HIGH-THROUGHPUT CATALYST DISCOVERY:

DEVELOPMENT OF A SELECTION PLATFORM FOR SCREENING DNA-ENCODED SMALL-MOLECULE CATALYST LIBRARIES IN ORGANIC SOLVENTS

by

KATHERINE DELANEY HOOK

(Under the Direction of Ryan Hili)

ABSTRACT

An amphiphilic DNA encoding platform was developed that allows the high-throughput evaluation of a large small-molecule library for a wide assortment of catalytic reactions in organic solvents. The cornerstone of this platform is a replicable, amphiphilic DNA that encodes a chemical library. Each member of the library is uniquely identifiable via a covalent attachment of encoded DNA (genotype) and small-molecule catalyst (phenotype). The amphiphilic nature of the DNA-encoded small-molecule library stems from the development of a PEG polymer that enables solubility in both aqueous and organic solvents. In vitro selection enables catalytically active members surviving the selection pressure – intermolecular bond formation of the DNA code to an affinity tag – to be separated from inactive members, PCR amplified and identified by high-throughput DNA sequencing. We explored the solubility, stability, and thermodynamics of PEGylated ssDNA in organic solvents, investigated the compatibility of PEGylated DNA as an encoding element during catalytic reactions in organic solvents, and developed methods to generate and screen a PEGylated DNA-encoded peptide library for catalytic activity in organic solvents.

INDEX WORDS: High-Throughput Catalyst Discovery, DNA-Encoded Library, PEGylated DNA, Amphiphilic DNA

# HIGH-THROUGHPUT CATALYST DISCOVERY: DEVELOPMENT OF A SELECTION PLATFORM FOR SCREENING DNA-ENCODED SMALL-MOLECULE CATALYST LIBRARIES IN ORGANIC SOLVENTS

by

KATHERINE DELANEY HOOK

BS, University of Tennessee at Chattanooga, 2012

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

DOCTOR OF PHILOSPHY

ATHENS, GEORGIA

2018

# HIGH-THROUGHPUT CATALYST DISCOVERY:

# DEVELOPMENT OF A SELECTION PLATFORM FOR SCREENING DNA-ENCODED SMALL-MOLECULE CATALYST LIBRARIES IN ORGANIC SOLVENTS

by

KATHERINE DELANEY HOOK

| | |
|---|---|
| Major Professor: | Ryan Hili |
| Committee: | Vladimir Popik |
| | Jeffrey Urbauer |

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

DEDICATION

For my Family: Liza, Patrick, Fran, Carey, Kee, Jim, Sally, Chris, Roxie and Olive.

ACKNOWLEDGEMENTS

To my Mom, I thank you for your unconditional encouragement, guidance and love throughout my academic and non-academic pursuits in life. You are a constant source of inspiration which has led me to become the woman I am today. To Carey and Kee, I appreciate your guidance, and for being positive role models for me to aspire.

To Patrick, I will never be able to put into words how appreciative I am of all your support throughout this process. You have been there for me since the start of this journey and it would have not been the same if you were not a part of it.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ..... v

LIST OF TABLES ..... ix

LIST OF FIGURES ..... x

LIST OF SCHEMES ..... xiii

CHAPTER

1 INTRODUCTION ..... 1

1.1 Summary ..... 1

1.2 Organocatalysis ..... 3

1.3 Combinatorial Synthesis ..... 4

1.4 DNA-Encoded Libraries ..... 5

1.5 DNA in Organic Solvents ..... 7

1.6 DNA Synthesis ..... 8

1.7 PCR ..... 10

1.8 REFERENCES ..... 13

2 SYNTHESIS, SOLUBILITY, AND THERMODYNAMICS OF PEGYLATED DNA IN ORGANIC SOLVENTS ..... 18

2.1 INTRODUCTION ..... 18

2.2 RESULTS AND DISCUSSION ..... 20

PEGylated DNA Synthesis ..... 20

Solubility of PEGylated DNA .....24
PEGylated DNA Architecture.....29
PCR Amplification of PEGylated DNA .....36
2.3 EXPERIMENTAL METHODS.....38
General Methods.....38
Synthesis of PEGylated Oligonucleotides .....38
PEG-DNA Solubility .....40
Thermal Melting of DNA .....41
2.4 REFERENCES .....43
3 PEGYLATED DNA AS AN ENCODING ELEMENT FOR CATALYST DISCOVERY.....49
3.1 INTRODUCTION .....49
3.2 RESULTS AND DISCUSSION.....55
3.2.1 Design and Synthesis of Amphiphilic DNA-Encoded Catalyst Molecular Architecture.....55
Aldol Reaction .....59
Friedel-Crafts Reaction.....83
3.3 EXPERIMENTAL METHODS.....89
Small Molecule Synthesis.....89
Post-Synthetic DNA Modification.....94
Catalysis Experiments.....95
3.4 REFERENCES .....100

4 MOCK SELECTION OF A PEGYLATED DNA ENCODED CATALYST IN ORGANIC SOLVENT ....113

4.1 INTRODUCTION ....113

4.2 RESULTS AND DISCUSSION ....114

Selection Procedure Overview....114

Selection Optimization with Biotinylated ‘Selection Control’ Libraries 120

Mock Selection of PEGylated DNA-Encoded Aldol Reaction Catalyst 129

High-Throughput Sequencing Analysis of Mock Aldol Selection....139

4.3 CONCLUSIONS....142

4.4 EXPERIMENTAL METHODS....143

4.4 REFERENCES ....151

APPENDICES

A S1. SUPPORTING INFORMATION....44

B S2: SUPPORTING INFORMATION....102

C S3: SUPPORTING INFORMATION....153

D ABBREVIATIONS ....159

LIST OF TABLES

Page

Table 2.1: Mass Spectrometry Results of Ta1 and PEGylated Ta1 ..... 23

Table 2.2: Dielectric Constant and UV Cutoff of Organic Solvents ..... 25

Table 2.3: qPCR Concentration Data of 40,000 MW PEG-DNA ..... 28

Table 2.4: Thermal Melting Temperature (Tm) Values (°C) ..... 32

Table 2.5: Oligonucleotide Sequences ..... 33

Table 3.1: Mass Spectrometry Results of Modified Oligonucleotides ..... 66

Table 3.2: Solvent Screen for DNA-Encoded Catalyst Activity ..... 79

Table 4.1: DNA Sequences ..... 116

Table 4.2: Selection Method Experiment Results ..... 125

LIST OF FIGURES

Page

Figure 1.1: General strategy for the high-throughput screening of catalysts using amphiphilic DNA-encoded libraries ..... 2

Figure 1.2: Common phosphoramidites used in DNA synthesis ..... 9

Figure 1.3: Automated solid-phase DNA synthesis products before purification ..... 9

Figure 2.1: Reverse-phase HPLC chromatogram of 40,000 MW PEGylated DNA ..... 22

Figure 2.2: PAGE characterization of PEG-DNA ..... 22

Figure 2.3: Absorption spectra of PEG-modified DNAs in organic solvents ..... 27

Figure 2.4: qPCR solubility of 40,000 MW PEGylated 48-nt DNA in various solvents ..... 28

Figure 2.5: PEGylated DNAs ..... 31

Figure 2.6: PEGylated dsDNA thermal melting curve ..... 31

Figure 2.7: PEGylated ssDNA with primer thermal melting curve ..... 32

Figure 2.8: Thermal melting of PEGylated dsDNA in 1,4-dioxane ..... 35

Figure 2.9: Thermal melting of PEGylated dsDNA in 1,2-dichloroethane ..... 36

Figure 2.10: PCR efficiency of PEGylated DNA ..... 37

Figure 3.1: Schematic for validation of PEGylated ssDNA as an encoding element ..... 50

Figure 3.2: Reaction types of catalytic reactions ..... 51

Figure 3.3: The copper(II)-catalyzed Friedel-Crafts alkylation reaction ..... 54

Figure 3.4: Molecular architecture of the DNA-encoded catalyst ..... 56

Figure 3.5: Solid phase DNA synthesis phosphoramidites ..... 58

Figure 3.6: Aldol Reaction scheme of the amphiphilic DNA-encoded catalyst with aldol substrate .....59
Figure 3.7: Optimization of streptavidin-Based EMSA with biotinylated 40,000 MW PEG-DNA positive control .....67
Figure 3.8: Aldol reaction with catalyst in solution – EMSA.....70
Figure 3.9: Catalyst in solution reaction solvent optimization – EMSA .....71
Figure 3.10: Reactant, catalyst, and reaction product structures. .....74
Figure 3.11: Streptavidin-mediated EMSA comparison of aldol .....74
Figure 3.12: Aldol reaction order of addition experiment .....76
Figure 3.13: Biotinylated aldehyde concentration optimization.....77
Figure 3.14: Aldol reaction time course experiment. .....78
Figure 3.15: EMSA experiments demonstrating selective catalysis.....81
Figure 3.16: Positive result verification of FC reaction with the catalyst in solution .....86
Figure 3.17: Molecular architectures of the Friedel-Crafts Cu(II) catalyst/ligand encoded PEG-DNA .....87
Figure 3.18: PEGylated Friedel-Crafts reaction and controls.....88
Figure 4.1: General scheme for in vitro mock selection of a DNA-encoded aldol catalyst .....114
Figure 4.2: Molecular architecture of the non-competitive N12 library.....117
Figure 4.3: General scheme of positive selection control.....121
Figure 4.4: Affinity pulldown selection method experiment.....124
Figure 4.5: Wash efficiency of streptavidin magnetic beads.....126
Figure 4.6: Positive selection control with increased washes.....127

Figure 4.7: Optimization of streptavidin affinity pull-down incubation conditions.....128

Figure 4.8: Restriction enzyme digestion of oligonucleotide sequence Ta_EcoRV .....130

Figure 4.9: PCR fidelity of synthetic oligonucleotides used in selection experiments. ..131

Figure 4.10: Aldol reaction catalyst selection with controls.....132

Figure 4.11: 1,000:1 Aldol reaction catalyst selection .....135

Figure 4.12: Time course experiment of aldol reaction catalysis selection.....136

Figure 4.13: 1,000:1 Aldol reaction catalyst selection .....138

Figure 4.14: Gel analysis of mock aldol selection resulting from a 500-fold dilution of DNA-encoded aldol catalyst.....139

Figure 4.15: Non-denaturing PAGE extraction of sequencing PCR samples .....140

Figure 4.16: High-throughput sequencing analysis of mock aldol selection resulting from a 2000-fold dilution of DNA-encoded aldol catalyst.....141

LIST OF SCHEMES

Page

Scheme 1.1: Purification of synthetic oligonucleotides.....10

Scheme 2.1: PEGylation of 5'-amino modified 48 nt ssDNA.....21

Scheme 3.1: DNA-templated cross-aldol reaction catalyzed by diprolinamide.....52

Scheme 3.2: Azido-PEG-ketone synthesis.....60

Scheme 3.3: Azido-PEG-aldehyde synthesis.....60

Scheme 3.4: Biotinylated aldehyde synthesis.....61

Scheme 3.5: Biotinylated ketone synthesis.....61

Scheme 3.6: Diproline coupling to amino-modified ssDNA.....63

Scheme 3.7: Copper-catalyzed azide-alkyne cycloaddition (CuAAC) of 3'-alkyne modified ssDNA and azido-PEG-benzaldehyde.....64

Scheme 3.8: Thiol-Michael addition of 40,000 MW PEG-maleimide to 5'-thiol modified ssDNA.....65

Scheme 3.9: Aldol reaction of reactant-tethered PEG-DNA.....69

Scheme 3.10: Azidoamide acyl-imidazole synthesis.....84

Scheme 3.11: Biotinylated indole synthesis.....84

CHAPTER 1

# INTRODUCTION

## 1.1 Summary

An amphiphilic DNA encoding selection methodology was developed[1] to enable the high-throughput evaluation of a sizable small-molecule library for catalytic reactions in organic solvents. The cornerstone of this platform is a replicable, amphiphilic DNA that encodes a chemical library. Each member of the library is uniquely identifiable via a covalent attachment of encoded DNA (genotype) and small-molecule catalyst (phenotype). The amphiphilic nature of the DNA-encoded small-molecule library stems from the development of a PEG polymer that enables solubility in both aqueous and organic solvents. *In vitro* selection enables catalytically active members surviving the selection pressure – intermolecular bond formation of the DNA code to an affinity tag – to be separated from inactive members, PCR amplified and identified by high-throughput DNA sequencing (Figure 1.1). The work discussed in this document aims to meet the following objectives: 1) Explore the solubility, stability, and thermodynamics of PEGylated ssDNA in organic solvents; 2) Investigate the compatibility of PEGylated DNA as an encoding element during catalytic reactions in organic solvents; 3) Develop methods to generate and screen a PEGylated DNA-encoded peptide library for catalytic activity in organic solvents.

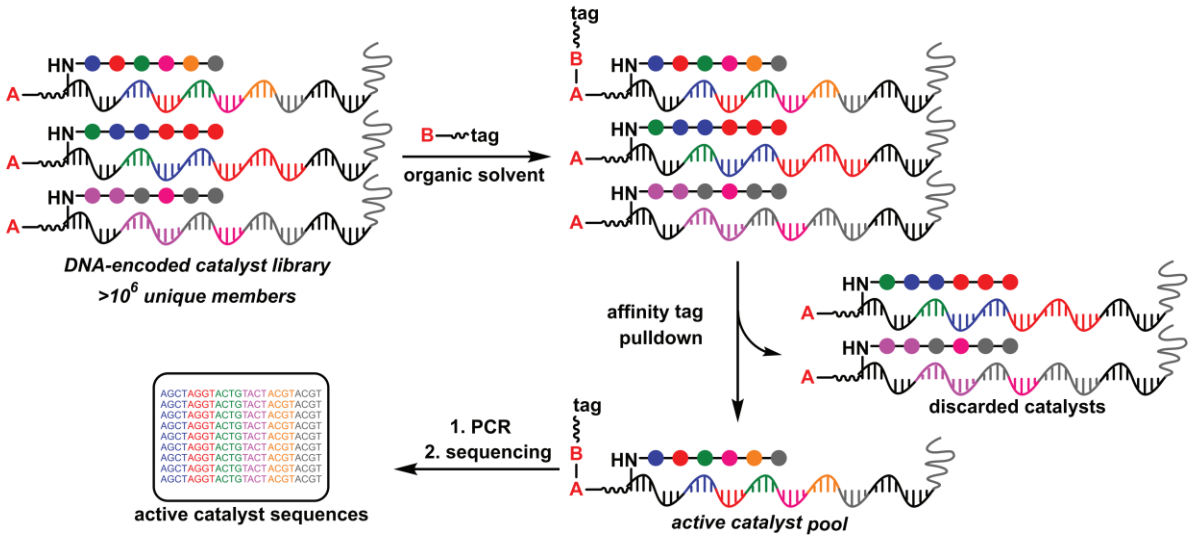



**Figure 1.1:** General strategy for the high-throughput screening of catalysts using amphiphilic DNA-encoded libraries.

The key innovation discussed in the following pages is a technology that brings the throughput of DNA-encoding into organic solvents, enabling future screening of large combinatorial libraries for small-molecule catalysts capable of bond forming reactions. This research is the first to apply a DNA encoding element that is soluble in organic solvents to the discovery of small-molecule catalysts. These studies advance fundamental understanding of the behavior of DNA in organic solvents, as well as provide a novel platform for high-throughput screening of libraries of small-molecule catalysts. We anticipate future use of this platform will lead to valuable insight into catalyst structure and mechanism, and make a significant impact on the field of organic chemistry.

## 1.2 Organocatalysis

Modern organocatalysis has rapidly evolved into an essential component of contemporary organic synthesis over the past decade. High-throughput synthesis and screening have dramatically changed the methods with which catalysis research is now conducted, demanding consideration of attractive features such as increased rates of innovation, cost effectiveness, reduced time to market, and an increased probability of success.[1] Small molecule organic catalysts, a broad class of molecules, boast a wide variety of chemical transformations and practical advantages over metal-based and macromolecular catalysts, including air and water stability, low cost, availability from renewable resources, and relative nontoxicity.[2] Synthetically useful transformations include generic modes of catalyst activation and induction through noncovalent interactions, such as hydrogen bond donation,[3] as well as covalent bond formation through iminium ion[4] and enamine[5] catalysis, among others. However, the development of more efficient and selective catalysts remains a significant challenge.

Organic small-molecule catalysts exist in many structural forms, ranging from a single amino acid, such as proline, to larger oligomeric frameworks, such as short peptide sequences. Within this broad family of catalysts, oligopeptides provide a particularly attractive scaffold to develop novel organic catalysts, as their rich functional diversity, dense hydrogen bond donor/acceptor network, rigid backbone, and relative ease of synthesis lend immense promise to their catalytic potential and practical utility. Indeed, with the development of synthetic combinatorial methods, the synthesis of libraries of oligomers comprising natural and unnatural peptides has become synthetically tractable, leading to the discovery of various peptide-like catalysts for a broad scope of transformations.[6]

Concurrently, the average molecular throughput of catalyst screening methods is insufficient to allow the adequate library diversity needed to produce multiple families or clusters of related catalysts to form a basis for new structure-activity relationships, a desirable goal for determining classes of catalytically viable molecules for more efficient discovery pathways.

### 1.3 Combinatorial Synthesis

Combinatorial synthesis, a technology for creating molecules *en masse* and testing them rapidly for desirable properties,[7] has increased sampling of the vast chemical space that organic small-molecules encompass, resulting in efforts to analyze the results from catalytic screens through the development of high-throughput screening methods.[8-11] These methods have led to the identification of novel synthetic pathways and greater insight and control of structure-activity relationships and chemical transformations, exemplifying some of the benefits offered by high-throughput catalyst screening.

While combinatorial synthesis can rapidly generate extensive chemical libraries ($>10^6$ unique members), the technology to screen such libraries is a major bottleneck of discovery efforts, with screening throughput typically limited to fewer than $10^4$ catalysts. There are currently five popular methods of high-throughput combinatorial library synthesis and screening, with the most notable being the one-bead-one-compound (OBOC)[8] strategy. This method involves chemical synthesis on the solid phase, rapidly generating a vast library using on-bead split-and-pool synthesis, and then reacting the library with the desired substrate. This *en masse* screening approach is typically followed by screening for immobilized, catalytically active molecules via illumination through fluorescence[12] or chemosensors.[13] A drawback to this method

is that, once screened, the beads are required to be selected manually or possibly by automated methods, such as flow cytometry.[14] This presents significant shortcomings such as complicated and time intensive protocols or expensive instrumentation. Additionally, on-bead screening incurs a significant depletion of the library as the compounds are consumed, requiring excessive amounts of starting materials which lead to increased cost of production, or limits this method to small quantities of library materials and lowered applicability for the range of use.

An alternative approach to screening consists of applying a 'selection pressure' to the library and removing the unreactive portion of the library so that only catalytically active members survive. Known as *in vitro* selection, noteworthy success has been shown when using this approach,[15-18] which offers significant advantages over the traditional screening approaches. The current research addresses the bottleneck in screening technologies for catalyst discovery with the development of a novel high-throughput selection-based screening platform for catalyst activity based on DNA encoding of chemical libraries in organic solvents.

### 1.4 DNA-Encoded Libraries

Inspired by the native encoding system of the cell, in which DNA (genotype) is transcribed into mRNA and further translated into peptides and proteins (phenotype), researchers began to develop encoded combinatorial display technologies, by biological and biochemical means, to identify functional members out of extensive collections of molecules. Various display technologies including phage display,[19] mRNA display,[20] ribosomal display,[21] and DNA display,[22] implement an *en masse* screening approach known as *in vitro* selection. This highly successful and influential approach was used for the discovery of catalytic biopolymers from libraries containing greater than $10^{14}$ members.[16] *In vitro* selection implements a 'selection

pressure' to remove catalytically inactive molecules from the library, thus leaving only the catalytically active species to identify. Selections offer significant advantages over traditional screening approaches including (*i*) a selection evaluates all molecules of a library simultaneously, regardless of library size[23-26] and (*ii*) selections are typically easier to execute as they do not require the spatial separation of library members nor sophisticated equipment. The overarching theme amongst these selection methods is that the biopolymer catalyst (phenotype) is spatially associated with its genetic code (genotype); thus, PCR amplification and DNA sequencing can reveal the identity of those biocatalysts that have performed the desired catalytic function to survive the selection pressure.

DNA-encoded libraries (DELs)[27] are composed of small organic molecules individually coupled to oligonucleotides serving as amplifiable identification barcodes. The DNA tagging procedure facilitates the construction and screening of compound collections of unprecedented size. The *in vitro* selection of small molecules from DNA-encoded libraries achieved significant momentum, particularly in the discovery of small molecule drugs,[17-18, 28] with library diversity rapidly increasing with the continual advancement of DNA sequencing methods. DNA-encoded libraries are typically screened by affinity-capture selections.[17, 29-32] Following washing cycles and PCR amplification of the DNA tags of the binding compounds, the identification of the enriched molecules is typically achieved by DNA sequencing of the amplicon using next-generation high-throughput sequencing technologies[33-36] and comparing the relative frequency of each DNA tag in the library before and after applying selection pressure.[32, 37-38]

The identification of bio-catalytically active small molecules from DNA-encoded chemical libraries requires the use of affinity-based selection strategies and suitable decoding techniques of the DNA sequences ('barcodes'). Generally, selections are performed by the

capture of affinity tagged compounds on a target protein immobilized on a solid support. The stringency of both capture and washing steps is critical for the outcome of affinity selections[39-42] Also, choosing the appropriate decoding strategy contributes considerably to the successful use of DNA-encoded chemical libraries.

### 1.5 DNA in Organic Solvents

Current literature on solvation of DNA into organic solvents is sparse.[43-46] Organic solutions are poor solvents for DNA, primarily attributed to a highly charged anionic phosphate backbone and hydrogen bonding framework of nitrogenous bases. Addition of organic solvent to solutions of oligonucleotides typically results in the formation of compact conformations and precipitation, depending on the concentration and length (base pairs) of the DNA. Solvents such as ethanol and isopropanol are often widely employed for purification of DNA by exploiting this behavior.[47] Because organic solvents have distinct properties of hydrogen bonding, polarity and hydrophobicity, the solvation of nucleic acids depends on the solvent used. The dielectric constant of these solutions is suggested to be the critical factor in determining the conformational behavior of oligonucleotides.[44]

Nucleic acids have been made more compatible with organic solutions by association or conjugation with hydrophobic groups such as quaternary ammonium salts,[48-52] alkylaryl phosphonates,[53] and long amphiphilic polymers.[54-57] Recently, Hermann[51-52] reported an organic-phase coupling strategy employing a DNA-surfactant conjugate. This technique is notable of possible use to the goals of our proposed research. The use of PEGylated DNA has been found able to act as a DNAzyme to catalyze the peroxidase ability of G-quadruplexes in organic

solvents.[21] The researchers found PEGylated DNA to be soluble in various organic solvents with a ssDNA of 21 nucleotides (nt) in length conjugated to a 20,000-molecular weight PEG moiety.

To date, *in vitro* selection processes must be performed in aqueous solutions to take advantage of DNA amplification and sequencing. We have found that attachment of an amphiphilic PEG polymer expands the solubility profile of single-stranded (ss)-DNA to include various anhydrous organic solvents (e.g., acetonitrile and 1,2-dichloroethane), a more favorable chemical environment for organocatalytic reactions.

## 1.6 DNA Synthesis

Standard oligonucleotide synthesis is performed on an automated DNA synthesizer. A DNA synthesizer allows for the rapid synthesis of oligonucleotides in a high yield of up to one hundred bases in length with >95 % coupling efficiency in less than one day. The oligonucleotides are formed by undergoing a series of reactions beginning on a chemical handle attached to a controlled pore glass (CPG) or polystyrene (PS) solid support.[58] Automated synthesis proceeds in a 3' to 5' direction with each new phosphoramidite reacting on the 5'-phosphate of the oligonucleotide. Nucleobase protecting groups eliminate possible side reactions on the amidite other than the phosphorylation sites (Figure 1.2).

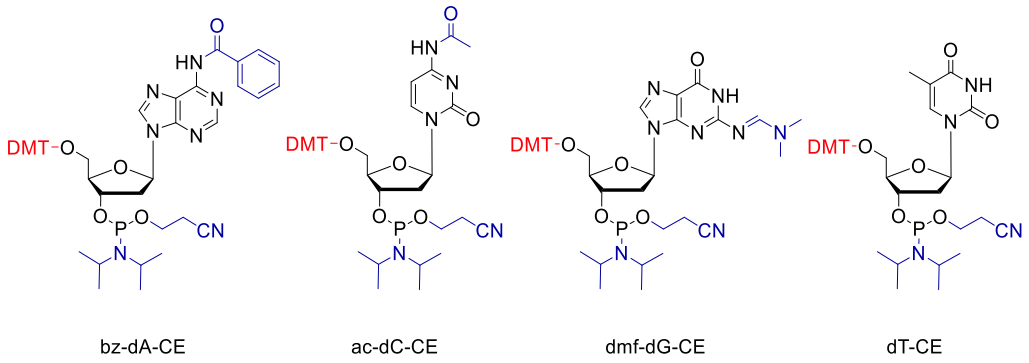



**Figure 1.2:** Common phosphoramidites used in DNA synthesis. Base-labile protecting groups (blue) and acid-labile protecting groups (red) are indicated.

Cleavage of the DNA from the solid support and deprotection of the nucleobase and phosphate are simultaneously performed in basic aqueous solution, typically 1:1 ammonium hydroxide: methylamine (AMA), to produce free DNA in solution containing the 5'-OH dimethoxy trityl (DMT) moiety left intact. The acid-labile DMT group allows full-length oligonucleotide synthesis product to be selectively isolated from incomplete, or truncated, DNA synthesis products (Figure 1.3). By taking advantage of the shift in polarity, the bulky aromatic group on the full-length product can be separated and purified by reverse-phase (RP)-HPLC.

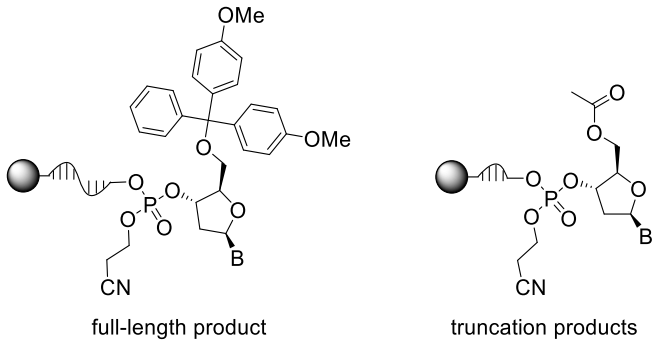
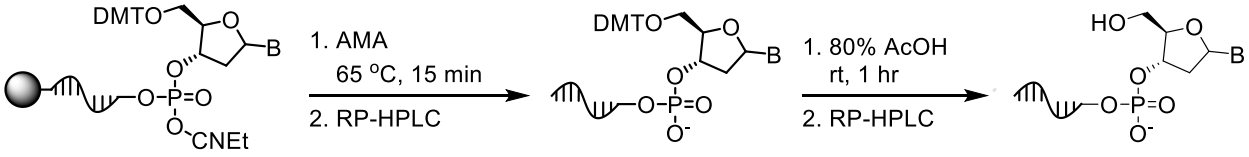



**Figure 1.3:** Automated solid-phase DNA synthesis products before purification.

Furthermore, DMT-deprotection in acidic solution (e.g., 80 % acetic acid in water) and subsequent purification by HPLC renders the oligonucleotide of interest desalted and ready for use or further post-synthetic modification. The full purification scheme is depicted in Scheme 1.1. Modification of the 3'- and 5'- end, as well as internal sequence modifiers, are used for post-synthetic transformations at specific sites on the oligonucleotide. These reagents are typically commercially available.




**Scheme 1.1.** Purification of synthetic oligonucleotides

## 1.7 PCR

Polymerase chain reaction (PCR) is a technique for creating exponential copies of DNA. PCR is a sensitive assay because it can generate enough copies to be analyzed using conventional laboratory methods using only trace amounts of DNA. Each PCR assay requires the presence of template DNA, primers, nucleotides, and DNA polymerase. The DNA polymerase is the critical enzyme that links individual nucleotides together, forming the PCR product. Nucleotides include adenine, thymine, cytosine, and guanine (A, T, C, G), comprising the four bases found in DNA. These act as the building blocks used by the DNA polymerase to create the resultant PCR product. The primers present in the reaction specify the exact DNA product to be amplified. The primers are short DNA fragments with a defined sequence complementary to the

target DNA that is to be detected and amplified, acting as an extension point for the DNA polymerase to build upon. The components are mixed in a test tube or well plate and then placed in a thermal cycler that allows repeated cycles of DNA amplification to occur in three necessary steps. The reaction solution is first heated to a high temperature to activate the Hot-Start Taq DNA polymerase, which also allows any hybridized strands to separate, a process termed denaturation. The temperature is subsequently lowered to allow the specific primers to bind to the target DNA segments, a process termed hybridization or annealing. Annealing between primers and the target DNA only occurs if they are complementary in sequence (e.g., A binding to T). The temperature is next raised again, at which time the DNA polymerase extends the primers by adding nucleotides to the developing DNA strand. With each cycle of these three steps, the number of copied DNA molecules increases exponentially.

