozyme pool and PCR primers are summarized in Table 1. In each round, the ligation step to attach the 3’-end of the oligonucleotide pool to the 5’-end of the DNA-anchored peptide substrate was performed using a DNA splint and T4 DNA ligase. The length of the 5’-thiol capture oligonucleotide used in selection alternated between short (odd-numbered rounds) and long (even-numbered rounds), to avoid separation of sequences whose conformation cause them to migrate to the same location as the capture standard.

Procedure for ligation step in round 1. A 35 μL sample containing 1 nmol of DNA pool, 850 pmol of DNA splint, and 750 pmol of 5’-phosphorylated DNA-peptide substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95°C for 1 min and cooling on ice for 5 min. To this solution, 4 μL of 10× T4 DNA ligase buffer (Fermentas) and 1 μL of 5 U/μL T4 DNA ligase (Fermentas) was added. The resulting solution was incubated at 37 °C for 12 h and separated by 8% PAGE.

Procedure for ligation step in subsequent rounds. A 16 μL sample containing the PCR-amplified DNA pool (~5–10 pmol), 25 pmol of DNA splint, and 50 pmol of 5’-phosphorylated DNA-hexa-/hepta-peptide substrate was annealed in 5 mM Tris, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 1 min and cooling on ice for 5 min. To this solution 2 μL of 10× T4 DNA ligase buffer (Fermentas) and 1 μL of 1 U/μL T4 DNA ligase (Fermentas) was added. The sample was incubated at 37 °C for 12 h and separated by 8% PAGE.
<table>
<thead>
<tr>
<th>Pool 3</th>
<th>Biased Pool (N₄₀)</th>
<th>D5228 5'-CGAAGTGCCGATGTAGTTTACCAGTTTCTGAGACCAATACGCATAAAGGAGCTGATCCTGGATGG-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pool 4</td>
<td>Pool (N₄₀)</td>
<td>D4925 5'-GGAATTAAGACTGAATTC-ATAGATGCTGATCCTGGATGG-3'</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5065 5'-GGACTACCTTTATCGATAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward Primer</td>
<td>D3987 5'-CGAAGAAGGCTCCTTC-3'</td>
</tr>
<tr>
<td>Pool 5</td>
<td>Pool (N₄₀)</td>
<td>D4638 5'-CGAATGAGACTGAATTC-ATAGATGCTGATCCTGGATGG-3'</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5066 5'-GGATATCTCGTTTAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward primer</td>
<td>D4628 5'-CGAAATGAGACTGAATTC-3'</td>
</tr>
<tr>
<td>Pool 6</td>
<td>Pool (N₄₀)</td>
<td>D4640 5'-CGAATGAGGGCTATTTTC-ATATGTCTTTTCAA</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5068 5'-GGACTATTGAAAGACATAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward primer</td>
<td>D4629 5'-CGAAATGAGCTAATTTTC-3'</td>
</tr>
<tr>
<td>Pool 7</td>
<td>Pool (N₄₀)</td>
<td>D4642 5'-CGAATGAGGTCAATTTTC-ATATGTCTTTTCAA</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5068 5'-GGACTATTGAAAGACATAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward primer</td>
<td>D4629 5'-CGAAATGAGCTAATTTTC-3'</td>
</tr>
<tr>
<td>Pool 8</td>
<td>Pool (N₄₀)</td>
<td>D4643 5'-CGAATGAGGTCAATTTTC-ATATGTCTTTTCAA</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5068 5'-GGACTATTGAAAGACATAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward primer</td>
<td>D4629 5'-CGAAATGAGCTAATTTTC-3'</td>
</tr>
<tr>
<td>Pool 9</td>
<td>Pool (N₄₀)</td>
<td>D4643 5'-CGAATGAGGTCAATTTTC-ATATGTCTTTTCAA</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5068 5'-GGACTATTGAAAGACATAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward primer</td>
<td>D4629 5'-CGAAATGAGCTAATTTTC-3'</td>
</tr>
<tr>
<td>Pool 10</td>
<td>Pool (N₄₀)</td>
<td>D4644 5'-CGAATGAGGTCAATTTTC-ATATGTCTTTTCAA</td>
</tr>
<tr>
<td></td>
<td>L substrate</td>
<td>D5068 5'-GGACTATTGAAAGACATAT-HEG-rA-3'</td>
</tr>
<tr>
<td></td>
<td>Reverse primer</td>
<td>D367 5'-AACAAACAACACXCCCATCAGGATCAGC-3'</td>
</tr>
<tr>
<td></td>
<td>Forward primer</td>
<td>D4629 5'-CGAAATGAGCTAATTTTC-3'</td>
</tr>
</tbody>
</table>