The PCR product is analyzed using polyacrylamide gel electrophoresis, which separates the DNA product based on the size and charge. The primary method used for visualization of DNA is staining the product using ethidium bromide dye, a small molecule which intercalates between the two strands of the DNA. However, due to the carcinogenic and serious health effects of ethidium bromide, its use has been reduced and replaced by SYBR$^{TM}$, a safe DNA gel stain.

The next generation advancement of the PCR technology was the development of RT-qPCR real-time qPCR technology allows quantification of PCR products in “real time” during each PCR cycle, yielding a quantitative measurement of PCR products accumulated during the course of the reaction. Real-time reactions are carried out in a thermocycler that permits measurement of a fluorescent detector molecule, which decreases post-processing steps and minimizes experimental error. This technique is most commonly achieved using fluorescence-based technologies.

In brief, like PCR, qPCR consists of a succession of amplification cycles in which the template nucleic acid is denatured, annealed with specific oligonucleotide primers, and extended to generate a complementary strand using a thermostable DNA polymerase. The exponential increase of amplicons (amplification products), in contrast with end-point PCR, can be monitored at every cycle (in real time) using a fluorescent reporter. The increase in fluorescence is plotted against the cycle number to generate the amplification curve, from which a quantification cycle Cq (often described as Ct for cycle threshold) value is determined. The Cq is the number of cycles for which the amount of fluorescence (hence, of the template) reaches a significant level that is higher than the background fluorescence. Therefore, the Cq value can be used to determine the initial concentration of target nucleic acid, serving as a basis for absolute or relative template quantification. All steps of (RT-)qPCR may introduce experimental errors. qPCR is a robust technique, but due to its high sensitivity, minimal variations can induce non-negligible differences in the results. RT-qPCR can be performed in triplicate (experimental replicates) to measure intra-assay variability, which follows a statistical distribution. Estimation of inter-assay variability is performed using a "reference" sample included in each experiment.

## 1.8 REFERENCES


1. Murphy, V.; Volpe Jr, A. F.; Weinberg, W. H., High-throughput approaches to catalyst discovery. *Current Opinion in Chemical Biology* **2003,** *7* (3), 427-433.

2. Jacobsen, E. N.; MacMillan, D. W. C., Organocatalysis. *Proceedings of the National Academy of Sciences* **2010,** *107* (48), 20618-20619.

3. Doyle, A. G.; Jacobsen, E. N., Small-Molecule H-Bond Donors in Asymmetric Catalysis. *Chemical Reviews* **2007,** *107* (12), 5713-5743.

4. Erkkilä, A.; Majander, I.; Pihko, P. M., Iminium Catalysis. *Chemical Reviews* **2007,** *107* (12), 5416-5470.

5. Mukherjee, S.; Yang, J. W.; Hoffmann, S.; List, B., Asymmetric Enamine Catalysis. *Chemical Reviews* **2007,** *107* (12), 5471-5569.

6. Davie, E. A. C.; Mennen, S. M.; Xu, Y.; Miller, S. J., Asymmetric Catalysis Mediated by Synthetic Peptides. *Chemical Reviews* **2007,** *107* (12), 5759-5812.

7. Borman, S., Industry is embracing the technology 'totally,' as researchers continue to advance the art of rapid synthesis and screening. *Chemical & Engineering News* April 6, 1996.

8. Lam, K. S.; Lebl, M.; Krchňák, V., The "One-Bead-One-Compound" Combinatorial Library Method. *Chemical Reviews* **1997,** *97* (2), 411-448.

9. Revell, J. D.; Wennemers, H., Identification of catalysts in combinatorial libraries. In *Creative Chemical Sensor Systems*, Springer: 2007; pp 251-266.

10. Revell, J. D.; Wennemers, H., Peptidic catalysts developed by combinatorial screening methods. *Current opinion in chemical biology* **2007,** *11* (3), 269-278.

11. Robbins, D. W.; Hartwig, J. F., A simple, multidimensional approach to high-throughput discovery of catalytic reactions. *Science* **2011,** *333* (6048), 1423-1427.

12. Krattiger, P.; McCarthy, C.; Pfaltz, A.; Wennemers, H., Catalyst–Substrate Coimmobilization: A Strategy for Catalysts Discovery in Split-and-Mix Libraries. *Angewandte Chemie International Edition* **2003,** *42* (15), 1722-1724.

13. Copeland, G. T.; Miller, S. J., A chemosensor-based approach to catalyst discovery in solution and on solid support. *Journal of the American Chemical Society* **1999,** *121* (17), 4306-4307.

14. Sklar, L. A.; Carter, M. B.; Edwards, B. S., Flow cytometry for drug discovery, receptor pharmacology and high-throughput screening. *Current opinion in pharmacology* **2007,** *7* (5), 527-534.

15. Kong, D.; Yeung, W.; Hili, R., In Vitro Selection of Diversely Functionalized Aptamers. *Journal of the American Chemical Society* **2017,** *139* (40), 13977-13980.

16. Wilson, D. S.; Szostak, J. W., In vitro selection of functional nucleic acids. *Annu Rev Biochem* **1999,** *68*, 611-47.

17. Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A., Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat Chem Biol* **2009,** *5* (9), 647-54.

18. Kleiner, R. E.; Dumelin, C. E.; Liu, D. R., Small-molecule discovery from DNA-encoded chemical libraries. *Chemical Society Reviews* **2011,** *40* (12), 5707-5717.

19. Smith, G. P.; Petrenko, V. A., Phage Display. *Chem Rev* **1997,** *97* (2), 391-410.

20. Takahashi, T. T.; Austin, R. J.; Roberts, R. W., mRNA display: ligand discovery, interaction analysis and beyond. *Trends Biochem Sci* **2003,** *28* (3), 159-65.

21. Schaffitzel, C.; Hanes, J.; Jermutus, L.; Pluckthun, A., Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. *J Immunol Methods* **1999,** *231* (1-2), 119-35.

22. Li, X.; Liu, D. R., DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew Chem Int Ed Engl* **2004,** *43* (37), 4848-70.

23. Hilvert, D., Genetic Selection as a Tool in Mechanistic Enzymology and Protein Design. In *Ernst Schering Research Foundation Workshop.*, Mulzer, J.; Bohlmann, R., Eds. Springer Berlin Heidelberg: 2000; pp 253-268.

24. Taylor, S. V.; Kast, P.; Hilvert, D., Investigating and Engineering Enzymes by Genetic Selection. *Angewandte Chemie International Edition* **2001,** *40* (18), 3310-3335.

25. Taylor, S. V.; Walter, K. U.; Kast, P.; Hilvert, D., Searching sequence space for protein catalysts. *Proc. Natl. Acad. Sci.* **2001,** *98* (19), 10596-601.

26. Dower, W. J.; Mattheakis, L. C., In vitro selection as a powerful tool for the applied evolution of proteins and peptides. *Current Opinion in Chemical Biology* **2002,** *6* (3), 390-398.

27. Kanan, M. W.; Rozenman, M. M.; Sakurai, K.; Snyder, T. M.; Liu, D. R., Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* **2004,** *431* (7008), 545.

28. Wrenn, S. J.; Weisinger, R. M.; Halpin, D. R.; Harbury, P. B., Synthetic ligands discovered by in vitro selection. *Journal of the American Chemical Society* **2007,** *129* (43), 13137-13143.

29. Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D., Encoded self-assembling chemical libraries. *Nature Biotechnology* **2004,** *22*, 568.

30. Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D., High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proceedings of the National Academy of Sciences* **2008,** *105* (46), 17670-17675.

31. Wichert, M.; Krall, N.; Decurtins, W.; Franzini, R. M.; Pretto, F.; Schneider, P.; Neri, D.; Scheuermann, J., Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation. *Nature Chemistry* **2015,** *7*, 241.

32. Franzini, R. M.; Neri, D.; Scheuermann, J., DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries. *Accounts of Chemical Research* **2014,** *47* (4), 1247-1255.

33. Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y.-J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L. I.; Jarvie, T. P.; Jirage, K. B.; Kim, J.-B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M., Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005,** *437*, 376.

34. Otto, T. D., Real-time sequencing. *Nature Reviews Microbiology* **2011,** *9*, 633.

35. Schuster, S. C., Next-generation sequencing transforms today's biology. *Nature Methods* **2007,** *5*, 16.

36. Martin, K.; Janet, K., High-throughput DNA sequencing – concepts and limitations. *BioEssays* **2010,** *32* (6), 524-536.

37. Buller, F.; Steiner, M.; Scheuermann, J.; Mannocci, L.; Nissen, I.; Kohler, M.; Beisel, C.; Neri, D., High-throughput sequencing for the identification of binding molecules from DNA-encoded chemical libraries. *Bioorganic & Medicinal Chemistry Letters* **2010,** *20* (14), 4188-4192.

38. Franzini, R. M.; Ekblad, T.; Zhong, N.; Wichert, M.; Decurtins, W.; Nauer, A.; Zimmermann, M.; Samain, F.; Scheuermann, J.; Brown, P. J.; Hall, J.; Gräslund, S.; Schüler, H.; Neri, D., Identification of Structure–Activity Relationships from Screening a Structurally Compact DNA-Encoded Chemical Library. *Angewandte Chemie* **2015,** *127* (13), 3999-4003.

39. Nielsen, J. B., S.; Janda, K. D. , Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993,** *115*, 9812-9813.

40. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A., Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc Natl Acad Sci U S A* **1993,** *90* (22), 10700-4.

41. Halpin, D. R.; Harbury, P. B., DNA Display II. Genetic Manipulation of Combinatorial Chemistry Libraries for Small-Molecule Evolution. *PLOS Biology* **2004,** *2* (7), e174.

42. Leimbacher, M.; Zhang, Y.; Mannocci, L.; Stravs, M.; Geppert, T.; Scheuermann, J.; Schneider, G.; Neri, D., Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2012,** *18* (25), 7729-37.

43. Ts'o, P. O.; Helmkamp, G. K.; Sander, C., Secondary structures of nucleic acids in organic solvents II. Optical properties of nucleotides and nucleic acids. *Biochimica et biophysica acta* **1962,** *55* (5), 584-600.

44. Mel'nikov, S. M.; Khan, M. O.; Lindman, B.; Jönsson, B., Phase behavior of single DNA in mixed solvents. *Journal of the American Chemical Society* **1999,** *121* (6), 1130-1136.

45. Rozenman, M. M.; Liu, D. R., DNA-Templated Synthesis in Organic Solvents. *ChemBioChem* **2006,** *7* (2), 253-256.

46. Rozenman, M. M.; Kanan, M. W.; Liu, D. R., Development and initial application of a hybridization-independent, DNA-encoded reaction discovery system compatible with organic solvents. *Journal of the American Chemical Society* **2007,** *129* (48), 14933-14938.

47. Piškur, J.; Rupprecht, A., Aggregated DNA in ethanol solution. *FEBS Letters* **1995,** *375* (3), 174-178.

48. Ijiro, K.; Okahata, Y., A DNA-lipid complex soluble in organic solvents. *Journal of the Chemical Society, Chemical Communications* **1992,** (18), 1339-1341.

49. Tanaka, K.; Okahata, Y., A DNA− Lipid Complex in Organic Media and Formation of an Aligned Cast Film1. *Journal of the American Chemical Society* **1996,** *118* (44), 10679-10683.
50. Sergeyev, V. G.; Pyshkina, O. A.; Lezov, A. V.; Mel'nikov, A. B.; Ryumtsev, E. I.; Zezin, A. B.; Kabanov, V. A., DNA Complexed with Oppositely Charged Amphiphile in Low-Polar Organic Solvents. *Langmuir* **1999,** *15* (13), 4434-4440.

51. Liu, K.; Zheng, L.; Liu, Q.; de Vries, J. W.; Gerasimov, J. Y.; Herrmann, A., Nucleic Acid Chemistry in the Organic Phase: From Functionalized Oligonucleotides to DNA Side Chain Polymers. *Journal of the American Chemical Society* **2014,** *136* (40), 14255-14262.

52. Chen, W.; Gerasimov, J. Y.; Zhao, P.; Liu, K.; Herrmann, A., High-Density Noncovalent Functionalization of DNA by Electrostatic Interactions. *Journal of the American Chemical Society* **2015,** *137* (40), 12884-12889.

53. Amberg, S.; Engels, J. W., Synthesis and Properties of Nonpolar DNA (Arylalkyl)phosphonates. *Helvetica Chimica Acta* **2002,** *85* (8), 2503-2517.

54. Kumar, M. R.; Bakowsky, U.; Lehr, C., Preparation and characterization of cationic PLGA nanospheres as DNA carriers. *Biomaterials* **2004,** *25* (10), 1771-1777.

55. Mok, H.; Park, T. G., PEG-assisted DNA solubilization in organic solvents for preparing cytosol specifically degradable PEG/DNA nanogels. *Bioconjugate chemistry* **2006,** *17* (6), 1369-1372.

56. Mok, H.; Kim, H. J.; Park, T. G., Dissolution of biomacromolecules in organic solvents by nano-complexing with poly(ethylene glycol). *Int J Pharm* **2008,** *356* (1-2), 306-13.

57. Abe, H.; Abe, N.; Shibata, A.; Ito, K.; Tanaka, Y.; Ito, M.; Saneyoshi, H.; Shuto, S.; Ito, Y., Structure Formation and Catalytic Activity of DNA Dissolved in Organic Solvents. *Angewandte Chemie International Edition* **2012,** *51* (26), 6475-6479.

58. Caruthers, M. H., Chemical synthesis of DNA and DNA analogs. *Accounts of Chemical Research* **1991,** *24* (9), 278-284.

CHAPTER 2

# SYNTHESIS, SOLUBILITY, AND THERMODYNAMICS OF PEGYLATED DNA IN ORGANIC SOLVENTS

## 2.1 INTRODUCTION

Previous studies by Ito[1] suggest the use of polyethylene glycol (PEG) conjugated to the end of a DNA strand enables solvation in organic solvents such as methanol, acetonitrile, 1,4-dioxane, and 1,2-dichloroethane (DCE). Interestingly, the study showed G-quadruplex structure and catalytic ability of the DNA was maintained when dissolved in methanol.[1] However, these studies show that base-paired structures are only maintained in various organic solutions when there are metal ions carried over from water solution followed by lyophilization and redissolvation in organic solvent. For the purposes of our study, secondary structure formation could be a potential limitation of the encoding DNA strand, as it only serves as an information carrier and is not intended to interact with the small-molecule catalysis to occur on the 3'-end of the encoding oligonucleotide. This is further explored in Chapter 3. Thus far, PEGylated DNA solubility in various polar, nonpolar, protic and aprotic organic solvents has been validated to a maximum length of 21 nucleotides (nt) when conjugated to 20,000 molecular weight (MW) PEG.[1]

Requirements of the proposed project within this document necessitate oligonucleotides longer than 21-nt; therefore, we hypothesized that the scope of PEG-DNA solubility could be expanded to meet the requirements of our encoding DNA with a 48-nt DNA conjugated with

longer, increased molecular weight PEG. The solubility of 48-nt ssDNA conjugated to different PEG polymers with 10,000 MW to 40,000 MW is examined to determine if PEG-DNA can be utilized in the study of catalytic ability of DNA-encoded peptide libraries in organic solvents.

In addition to solvation of PEGylated single-stranded DNA in organic solvents, the solubility and stability of double-stranded PEGylated DNA was studied in order to determine the optimal screening platform for catalysis. DNA duplex formation between two complementary ssDNA strands is thermodynamically favored in various aqueous buffers due mainly to the hydrophobic effect, which drives assembly by nucleobase stacking; cation-mediated shielding of the electrostatic repulsion between the two phosphate backbones; and interstrand hydrogen-bonding, once the duplex is assembled[2]. In organic solvents, the hydrophobic effect is no longer present, and the electrostatic repulsion is greatly enhanced;[3] however, hydrogen bonding is generally more favored as the polarity of the solvent decreases.[4] Thus, in the case of double-stranded PEGylated DNA libraries, duplex stability will need to be determined in organic solvents.

The double-stranded portions of the PEGylated DNA library may come in the form of a short primer-length duplex formed by annealing reactants or as longer duplexes formed following a primer extension by polymerase. Stable secondary structure formation within ssDNA has been directly observed in organic solvents,[1] and indirect evidence to support the stability of duplex DNA in anhydrous organic solvents has been reported.[5] However, contradicting reports have also been made, including greatly reduced thermal melting ($T_m$) temperatures for DNA duplexes in glycerol, with no duplex formation being observed in DMSO or methanol.[6] Furthermore, other investigators have extrapolated the $T_m$ temperature for a 22-mer dsDNA to be -6.2 °C in 100% DMF.[7] These inconclusive reports suggest that there remains

considerable work required to better understand the thermodynamics of DNA in organic solvents.

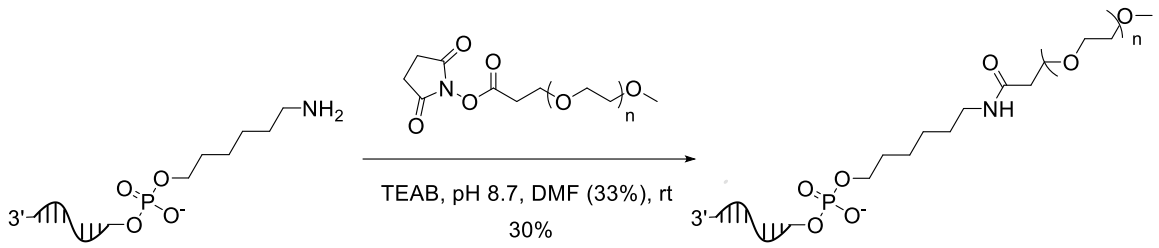
## 2.2 RESULTS AND DISCUSSION

### 2.2.1 PEGylated DNA Synthesis

Synthesis of oligonucleotides conjugated to various molecular weight polyethylene glycol (PEG) polymers was performed to characterize their solubility and stability in various organic solvents. Since the size of polyethylene glycol (PEG) necessary to enable the solubilization of a 48-nucleotide ssDNA in various organic solvents was unknown, PEG-NHS esters of 10,000, 20,000, 30,000 and 40,000 molecular weight were conjugated to the 5'-end of the oligonucleotide.

Before post-synthetic modification with PEG, oligonucleotide synthesis was performed on an automated DNA synthesizer. The ssDNA encoding element for the selection platform requires both an encoding region that specifies the catalyst, and two primer regions for amplification. As a model ssDNA length, we chose 48 nucleotides (nt), which accommodates two 18-nt primer sites and a 12-nt catalyst-encoding region; a 12-nt region can encode greater than 16 million unique molecules by established split-and-pool tandem DNA/small-molecule synthesis methods.[8] A 48-nucleotide (nt) oligomer sequence was designed to contain minimal secondary structure and to minimize the possible catalytic activity of the DNA strand itself.[9] Modified phosphoramidites for functionalization of the 3'- and 5'- end, as well as internal sequence modifiers, are used for post-synthetic transformations at specific sites on the oligonucleotide. We employed a 5'-amino modifier phosphoramidite to enable post-synthetic

modification of the oligonucleotide with PEG. The 5'-amine modified oligonucleotide was acylated with a PEG-*N*-hydroxysuccinimide (NHS) ester in aqueous solution (Scheme 2.1).




**Scheme 2.1:** PEGylation of 5'-amino modified 48 nt ssDNA

Separation and purification of the PEGylated DNA were performed by HPLC. Separation from the starting material is possible due to the lipophilic nature PEG moiety, which increases the retention time of the product (Figure 2.1). As seen from the HPLC chromatogram, the non-polar PEG increases the retention time of the oligonucleotide considerably. The 5'-amino modified 48mer has a retention time of 7.2 minutes, and 5'-40,000 PEG 48mer has a retention time of 21.5 minutes.

In addition to HPLC shift of retention time, characterization of PEG-DNA final products is performed by gel-shift assay using polyacrylamide gel electrophoresis (PAGE). PEG-DNA products were characterized by both denaturing PAGE and non-denaturing or native PAGE. Urea-PAGE or denaturing urea polyacrylamide gel electrophoresis employs 6-8 M urea, which denatures secondary DNA structures and is used for their separation in a polyacrylamide gel matrix based on the molecular weight. Urea disrupts secondary structure in DNA and is typically used for single-stranded DNA analysis. However, urea also disrupts the structure of enzymes or

proteins, which are commonly employed in DNA binding assays. Therefore, both denaturing and native PAGE was used to characterize the PEGylated DNA (Figure 2.2).

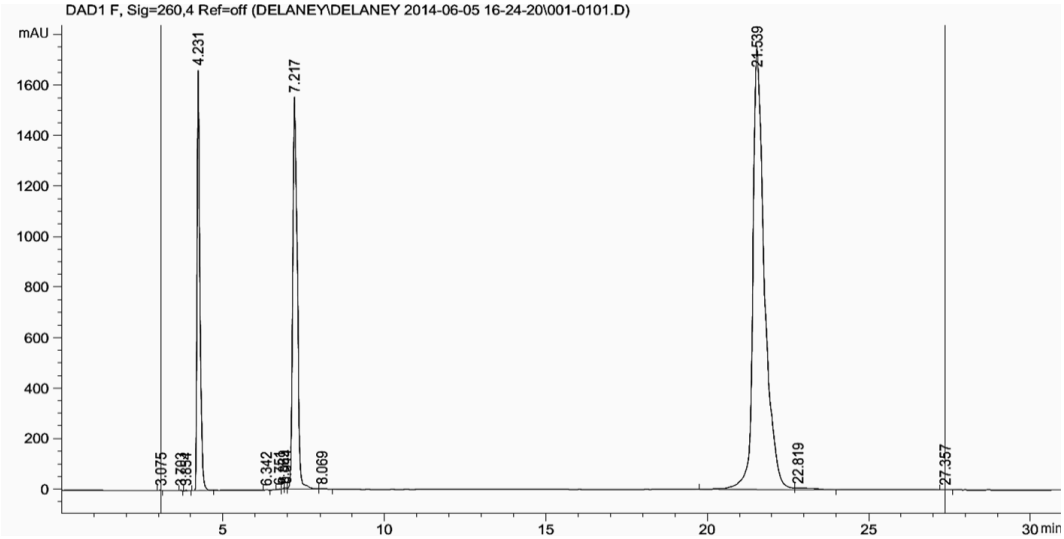



**Figure 2.1:** Reverse-phase HPLC chromatogram of 40,000 MW PEGylated DNA

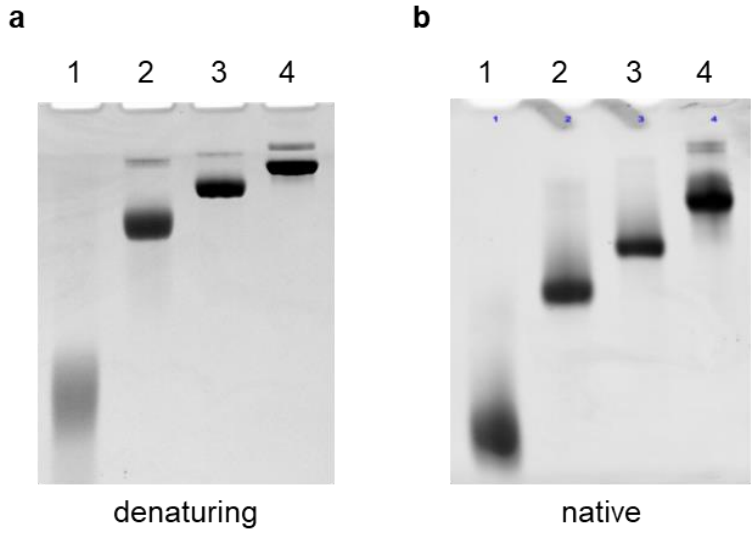



**Figure 2.2:** PAGE characterization of PEG-DNA. a) 5% denaturing PAGE (150 V, 2 hrs), EtBr stain. b) 5% non-denaturing PAGE (150 V, 80 min), EtBr stain. Lane 1: PEG 10,000-Ta1. Lane 2: PEG 20,000-Ta1. Lane 3: PEG 30,000-Ta1. Lane 4: PEG 40,000-Ta1.

The PEGylated products cause a very noticeable decrease in migration on a gel relative to non-PEGylated DNA. Consequently, the minimum required runtime of electrophoresis was extended to two hours on 5% gels for the different molecular weight PEG-DNAs. However, the difference in molecular weight is observed by the difference in migration rates of the samples.

Finally, matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was used to confirm the mass accuracy of the purified product (Table 2.1). Full spectra can be found in the Supplementary Figures at the end of this chapter. It is important to note the PEG moieties are of average molecular weight and the range in final product masses is characterized as a broad mass signal. It is also important to note that large PEG modification on an oligonucleotide should be the last post-synthetic modification performed when creating multi-functionalized oligonucleotides. Mass analysis of any post-synthetic modification after PEGylation will be challenging to measure accurately.

**Table 2.1:** Mass Spectrometry Results of Ta1 and PEGylated Ta1

| Entry | Sample | Calculated Mass (Da) $[M-H]^-$ | Experimental (Da)[a] $[M-H]^-$ |
|---|---|---|---|
| 1 | Ta1-$NH_2$ | 14963.7 | 14,963.5[b] |
| 2 | Ta1-10,000 PEG | 24,963 | 25,378 |
| 3 | Ta1-20,000 PEG | 34,963 | 36,237 |
| 4 | Ta1-30,000 PEG | 44,963 | 43,884 |
| 5 | Ta1-40,000 PEG | 54,963 | 54,753.3 |

[a] Experimental mass determined by MALDI (Negative-Mode) unless otherwise indicated.
[b] Experimental mass determined by ESI-MS (Negative-Mode).

### 2.2.2 Solubility of PEGylated DNA

We prepared a model 5'-amino modified 48-nt ssDNA sequence, which we conjugated to PEG-*N*-hydroxysuccinimide (PEG-NHS) esters ranging in average mass from PEG 10,000 to PEG 40,000. To determine if attaching a PEG polymer to ssDNA influenced its solubility in organic solvents, a 5.0 μM solution of the PEGylated DNA was prepared in various solvents and analyzed by UV-Vis spectrophotometry. Polyethylene glycol is a synthetic polymer that is soluble in water and most organic solvents, depending on size. When coupled to a hydrophilic ssDNA, the molecule becomes amphiphilic. The solubility of the oligomer was found to correlate with the molecular weight of the attached polyethylene glycol moiety. Thus, the PEG-DNA becomes more soluble as the molecular weight of the conjugated PEG increases.

The solubility of PEG-DNA was studied in various aqueous and organic solvents including methanol (MeOH), 1,2-dichloroethane (DCE), 1,4-dioxane, acetonitrile (MeCN), heptane and 1-octanol. Dissolution of the ssDNA into organic solvents that do not absorb light at 260 nanometers (nm), the wavelength maximum of DNA, were characterized by UV-Vis spectrophotometry; whereas solvents that have an overlapping absorbance in this wavelength region (e.g., DMF and toluene) were quantified by real-time or quantitative (q)-PCR. The polarity of solvents explored ranged from polar to nonpolar (Table 2.2). Since PEG is not soluble in aliphatic hydrocarbons, heptane was used as a negative control.

**Table 2.2:** Dielectric Constant and UV Cutoff of Organic Solvents [10-11]

| **Solvent** | **ε** | **$\lambda_c$ (nm)** |
|---|---|---|
| Water | 80.10 | 191 |
| Acetonitrile | 36.64 | 190 |
| Methanol | 33.0 | 210 |
| 1,2-Dichloroethane | 10.42 | 226 |
| 1,4-Dioxane | 2.22 | 215 |
| Heptane | 1.92 | 197 |
| 1-Octanol | 10.30 | 220 |
| Dimethyl sulfoxide | 47.24 | 265 |
| *N*,*N*-Dimethylformamide | 38.25 | 270 |
| Toluene | 2.38 | 286 |

ε: dielectric constant (relative permittivity); $\lambda_c$: cutoff wavelength, below which the solvent absorption becomes excessive.

First, the concentration of PEG-DNA in deionized water was confirmed by measurement of absorbance at 260 nm ($A_{260}$). Equal amounts of each oligonucleotide were aliquoted and lyophilized. Lyophilization allows removal of water through the process of sublimation from a frozen solid sample. Resuspension in anhydrous organic solvent and incubation at room temperature overnight allow maximum dissolution when vortex or rotation perturb the sample. The PEG-DNA samples were centrifuged to ensure suspended particulates were compacted and a pure solution was obtained before dilution and absorbance measurement. Aliquots of the samples were diluted by volume to ensure accurate absorbance values were observed.

Absorbance measurements were performed using UV-Vis spectrophotometry for the various PEG-DNA molecules synthesized in organic solvents of interest (Figure 2.3). PEG 10,000 was unable to facilitate solubility of the 48 nt ssDNA into any solvents except water and methanol (Figure 2.3a). We began to observe partial solubility of PEGylated DNA in 1,2-dichloroethane (DCE) and acetonitrile (MeCN) at PEG weights of 20,000 Da (Figure 2.3b);

however, a general solubility profile was observed when using PEG 40,000 (Figure 2.3d). Importantly, PEG 40,000 enabled solubility in a range of organic solvents while maintaining excellent solubility in water. The only family of organic solvents tested that was not capable of dissolving the PEGylated DNA was the long aliphatic hydrocarbon solvents, including heptane and 1-octanol.

Since solvent effects can influence nucleobase absorbance,[5] we independently quantified solubility using quantitative PCR (qPCR). qPCR uses real-time fluorescence to measure the quantity of DNA present at each cycle during a PCR. A double-stranded DNA binding dye (e.g., SYBR® Green) generates the fluorescent signal. At a point where the qPCR fluorescence signal is detectable over the background fluorescence, a quantification cycle, or $C_q$ value, can be determined. $C_q$ values can be used to calculate absolute target quantities in reference to an appropriate standard curve, derived from a series of known DNA dilutions.

The solubility of 40,000 MW PEG-DNA was evaluated for solvents of interest that could not be quantified by UV-Vis due to overlapping absorption at 260 nm, the wavelength maximum of DNA (Figure 2.4 and Table 2.3). The solvents evaluated include DMSO, toluene, and DMF in addition to deionized water, 1,2-dichloroethane, and heptane. The latter three solvents were evaluated to compare the efficiency of the qPCR quantitation to the results acquired by UV-Vis spectrophotometry.

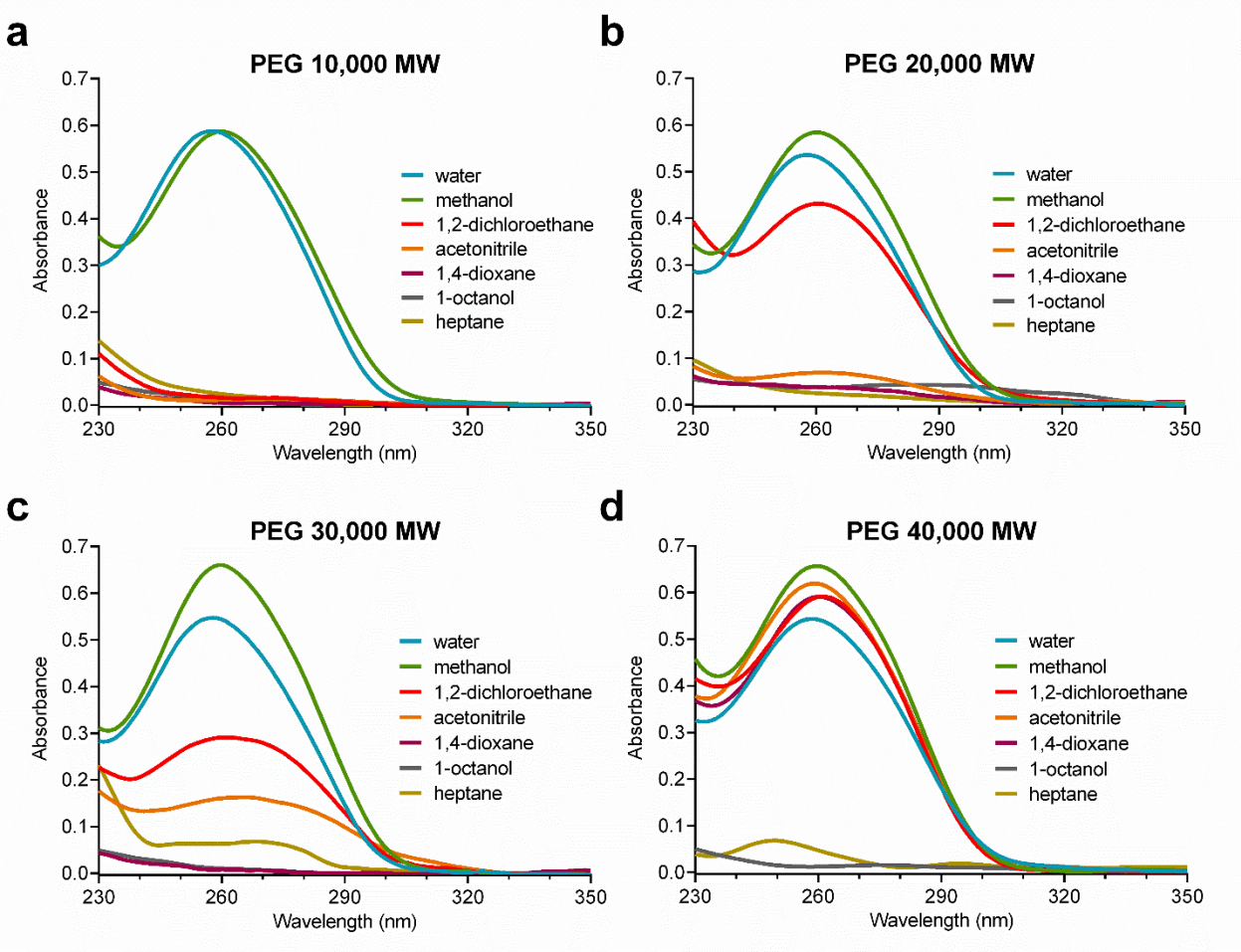



**Figure 2.3:** Absorption spectra of PEG-modified DNAs (PEG-DNAs) in organic solvents. Amino-modified 48-nt DNA was conjugated with PEG of different lengths. Molecular weight of PEG was 10,000 (a), 20,000 (b), 30,000 (c) and 40,000 (d). The sequence of DNA was Ta1 (see DNA sequence table). The PEG-modified DNA was dissolved at a concentration of 5 μM in deionized water (blue), methanol (green), 1,2-dichloroethane (red), acetonitrile (orange), 1,4-dioxane (purple), 1-octanol (grey) or heptane (yellow) after lyophilization. Following low-speed centrifugation, the solution was diluted to 1 μM with the appropriate solvent, and absorbance spectra were measured.

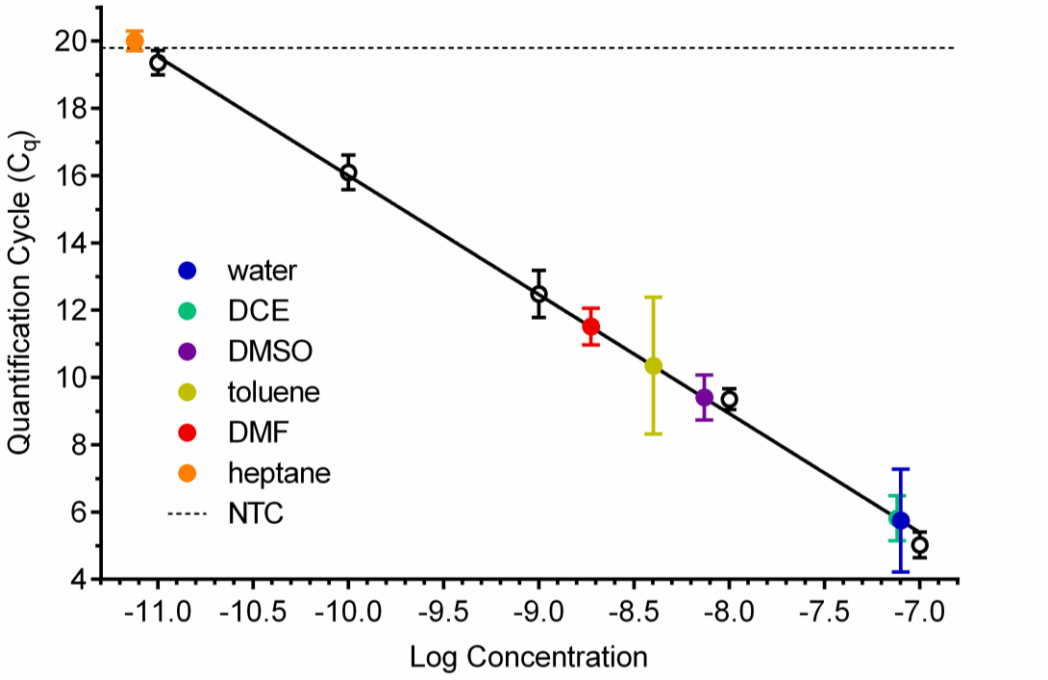



**Figure 2.4:** qPCR solubility of 40,000 MW PEGylated 48-nt DNA in various solvents (DMSO, DMF, toluene, 1,2-dichloroethane, water, and heptane). Samples dissolved at 1.0 μM and diluted 10-fold for amplification. No template control (NTC) $C_q$ shown as the dotted line at the top of the graph. $C_q$ values analyzed by instrument software. Log concentration determined by a linear fit to the standard dilution curve (black).

**Table 2.3.** qPCR Concentration Data of 40,000 MW PEG-DNA

| **Solvent** | **Concentration (μM)**[a] |
|---|---|
| Water | 0.8 ± 0.1 |
| DCE | 0.7 ± 0.1 |
| DMSO | 0.08 ± 0.02 |
| DMF | 0.02 ± 0.01 |
| Toluene | 0.04 ± 0.01 |
| Heptane | 0[b] |

[a] concentration determined by the fit to a standardized concentration amplification curve (green).
[b] no amplification seen.

When dissolved at a concentration of 1.0 μM, deionized water and 1,2-dichloroethane (DCE) were found to have a concentration of 0.8 ± 0.1 μM and 0.7 ± 0.1 μM, respectively. These solvents were expected to show the complete dissolution of PEG-DNA, as previously recorded when quantified by UV-Vis spectrophotometry at a concentration of 5.0 μM. As seen from the raw fluorescence data, the concentration of these samples is near the upper threshold of the amplification signals detected. It was theorized that the initial concentration was too high for the complete detection of initial exponential amplification resulting in an incomplete signal to be analyzed. Since the $C_q$ values are derived from the amplification curve, the resulting concentration data is potentially undercalculated for these samples. Therefore, the water and DCE concentrations detected may not correspond to the maximum dissolution of the samples.

Additionally, the experimental concentrations of 40,000 MW PEG-DNA in DMSO, DMF, and toluene were calculated to 0.08 ± 0.02 μM, 0.02 ± 0.01 μM and 0.04 ± 0.01 μM, respectively. The concentration of dissolved PEG-DNA in these solvents is significantly reduced, by an order of magnitude, in comparison to the samples dissolved in water or DCE. The concentration values measured in these solvents are more reliable as they are detected within the range standard dilution curves.

### 2.2.3 PEGylated DNA Architecture

In addition to solvation of PEGylated single-stranded DNA in organic solvents, the solubility of double-stranded, PEGylated DNA was studied to determine the optimal architecture for the DNA-encoded catalyst screening platform. The architectures of three main DNA hybridization structures of PEGylated DNA were examined: ssDNA, ssDNA with primer, and ds DNA (Figure 2.5). Nucleic acid hybridization is detected by altering the temperature of the

sample and monitoring the change in absorbance at 260 nm. The melting temperature ($T_m$) is known as the temperature at which half of the dsDNA has denatured and became ssDNA. $T_m$ depends on the length of the DNA, GC content, and buffer in which the DNA is measured. Absorption of dsDNA is lower than its dissociated ssDNA due to the hyperchromic effect. Base pair stacking and hydrogen bonding of duplex DNA limits the resonance of aromatic rings; therefore, when the duplex dissociates the absorbance intensity increases and can be detected.

The melting curves of reverse complement and primer annealed 48mer PEG-DNA were measured and analyzed to determine the temperature of duplex dissociation in aqueous buffer. Melting and annealing curves of 40,000 MW PEG-DNA hybridized to the 48nt reverse complement, and 18nt reverse primer, in aqueous buffer was recorded. Analysis of both curves was performed, noting some limitations explained below. Melting curve $T_m$ can be inaccurate because the initial hybridization of the sample is not measured. It is possible all the strands have not hybridized if not incubated at the right temperature or amount of time.

Furthermore, the initial state of the DNA at the beginning of annealing curve measurement is more likely to be dissociated entirely at the higher temperature required for the measurement. One limitation of annealing $T_m$ measurement is evaporation of the solvent and, therefore, decrease in concentration can occur. The $T_m$ can be extrapolated three different ways: horizontal intercept, sloping intercept, and inflection point. The shape of the melting curve determines which fitting method is most appropriate (Table 2.4).

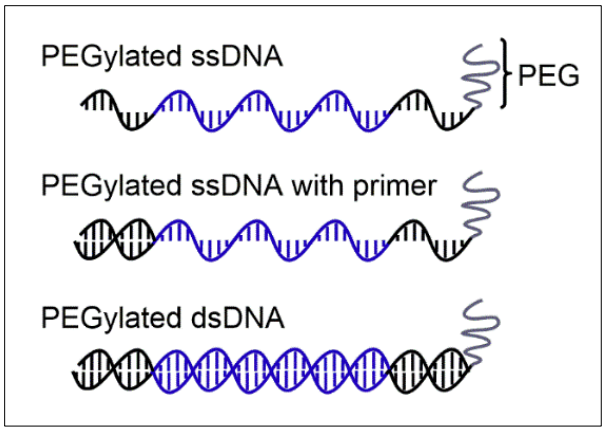



**Figure 2.5:** PEGylated DNAs

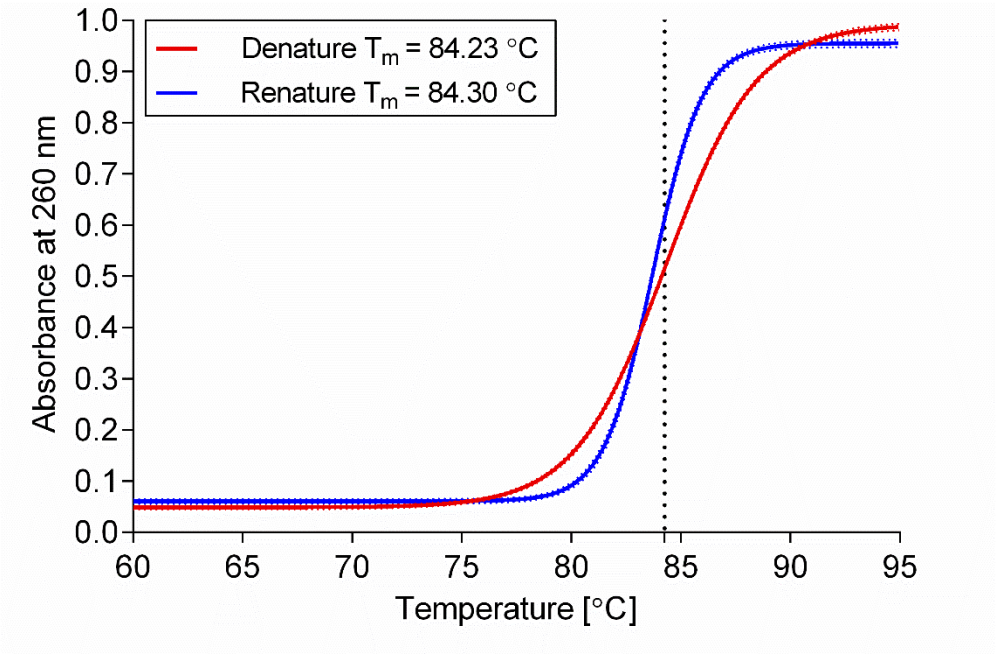



**Figure 2.6:** PEGylated dsDNA thermal melting curve

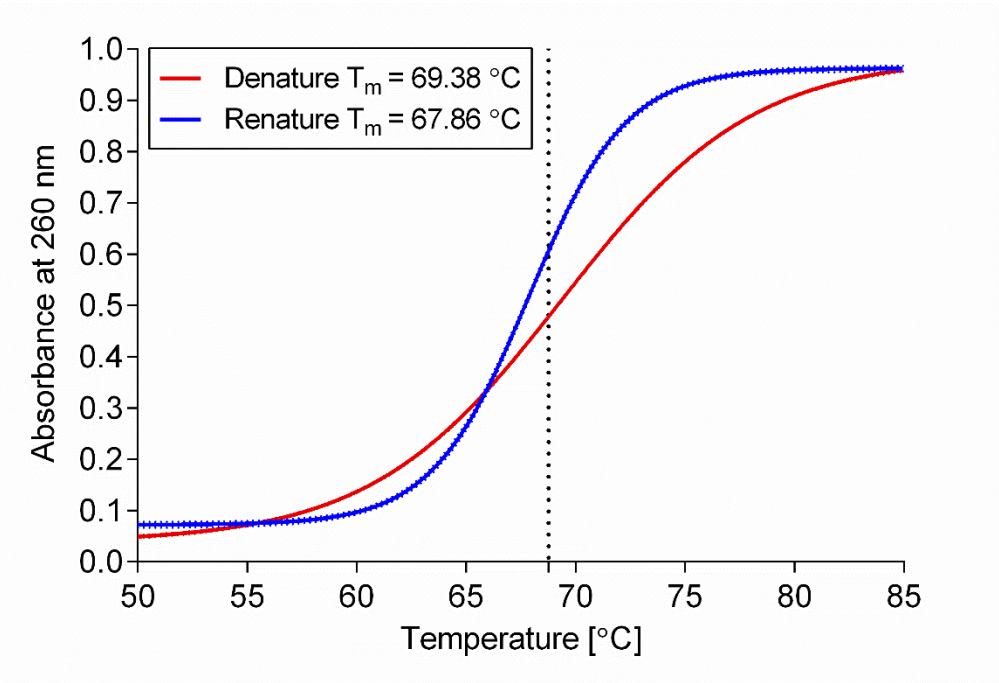



**Figure 2.7:** PEGylated ssDNA with primer thermal melting curve

**Table 2.4:** Thermal Melting Temperature ($T_m$) Values (°C) [a]

| **Ta1/Ta1RC** | **Fitting Method** | | |
|---|---|---|---|
| | Sloping Intercept | Horizontal Intercept | Inflection |
| Melt - $T_m$ Observed (°C) | **84.23** | 84.30 | 85.10 |
| Anneal - $T_m$ Observed (°C) | 91.06 | **84.30** | 84.40 |
| % GC | 60% | 60% | 60% |
| **Ta1/PRa** | | | |
| | Sloping Intercept | Horizontal Intercept | Inflection |
| Melt - $T_m$ Observed (°C) | **69.38** | 69.21 | 70.4 |
| Anneal - $T_m$ Observed (°C) | 69.13 | **67.85** | 68.85 |
| % GC | 66.7% | 66.7% | 66.7% |

[a] measurements recorded at 1.0 µM concentration in 165 mM Na buffer

**Table 2.5:** Oligonucleotide Sequences

| | Sequence (5' → 3') |
|---|---|
| Ta1 | CGTACGGTCGACGCTAGC ATGTCCAGTTAG CACGTGGAGCTCGGATCC |
| Ta1RC | GGATCCGAGCTCCACGTG CTAACTGGACAT GCTAGCGTCGACCGTACG |
| PRa | GGATCCGAGCTCCACGTG |

The melting and annealing curves of PEGylated dsDNA with reverse complement (Ta1/Ta1RC) were analyzed, and melting temperature ($T_m$) was predicted to be 84.23°C, and 84.30°C, respectively. The melting and annealing curves of PEGylated DNA with primer complement (Ta1/PRa) were analyzed, and melting temperature ($T_m$) was found to be 69.38°C, and 67.85°C, respectively. Both sets of curves were analyzed using a sloping intercept fit on the melting curve and horizontal intercept for the annealing curve. The lines of the melting curves have noticeable sloping before and after dissociation. Therefore, sloping intercept fitting model produces the most accurate $T_m$ value for both melting curves. The annealing curves have less sloping before and after dissociation; therefore, the horizontal intercept fitting method was used to determine melting temperature.

Following $T_m$ measurements of the PEGylated DNA in aqueous buffer, we first attempted to replicate the thermal melting experiments in 1,4-dioxane and 1,2-dichloroethane. The PEGylated dsDNA (Ta1/Ta1RC) was attempted first due to its higher melting temperature relative to the primer annealed architecture. This action was undertaken due to previous reports suggesting significantly reduced thermal melting ($T_m$) temperatures for DNA duplexes in glycerol, with no duplex formation observed in DMSO or methanol.[6] Additionally, researchers have extrapolated the $T_m$ temperature for a 22-mer dsDNA to be -6.2 °C in 100% DMF.[7]

Although stable secondary structure formation within ssDNA has been directly observed in organic solvents[1] and indirect evidence to support the stability of duplex DNA in anhydrous organic solvents has been reported,[5] we were unable to obtain reliable data in the solvents tested. First, the thermal melting of dsDNA dissolved in 1,4-dioxane was recorded (Figure 2.8). It is important to note that the melting point of 1,4-dioxane is 11.8 °C, which contributes to the inaccurate absorbance measurements while the solvent is in a solid state. Analyzing the curve from 13 – 35 ºC, the absorbance measurements frequently increase and decrease multiple times as the temperature increases. The most notable change is a small increase in absorbance from 23-24 °C potentially indicating dissociation of the duplex DNA; however, the shift in absorbance is insignificant for the amount DNA in the sample, and an almost equal decrease in absorbance is seen directly afterward.

Additionally, thermal melting measurements were recorded for the PEG-dsDNA in 1,2-dichloroethane (Figure 2.9). The initial measurement shows an increasing absorbance; however, the temperature threshold of the instrument prevents analysis at a lower temperature. The rest of the curve shows no evidence of a sharp increase in absorbance. There are multiple explanations as to why thermal melting was not detected for the PEGylated DNA architectures in the organic solvents examined: 1 - the ssDNA counterparts were unable to maintain stable secondary structure in the organic solvent, 2 – the $T_m$ of the hybridized DNA is at or below the freezing point of the solvent of interest or 3 – below the temperature output threshold of our Peltier temperature controller. Due to the uncertainty of these results we chose the single-stranded PEG-DNA architecture to be best suited for the DNA-encoded library molecular architecture.

Additionally, based on our limited data collected, the dsDNA architecture was either unstable or had a melting point less than 25 °C in two of the organic solvents of interest to be

used for catalysis reactions. Furthermore, performing catalysis reactions at a temperature less than 25° C was found to slow the rate of catalysis in the aldol reaction. Therefore, it was concluded that a duplex architecture would not be favorable for the reaction rate in further studies.

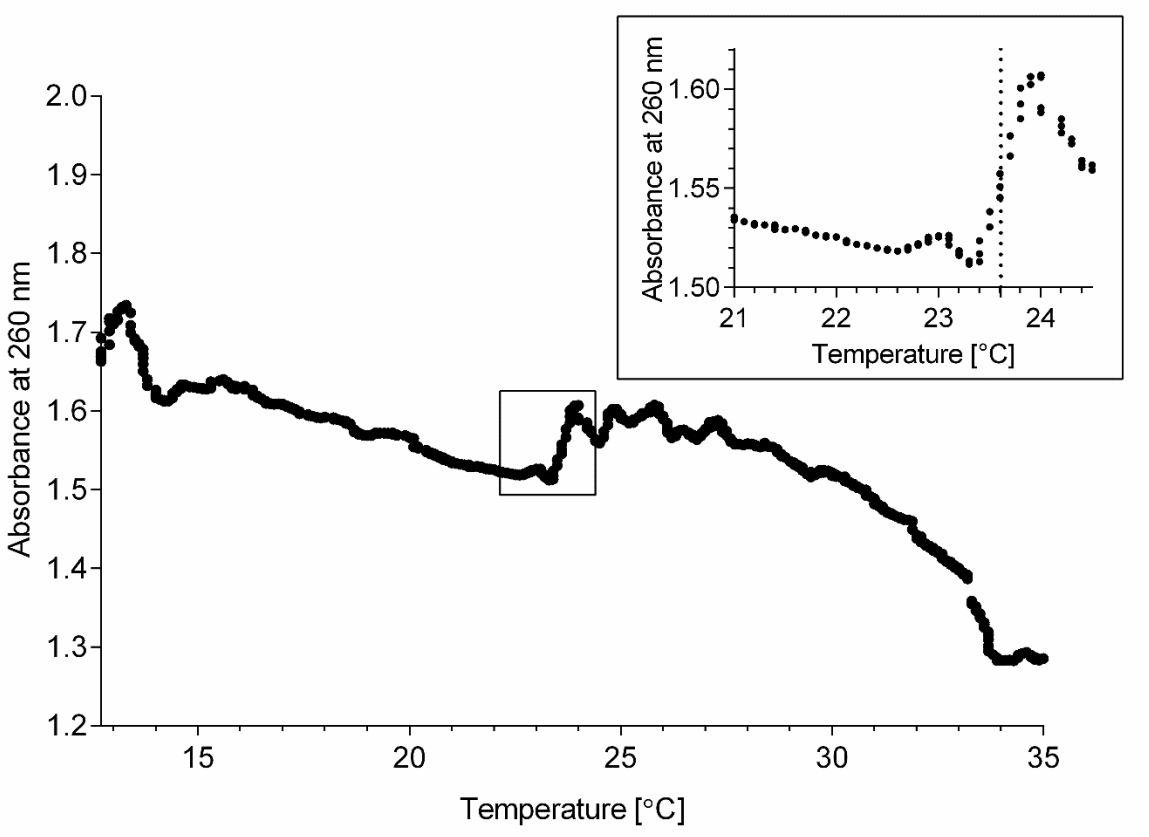



**Figure 2.8:** Thermal melting of PEGylated dsDNA in 1,4-dioxane

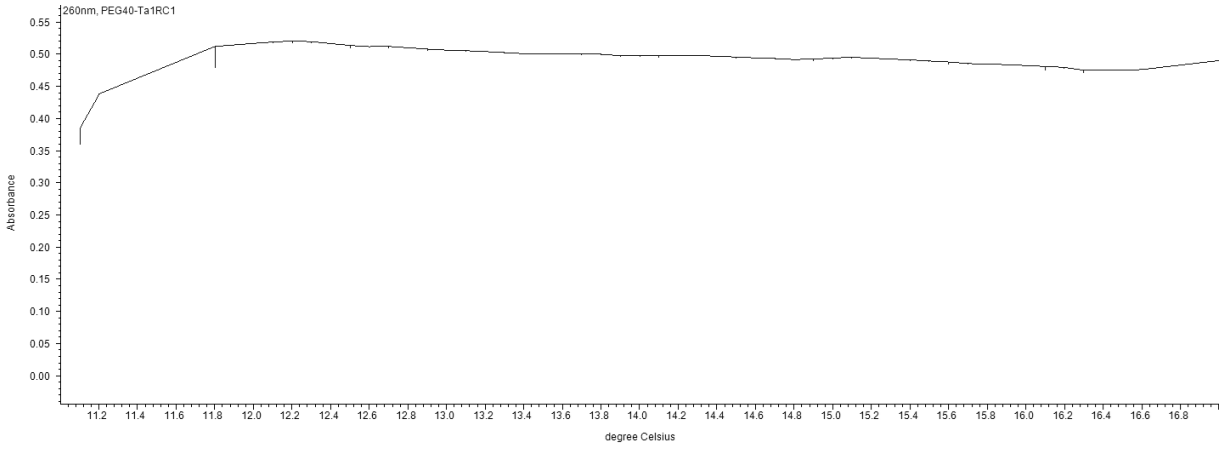



**Figure 2.9:** Thermal melting of PEGylated dsDNA in 1,2-dichloroethane

### 2.2.4 PCR Amplification of PEGylated DNA

Selection of catalysts capable of catalyzing a covalent bond formation to an affinity tag from an encoded library entails affinity purification of catalytically active molecules and PCR amplification of selection survivors for subsequent sequencing and identification. Amplification of catalytically active library members is essential to produce enough sample for sequencing identification or subsequent rounds of selection. As our oligonucleotides contain very large MW PEG moieties, it was critical to determine whether the polymerase activity will be affected/impeded. A PCR experiment was performed on the 10,000 and 40,000 MW PEG-DNA to determine if the attached PEG would affect polymerase activity (Figure 2.10).

The experiment included a No Template Control (NTC), in which the sample does not contain the template strand. This control is used to show there is no amplification seen without the template strand and, therefore, no contamination of the PCR reaction solution. The NTC was performed by amplifying a sample containing only primers and polymerase, without the template, for 18 cycles alongside the standard PCR procedure.

The gel image of the PEGylated DNA PCR products displays that the added bulk of the PEG group does not appear to hinder the activity of the polymerase noticeably. The intensity of the product DNA band increases with each set of cycles indicating that the concentration of DNA is increasing and, therefore, is amplified. Negligible template amplification of the NTC is apparent after 18 cycles, demonstrating that the polymerase is indeed not substantially hindered by the attached PEG group and that this amplification is not a result of template-contaminated primers. Overall, these results indicate that it will be possible to amplify the selection survivors of our amphiphilic DNA-encoded library without cleavage of the soluble moieties (PEG) after affinity separation.