**Table 1.** The sequences of the oligonucleotide pools, DNA anchor sequences, and primers used in selections. All L substrates contain binding arms corresponding to their respective pool. HEG denotes a hexa(ethylene glycol) spacer (see “Synthesis of DNA-Anchored Peptides”).
Procedure for selection step in round 1. Each selection experiment was initiated with 200 pmol of the ligated N₄₀ or N₃₀ pool. A 20 μL sample containing 200 pmol of ligated N₄₀ or N₃₀ pool was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95°C for 1 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 40 μL total volume containing 70 mM HEPES, pH 7.5, 1 mM ZnCl₂, 20 mM MnCl₂, 40 mM MgCl₂, and 150 mM NaCl. Selection PA1 was instead in 1 mM Zn²⁺ and 3 μM of each Ce³⁺, Eu³⁺, and Yb³⁺. All samples were incubated at 37 °C for 14 h.

Procedure for selection step in subsequent rounds. A 10 μL sample containing ligated pool was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95 °C for 1 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, 1 mM ZnCl₂, 20 mM MnCl₂, 40 mM MgCl₂, and 150 mM NaCl. Selection PA1 was instead in 1 mM Zn²⁺ and 3 μM of each Ce³⁺, Eu³⁺, and Yb³⁺. All samples were incubated at 37 °C for 14 h.

Procedure for preparation of 5'-thiol capture oligonucleotide. The 5'-thiol capture oligonucleotide was prepared by DTT reduction of the precursor oligonucleotide (prepared using the Glen Research thiol modifier C₆ S-S) immediately prior to each use. To accomplish this deprotection, a 50 μL sample containing 2 to 5 nmol of precursor oligonucleotide in 100 mM HEPES, pH 7.5, and 25 mM DTT was incubated at 37 °C for 2 h. To remove excess DTT, the sample was precipitated by adding 220 μL of water, 30 μL of 3 M NaCl, and 900 μL of ethanol.

Procedure for capture step in round 1. To each 40 μL selection sample, 6 μL of 0.5 M EDTA, pH 8.0 (3 μmol) was adding as a chelating agent. The DNA was precipitated by adding 224 μL of water, 30 μL of 3 M NaCl, and 900 μL of ethanol. The precipitated sample was brought to 40 μL total volume containing 300 pmol of capture splint and 500 pmol of short 5'-thiol capture
oligonucleotide in 100 mM Na$_2$CO$_3$, pH 9.2, and incubated at 37 °C for 12 h. The sample was loaded directly on an 8% polyacrylamide gel.

**Procedure for capture step in subsequent rounds.** To each 20 μL selection sample, 3 μL of 0.5 M EDTA, pH 8.0 (1.5 μmol) was added. Each sample was precipitated by the addition of 247 μL of water, 30 μL of 3 M NaCl, and 900 μL of ethanol. The precipitated sample was brought to 10 μL total volume containing 20 pmol of capture splint, 100 mM Na$_2$CO$_3$, pH 9.2, and 100 pmol of the either the long 5’-thiol capture oligonucleotide (odd rounds) or the short 5’-thiol oligonucleotide (even rounds). The solution was incubated at 37 °C for 12–14 h and loaded directly on 8% polyacrylamide gel.