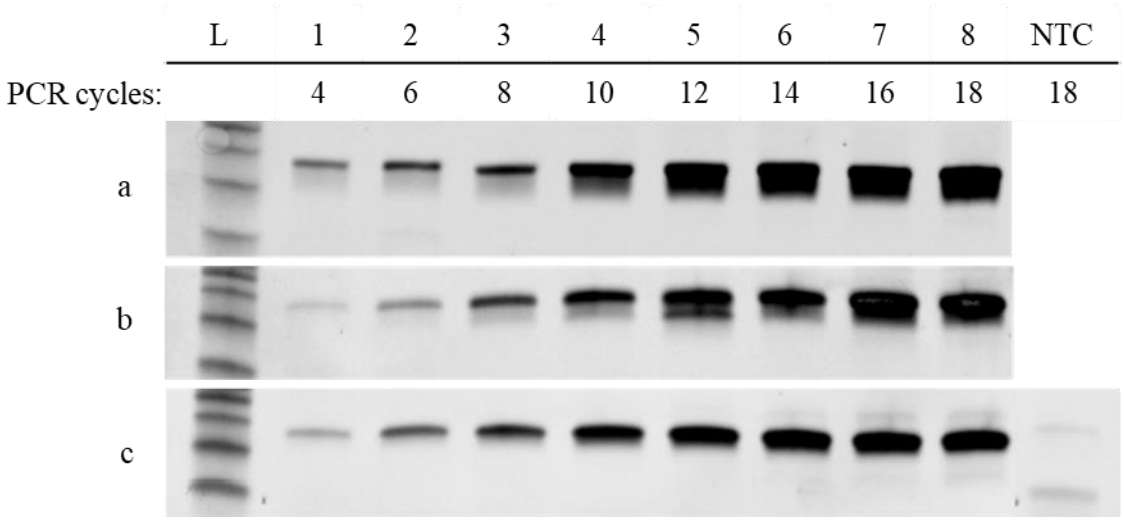



**Figure 2.10:** PCR efficiency of PEGylated DNA. a) Standard 48 mer. b) 10,000 MW PEG-Ta1. c) 40,000 MW PEG-Ta1. NTC: No-template control. All samples were run on 10% TBE non-denaturing gel at 150 V for 40 minutes and stained with EtBr.

## 2.3 EXPERIMENTAL METHODS

### General Methods

5’-Amine modified oligonucleotides were synthesized on polystyrene (PS) Universal Solid Support III (Glen Research, Cat No. 26-5110) with a 12-port MerMade automated DNA synthesizer (Bioautomation Corporation). A, T, C, and G phosphoramidites were purchased from Bioautomation Corporation. Modified phosphoramidites were purchased from Glen Research. PEG-NHS ester reagents were purchased from JenKem Technology. Oligonucleotides were purified by reverse phase HPLC on Hydrosphere C18 column (YMC). HPLC purification was performed using a [5% acetonitrile in 0.1 M TEAA, pH 7] to [90% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. All solvents used for solubility measurements were of HPLC grade and purchased from Sigma-Aldrich. Solvents were dried over activated 3 Å molecular sieves. Absorption spectra were recorded with BioEvolution260 UV-Vis spectrophotometer (Thermo Scientific). Samples were detected using a 10mm path length cuvette and scanning range of 350-220 nm wavelength. Thermo INSIGHT software was used to visualize the absorbance data. Absorption and qPCR spectra data were exported and graphed using GraphPad Prism. PCR reactions were performed using a Bio-Rad T100 thermocycler. qPCR reactions were performed with a Bio-Rad MiniOpticon$^{TM}$ Real-Time PCR system. All water used in experiments was deionized and nuclease-free.

### 2.3.1 Synthesis of PEGylated Oligonucleotides

#### 5’-Amino Modified 48-nt Oligomer

The DNA template for PEG-DNA solubility experiments was synthesized using the 5’-MMT-amino modifier C6 phosphoramidite (Glen Research, Cat No. 10-1906). The PS solid support

after a 1.0 μmol-scale automated oligonucleotide synthesis performed with a DMT-on synthesis program, were washed with acetonitrile (1 mL x 3) and dried over vacuum. The modified oligonucleotide cleavage from solid support, phosphate deprotection and base deprotection was performed in 1:1 ammonium hydroxide: methylamine (AMA) for 10 min at 65 ºC. The supernatant was filtered from the support and concentrated to 100 μL in centrifugal vacuum after addition of tris(hydroxymethyl)aminomethane (TRIS) base (25 mg) to ensure a basic pH of solution was maintained. The cleaved and deprotected oligonucleotide was purified by reverse-phase HPLC using a DMT-on program and lyophilized overnight to yield full-length product with 5'-MMT protection. Monomethoxy trityl (MMT) was deprotected in 20% glacial acetic acid for 1 hour at room temperature. MMT deprotection solution was lyophilized, resuspended in DI water and purified by reverse phase HPLC using a DMT-off program. The product fractions are lyophilized and resuspended in DI water to yield full length 5'-C6-amino modified 48-nt DNA (10% yield, 95.4% coupling efficiency). ESI -MS $[M-H]^-$ Calculated: 14963.7, Found: 14963.5.

**PEGylation of 5'-Amino Modified Oligonucleotides**

5'-amino modified 48-nt DNA (50 μL, 500 μM stock) [0.167 mM reaction concentration] dissolved in triethylammonium bicarbonate (TEAB) buffer (pH 8.7, 50 μL, 177 mM stock) [88 mM] was added to either 10,000, 20,000, 30,000 or 40,000 MW PEG-NHS ester dissolved in 50 μL of dimethylformamide (DMF) [4.17mM]. The reaction solution was vortexed at high speed for 30 min and reduced to medium speed for 4 hours at room temperature. The product was quenched with 500 mM TRIS-HCl (50 μL) and vortexed for 10 minutes. The solution was then frozen and lyophilized to remove solvent. The crude product was purified by reverse-phase HPLC using a gradient of 10-60% acetonitrile in 0.1M TEAA buffer, pH 7.1. Product fractions

were frozen and lyophilized to yield PEG-modified DNA. HPLC – Figure 2.3. MALDI MS: Experimental – Table 2.1. PAGE – Figure 2.2.

### 2.3.2 PEG-DNA Solubility

**Measurement of PEGylated DNA by UV-Vis Spectrophotometry**

500 pmol of 10,000-40,000 MW PEGylated-DNA in water was dispensed into Eppendorf tubes. The samples were dried by lyophilization. 100.0 μL of water, methanol, acetonitrile, 1,4-dioxane, 1,2-dicholorethane, heptane, or 1-octanol was added to dry PEG-DNA and vortexed overnight to ensure complete dissolution. Samples were mildly centrifuged to compact any undissolved suspended particulates. 5.0 μL of PEG-DNA solution was diluted with 95.0 μL of respective solvent to make the samples 1.0 μM. Absorbance was measured at 260 nm. The respective dissolving solvent was used as the Blank prior to DNA measurement.

**Measurement of PEGylated DNA solubility in UV-active solvents by qPCR**

A 10-fold standard dilution curve of aqueous phase oligo was prepared and tested in tandem with duplicate samples of PEG 40,000-DNA in water, DMSO, DMF, 1,2-dichloroethane, toluene, and heptane. The initial sample of DNA contained a 5 μM solution (500 pmol in 100 μL solvent). Once dissolved and gently centrifuged to compact any solid suspension, 5.0 μL was removed, evaporated in speed vacuum, and dissolved in 10 μL of water [0.1 μM]. The sample was serially diluted by adding 10 μL sample to 90 μL of water to a concentration of 1 nM. 1.0 μL of this solution was combined with SsoAdvanced polymerase (Bio-Rad). Thermocycling and fluorescence measurements were performed on MiniOpticon$^{TM}$ Real-Time PCR system (Bio-Rad).

### 2.3.3 Thermal Melting of DNA

**Thermal Melting in Aqueous Solution**

48mer Ta1-PEG 40K was incubated with either the reverse compliment (Ta1RC) or 3'-end 18mer reverse primer (PRa) in 15 mM sodium citrate, 150 mM NaCl, pH 7.0 buffer (10 μM) at 90 °C for 5 minutes and slowly cooled to room temperature overnight. Solution was diluted to 1 μM with 1 mL of buffer solution. 500 μL was used to measure the change in absorbance at 260 nm with temperature ranging from 50 – 85 °C (Ta1/PRa) or 60 – 95 °C (Ta1/Ta1RC). The temperature was increased at a rate of 0.4 °/min until maximum temperature was met and held for 15 min. The temperature was then decreased at the same rate and held for 15 min at the minimum temp.

**Thermal Melting in Organic Solvents**

Ta1-PEG40K template (1.0 nmol, 10 μM) and reverse complement (Ta1RC) (1.0 nmol, 10 μM) were dissolved aqueous buffer containing NaCl (100 mM) and sodium phosphate (10 mM). The sample was heated to 90ºC for 5 minutes and slowly cooled to room temperature overnight. The annealed solution was then lyophilized to remove water. The dried dsDNA was dissolved in 500 μL of 1,2-dicholorethane or 1,4-dioxane to make a 2.0 μM solution. The samples were mildly vortexed for 5 hours. The samples were centrifuged at 1600 RCF for three minutes. 250 μL of the annealed DNA in solvent was removed and diluted to 1 μM. 500 μL of solution was added to a pre-blanked cuvette and heated at 0.5 ºC per minute from 10-25 ºC or 0-35 ºC for 1,2-dichloroethane and 1,4-dioxane, respectively. Absorbance at 260 nm was measured every 10 seconds.

### 2.3.4 PCR of PEGylated Oligonucleotides

0.1 pmol of PEG-DNA (10,000, 20,000, 30,000, or 40,000 MW PEG) (1.0 μL 0.1 μM stock) was added to PCR mix containing 30 pmol of forward primer (PFa) (1.5 μL, 20 μM stock) (1.5 μM), 30 pmol of reverse primer (PRa) (1.5 μL, 20 μM stock) and water (6.0 μL). Solutions were added to an 8-well PCR strip. Hot Start Taq PCR Mastermix 2X (10 μL) was added last with pipette mixing. PCR primer concentration was 1.5 μM and total reaction volume was 20.0 μL The PCR reactions were performed with the following program:

Step 1: 3 minutes, 95°C
Step 2: 30 seconds, 95°C
Step 3: 3 seconds, 58°C
Step 4: 1 minutes, 72°C
GO TO Step 2, 17 times
Step 5: 7 minutes, 68°C
Step 6: infinity, 4°C

Reactions were stopped and stored on ice at the end of cycle 4, 6, 8, 10, 12, 14, 16, and 18.

**Non-Denaturing PAGE of PEG-DNA PCR Products**

5 μL of each PCR reaction were mixed with 1 μL of 6X non-denaturing loading dye and then run on a 10% non-denaturing PAGE gel at 150 V for 40 minutes. The gel was stained with ethidium bromide and imaged.

## 2.4 REFERENCES


1. Abe, H.; Abe, N.; Shibata, A.; Ito, K.; Tanaka, Y.; Ito, M.; Saneyoshi, H.; Shuto, S.; Ito, Y., Structure Formation and Catalytic Activity of DNA Dissolved in Organic Solvents. *Angewandte Chemie International Edition* **2012,** *51* (26), 6475-6479.

2. Rauzan, B.; McMichael, E.; Cave, R.; Sevcik, L. R.; Ostrosky, K.; Whitman, E.; Stegemann, R.; Sinclair, A. L.; Serra, M. J.; Deckert, A. A., Kinetics and Thermodynamics of DNA, RNA, and Hybrid Duplex Formation. *Biochemistry* **2013,** *52* (5), 765-772.

3. Ralf, G., Solvents and Solvent Effects in Organic Chemistry. 4th Ed. By Christian Reichardt and Thomas Welton. *Angewandte Chemie International Edition* **2011,** *50* (48), 11289-11289.

4. L., C. J.; A., H. C.; R., L. C. M.; Alejandro, P. V.; G., V. J., Solvent Effects on Hydrogen Bonding. *Angewandte Chemie International Edition* **2007,** *46* (20), 3706-3709.

5. Rozenman, M. M.; Liu, D. R., DNA-Templated Synthesis in Organic Solvents. *ChemBioChem* **2006,** *7* (2), 253-256.

6. Bonner, G.; Klibanov, A. M., Structural stability of DNA in nonaqueous solvents. *Biotechnology and bioengineering* **2000,** *68* (3), 339-344.

7. Sen, A.; Nielsen, P. E., On the stability of peptide nucleic acid duplexes in the presence of organic solvents. *Nucleic acids research* **2007,** *35* (10), 3367-3374.

8. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A., Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc Natl Acad Sci U S A* **1993,** *90* (22), 10700-4.

9. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **2003,** *31* (13), 3406-3415.

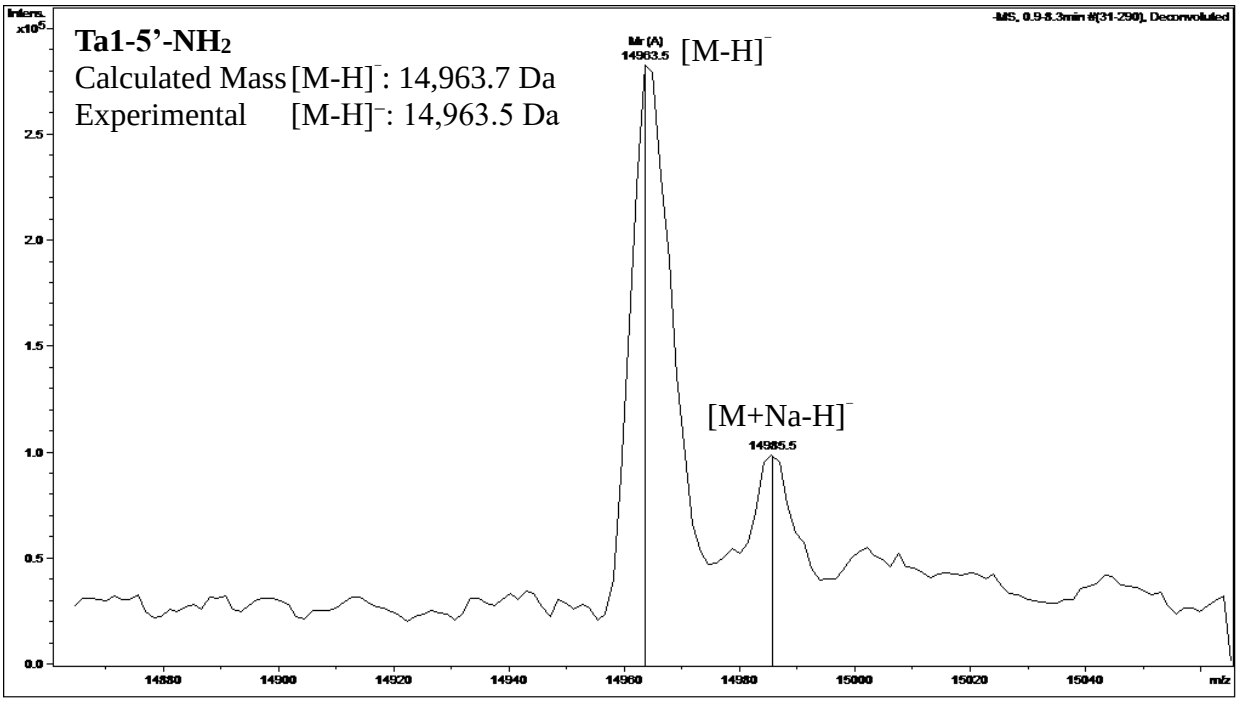
10. Bruno, T. J.; Svoronos, P. D., *CRC handbook of basic tables for chemical analysis*. CRC press: **2010**.

11. Wohlfarth, C., Selection and arrangement of static dielectric constants of pure liquid substances and their binary liquid mixtures. In *Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures: Supplement to Volume IV/17*, Lechner, M. D., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, **2015**; pp 1-2.

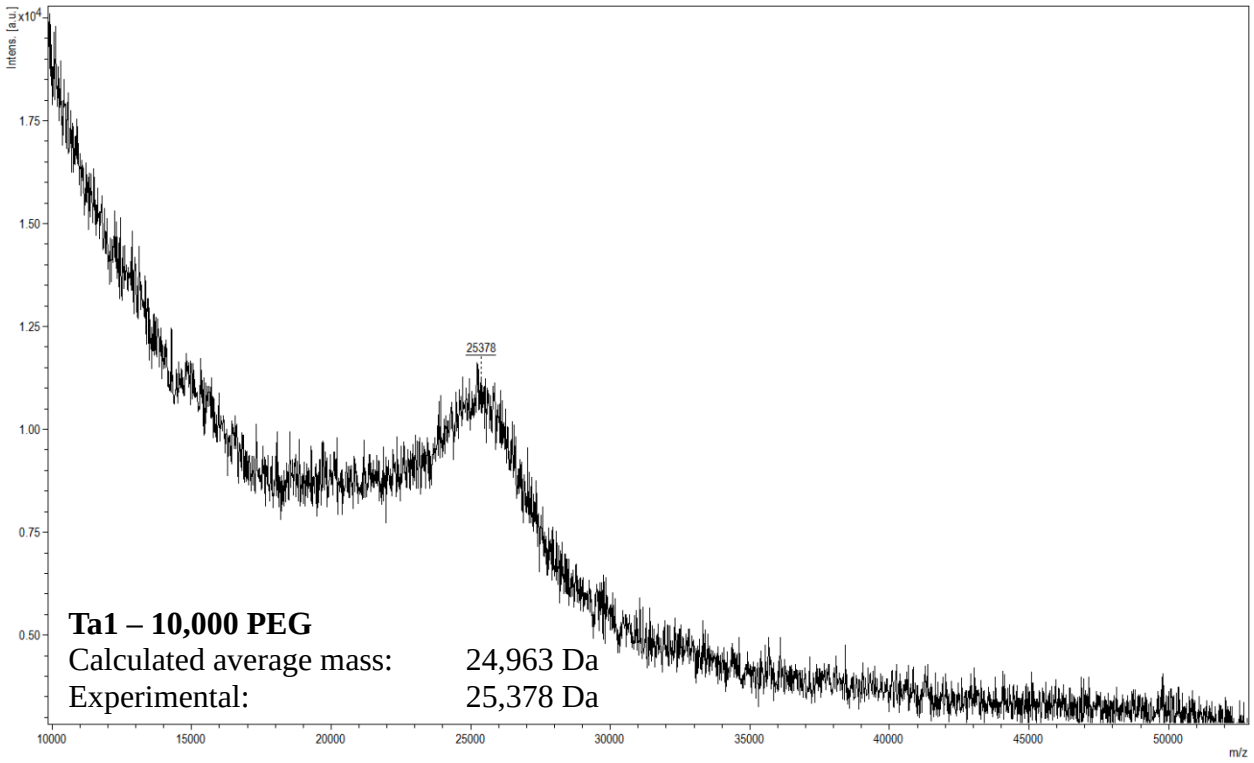
## S1. SUPPORTING INFORMATION

### S1.1 SELECTED SPECTRA

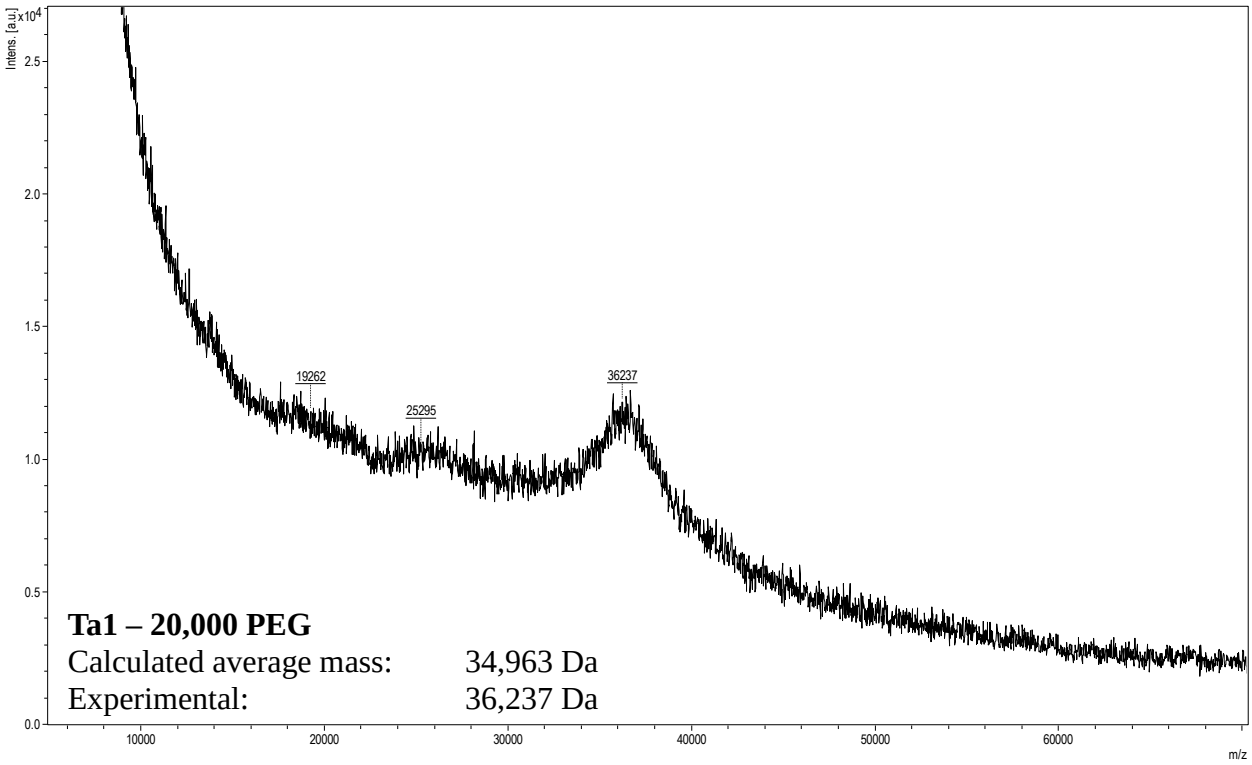
#### S1.1.1 Negative Mode ESI Mass Spectrum of Ta1-5'-$NH_2$



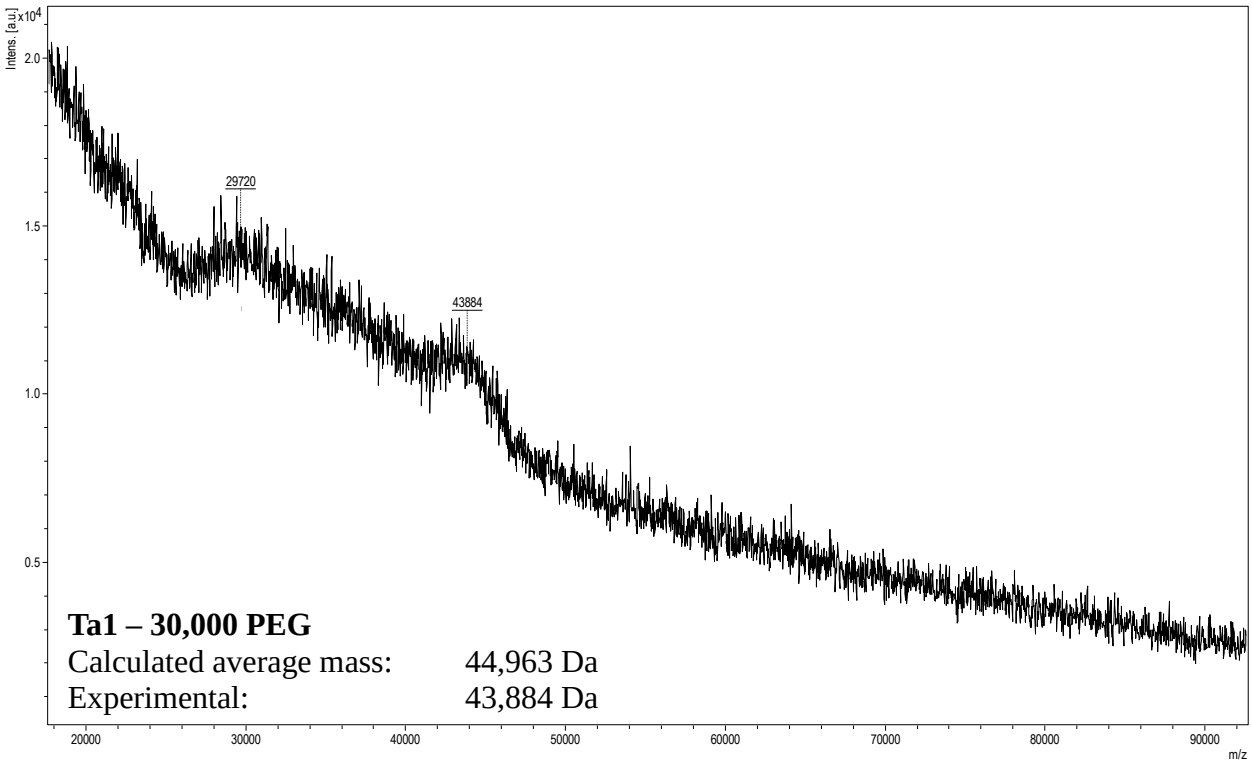
### S1.1.2 MALDI Mass Spectrum of 10,000 MW PEG-Ta1



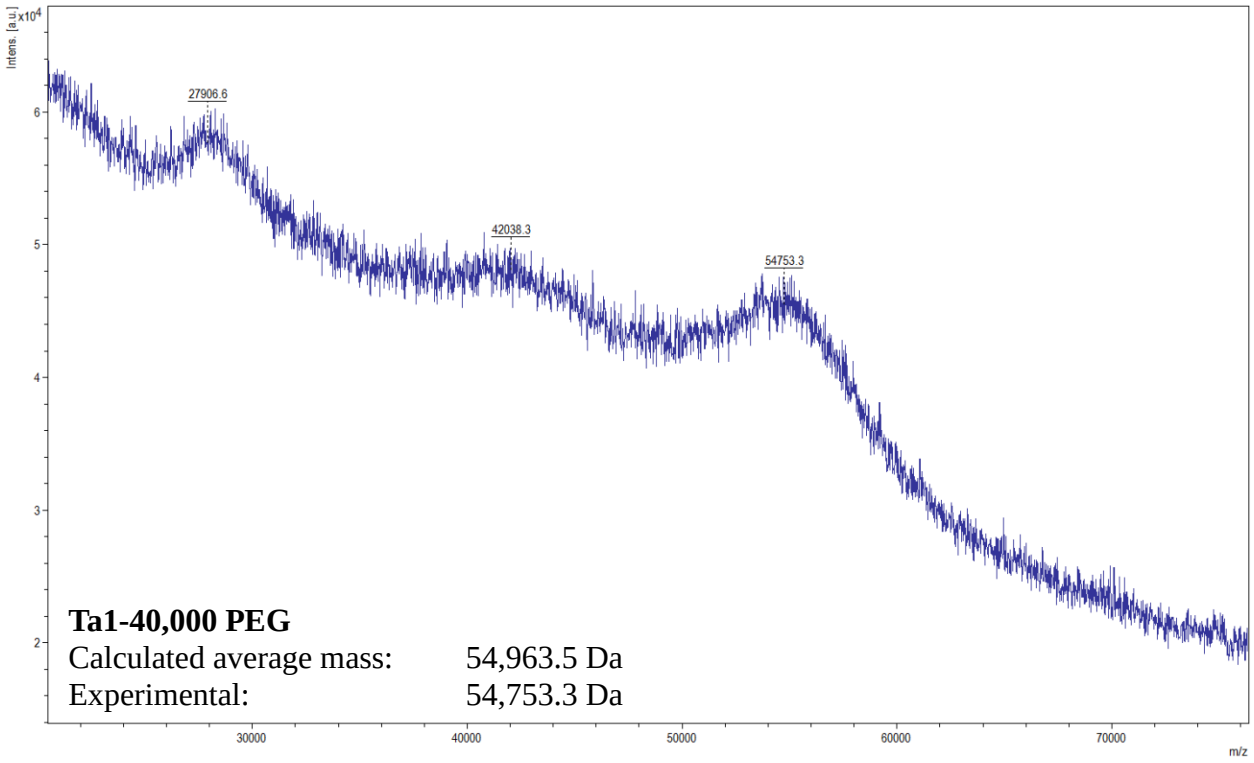
### S.1.1.3 MALDI Mass Spectrum of 20,000 MW PEG-Ta1



### S.1.1.4 MALDI Mass Spectrum of 30,000 MW PEG-Ta1



### S.1.1.5 MALDI Mass Spectrum of 40,000 MW PEG-Ta1



CHAPTER 3

# PEGYLATED DNA AS AN ENCODING ELEMENT FOR CATALYST DISCOVERY

## 3.1 INTRODUCTION

The influence of PEGylated DNA on small molecule catalysis is of interest to the objective of developing a DNA-encoded catalyst library for in vitro selection in organic solvents. We have previously shown a 48-nt ssDNA coupled to 40,000 MW PEG to be soluble in various organic solvents (MeOH, DCE, MeCN, 1,4-dioxane) in addition to water.[1] Compatibility of the PEGylated DNA to be used as a viable encoding element was confirmed by studying the scope of DNA-linked reactants for various chemical reactions with known small-molecule catalysts in organic solvents. The scope of chemistry permitted on PEGylated DNA was studied to identify limitations in reaction type or catalytic mechanism.

Determination of whether PEGylated DNA was compatible with various reactions in organic solvents, a modified oligonucleotide with one substrate (*Reactant A*) covalently attached to the PEGylated DNA was incubated with another substrate (*Reactant B*) and a catalyst in an organic solvent (Figure 3.1a). Additionally, *Reactant B* displays a molecular tag, such as a fluorophore, clickable functional group, or biotin to monitor reaction efficiency. A well-characterized catalyst, chosen from the abundant information that exists in the literature, was chosen for each reaction studied to catalyze the desired bond formation between *Reactant A* and *B.* Reaction success results in the attachment of the affinity tag to the DNA by means of carbon-carbon bond formation.

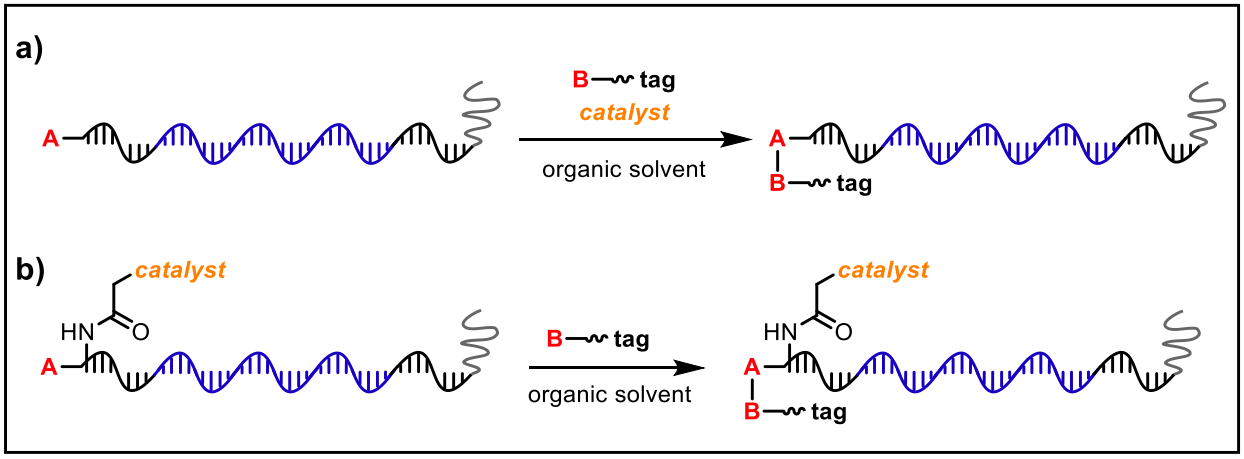



**Figure 3.1:** Schematic for validation of PEGylated ssDNA as an encoding element with reactant bound on the 3'-end and (a) catalyst in solution or (b) catalyst covalently attached proximally to 3'-end.

Furthermore, following the confirmation of successful catalysis containing reactant-bound PEGylated DNA and free catalyst in solution, we sought to explore reactions with a small-molecule catalyst tethered to the PEGylated DNA in addition to *Reactant A* (Figure 3.1b). Our data suggest that the optimal display of *Reactant A* and the catalyst is for *Reactant A* to be covalently linked to the 3'-end of the DNA, proximal to the catalyst, while the affinity-tagged *Reactant B* is present in solution in mM concentrations. We demonstrate the desired bond formation can occur between *Reactant A* and *Reactant B* in an organic solvent when the catalyst is covalently attached to the encoding DNA strand with assistance of a large polyethylene glycol (PEG) unit at the 5' end to increase solubility.

Proof-of-principle reactions examined include the aldol reaction[2] and Friedel-Crafts acylation[3] (Figure 3.2). Well-characterized catalysts have been discovered for these processes, which serve as positive controls for this discovery platform, and importantly are readily conjugated to PEGylated DNA. Fundamental experiments were also performed to examine the

time course of the reaction, ideal concentrations of DNA-bound catalyst and reactants, and solvent effects. These experiments, explained in detail in the following chapter, confirm the potential of this approach, provide useful information regarding optimal linker length and composition to attach the catalyst and substrate to the PEGylated DNA.

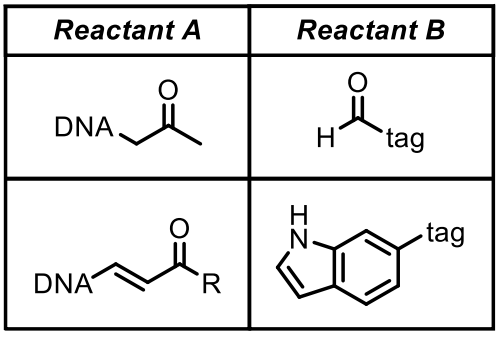



**Figure 3.2:** Reaction types of catalytic reactions

Characterization of the catalytic aldol reaction was performed using a streptavidin-based electrophoretic mobility shift assay (EMSA). Catalytic reaction success differentiates the desired product via biotin installment, and incubation with streptavidin allows visual comparison of unreacted starting material and successfully catalyzed reaction products by a shift in mobility using native polyacrylamide gel electrophoresis (PAGE).

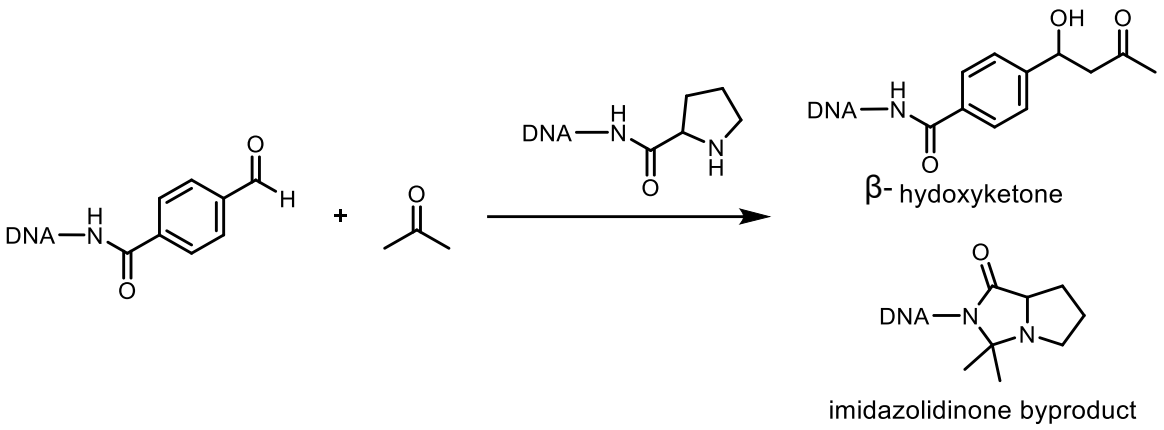
Streptavidin is a tetrameric protein isolated from the actinobacterium *Streptomyces avidinni*, and has a molecular weight of 60,000 g/mol (MW = 4 x 15,000).[4] Streptavidin is remarkable for its ability to bind up to four molecules of *d*-biotin with unusually high affinity (dissociation constant $K_d = 10^{-15}$ M).[5] The ability of streptavidin to bind derivatized forms of

biotin has led to its widespread use in diagnostic assays that require formation of an essentially irreversible and specific linkage between biological macromolecules.[6-7]

### 3.1.1 Aldol Reaction

Interested in the potential of small peptide catalysts[8] and encouraged by the reported success of DNA-templated aldol reactions catalyzed by proline-modified ssDNA in aqueous solvents, we implemented the secondary amine catalyzed aldol reaction between a ketone and benzaldehyde as our model reaction. We first used this system to ensure the required catalytic bond formation could occur with the PEG modification on the 3'-reactant tethered ssDNA.

The aldol reaction results in a carbon-carbon bond formation that occurs via enamine catalysis. Enamine attack of the aldehyde results in the formation of an iminium ion which is subsequently hydrolyzed to form the β-hydroxyketone product. Diproline was used to catalyze the aldol reaction when experimenting with the DNA architecture with catalyst tethered to PEG-DNA. The dipeptide was employed, as opposed to the single amino acid, due to concerns of possible reaction byproducts that decrease the catalytic turnover of proline and yield of our product of interest. As reported by Marx,[9] proline can undergo condensation with the ketone substrate to render the catalyst inactive (Scheme 3.1). The researchers overcame this shortcoming by constructing an oligonucleotide with the diprolinamide moiety. This implementation led to less imidazolidinone formation due to the absence of a γ-amide proton to the N-atom involved in enamine formation.




**Scheme 3.1:** DNA-templated cross-aldol reaction catalyzed by prolinamide. Adapted from [9]

### 3.1.2 Friedel Crafts Reaction

Inspired by the work performed to evolve a deoxyribozyme for the Friedel-Crafts reaction between acyl imidazole and 5-methoxy indole, the copper(II)-catalyzed Friedel-Crafts reaction was incorporated as an additional reaction scheme to broaden the reaction scope of the selection platform. Friedel-Crafts alkylation[3] is a carbon-carbon bond forming reaction between electron-rich aromatics and electron deficient alkenes. Lewis-acid catalysis enables 1,2- or 1,4-addition depending on the reaction conditions and reactants employed. We envisioned this reaction could be imported into our DNA-encoded catalyst selection platform, whereby a bipyridine ligand is conjugated to a modified nucleobase of a PEGylated oligonucleotide with a uniquely encoded sequence and coordinated with Cu(II) to form the active catalyst. An affinity-tagged electron-rich substrate is combined with the encoded catalyst ssDNA modified with a Cu(II) coordinating *Reactant A*. The proposed catalytic cycle of the Friedel-Crafts alkylation reaction of α,β-unsaturated-2-acylimidazole with 5-methoxyindole is shown below (Figure 3.3).

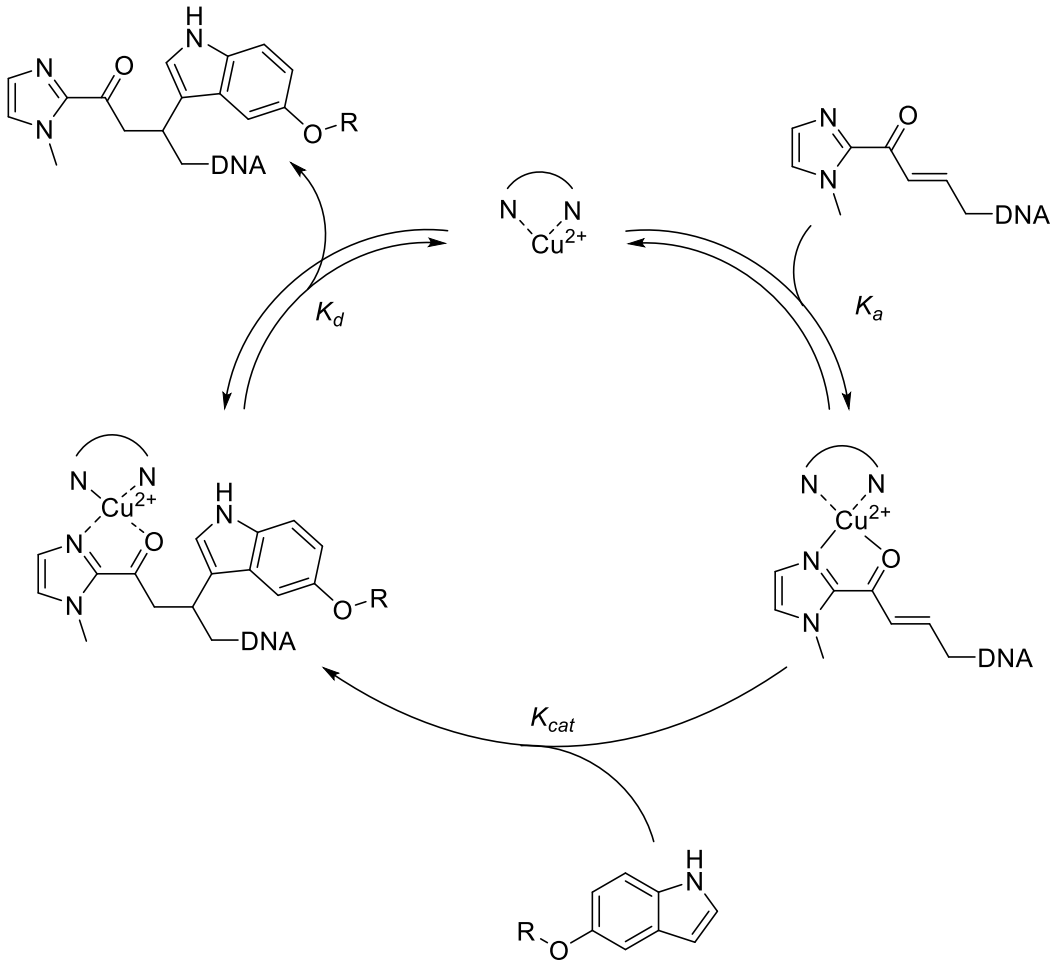



**Figure 3.3:** The copper(II)-catalyzed Friedel-Crafts alkylation reaction of α,β-unsaturated-2-acylimidazole with 5-methoxyindole. Adapted from[10]

The indole substrate will be used as the electron donating aromatic in the Friedel-Crafts reaction. It will bind with the copper coordinated substrate tethered to DNA. The chalcone derivative tethered to the DNA can coordinate to activated copper (II) catalyst, thereby rendering the α,β-unsaturated bond electron deficient and susceptible to nucleophilic attack.

## 3.2 RESULTS AND DISCUSSION

### 3.2.1 Design and Synthesis of the Amphiphilic DNA-Encoded Catalyst

Having determined the optimal PEG length for conjugation to a 48 nt ssDNA to enable solubility in organic solvents, we next investigated the effect of small-molecule catalysis on these amphiphilic PEG-DNAs in a variety of organic solvents. We designed a DNA architecture that would accommodate the catalyst site, the reactant site, and a PEGylation site, and could be readily synthesized by solid-phase DNA synthesis. The oligonucleotide displays three orthogonal modification handles to enable catalysis reactions in which the small-molecule catalyst and *Reactant A* are both covalently attached to the encoding DNA strand (Figure 3.4). We reasoned that the PEGylation site should be distal to the catalyst and reactant site to minimize interference on catalysis by the PEG chain. We also incorporated a long spacer between the catalyst site and the reactant site to permit sufficient flexibility for the catalyst to engage the substrate comfortably. A 3'-alkynyl group was chosen to permit ready conjugation of different aldol reactants by copper-catalyzed click reaction.[11] A flexible PEG spacer was used to separate the DNA-linked reactant and catalyst conjugated to a modified thymidine base; This was followed by a 3'-end primer-binding site, a 12 nt encoding region, and a 5'-end primer-binding site. At the 5'-terminus was installed a thiol, which was used for conjugation to PEG 40,000 maleimide.

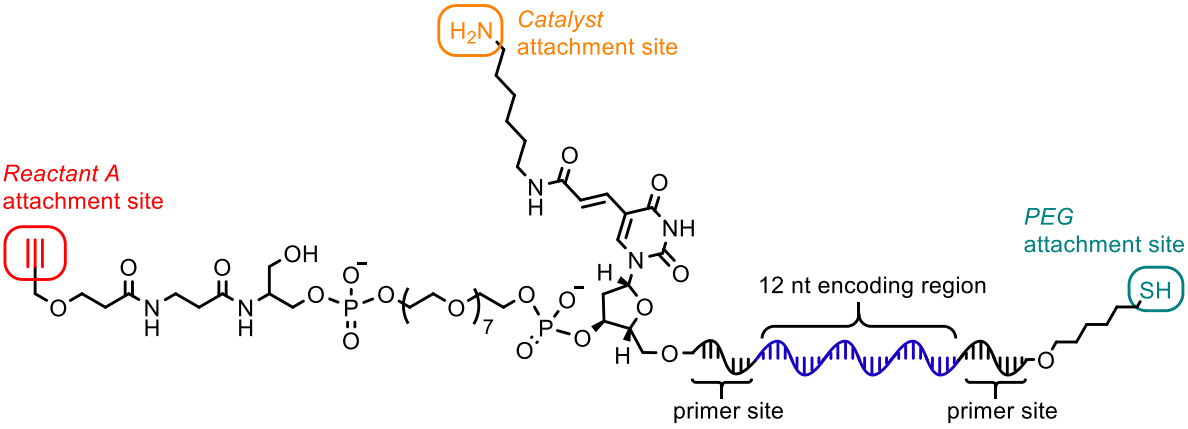



**Figure 3.4:** Molecular architecture of the DNA-encoded catalyst

DNA synthesis was performed with standard DNA phosphoramidites of the four canonical DNA bases adenine, thymine, cytosine, and guanine (A, T, C, and G) in addition to four commercially available modifier phosphoramidites for orthogonal post-synthetic modification of the ssDNA (Figure 3.5). The 3'-alkynyl linkage was incorporated into the ssDNA architecture using a 3'-alkyne serinol modified CPG solid support (**3'AlkSer**, Figure 3.5 b), followed by coupling of an 18 atom PEG linker, **Sp18**. Optimal linker length requires the reactant and catalyst to be within sufficient proximity to interact with each other, maintain flexibility in the system, and minimize potential interactions with the DNA itself. The possibility of interference by the structure of DNA during reactions between substrates is minimized by using long and flexible substrate–DNA linkers.[12] Next, the hexylamino-modified thymine base (**AmC6dT**) was coupled to incorporate the primary amine handle for post-synthetic modification with a carboxylic acid functionalized small-molecule catalyst, followed by an 18 nt forward primer, 12 nt encoding sequence, and 12 nt reverse primer region using **bz-dA-CE**, **ac-dC-CE**,

**dmf-dG-CE**, and **dT-CE** (Figure 3.2a). Finally, a 5'-disulfide modifier was added to the ssDNA for the addition of the 40,000 MW PEG-maleimide moiety after thiol activation.

Cleavage from the solid support and deprotection of the hydrophobic protecting groups found on the DNA bases and phosphate linkages was performed in AMA, and the DMT-OH protected ssDNA was purified by HPLC to recover full-length product. Subsequent DMT deprotection in acetic acid solution and additional purification by HPLC afforded the final desired synthetic oligonucleotide. Final ssDNA products were analyzed and confirmed by the shift in retention time on HPLC, mobility shift by PAGE, and analysis by mass spectrometry.

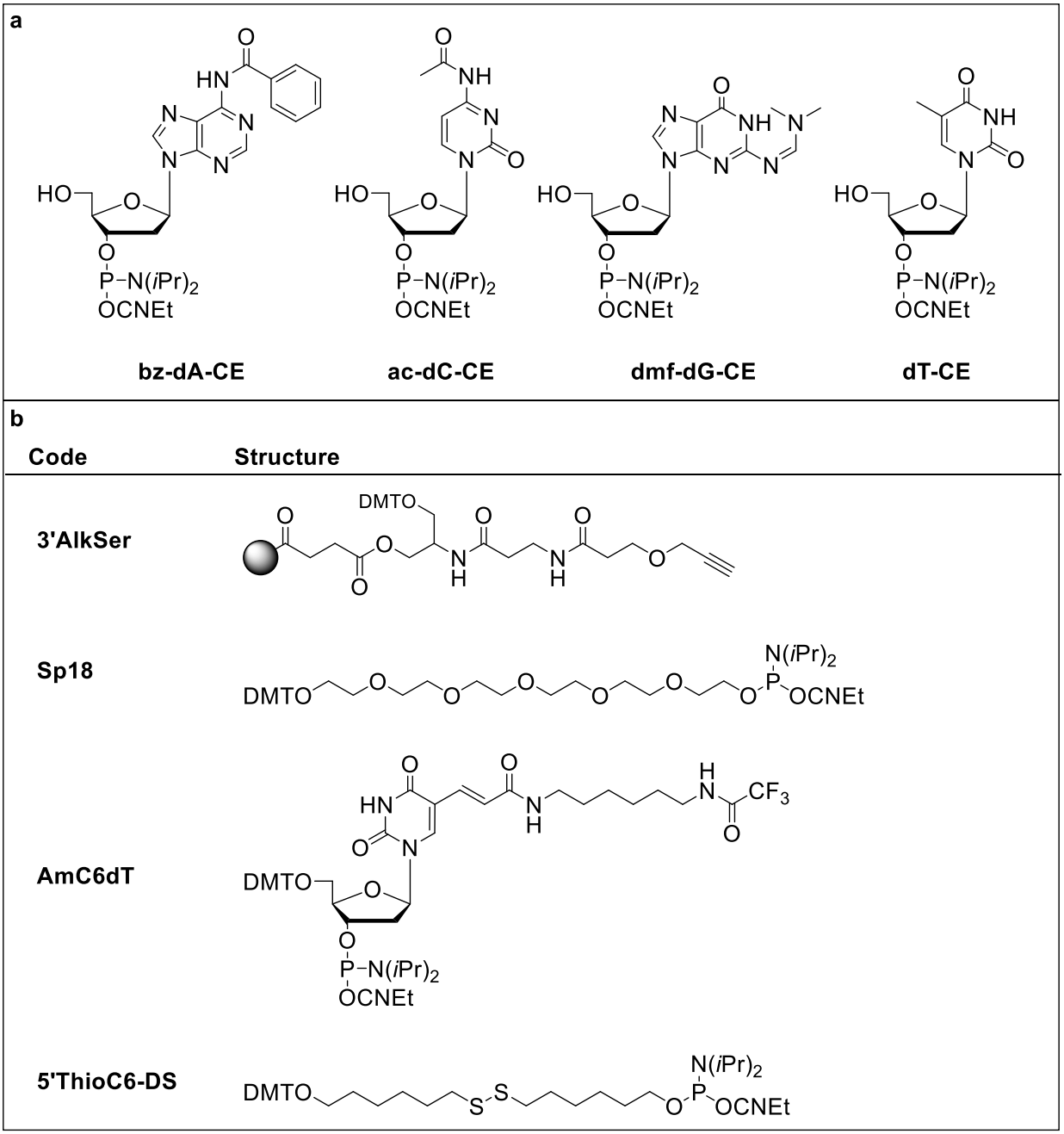



**Figure 3.5:** Solid phase DNA synthesis phosphoramidites. (a) Canonical nucleobase phosphoramidites used for the encoding sequence. (b) Modified phosphoramidites to accommodate orthogonal DNA modification.

.

### 3.2.2 Aldol Reaction

The secondary amine catalyzed aldol reaction between a ketone and an aldehyde was implemented as our first model system to determine the effect of PEGylated DNA on the aldol reaction catalyzed by diprolinamide (Figure 3.5). Before performing the catalysis experiments with PEG-DNA, the substrates necessary to make oligonucleotide (**1**) and biotinylated reactant (**2**) were synthesized.

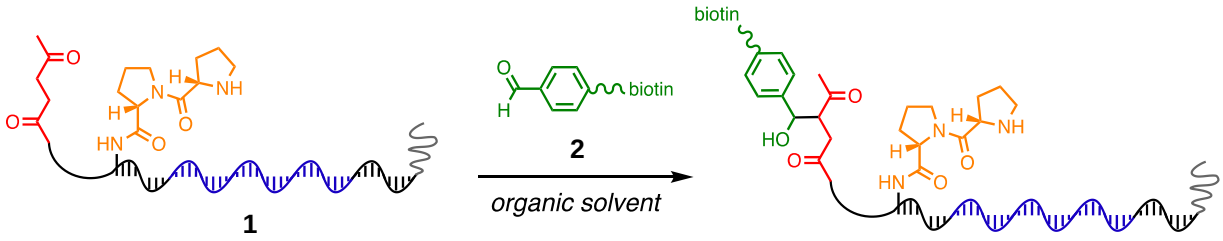
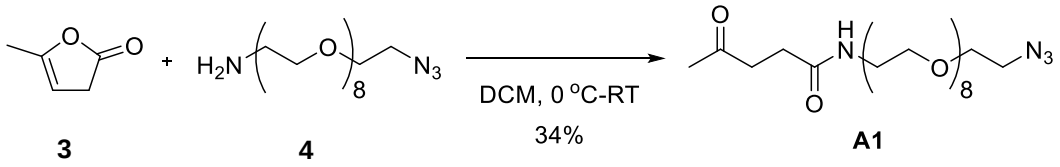



**Figure 3.6:** Aldol Reaction scheme of the amphiphilic DNA-encoded catalyst with aldol substrate. Ketone aldol reactant is shown attached to DNA.

An azido-PEG-ketone (**A1**) was synthesized to serve as *Reactant A* (Scheme 3.2). α-Angelica lactone (**3)** undergoes a ring-opening reaction in the presence of the azido-PEG-amine (**4)** to afford the azido-PEG-ketone (**A1**). Biotinylated benzaldehyde (**B**$_1$) was synthesized (Scheme 3.3) by amide bond formation between the carboxylic acid of 4-formylbenzadehyde (**5**) of the amino group of biotin-PEG-amine (**6**) using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide·HCl (EDC·HCl) and N-hydroxysuccinimide (NHS) in DMF/MeCN. Carbodiimide conjugation works by activating the carboxyl group for direct reaction on the primary amine via amide bond formation. EDC couples NHS to the carboxylic acid to form an NHS ester that is readily displaced by the nucleophilic attack from the primary amino group once added to the

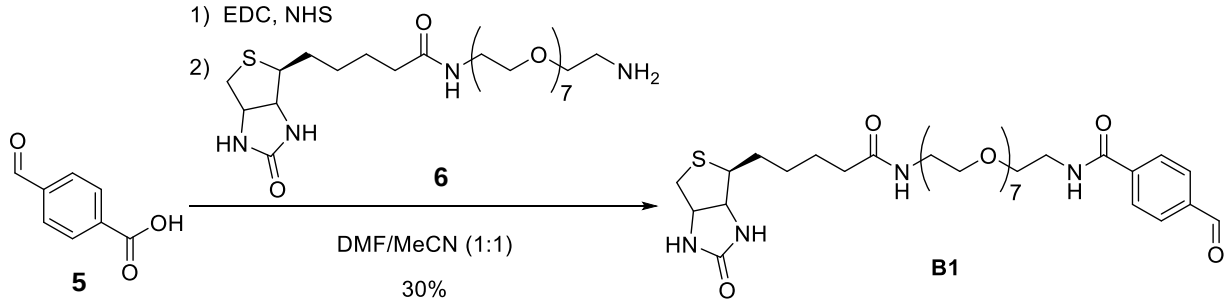
reaction mixture. The primary amine forms an amide bond with the original carboxyl group, and an EDC byproduct is released as a soluble urea derivative.

We decided to use the ketone and aldehyde reactants in both possible positions in the catalysis reaction. Therefore, an azido-PEG-aldehyde (**A2**) was synthesized to test the reactivity compared to the ketone (Scheme 3.4). The reaction was also performed using EDC/NHS coupling chemistry as explained above. Biotinylated ketone (**B2**) was synthesized by α-angelica lactone (**3**) ring-opening reaction, as previously explained, in the presence of biotin-PEG-amine (**6**) (Scheme 3.5).



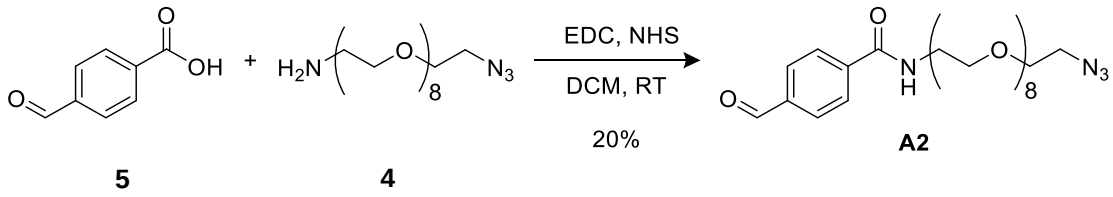

**3** **4** **A1**

**Scheme 3.2:** Azido-PEG-ketone synthesis (**A1**)

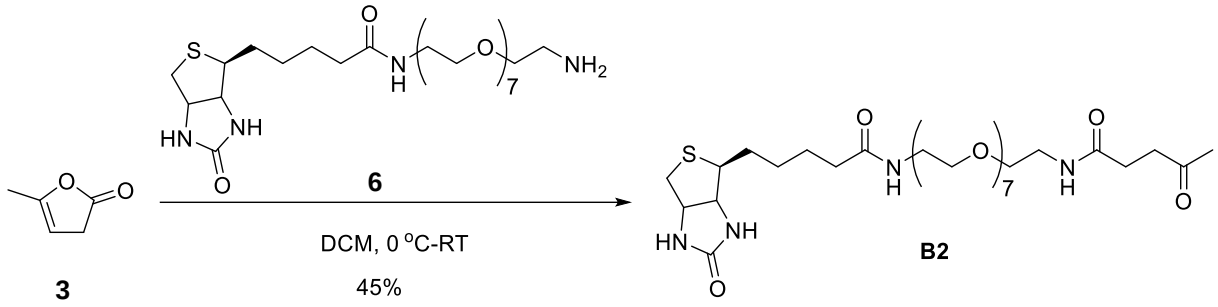



**5** **6** **B1**

**Scheme 3.3:** Azido-PEG-aldehyde synthesis (**B1**)




**Scheme 3.4:** Biotinylated aldehyde synthesis (**A2**)




**Scheme 3.5:** Biotinylated ketone synthesis (**B2**)

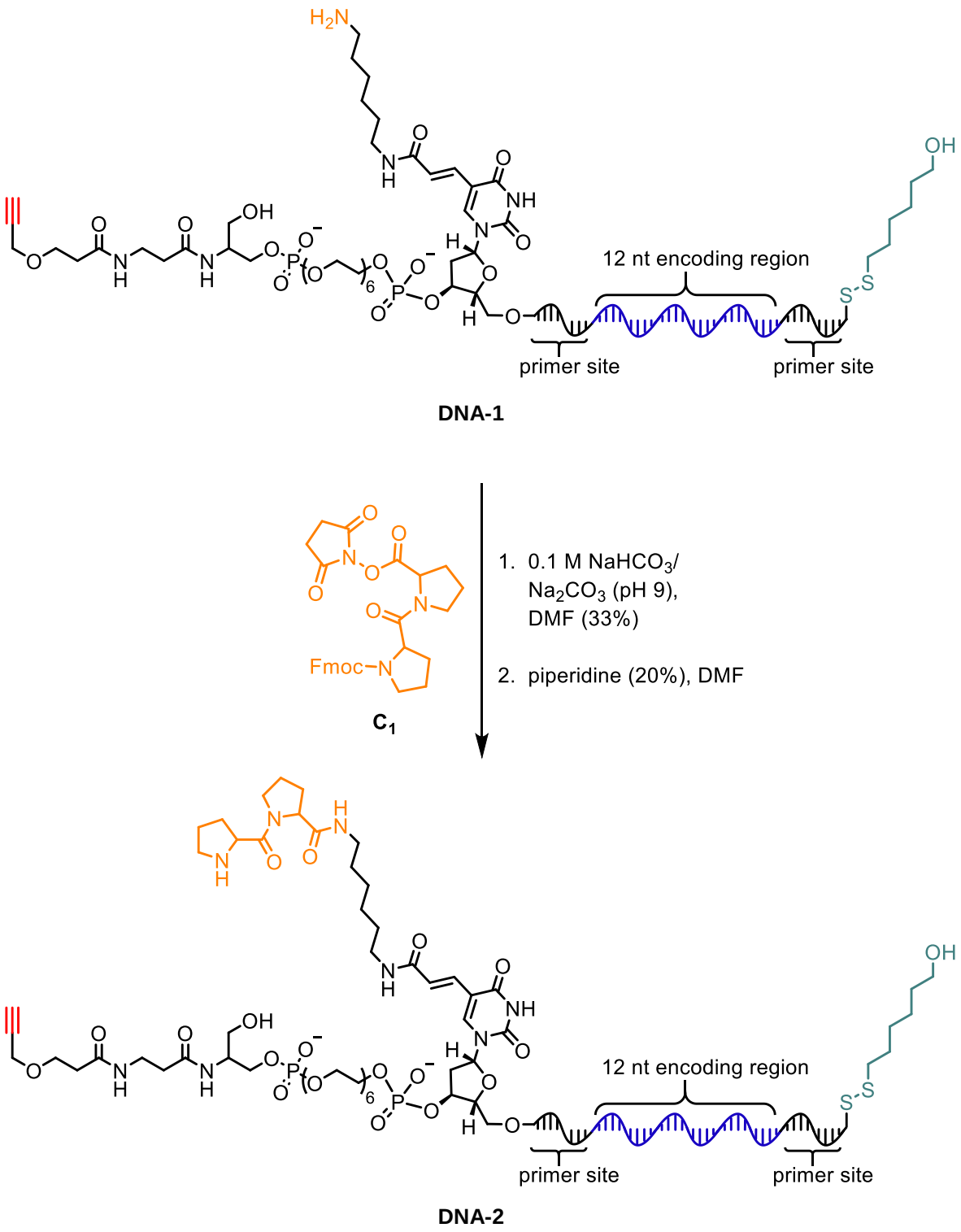
Following synthesis of the necessary reactants for the aldol reaction, synthesis of the DNA catalyst architecture was performed. The first post-synthetic modification made to the ssDNA architecture was the addition of Fmoc-Pro-Pro NHS ester (**C1**) to the amine modified thymine base adjacent to the 3'-end, **DNA-1** (Scheme 3.6). Fmoc-pro-pro-OH was first converted into the activated NHS ester for more efficient amide bond formation. The amide coupling reaction was performed in DMF with 0.1 M sodium bicarbonate solution at pH 9. At this pH only primary amines are reactive, therefore reducing the possibility of exocyclic secondary amine reactions on the nucleobases. Fmoc deprotection was performed in a 20% piperidine solution to afford the free amino group of the diproline catalyst (**DNA-2**).