**Procedure for PCR.** Both a 10-cycle and a 30-cycle PCR reaction were performed at every round of selection. The initial 100 μL solution for 10-cycle PCR contained 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$_4$)$_2$SO$_4$, 10 mM KCl, 2 mM MgSO$_4$, and 0.1% Triton X-100]. The 10-cycle PCR program was as follows: 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 via phenol/chloroform extraction. Next, a 50 μL solution was prepared for 30-cycle PCR, which contained 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 $\alpha$-$^{32}$P-dCTP (800 Ci/mmol), and 5 μL of 10× Taq polymerase buffer. The 30-cycle PCR program was as follows: 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 products from this reaction were separated by 8% PAGE. Because the reverse primer contains a non-amplifiable spacer which stops the action of Taq polymerase, the forward and reverse PCR products are readily separable.
RESULTS AND DISCUSSION

Selections for the Identification of Sequence-General pSer and pThr Lyase Deoxyribozymes

*Initial selection design.* The inability of DhaDz1, the previously identified pSer lyase deoxyribozyme, to act on all possible pSer-containing substrates underscores the importance of identifying new lyase deoxyribozymes able to act on substrates with a diversity of side-chains. Because the selection which identified DhaDz1 was performed with DNA-anchored AAASPAA, no selection pressure was applied that would enrich the DNA sequences whose active site accommodated more chemically diverse side-chains. Additionally, characterizations of DhaDz1 and DhaDz2 revealed that they were unable to perform catalysis on an alanine-rich substrate, like that used in selection except that they contained pThr instead of pSer. DhaDz1/DhaDz2 lack pThr lyase activity because there was no selection pressure imposed for them to catalyze β-elimination on a pThr-substrate. However, the inability of DhaDz1 to work on substrates like those mentioned above does not reflect the abilities of DNA enzymes not yet identified. By applying selection pressure specific to the desired activity during selection, deoxyribozymes which overcome the limitations of others can be identified. Thus, deoxyribozymes which act in a peptide-sequence-independent manner could be enriched by presenting a diverse class of peptide sequences to the pool of DNA sequences as substrates, and pThr lyase deoxyribozymes could be enriched by presenting the pool with a pThr-substrate.
Figure 1. Generic scheme for in vitro selection of pSer/pThr lyase deoxyribozymes, where R = H for pSer and R = CH₃ for pThr. A DNA-anchored pThr/pSer-containing peptide is covalently attached to the 3′-end of the DNA pool through splint ligation. The ligated sequences are incubated with metal ions according to the individual selection. DNA-catalyzed elimination of phosphate from the pThr/pSer on the DNA-anchored substrate leads to the formation of Dhb/Dha, which is then captured by a 5′-thiol oligonucleotide through a splinted reaction. The gray lines represent primers used during PCR amplification (the reverse on the left, and the forward on the right), which bind to fixed sequences flanking the randomized portion (see Table 1). Not depicted in this image is the hexa(ethylene glycol) [HEG] tether linking the 3′-end of the DNA anchor to the peptide substrate.

Inspired by these considerations, new selection experiments were designed to identify pSer/pThr lyases which can act on substrates with chemically-distinct side-chains, or that are “sequence general.” Peptides AKKXKKA, AEEEXEA, AFFXFFA (where X = pSer, pThr, or Dhb) and AAA(Dhb)AA were synthesized and conjugated to a DNA anchor by reductive amination. The initial “alternation strategy” employed in this selection was to maintain the identity of pSer or pThr through all rounds of selection and alternate the surrounding residues during each round of selection. An example pThr lyase selection would alternate between AKKTₚKKA, AEETₚEEA, and AFFTₚFFA, thereby forcing DNA enzymes to accommodate positive charges.
(Lys), negative charges (Glu), and aromatic groups (Phe) to survive the selection pressure. The selections cover 8 randomized pools of oligonucleotides, where randomized regions 40 nucleotides in length (N_{40}) are flanked by binding arms unique to each pool (Figure 3).

**Refinement of the capture strategy.** An essential component of initiating new in vitro selection experiment is the development and refinement of a “capture” reaction by which the catalytically active sequences can be separated from the noncatalytic or undesired sequences. In this case, the desired sequences to be captured are those which catalyze the formation of Dha and Dhb. Thus, the capture mechanism was largely taken from that used in the identification of DhaDz1 and DhaDz2. This capture mechanism relies on performing a Michael-type addition on the electrophilic residue by a 5′-thiol oligonucleotide, leading to an increase in mass and thus an upward shift during denaturing PAGE. The yield of this reaction is ~30-40%, which is clearly sufficient to identify deoxyribozymes, as selections employing this capture method successfully identified the aforementioned pSer lyase deoxyribozymes. With this framework in hand, we first validated the successful capture of a Dhb-containing standard, AAA(Dhb)AA.

**Figure 2.** Schematic of the splinted capture reaction for selections. The DNA splint is complimentary to the 3′-end of the DNA anchor and the 5′-thiol capture oligonucleotide, thus placing the nucleophilic thiol near the dehydrated residue. Addition of the 5′-thiol oligonucleotide to Dha or Dhb allows an upward PAGE shift exclusive to the active sequences.
Figure 3. The initial selection scheme for identification of general pSer/pThr lyase deoxyribozymes. Selections MA1, LR1, and LS1 seek to identify pSer lyase deoxyribozymes, while MB1, LT1, LV1, LW1, LX1, and LY1 seek to identify pThr lyase deoxyribozymes. Selection MA1 is a reselection, where the sequence is biased towards the DhaDz1 deoxyribozyme (9JL113). While MB1 similarly biases the pool towards DhaDz1, it acts on a pThr-containing substrate, thus trying to evolve DhaDz1 to have pThr lyase activity. LS1 and LX1 incorporate the alternating substrate strategy, where the DNA-anchored peptide ligated to the pool alternates every round between K-, F-, and E-rich peptides. All selections were in 40 mM Mg$^{2+}$, 20 mM Mn$^{2+}$, and 1 mM Zn$^{2+}$ at pH 7.5.
Inspired by the selections which identified DhaDz1 and DhaDz2, the new selections use the short 5′-thiol capture oligonucleotide during even-numbered rounds of the selection process, and the long 5′-thiol capture oligonucleotide during odd-numbered rounds. This is important, as DNA sequences may fold in such a manner such that they migrate along with the captured products. By migration of the captured product within the gel, these undesired sequences will be enriched in the population if they too only alternate their migration. During the selections which identified DhaDz1 and DhaDz2, Dr. Chandrasekar observed that use of the short 5′-thiol oligonucleotide resulted in a higher capture yield than when using the long 5′-thiol oligonucleotide, which was attributed to inefficient coupling of the thiol phosphoramidite and poor resolution during 12% PAGE. To remedy this, we instead prepared the long 5′-thiol oligonucleotide by splint ligation of the short 5′-thiol oligonucleotide (purified by 20% PAGE) with an unmodified oligonucleotide sequence (Figure 2). Capture reactions employing this new 5′-thiol oligonucleotide have a comparable yield to those employing the short 5′-thiol oligonucleotide (~30-40%).

Assaying the capture of substrates with varying side-chains. To assess the viability of the newly synthesized Phe, Glu, and Lys-rich peptides as substrates for selection, the capture yields of the control Dhb-containing peptides were tested. While both the Dhb-containing F-rich and K-rich peptide exhibited viable capture yield >30%, the E-rich peptide was poorly captured (≤ 10%), insufficient for use in selection. This poor yield may be explained by charge repulsion between the negatively charged thiol nucleophile and the negatively charged glutamic acid (Glu) residues flanking the electrophilic site, thereby preventing Michael-type addition.

To test whether the pSer-and pThr-containing A-, F-, K-, and E-rich peptides could be captured, the DNA-anchored phosphorylated substrates were incubated under capture conditions.
Unfortunately, K-rich peptides with either pSer or pThr demonstrated >15% capture yield (figure 4), potentially due to the S_N1 displacement of pSer/pThr with the thiol oligonucleotide. The high yield of this undesired reaction means that more than 15% of the total population of sequences will be enriched independently of their ability to catalyze the desired reaction. Given this significant “headwind” for selection using K-rich peptides and the low capture yield for the standard E-rich substrate, the selections were modified to avoid including the K-rich or E-rich peptides.