The 3'-alkynyl linkage was incorporated on the ssDNA for conjugation of different aldol reactants by copper-catalyzed azide-alkyne cycloaddition (CuAAC).[9] This reaction is widely

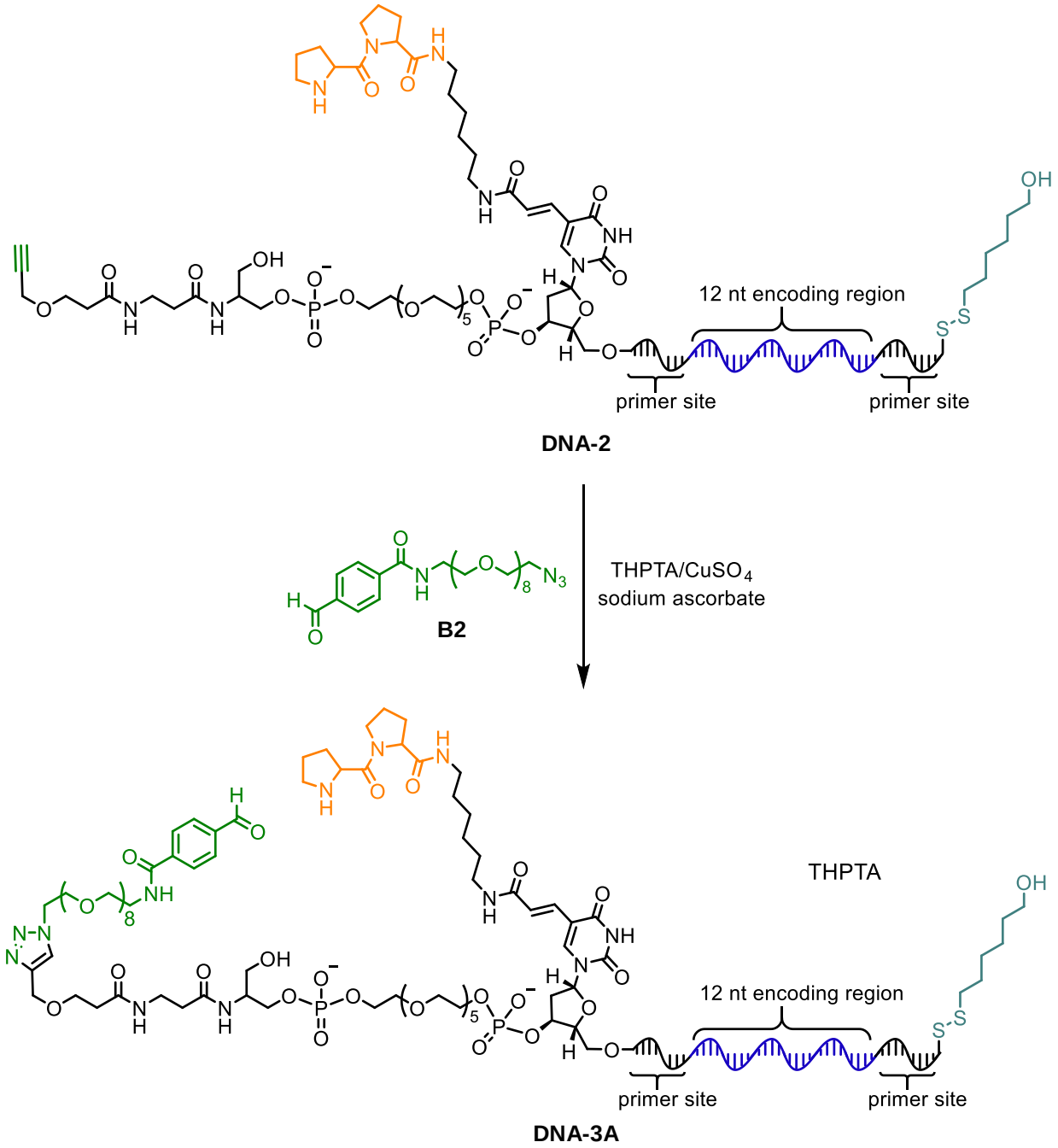
used in chemical biology applications due to its effectiveness in richly functionalized environments, including reactivity in aqueous systems.[13] CuAAC chemistry was used to attach an azido-PEG-ketone (**A1**) to the 3'-alkyne modified ssDNA (**DNA-2**). THPTA was used in the reaction as an accelerating ligand[14] and ascorbic acid as a reducing agent to afford the benzaldehyde- (**DNA-3A**) or the ketone- (**DNA-3K**) coupled diproline-modified DNA (Scheme 3.7).

The 5'-end thiol modification on the ssDNA allows 40,000 MW PEG-maleimide coupling by thiol-Michael addition chemistry (Scheme 3.8). First, the disulfide is reduced to the thiol using the reducing agent dithiothreitol (DTT), commonly known as Cleland's Reagent,[15] in sodium phosphate buffer at a pH of 8.5. Reduction of the disulfide bond on the ssDNA occurs via two sequential thiol-disulfide exchange reactions with DTT, which forms a stable six-membered ring with an internal disulfide bond. The DTT reduction must be performed at a basic pH (pH > 7) so that the negatively charged thiolate is formed. Next, buffer exchange and DTT removal from the solution was performed using a centrifugal gel filtration column. Due to the propensity of the activated thiol to form diols, 40,000 MW PEG-maleimide was immediately added following buffer exchange to afford the aldehyde-propro-PEG DNA (**DNA-5A**) or ketone-propro-PEG DNA (**DNA-5K**). The thiol-Michael addition reaction occurs via nucleophilic attack on the electron deficient carbon-carbon double bond of the PEG-maleimide in a sodium phosphate buffer at pH 6 and occurs in the absence of a catalyst.[16]

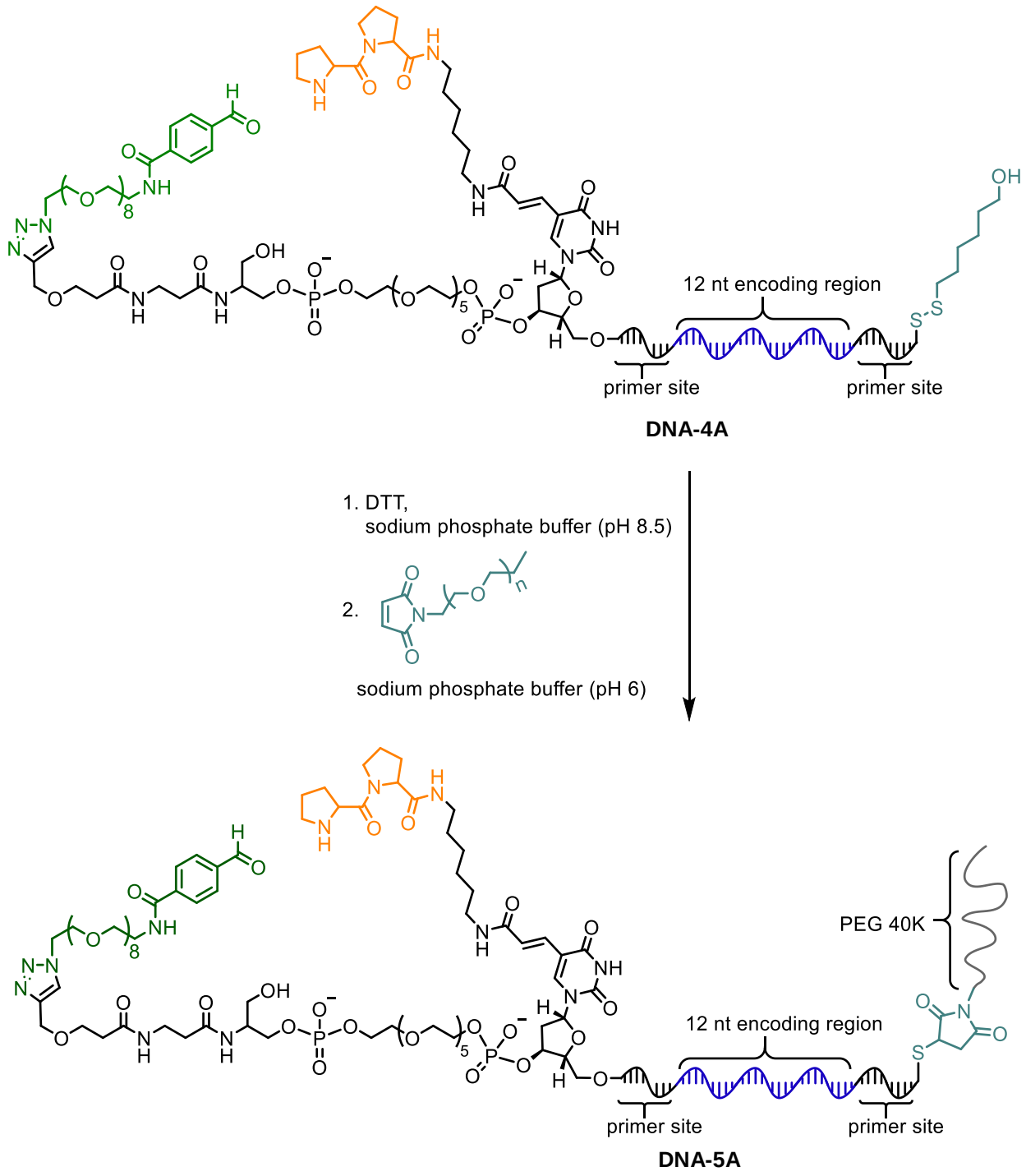
Following each post-synthetic modification of the ssDNA architecture, the reactions were purified by HPLC and the isolated product of interest was characterized by the shift in retention time on HPLC and analysis by mass spectrometry (Table 3.1).




**Scheme 3.6:** Diproline (**C1**) coupling to amino-modified ssDNA. Reaction centers are colored orange.

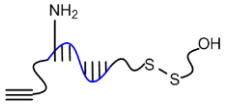
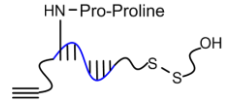
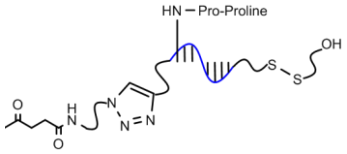
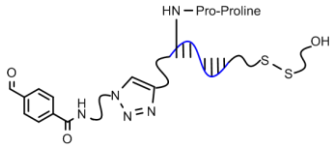



**Scheme 3.7:** Copper-catalyzed azide-alkyne cycloaddition (CuAAC) of 3'-alkyne modified ssDNA (**DNA-2**) and azido-PEG-benzaldehyde (**B1**) or azido-PEG-ketone (**A1**, not shown). Reaction centers are colored green.




**Scheme 3.8:** Thiol-Michael addition of 40,000 MW PEG-maleimide to 5'-thiol modified ssDNA (**DNA-3A**) or with the ketone (**DNA-3K**, not shown). Reaction centers are colored light blue.

**Table 3.1:** Mass Spectrometry Results of Modified Oligonucleotides

| Oligonucleotide Code[a] | | Calculated Mass | LC-ESI-MS |
|---|---|---|---|
| **DNA-1** (YAm48DS) |  | 16,181.96 | 16,185.6 (+3.64) |
| **DNA-2** (YPP48DS) |  | 16,377.07 | 16,376.0 (-1.07) |
| **DNA-3K** (KPP48DS) |  | 16,931.69 | 16,914.0 (+0.31) |
| **DNA-3A** (APP48DS) |  | 16,947.71 | 16,948.1 (+0.93) |

[a]Oligonucleotide Code: Y – alkyne, Am – amino, K – ketone, A – aldehyde, PP – diproline, DS – disulfide.

## Validation of Streptavidin-Based EMSA with Biotinylated PEG-DNA

Before the catalyst experiments, the ability to visualize streptavidin binding by gel shift was confirmed. A 3'-biotinylated oligonucleotide was synthesized with a 5'-amine and PEGylated with a 40,000 MW PEG NHS-ester (**DNA-bt**, Figure 3.7a). This affinity tag labeled oligonucleotide was used as the positive control for all the streptavidin-based electrophoretic mobility shift assay (EMSA) experiments performed as a comparison of streptavidin-bound PEGylated ssDNA movement during gel electrophoresis. Various equivalents of streptavidin were incubated with ten picomoles of the biotinylated DNA and characterized on a 5% acrylamide non-denaturing gel (Figure 3.7b).

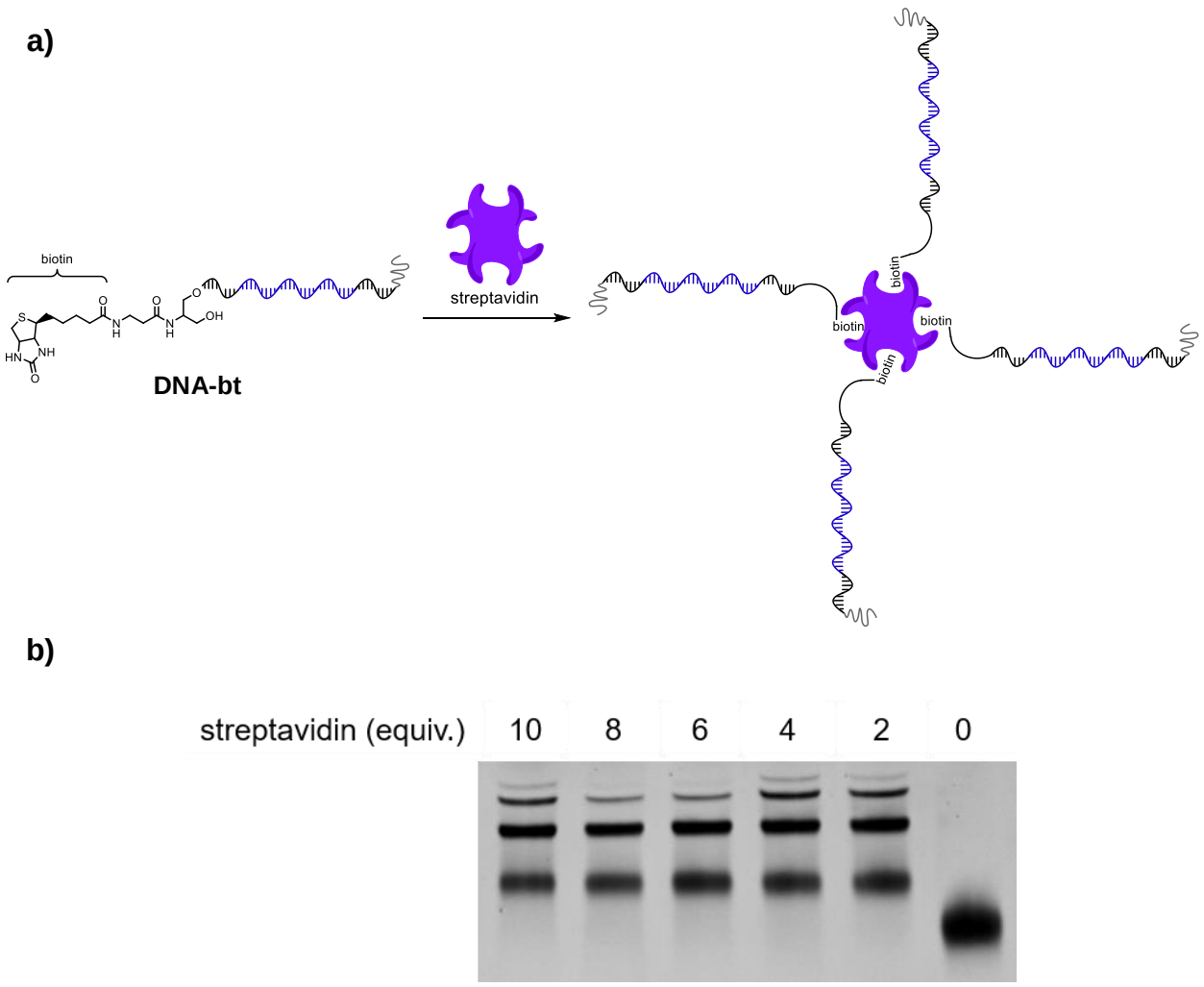



**Figure 3.7**: Streptavidin binding of biotinylated PEG-DNA (**DNA-bt**). a) Schematic of streptavidin binding to biotinylated PEG-DNA b) Optimization of streptavidin-based EMSA with biotinylated 40,000 MW PEG-DNA positive control.

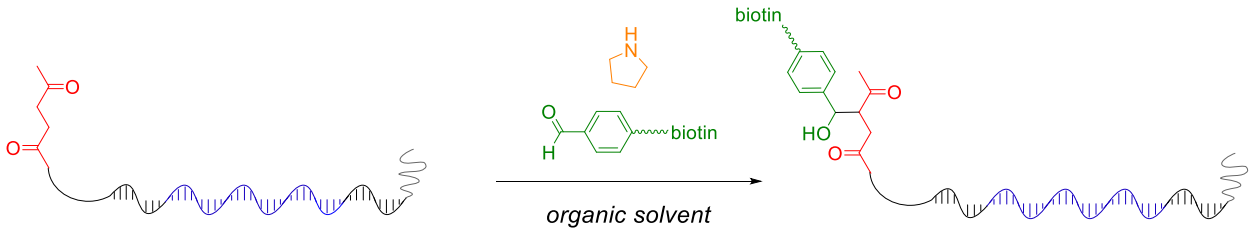
The results of the positive control streptavidin binding assay show that there is not a noticeable difference when increasing amounts of streptavidin are added to the binding solution. Therefore, it was decided to use 2-4 equivalents of streptavidin in subsequent streptavidin binding assays, as this result demonstrated the most visible bands of all four binding combinations. These results confirm the viability of this assay to be used for characterization of the catalytic aldol reaction, as catalytic reaction success differentiates the product via biotin

installment. After incubation of the reaction mixture with streptavidin, a mobility shift between unreacted starting material and successfully catalyzed reaction products using native gel electrophoresis can be visualized.

## Compatibility of PEGylated DNA as an Encoding Element for Catalytic Reactions in Organic Solvents

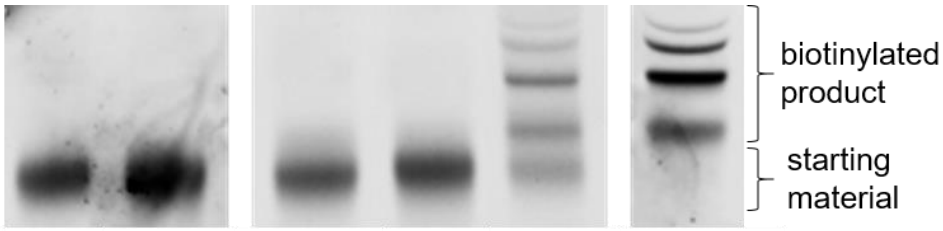
*Free Catalyst in Solution*

The compatibility of PEGylated DNA to be used as an encoding element and reactant scaffold in organic solvents was first tested by evaluating its ability to facilitate successful covalent bonding in the presence of a solution-based catalyst via the Aldol reaction. Before experiments with the diproline catalyst tethered to the DNA, we first used this system to determine if the required catalytic bond formation could occur with the presence of PEG modification on the ssDNA. A modified oligonucleotide was synthesized the ketone substrate (*Reactant* A) covalently attached to the 3'-end of the PEGylated DNA without a catalyst. Incubation with the reaction partner (*Reactant B*) and catalyst at mM concentrations in solution was performed to determine if the PEGylated DNA is compatible with various reactions in organic solvents (Scheme 3.9).




**Scheme 3.9:** Aldol reaction of reactant-tethered PEG-DNA and biotinylated aldehyde with pyrrolidine (catalyst) in solution.

The aldol reaction was performed with the ketone-tethered PEG-DNA and biotinylated aldehyde in the presence of an amine catalyst, pyrrolidine free in solution. We first tested the reaction in DMF and MeCN to determine if successful reaction products could be visualized in these solvents (Figure 3.8). Following purification and streptavidin binding, the reaction contents were visualized by staining the gel with ethidium bromide (EtBr) and exposure to UV light. Images were captured using a Bio-Rad Imaging system. The exposure intensity of the bands was measured by the imaging software standard with the imaging system. The relative intensity of each band is used to determine the quantity of DNA relative to other bands and lanes on the gel. The product yield of the reaction was determined by comparing the amount of biotinylated DNA catalysis product to non-biotinylated DNA starting material.




| | 1 | 2 | 3 | 4 | 5 | 6 (POS) |
|---|---|---|---|---|---|---|
| ketone on DNA | + | + | – | + | + | |
| biotin-aldehyde | – | + | + | + | + | |
| pyrrolidine | – | – | + | + | + | |
| streptavidin | + | + | + | + | + | |
| reaction solvent | – | MeCN | MeCN | DMF | MeCN | |

**Figure 3.8:** Aldol reaction with the catalyst in solution – EMSA. 5% Native PAGE. Lane 5 streptavidin bound, biotinylated reaction product (70% yield). Reaction conditions: PEG-DNA (1 μM) was dissolved in either DMF or acetonitrile (MeCN). When appropriate, biotinylated aldehyde (20 mM) and pyrrolidine (200 mM) were added. Reactions were vortexed for 5 days, purified by Centri-Sep column, incubated with streptavidin, followed by 5% native PAGE.

The products of the aldol reaction with ketone-tethered PEGylated DNA, biotinylated aldehyde, and pyrrolidine in DMF and MeCN can be seen in lanes 4 and 5 (Figure 3.8). These results show that the catalyzed carbon-carbon bond formation proceeds in MeCN with a reaction yield of 70%. A reaction product was not seen when the reaction was carried out in DMF. Importantly, to rule out possible false positive results two control experiments were carried out alongside the experimental reaction. First, to ensure the reaction would not proceed without a secondary amino catalyst present, the reaction was performed without addition of the catalyst, pyrrolidine, in solution (Figure 3.8, lane 2). It can be seen from this lane that there are no biotinylated products visible. It was concluded that in these reaction conditions the reaction does not proceed without the presence of the catalyst in solution. Second, potential side reactions of

the biotinylated aldehyde with secondary amino groups on the DNA bases were controlled for by exclusion of the ketone reactant on the DNA (Figure 3.8, lane 3). Again, biotinylated product bands are not visible, which allows us to rule out the DNA bases as a reactive site for the biotinylated aldehyde in these conditions. Additionally, there is no evidence to suggest binding affinity of the starting material oligonucleotide to streptavidin when subjected to the protein binding conditions (Figure 3.8, lane 1).

Following visualization of reaction products in MeCN, we next sought to determine other solvents that would allow the reaction to proceed successfully. The reaction was performed in a variety of solvents, in which the PEGylated DNA was found to be soluble, to determine optimal solvent conditions. The reaction products were incubated with streptavidin and reaction yields were determined by gel shift assay (Figure 3.9).

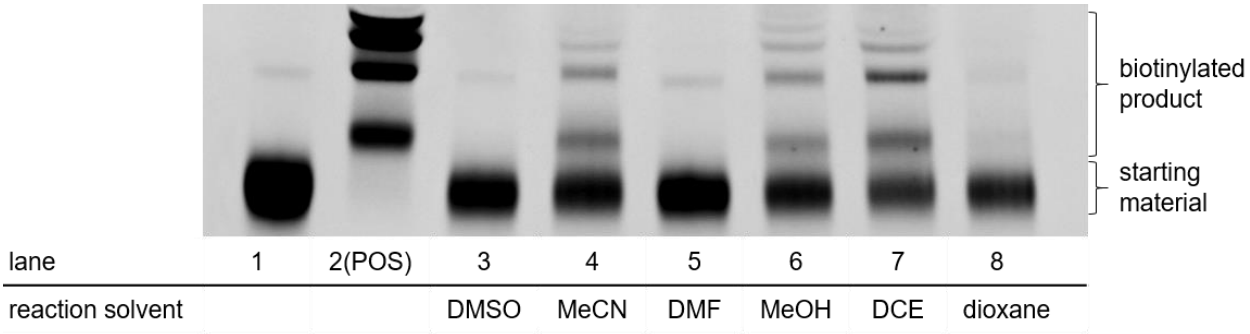



**Figure 3.9:** Catalyst in solution reaction solvent optimization – EMSA. Aldol reaction.

These results indicate that the aldol reaction with pyrrolidine as a catalyst in solution is more efficient in acetonitrile, methanol, and 1,2-dichloroethane (Figure 3.9, lanes 4, 6 and 7). There is a faint streptavidin bound band for dioxane as well; however, for DMF and DMSO the

reaction conditions were not favorable for product formation. Overall, the aldol reaction success in acetonitrile and reasonable desired product yield with the catalyst in solution provides initial evidence for the use of PEGylated DNA as an encoding element.

*Catalyst-Tethered PEG-DNA*

Following the success of the reactant-bound DNA with the free catalyst in solution, catalysis reactions with the small-molecule catalyst covalently attached to the 3'-end of the PEGylated DNA were examined (Figure 3.10 a). The ability of a DNA-conjugated pro-proline to catalyze the aldol reaction between a DNA-tethered ketone and biotinylated aldehyde or vice versa was explored. Previous experiments of the '*in trans'* version of this experiment demonstrate successful catalysis product between the ssDNA template and biotinylated aldol reactant; that is, a ketone-conjugated PEG-DNA can undergo carbon-carbon bond formation with a biotinylated-aldehyde in the presence of an external catalyst, pyrrolidine. The present study removes the need for the external pyrrolidine catalyst with the presence of a diprolinamide catalyst directly conjugated to the DNA scaffold itself.

Our overall goal for *in vitro* selection of active DNA-encoded catalysts from a library of potential DNA-encoded small-molecule catalysts is first realized by proving the principle aldol reaction is catalyzed by diproline tethered PEG-DNA. Since the catalyst-selection system for bond-forming reactions can have either reactant immobilized on DNA, we chose to examine the effect of aldol substrate identity conjugated to the DNA-encoded catalyst. Reactant structures for the diproline catalyzed aldol reaction are shown in rows (b) and (c) of Figure 3.10. Two reactant position scenarios exist, one in which the enolizable ketone is conjugated to the 3'-end of the

PEGylated DNA (**$A_1$**) and a biotinylated aromatic aldehyde is in solution (**$B_1$**), or the aromatic aldehyde is DNA-linked (**$A_2$**), and the biotinylated ketone (**$B_2$**) is in solution.