Figure 4. Summary of PAGE-shift capture assays performed on substrates synthesized for the MA1-LY1 selections. Each sample was incubated in selection conditions and the resulting product was incubated in 100 mM Na_2CO_3, pH 9.2, at 37 °C for 12 h. (A) Capture of the AAAS^9AA, AAAT^9AA, and the AAA(Dhb)AA peptides, each of which was covalently attached to a HEG-rA oligonucleotide and ligated to the corresponding DNA pool (Table 1). No background activity was observed for the alanine-rich phosphopeptide substrates. (B) Capture of the AKKS^9KKA, AKKT^9KKA, and the AKK(Dhb)KKA peptides, each of which was covalently attached to a HEG-rA oligonucleotide and ligated to the corresponding DNA pool (Table 1). Significant background activity >15% was observed for both lysine-rich phosphopeptide substrates. (C) Capture of the AFFS^9FFA, AFFT^9FFA, and the AFF(Dhb)FFA peptides, each of which was covalently attached to a HEG-rA oligonucleotide and ligated to the corresponding DNA pool (Table 1). A modest background reaction was observed for the pSer-containing phosphopeptide, but no background was observed for the pThr-containing substrate. Not shown are GIT^9PVAF, GI(Dhb)PVAF, AEES^9EEA, AEET^9EEA, and the AEE(Dhb)EEA substrates, each of which had negligible yields.
Capture progression of MA1 and MB1. Concurrently with assaying the capture yields of the selection substrates, two selection experiments were performed. The first selection, MA1, was a reselection of the DhaDz1 enzyme; every base in the randomized pool where every nucleotide has a 75% chance to be identical to the nucleotide at that position in DhaDz1 (Table 1). While MA1 kept the substrate used in the selection of the parent sequence (AAAS\textsuperscript{P}AA), MB1 changed the substrate to AAAT\textsuperscript{P}AA and sought to evolve DhaDz1 such that it could act as a pThr lyase. At round 8, activity was observed in the MA1 selection, and the active sequences were cloned. Unfortunately, the resulting sequences were identical to the parent sequence. This outcome is likely due to contamination of the pool by sequences of the parental enzyme, as DhaDz1 has identical binding arms to the pool used in selection. No activity was observed by round 9 for the MB1 selection (Figure 5).

![Figure 5](image-url) Progression of the MA1 and MB1 in vitro selection experiments. MA1 was cloned at round 8 but only yielded deoxyribozymes identical to DhaDz1 (9JL113), or with single point-mutants. This is likely due to contamination of the selection pool by DhaDz1, which has the appropriate PCR primer binding sites to enable amplification, as well as the pool binding arms require for ligation.
Finalized selection design. The finalized selection scheme (Figure 6) aims to identify sequence-general pSer and pThr lyases while recognizing the practical limitations of what substrates can be used in selection. The selection experiments encompass 7 randomized pools, which have a random region length of 40 nucleotides (N40), apart from the 30-nucleotide random region length (N30) pool used in the PA1 selection. Because there are 4^{30} possible N_{30} sequences as opposed to the 4^{40} possible N_{40} sequences, N_{30} selections explore a higher proportion of the possible sequences.

The QM1 selection is a reselection which has Pool 3 binding arms, as opposed to the Pool 4 binding arms used in the MA1 selection, thereby eliminating the risk for contamination. Because the selection mechanics remain almost identical to that used in a successful selection, the QM1 serves as the functional equivalent to a positive control. This is amplified due to the predisposition of the randomized region towards the parent sequence. The PB1 selection seeks to identify pThr lyases using lanthanides as cofactors, inspired by previous identification of lanthanide-dependent deoxyribozymes.\textsuperscript{42} The remaining selections seek to identify pSer/pThr lyases using one of the viable peptide substrates alone, or by the alternating method, where F-rich substrates are used on odd selection rounds, and A-rich substrates on even numbered rounds.
Figure 6. Finalized selection schemes for general pSer/pThr deoxyribozymes, incorporating the practical restrictions assayed above. Selections QM1, LR1 and LS1 seek to identify pSer lyase deoxyribozymes, while PA1, PB1, LV1, LW1, and LX1 seek to identify pThr lyase deoxyribozymes. QM1 is a reselection of DhaDz1. Selections LS1 and LX1 employ the alternating strategy to achieve generality, where the substrate ligated to the DNA pool alternates between an F-rich (odd rounds) and A-rich (even rounds) peptide. All selections were performed in 40 mM Mg\(^{2+}\), 20 mM Mn\(^{2+}\), 1 mM Zn\(^{2+}\) at pH 7.5, except for PA1, which is instead performed in 1 mM Zn\(^{2+}\) and 3 μM of each Ce\(^{3+}\), Eu\(^{3+}\), and Yb\(^{3+}\) at pH 7.5. Additionally, the loop in PA1 is sequestered by a homologous oligonucleotide for rounds 1, 2, 4, 6, 8, etc.
Selection progression. The selections described above were performed through round 10 (figure 7). For all selections, no activity was observed above the level of the background reaction. Because the reselection of DhaDz1 failed, this put into question the proper execution of the selection experiments, whether by manual error or some other means. To put aside any uncertainty about the purity of the selection substrates and the capture oligonucleotides, the selection substrates and the two 5′-thiol capture oligonucleotides were resynthesized, and tested in a manner identical to that described above. The selections have been reinitiated starting from Round 1 and are currently in progress.