We subjected both the ketone-conjugated DNA and the aldehyde-conjugated DNA architectures to the catalysis reaction in DCE. (Figure 3.11). EMSA analysis showed that catalytic bond formation proceeded with both architectures with reasonable desired product yield. The catalysis experiment was performed by incubating the aldehyde or ketone-pro-pro-PEG oligo with the biotinylated ketone or aldehyde, respectively, for five days at room temperature in 1,2-dichloroethane. A positive control biotinylated PEG-DNA of the same length was used for comparison to full product conversion. A longer reaction time (5 days) was permitted as well as immediate post-purification by a gel separation column, allowing the ssDNA starting material and product to be separated from the excess biotinylated reactant (*Reactant B*) which competes with successful reaction products for streptavidin binding, decreasing the binding efficiency of biotinylated ssDNA products and subsequent perceived decreased yield by visualization with PAGE. After purification, samples were incubated with four equivalents of streptavidin, run on a 5% non-denaturing PAGE gel and imaged with ethidium bromide.

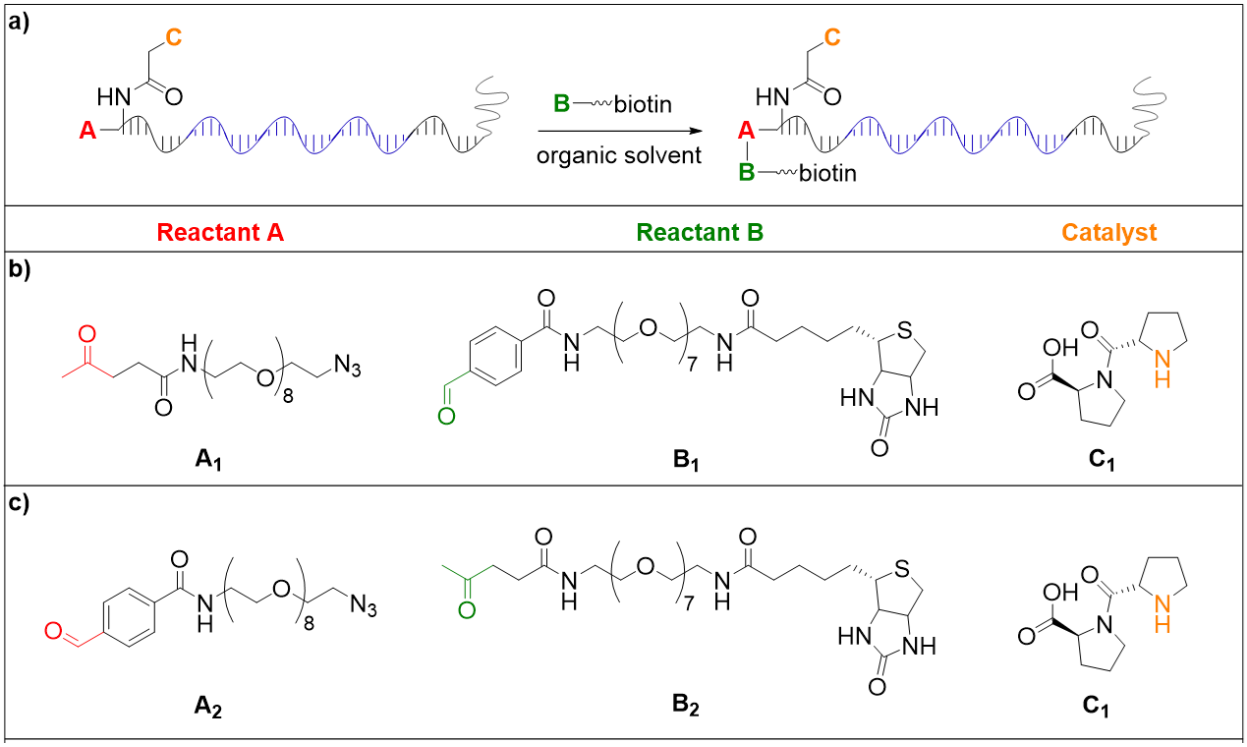



**Figure 3.10:** Aldol reactant, catalyst, and reaction product structures

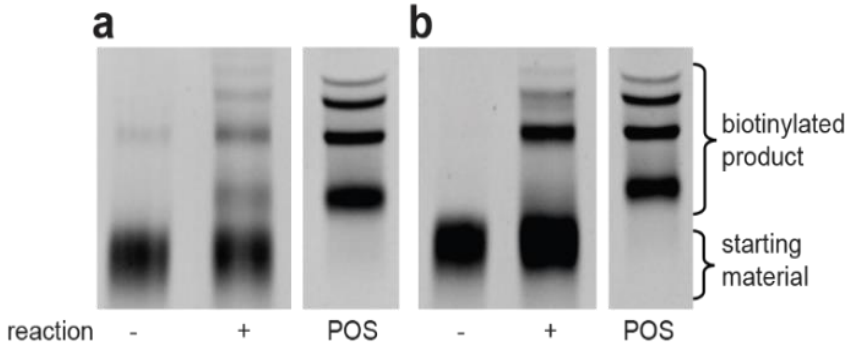



**Figure 3.11:** Streptavidin-mediated EMSA comparison of aldol catalysis with either a) aldehyde reactant on DNA with the biotinylated ketone in solution and b) ketone on DNA with the biotinylated aldehyde in solution. Reactions were performed for five days in DCE. POS = positive control, biotinylated DNA.

## Aldol Reaction Optimization with Catalyst on DNA

*Reactant/Solvent Order of Addition Experiment*

To gain insight as to the accessibility of the reactant and catalyst tethered to the PEGylated DNA, the order in which catalysis reaction reagents and solvent were added to the reaction mixture was examined (Figure 3.12). We hypothesized that the PEG moiety on the ssDNA could be aggregating with other PEG chains and possibly form a micellar-type structure in which the hydrophilic DNA orients into the core of the macrostructure surrounded by PEG when exposed to certain organic solvents. If the PEG-DNAs are not separate entities in solution, this would potentially affect the accessibility of the 3'-end conjugated reactant and proline-proline catalyst and its ability to react with the biotinylated reagent in solution.

To determine if this limitation could be overcome, the addition of the biotinylated reactant was performed in two separate ways. First, the PEG-DNA was evaporated to dryness and the biotinylated reactant, dissolved in acetonitrile, was added to the dry pellet. We hypothesized addition of the small-molecule reactant before solubilizing the DNA in organic solvent could overcome the possible steric inaccessibility of the functionalized PEG-DNA caused by such aggregates, and that the biotinylated reagent added to dry PEG-DNA would be more likely to be interspersed throughout the reaction solution. Additionally, a separate experiment was performed in which the PEGylated DNA was dissolved in acetonitrile before addition of the biotinylated reactant to examine the difference of reaction yield in comparison to the first experiment. The objective of this experiment was to determine whether the aggregate formation is affecting the reaction efficiency of the catalyst-tethered PEGylated oligos.

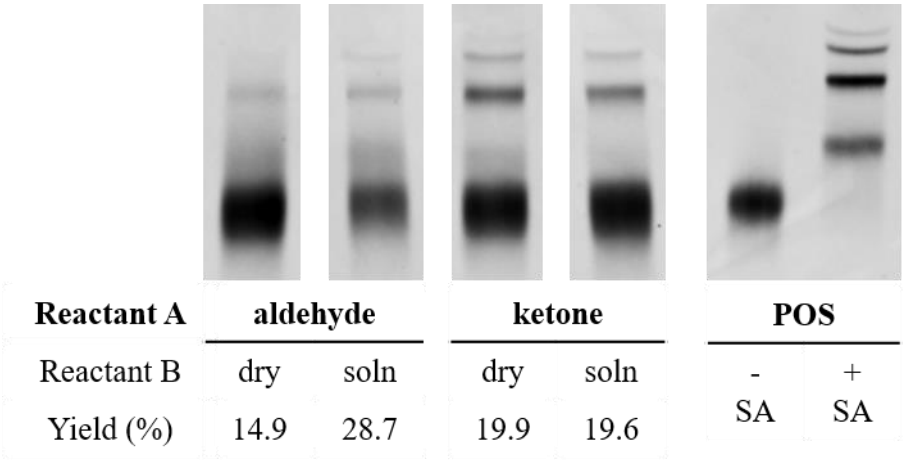



**Figure 3.12:** Aldol reaction order of addition experiment. 5% native PAGE. 90 min. Reactant A is covalently bound to the 3'-end of the PEGylated oligonucleotide (0.5 μM) and biotinylated reactant (40 mM) in MeCN. POS is the positive control biotinylated PEG-DNA.

*Biotinylated Aldehyde Reaction Concentration Optimization*

Reactant and catalyst tethered PEG-DNA (ketone-propro-PEG-48mer) was reacted with various concentrations of biotinylated aldehyde to determine optimal reaction conditions. The reaction was incubated at room temperature for four days in DMSO, followed by purification, streptavidin incubation, and visualization by native PAGE (Figure 3.13).

The gel image shows that with increasing concentration of biotinylated reagent the yield stays relatively constant with yields of 22%, 19%, and 18% for reaction mixtures of 0.5 μM ketone-propro-PEG-DNA with 10, 20, and 30 mM biotinylated reactant, respectively (Figure 3.13, Lanes 3, 4, and 5). It is important to note that while there are shifted bands that follow the 4-band pattern of streptavidin binding, their migration is different from the positive control. A possible explanation for this occurrence is attributed to the different sequences used for positive control and catalyst ssDNA.

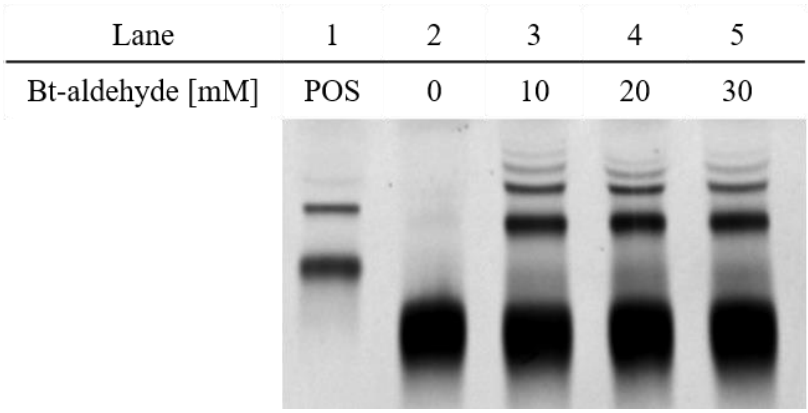
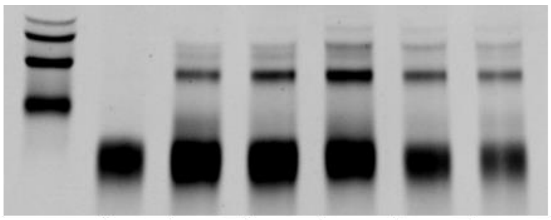


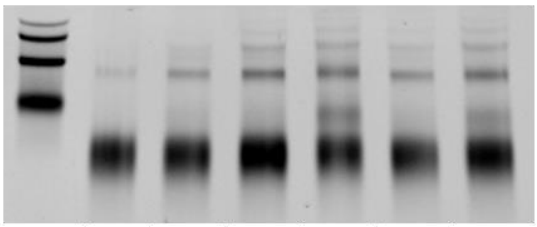

**Figure 3.13:** Biotinylated aldehyde concentration optimization – EMSA. Aldol Reaction performed was with ketone and diproline tethered PEG-DNA (0.5 μM) in MeCN.

The optimal concentration of biotinylated reactant in the aldol reaction was previously determined to be 10 mM. With this concentration, a time course experiment was set up to determine the length of time needed to produce an adequate yield of desired product. Catalyst-tethered PEG-oligo containing either a ketone or an aldehyde functionality was reacted with a biotinylated reactant in DCE for 1-5 days. Following the reaction, the crude product was purified, incubated with streptavidin, and visualized by 5% native PAGE (Figure 3.14).

These results indicate that the greatest yield of the aldol product is produced after three days. The reaction product can be seen after one day of incubation at room temperature and increases in yield until day 3. The yield slightly decreases on day 4 and increases again on day 5. Overall, both aldol product yields are most significant on day three with aldehyde-propro-PEG-oligo and ketone-propro-PEG-oligo incubated with biotinylated ketone or aldehyde, respectively. However, after five days of reaction intermediate or side product bands are at their lowest levels.



| (ketone) lane | 1(POS) | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| time (days) | - | 0 | 1 | 2 | 3 | 4 | 5 |
| yield (%) | | | 22.0 | 20.3 | 28.8 | 23.0 | 27.5 |



| (aldehyde) lane | 1(POS) | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| time (days) | - | 0 | 1 | 2 | 3 | 4 | 5 |
| yield (%) | | | 39.5 | 51.1 | 69.2 | 49 | 62.4 |

**Figure 3.14:** Aldol reaction time course experiment. Ketone-propro-PEG-oligo and biotinylated aldehyde in DCE. 5% native PAGE. 150 V. 90 min. EtBr stain.

**DNA-Encoded Catalyst Activity in Various Organic Solvents**

The catalytic activity of the DNA-encoded diproline catalyst was tested in various organic solvents and found to have the highest activity in 1.2-dichloroethane (Table 3.2).

**Table 3.2:** Solvent Screen for DNA-Encoded Catalyst Activity

| entry | solvent | yield [a] |
|---|---|---|
| 1 | DMSO | <5 % |
| 2 | DMF | <5% |
| 3 | $H_2O$ | 8% |
| 4 | 1,4-dioxane | 10% |
| 5 | MeOH | 27% |
| 6 | MeCN | 27% |
| 7 | DCE | 43% |

[a] ketone-catalyst DNA (0.5 μM), biotinylated benzaldehyde (500 μM), reactions were shaken for five days at room temperature. The yield was determined by PAGE.

*Catalysis Test with Optimized Reaction Conditions*

Catalyst-tethered reactant-PEG DNA was reacted with the biotinylated substrate for 5 days and analyzed by streptavidin-based EMSA. Visualization of the gel under UV light after staining with EtBr shows that the biotin reactant is reacting with the PEG-DNA, confirming the viability of these reagents to be used in the catalysis selection experiments.

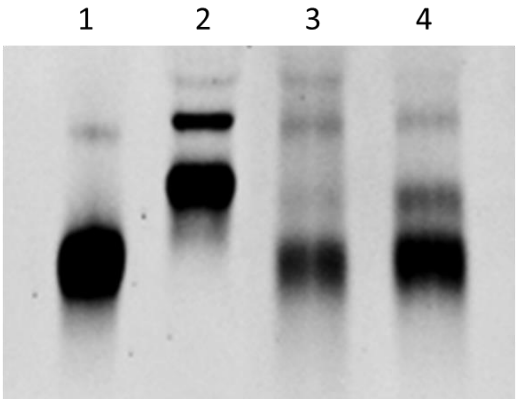
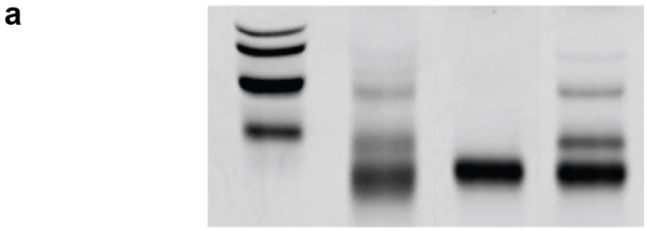


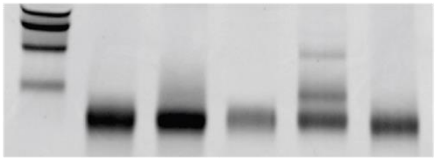

**Figure 3.14:** Aldol reaction catalysis with optimized conditions. 5% non-denaturing PAGE – 2hrs, 150 V. 1. Biotin-PEG DNA - no streptavidin. 2. Biotin-PEG DNA – streptavidin bound. 3. Aldehyde-ProPro-PEG DNA – aldol reaction. 4. Ketone-ProPro-PEG DNA – aldol reaction

***Controlling for Limitations***

Several issues that might diminish the enrichment of the known aldol catalyst during the selection are: i) DNA bases react to form stable covalent adducts with the biotinylated aldol reactants; ii) DNA catalyzes the aldol reaction and results in non-specific biotinylation of DNA; iii) the catalyst forms stable covalent adducts with the biotinylated aldol reactant; iv) inter-strand catalysis results in biotinylation of inactive library members. However, these issues are testable and controllable. To address issues *i-iii*, control experiments were designed to demonstrate that catalysis and bond-formation with biotinylated reactant only happen when the DNA molecule has both the catalyst and the aldol reactant attached (Figure 3.15).



| a | 1 (POS) | 2 | 3 | 4 |
|---|---|---|---|---|
| catalyst on DNA | - | + | - | - |
| ketone on DNA | - | + | + | + |
| biotin-aldehyde | - | + | + | + |
| free catalyst | - | - | - | + |
| streptavidin | + | + | + | + |



| b | 1 (POS) | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| catalyst on DNA | - | + | + | + | + | + |
| ketone on DNA | - | - | - | + | + | + |
| biotin-aldehyde | - | + | + | + | + | - |
| streptavidin | + | - | + | - | + | + |

**Figure 3.15:** EMSA experiments demonstrating selective catalysis. a) Aldol reaction depends on the presence of a catalyst on DNA (Lane 2 vs. Lane 3). Aldol reaction is rescued by the addition of 1 mM pyrrolidine as a catalyst (Lane 4). b) EMSA shift requires attached substrate (Lane 3 vs. Lane 5) and requires aldol reactant in solution (Lane 5 vs. Lane 6). Reaction conditions: DNA (0.5 μM) was dissolved in of DCE. Biotinylated benzaldehyde (500 μM) was added, and the reaction was shaken for 5 days at room temperature; pyrrolidine (1 mM) was added when applicable. POS = positive control, biotinylated DNA.

Issues *i* and *ii* examine the scenarios of DNA bases reacting to form stable covalent adducts with the biotinylated aldol reactants, and DNA catalyzing the aldol reaction resulting in non-specific biotinylation of DNA. Both these scenarios are addressed by a reaction in which the DNA-tethered catalyst was excluded from the ssDNA architecture (Figure 3.15a, lane 3). The catalyst can be confirmed to be present by the addition of an external aldol catalyst, pyrrolidine,

in solution (Figure 3.15a, lane 4). These scenarios can be excluded as reasoning for a positive catalysis result.

Issue *iii* addresses the possibility of the catalyst forming stable covalent adducts with the biotinylated aldol reactant. This scenario was tested by reacting the ssDNA catalyst without the presence of the reactant tethered to the DNA (Figure 3.15b. lane 3). As seen from the gel image, there is no visible biotinylated product which confirms that biotinylated reactant and catalyst interaction is not falsely displaying the positive catalysis result.

### 3.2.3 Friedel-Crafts Reaction

We envisioned that a single-stranded PEGylated DNA equipped with a 3′-linked acyl imidazole (**A3**) and Copper(II) bipyridine-modified nucleobase (**C2**), adjacent to the 3'-end, could be incubated with biotinylated indole (**B3**), and DNAs that form a covalent bond with **B3** could be isolated via streptavidin-coated magnetic beads (Figure 3.16 d).

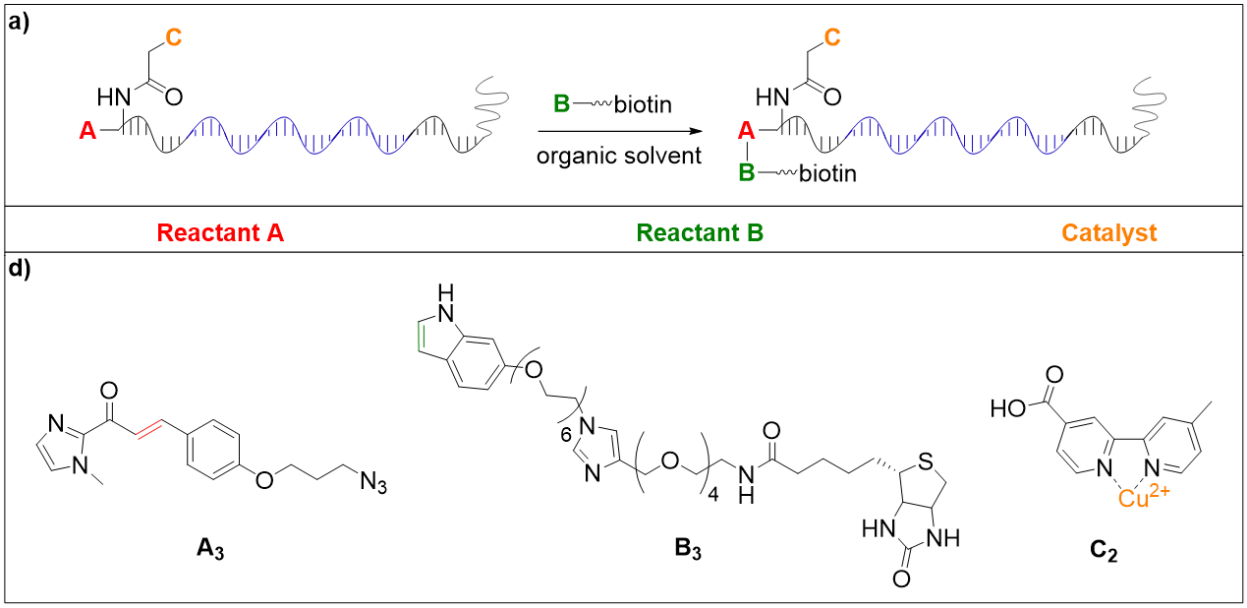
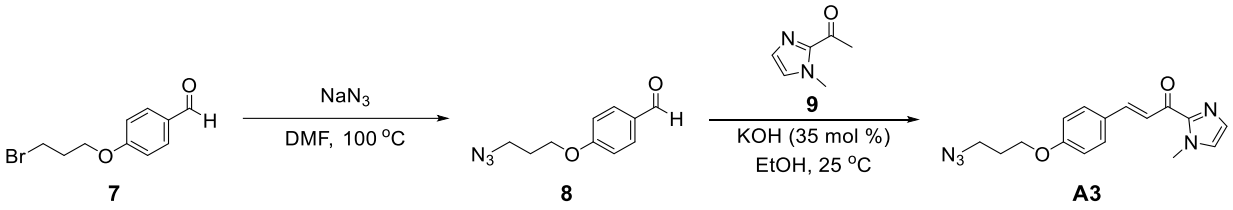


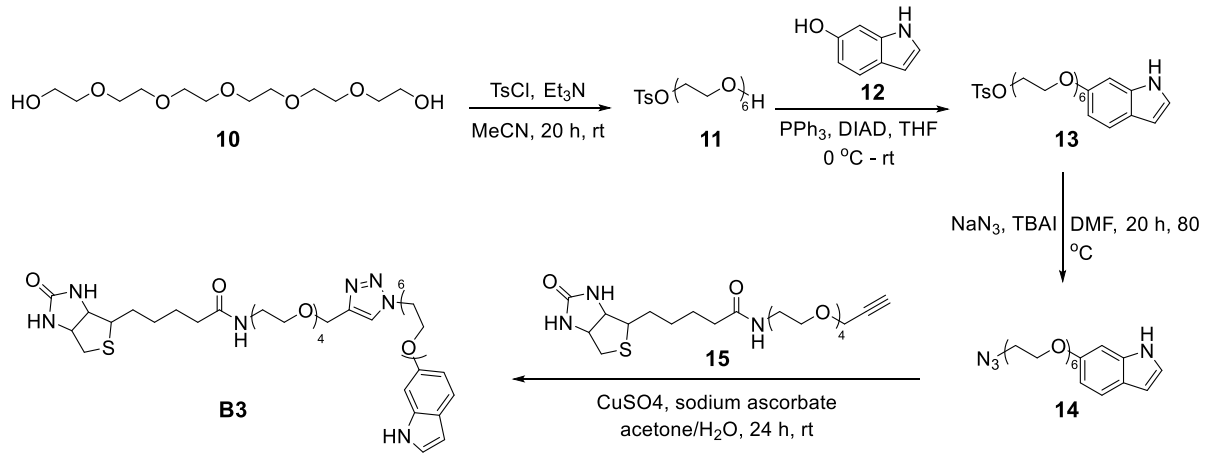

**Figure 3.16:** Aldol reactant, catalyst, and reaction product structures

Toward this end, we synthesized azido-amide acyl imidazole (**A3**) (Scheme 3.10) and conjugated this compound to a 5′-alkyne modified DNA by CuAAC reaction in the same manner as explained in the aldol reaction catalyst-encoded DNA. The conjugation product was purified by HPLC and confirmed by matrix-assisted laser desorption/ionization mass spectrometry. The biotinylated-PEG-indole (**B3**) was synthesized by Mitsunobu[17] reaction between the free alcohol

of a mono-tosylated hexamethylene glycol (**11**) and 5-hydroxy indole (**12**) in the presence of diisopropyl azodicarboxylate (DIAD) and triphenylphosphine ($PPh_3$). Following product isolation of the tosylated hexapeg indole (**13**), this compound was converted to the azide with sodium azide and tetrabutylammonium iodide (TBAI). Alkyne-Modified biotin (**15**) was conjugated to the azido-hexapeg-indole (**14**) using standard CuAAC conditions to afford the biotinylated indole (**B3**). (Scheme 3.11)




**Scheme 3.10:** Azidoamide acyl-imidazole synthesis (**A3**)




**Scheme 3.11:** Biotinylated indole synthesis (**B3**)

**Catalyst in Solution**

The Cu(II)-catalyzed Friedel-Crafts proof of concept reaction employs a 5'-PEGylated oligonucleotide modified with the aza-chalcone derivative on the 3' end. Copper(II) nitrate was complexed with a bipyridine ligand to use as the catalyst in solution. The biotinylated indole reactant serves as the nucleophile to the α,β-unsaturated ketone.

When using the bipyridine-copper nitrate as the active catalyst in this reaction, the biotinylated product is visualized in the streptavidin gel shift assay. We first employed this reaction in aqueous MOPS buffer to confirm the occurrence of carbon-carbon bond formation. The overall yield was 25 % in 24 hours at room temperature when the amount of catalyst introduced was 100 equivalents. When the amount of catalyst is only 30 mol%, no visible product can be seen (Figure not shown).

Following the PEGylation reaction, there were three visible peaks during HPLC purification. Due to the mass inconsistency and substantial molecular weight of the PEG moiety, it is impossible to characterize which fraction is the product of interest by mass spectrometry. Therefore, all three fractions were tested for reactivity.