![Figure 7](image_url)

**Figure 7.** Selection progression for in vitro selection experiments QM1-LX1. (A) Capture yield of the N₃₀ selection, PA1, compared with the N₃₀ capture standard. No activity was observed through round 10. (B) Capture yield of the N₄₀ selections QM1, PB1, LR1, LS1, LV1, LW1 and LX1, compared with the N₄₀ capture standard. No activity was observed through round 10.
Applications in Phosphoproteomics

Sequence-general pSer and pThr deoxyribozymes have the potential to act as tools for phosphoproteomics enrichment. While DhaDz1 and DhaDz2 can tolerate some substrate diversity, many possible peptides cannot be modified and thus not enriched in a phosphoproteomics study. A single deoxyribozyme that is completely sequence-general, or a mixture of deoxyribozymes that can collectively act on any substrate, would be able to enrich phosphate-containing peptides in an indiscriminate manner, unlike current enrichment technology such as IMAC (immobilized metal affinity chromatography), which preferentially enriches acidic peptides.43

Two proteins were chosen as targets as a proof-of-concept study of the ability of pSer lyase deoxyribozymes to act as phosphoproteomics tools. The two proteins, chicken albumin and β-casein, explore a different amount of possible phosphorylation states: chicken albumin has two pSer residues (mostly phosphorylated; two equivalents of phosphates relative to the protein), while β-casein can have up to five phosphorylated sites but exhibits substantial variability in phosphorylation state (0-5 phosphate equivalents). Additionally, chicken albumin contains a glycosylation site, which thus provides an avenue to demonstrate the chemoselectivity of potential enzymes. Dr. Chandrasekar tested DhaDz1 with both the full-lengthed and fragmented substrates and observed negligible yields. Future pSer and pThr lyase deoxyribozymes can be tested on tryptic digests of these proteins, and the amount of phosphates released can be obtained through use of the malachite green phosphate assay.34
### CONCLUSIONS

The work described in this thesis seeks to expand the catalytic ability of DNA. Identification of the DhaDz1 and DhaDz2 deoxyribozymes demonstrated the fundamental ability of DNA to catalyze the conversion of pSer to Dha, and thereby suggested experimental routes to elucidate further chemical functional abilities of DNA.

DhaDz1 has been shown to act on some, but not all, of the tested pSer-containing substrates. The inability of this DNA enzyme to perform catalysis on pSer residue embedded within any peptidic context greatly limits the scope of its applicability. Additionally, no deoxyribozyme can currently act as a pThr lyase, catalyzing the conversion of pThr to Dhb. To this.
end, a set of selections which seeks to identify general pThr and pSer lyases through a substrate-alternation strategy is underway. To develop a viable selection strategy, the capture yield of potential selection substrates was extensively assayed. These experiments showed that substrates containing negatively charged glutamic acid residues were unable to be captured. Additionally, they revealed that substrates containing positively charged lysine residues lead to significant non-enzymatically catalyzed phosphate elimination. These selection experiments were performed through ten rounds of the selection process with no observable catalytic activity. However, since the reselection experiment failed, the lack of observed activity is likely a consequence of manual error at some point of the selection cycle or some other means, rather than the selection design itself.

Any identified pThr lyase deoxyribozymes will broaden the catalytic scope of DNA. Additionally, the practical applications of new enzymes identified through these selections can be evaluated. One such application is the use of these enzymes as potential phosphoproteomics tools, which can be assayed through methodology described here. These experiments are already designed, but there is a wealth of other applications that could be explored. pSer and pThr lyase deoxyribozymes have extraordinary potential, including usage in preparation of natural products, creation of site-specific dehydrated residues for diversification via Michael-type additions, as well as a phosphoproteomics enrichment tool.
REFERENCES