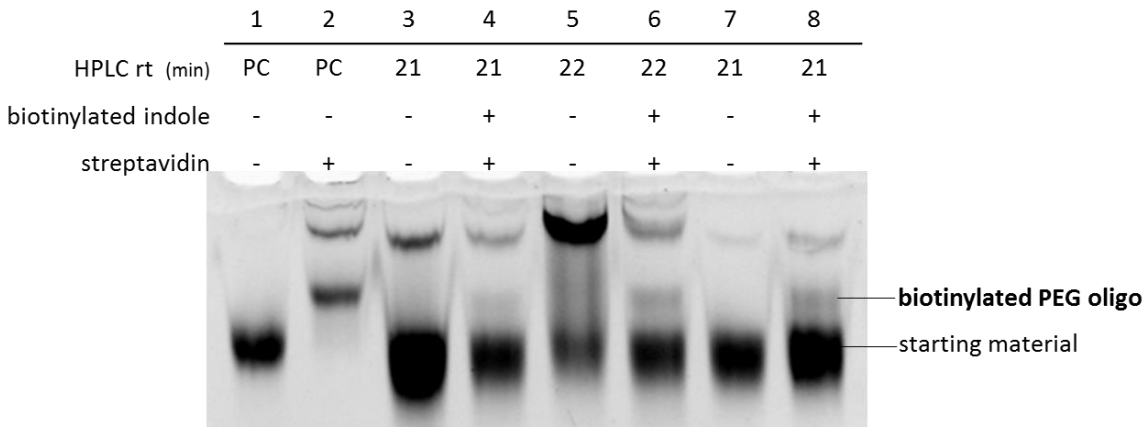
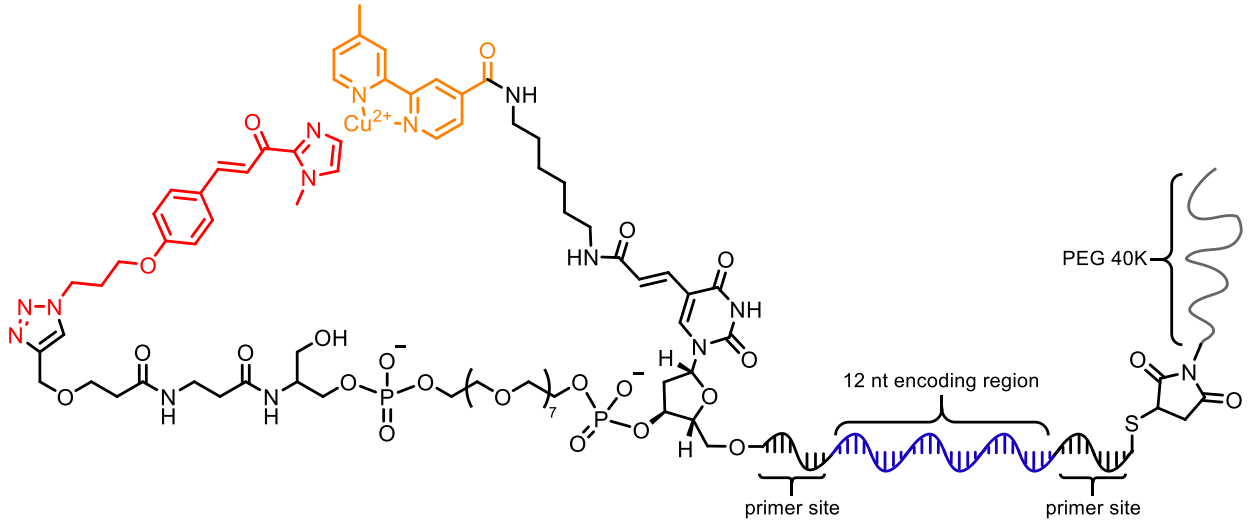


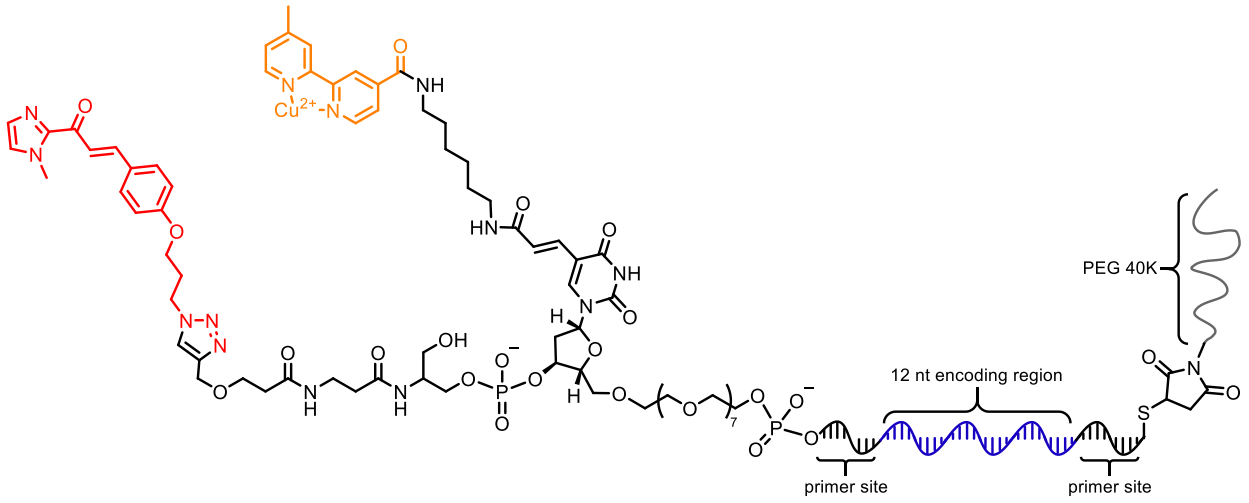

**Figure 3.16:** Positive result verification of FC reaction with the catalyst in solution. 1.5 μM PEG-DNA, 4 mM biotin-indole substrate, 150 μM $Cu(bpy)(NO_3)_2$, 20 mM MOPS, pH 6.5, 20% MeCN. 24 h, rt.

## Catalyst on DNA

Two different architectures were synthesized to examine the differences in reactivity between Architecture A when catalyst and reactant are separated by a PEG spacer, and Architecture B, where catalyst and reactant are separated from the DNA by the PEG spacer (Figure 3.17). Copper nitrate was chelated to the bipyridine moiety on the PEG DNA, to ensure there was no free copper in solution, and subsequently purified by Centri-sep column to remove all salts and small molecules. The initial attempt of the alkylation reaction with bpy-Cu(II) catalyst tethered to PEG-DNA was performed (Figure 3.18).




**Architecture A**




**Architecture B**

**Figure 3.17:** Molecular architectures of the Friedel-Crafts Cu(II) catalyst/ligand encoded PEG-DNA

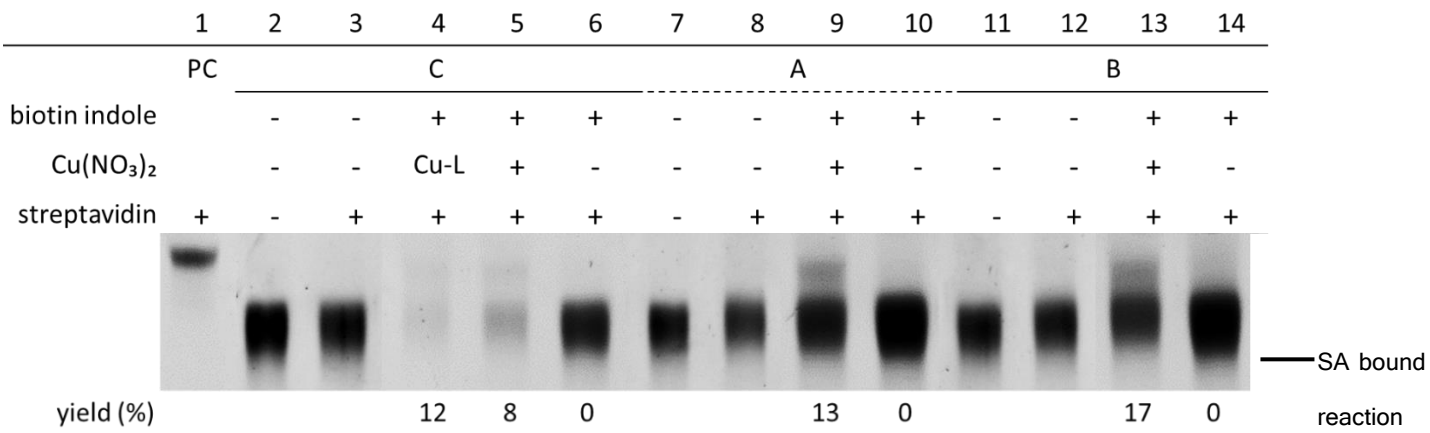



**Figure 3.18:** PEGylated Friedel-Crafts reaction and controls – EMSA
Lane 1- streptavidin bound positive control. Lanes 2/3 – architecture C (no ligand) oligo starting material. Lane 4 –copper ligand in solution. Lane 5 – reaction with copper nitrate (no ligand). Lane 6 – no Cu. Lanes 7/8 – Architecture A starting material oligo. Lane 9 – reaction with Cu(II). Lane 10 – without Cu(II). Lanes 11/12 – Architecture B was starting material oligo. Lane 13 – reaction with Cu(II). Lane 14 – without Cu(II). Reaction conditions: 1.0 μM oligo (20 pmol), 1.0 nmol $Cu(NO_3)_2$ (if present)/ 2.0 nmol Cu(bpy) (if present), 0.3 mM biotinylated indole (5.32 nmol), 20 mM MOPS buffer pH 6.5, 20 % MeCN, 25 h, rt.

The Friedel-Crafts reaction gel shift assay for PEGylated DNA shows that reaction product is only visualized when copper nitrate is present to coordinate to the bipyridine ligand covalently linked to the oligonucleotide (Lanes 9 and 13, Figure 3.18). Architecture C, with no catalyst linked to the DNA, shows that reaction product can be seen when both Cu-bpy and free copper nitrate are in solution (Lanes 4 and 5). It is important to note that only one streptavidin binding product band was analyzed, due to by-product formation that migrated alongside the upper three SA-bound bands. These were omitted from the gel image for clarity.

## 3.3 EXPERIMENTAL METHODS

### 3.3.1 Small Molecule Synthesis

**Aldol Reactants**

**Biotinylated ketone**

α-Angelica lactone (8.92 μL, 0.1 mmol) dissolved in 200 μL $CH_2Cl_2$ was cooled to 0 °C. Upon addition of Biotin-PEG-amine (59.5 mg, 0.1 mmol) the reaction was warmed to room temperature and stirred overnight. Crude reaction mixture was directly loaded onto silica gel column and eluted with 5-10% MeOH in $CH_2Cl_2$. Product fractions were collected and concentrated *in vacuo* to yield 20.8 mg of biotin-PEG-ketone as a yellow oil (30 % yield). $^1$H-NMR (400 MHz, $CDCl_3$): δ (ppm) 4.61 (m, 1H), 4.42 (m, 1H), 3.56-3.64 (m, 28H), 3.35-3.45 (m, 3H), 3.17-3.20 (m, 2H) 2.91-2.95 (m, 1H), 2.80-2.84 (m, 3H), 2.47 (t, 2H,), 2.29-2.32 (m, 2H), 2.18 (s, 3H), 1.6-1.8 (m, 4H), 1.45-1.48 (m, 2H). ESI-MS: $C_{31}H_{56}N_4O_{11}SCl^-$ $[M\text{-}Cl]^-$ calculated 727.34, found 727.3.

**Biotinylated aldehyde**

4-formyl benzoic acid (26.4 mg, 0.18 mmol, 2.1 eq.) and EDC-HCl (35.5 mg, 0.19 mmol, 2.2 eq.) were weighed into 10 mL round bottom flask and purged with nitrogen gas. To the biotin-PEG-amine (50 mg, 0.08 mmol) was added 500 μL of DMF/MeCN (1:1) and stirred until dissolved. The biotin-PEG-amine solution was slowly added to the reaction vial followed by 1.18 mL of DMF/MeCN for a total volume of 1.68mL. The reaction was stirred at room temperature for 3 hr. (reaction completeness monitored by TLC). The solvent was removed *in vacuo*; Crude product was dissolved in DCM and loaded onto silica gel column and eluted with 0-10% MeOH in $CH_2Cl_2$. Product fractions were collected and concentrated *in vacuo* to yield 20.90 mg (0.0288 mmol, 34% yield) of biotin-PEG-aldehyde as a yellow oil. Rf = 0.36 in 10%

MeOH/DCM. $^{1}$H-NMR (400 MHz, $CDCl_3$): δ (ppm) 10.06 (s, 1H), 7.92-7.99 (m, 4H), 7.15 (br. s, 1H), 3.36-3.67 (m, 36 H). ESI-MS: $C_{34}H_{54}N_4O_{11}$ $[M + Na]^+$ calculated 749.34, found 749.4.

**Azido-PEG-aldehyde**

Azido-PEG-amine (110 mg, 0.25 mmol, 0.5M) was added to 4-formyl benzoic acid (75 mg, 0.5 mmol, 2.0 eq.), N-hydroxysuccinimide (57.5 mg, 0.5 mmol, 2.0 eq.), and EDC (100.7 mg, 0.525 mmol, 2.1 eq.) in 0.5 mL DCM. Reaction mixture was stirred at room temperature for 5 hr., diluted with 5 mL of DCM and 5 mL of 1M HCl. Organic layer was separated, dried with magnesium sulfate, and concentrated on rotary evaporator. Mixture was purified by silica gel chromatography (DCM to 10% MeOH in DCM) to yield a yellow oil (25.64 mg, 19% yield). $^{1}$H-NMR (400 MHz, $CDCl_3$): δ (ppm) 10.06 (s, 1H), 7.92-7.99 (m, 4H), 7.15 (br. s, 1H), 3.36-3.67 (m, 36 H). ESI-MS: $C_{26}H_{42}N_4O_{10}$ $[M+H]^+$ calculated 571.29, found 573.2.

**Azido-PEG-ketone**

α-Angelica lactone (22.5 μL, 0.25 mmol) dissolved in 0.5 mL of $CH_2Cl_2$ was cooled to 0 °C. Azido-PEG-amine (0.25 mmol) was added to the solution and allowed to warm to room temperature. Reaction was stirred under nitrogen overnight. Crude reaction mixture directly loaded into silica gel column and purified using a solvent gradient of 0-10% MeOH in EtOAc to afford the product (60 mg, 44.7% yield). $^{1}$H-NMR (400 MHz, $CDCl_3$): δ (ppm) 6.26 (br. s, 1H), 3.37-3.68 (m, 36 H), 2.78 (t, 2H), 2.42-2.45 (t, 2H), 2.17 (s, 3H). ESI-MS: $C_{23}H_{44}N_4O_{10}$ $[M+H]^+$ calculated 537.31, found 537.2.

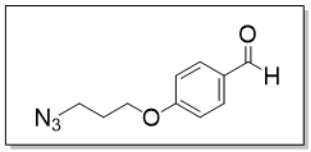
**Fmoc-proline NHS ester**

10 mg of N-hydroxysuccinimide was dissolved in DCM (350 μL, 0.25 M) along with 1.0 equiv. Fmoc-proline (29.31 mg, 86.9 μmol) and 2.0 equiv. of EDC (26.98 mg, 173.8 μmol). The reaction was stirred for 5 hrs. at room temperature under $N_2$ atmosphere. Reaction mixture was

diluted with DCM (1 mL), washed with 0.01 M HCL (5 mL), brine, dried with magnesium sulfate, and concentrated *in vacuo* to afford product as a white foam (35.57 mg, 94% yield). Product was used directly for coupling reaction.

**Friedel Crafts Reactants**

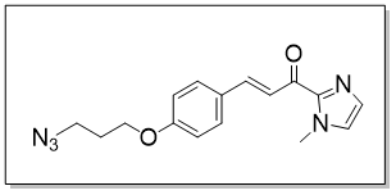
**4-(3-azidopropoxy)benzaldehyde**



A mixture of 4-(3-bromopropoxy)benzaldehyde (2.06 mmol, 0.5 g) and $NaN_3$ (5.15 mmol, 0.335 g) in 10 mL DMF was heated at 100 °C overnight under a nitrogen gas atmosphere. The mixture was poured into water and extracted with $CHCl_3$ several times. Removal of solvent gave the pure azide compound. Yellow oil (307 mg, 72.5%).

$^1$H NMR (400 MHz, $CDCl_3$): $\delta$ 9.88 (s, 1H), 7.83 (dt, 2H), 7.00 (dt, 2H), 4.13 (t, 2H), 3.54 (t, 2H), 2.05–2.12 (m, 2H). $^{13}$C NMR (400 MHz, $CDCl_3$): $\delta$ 190.74, 163.63, 131.99, 130.11, 114.72, 64.86, 48.02, 28.59.

**3-(4-(3-azidopropoxy)phenyl)-1-(1-methyl-1H-imidazol-2-yl)prop-2- en-1-one**

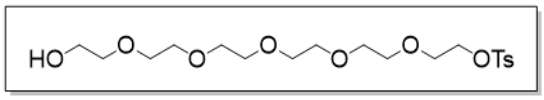


To a solution of 2-acetyl-1-methylimidazole (395.4 mg, 1.27 mmol) in EtOH (10.0 mL) was added the azido-benzaldehyde derivative (260.0 mg, 1.27 mmol) and a catalytic amount of KOH (25 mg). The solution was stirred for 20 h then transferred to a separatory funnel. The reaction mixture was concentrated, saturated NaCl (50.0 mL) was added and the mixture was extracted with EtOAc (3 x 30 mL). The combined organic extracts were dried over sodium sulfate, filtered, and concentrated in vacuo. The resulting residue was recrystallized in EtOH to provide the chalcone (280 mg, 71%) as a yellow solid.

$^{1}$H NMR (400 MHz, $CDCl_3$): δ 7.95 (d, J = 18.0 Hz, 1H), 7.78 (d, J = 18.0 Hz, 1H), 7.64 (d, J = 8.0 Hz, 2H), 7.20 (s, 1H), 7.06 (s, 1H), 6.9 (d, J = 8.0 Hz, 2H), 4.08 (m, 5H), 3.52 (t, J = 6.0 Hz, 2H), 2.06 (m, 2H). $^{13}$C NMR (400 MHz, $CDCl_3$): $\delta$ 180.45, 160.56, 144.05, 143.02, 130.42, 129.03, 127.83, 126.97, 120.31, 114.37, 64.51, 48.04, 36.24, 29.59, 28.59.

Mass Spectrometry: MS-ESI (m/z): Calcd for $C_{16}H_{17}N_5O_2$ $[M+H]^+$ , 312.14. Found 312.1.

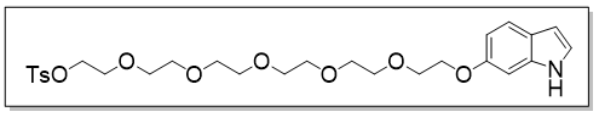
**Mono tosylated hexa PEG**




Hexaethylene glycol (1.0 g, 3.54 mmol) and triethylamine (0.493 mL, 3.54 mmol) were dissolved in acetonitrile (15 mL). Toluene –p-sulfonyl chloride (0.675 g, 3.54 mmol) was dissolved in acetonitrile (10 mL) and added dropwise over 20 min and kept stirring at room temperature for 24h. The solution was concentrated under vacuum and diluted with water and extracted with dichloromethane (50 mL). The organic layer was dried over $Na_2SO_4$ and then concentrated in vacuum. The crude product was purified by flash column chromatography using acetone: chloroform (30:70) as an eluent to provide a thick oil (486.9 mg, 31.5%).

Mass Spectrometry: MS-ESI (m/z): Calcd for $C_{19}H_{32}O_9S$ $[M+H]^+$ , 437.18. Found 437.1.

**Tosylated hexapeg-indole**

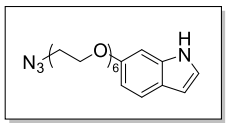



In a 25 mL round bottom flask $PPh_3$ (0. 290 g, 1.1 mmol) and DIAD (0.220 g, 1.1 mmol) were taken in dry THF (5.0 mL) under argon atmosphere and cooled to 0°C. To this solution 5-hydroxyindole (0.133 g, 1.0 mmol) and mono tosylated hexa PEG (0.436 g, 1.0 mmol) in THF (5.0 mL) was added slowly and stirred at room temperature for 24 h. The reaction mixture was concentrated under vacuum and purified by column chromatography using ethyl acetate: hexanes

(60: 40) then with methanol /dichloromethane (1: 99) on a second column to produce a yellow oil (163 mg, 18 % yield)

Mass Spectrometry: MS-ESI (m/z): calculated = 550.22 $(M-H)^-$. Found (-ESI): 550.0 (M-H)

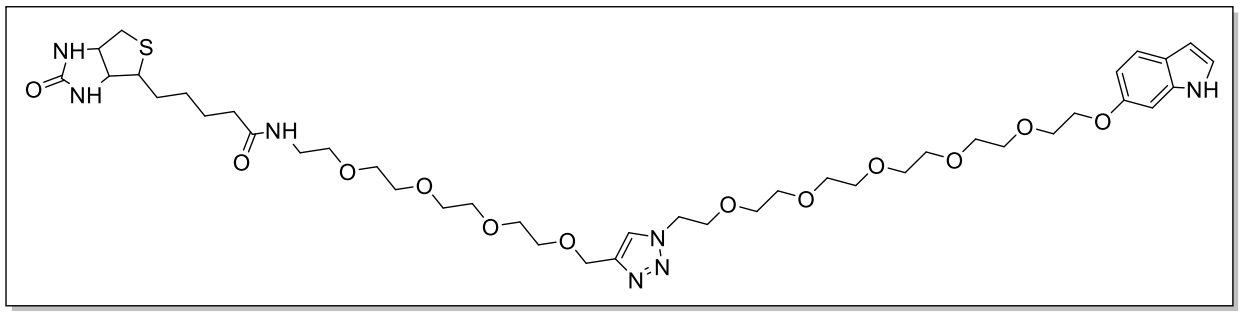
**Azido-hexapeg-indole**



Tosylated-HEG-indole(2) (0.225 mmol) was dissolved in 1.125 mL of anhydrous DMF and transferred to a dry 10 mL RBF. Sodium azide (1.125 mmol) and tetrabutylammonium iodide (TBAI) (8 mg) were added to the solution. The reaction was stirred for 20.5 at 80 ºC under inert gas. Crude mixture was washed with brine and extracted with EtOAc (2x15 mL), dried with sodium sulfate, and solvent removed under vacuum. Crude red oil was purified by column chromatography with DCM as eluent to provide a yellow oil (79.54 mg, 83.6% yield).

Mass Spectrometry: MS-ESI (m/z): calculated = 421.2 $(M-H)^-$. Found (-ESI): 421.1 (M-H)

**Biotinylated indole**



Azido-PEG-indole (55 mg, 0.13 mmol) and biotin-$PEG_4$-alkyne (59 mg, 0.130 mmol) were dissolved in acetone (565 µL). Copper sulfate tetrahydrate (3.25 mg, 0.013 mmol) was dissolved in 280 µL of water to make 12 mM solution and added to the reaction vial. Sodium ascorbate (5.15 mg, 0.026 mmol) was dissolved in 280 µL of water and added to the reaction last. The reaction was stirred for 24 h at room temperature under inert gas.

Mass Spectrometry: MS-ESI (m/z): calculated = 880.44 $(M+H)^+$. Found (+ESI): 880.4 (M+H)

### 3.3.2 Post-Synthetic DNA Modification

**Aldol Reaction DNA**

**Proline-proline coupling on amino-modified oligonucleotide**

Fmoc-pro-proline NHS ester (10 mg/ml, ~20 mM in DMF) was prepared fresh and added to amino-modified oligo (20 nmol, 500 μM) solution in 10 X conjugation buffer ($NaHCO_3/Na_2CO_3$, pH 9, 1 M). Mixture was vortexed and incubated at room temperature for 2 hr. Crude mixture was evaporated to dryness, purified by size exclusion column and evaporated to dryness again. Product was Fmoc deprotected in 20 % piperidine/DMF for 10 min at room temperature and evaporated to dryness, deprotection repeated. Product was purified by HPLC on Hydrosphere C18 column to yield the diprolinamide bound oligonucleotide (6.26 nmol, 31.3 % yield, rt = 17.8 min).

**Azido-ketone (4) or benzaldehyde (3) reactant click conjugation**

Alkyne oligo (25 nmol, 500 μM) was combined with 10 mM azido-PEG-ketone (1000 nmol, 40 eq.), THPTA/$CuSO_4$ (pre-chelated, 100 mM/50 mM, 625 nmol, 25 eq.), and sodium ascorbate (100 mM, 1250 nmol, 50 eq.) and rotated for 1hr at room temperature. Reaction mixture was filtered through gel column to remove salts and copper sulfate and purified by RP-HPLC on hydrosphere column (rt = 19.9 min) to produce 12.78 nmol (51.1% yield).

**PEGylation of 5'-disulfide oligonucleotides**

Disulfide oligo (11 nmol) was incubated with 100 mM DTT (125 μL) in sodium phosphate buffer (50 mM, pH 8.5) for 1 hr. Residual salts and DTT were removed by Princeton separation column preloaded with phosphate buffer (50 mM, pH 6.0). Activated sulfhydryl oligo was immediately combined with 40K PEG maleimide (130 eq., 20 mM in pH 6 phosphate buffer),

vortexed, and incubated at 4 °C overnight. Reaction mixture was purified by RP-HPLC on hydrosphere column (rt = 21 min) to produce 3.59 nmol (32.6 % yield).

**Friedel Crafts DNA**

**Bipyridine coupling to amine modified DNA**

Equal volumes of 4'-methyl-2,2'-bipyridine-4-carboxylic acid (900 mM, 33 μL), EDC HCl (900 mM, 33 μL), and N-hydroxysuccinimide (900 mM, 33 μL) in DMF were combined and shaken (1h, rt). The crude NHS ester was added in two portions (50 μL each) to a solution of amino-modified oligo (50 nmol, .5 mM) and 23 μL 10X conjugation buffer ($NaHCO_3/Na_2CO_3$, pH 9, 1M), The reaction was incubated at rt for 2h. Small molecules removed by Princeton Separations Centri-Sep Column, and purified by RP-HPLC

**CuCAAC reaction of azido-acylimidazole to alkyne oligo**

Alkyne oligo (10 nmol, 500 μM) was combined with 10 mM azido-acylimidazole (500 nmol, 50 eq.), THPTA/CuSO4 (pre-chelated, 100 mM/25 mM, 625 nmol, 25 eq.), and sodium ascorbate (100 mM, 500 nmol, 50 eq.) and rotated for 1hr at room temperature. Reaction mixture was filtered through gel column to remove salts and copper sulfate and purified by RP-HPLC on hydrosphere column (M: DELANEY-DMT-OFF-AT48, rt = 27 min).

### 3.3.3 Catalysis Experiments

**Aldol Reaction**

**Catalyst in solution-aldol reaction**

10 pmol of ketone-Ta1-PEG was evaporated to dryness. Biotinylated aldehyde (25 μL, 20 mM in DCE) was added to oligo and evaporated to dryness. 8 μL of MeCN or DMF was added to dried reactants, followed by pyrrolidine (2 μL, 50 mM). The reaction was incubated at 25 ºC

overnight, evaporated to dryness, dissolved in 50 μL of $H_2O$, and purified by Princeton separation column. Product was evaporated to dryness, dissolved in 2.8 μL of $H_2O$ and 2.2 μL of streptavidin in buffer (10 mg/mL). Mixture was incubated at room temperature for 45 minutes, 1 μL of 6X loading dye added, and subjected to 5% native PAGE at 150 V for 85 min. Gel was stained with EtBr and imaged.

**Catalysis on DNA-aldol reaction**

Five separate vials of aldehyde or ketone – propro-PEG-oligo (20 pmol each) were evaporated to dryness from aqueous solution and dissolved into 36 μL of organic solvent (MeCN, DCE, or dioxane) (0.5 μM). Each solvent was used to make a 20 mM solution of the biotinylated ketone and aldehyde and 4,000 equivalents were added to the oligo-organic solvent solution (4 μL, 80 nmol, 2.0 mM). The positive catalyst control was performed in the same manner stated above, except 28 μL of solvent was used to dissolve the oligo in addition to pyrrolidine (800 nmol, 8 μL, 100 mM in MeCN) and the same amount of biotinylated reactant. Reactions were incubated at 25 ºC for 5 days, evaporated to dryness, dissolved in water, and excess biotinylated reactant was removed by Princeton gel separation column. Product was evaporated to dryness incubated with streptavidin (2.22 μL, 2 eq., 18 μM in buffer) for 30 min at room temperature. Mixtures were diluted to 5 μL with water and 1.0 μL 6X loading dye, and run for 85 minutes on a 5% TBE non-denaturing PAGE, stained with EtBr, and imaged.

**Catalyst on DNA – solvent optimization**

Ketone-propro-Ta_EcoRV-PEG (10 pmol) was evaporated to dryness in speed vacuum. Biotinylated aldehyde (30 μL, 20 mM in DCE) was added to oligo and mixture was evaporated to dryness. Dried reactants were suspended in 20 μL of various anhydrous solvents, vortexed for 45 min, and incubated overnight at 25 ºC. The crude reaction mixture was evaporated to dryness,

purified by Princeton separation column, incubated with streptavidin in buffer, and visualized by PAGE.

**Reaction concentration optimization**

Ketone-propro-PEG-DNA (0.5 μM) and biotinylated aldehyde (10 mM, 20 mM, and 30 mM) were dissolved in DMSO and mildly vortexed at room temperature for 4 days. The reaction was evaporated to dryness, dissolved in 100 μL DI water, and purified by Princeton gel separation column for removal of excess biotinylated aldehyde. The product was evaporated to dryness and incubated with streptavidin (2.22 μL, 2 eq., 18 μM in buffer) for 30 min at room temperature. The streptavidin binding mixture was mixed with gel loading dye and run with 5% native PAGE, stained with EtBr and imaged.

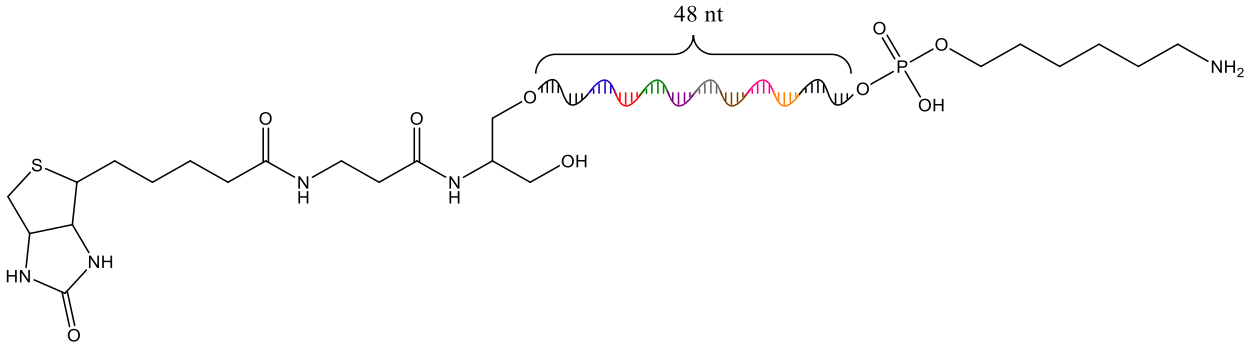
**Time course experiment**

Five separate tubes with 10 pmol of ketone or aldehyde-pro-pro-Ta_EcoRV-PEG oligo each were evaporated to dryness in speed vacuum. 10 μL of 40 mM biotinylated aldehyde or ketone in anhydrous DMSO was added to the dried oligo. The solutions were vortexed for 30 minutes and diluted to 0.5 μM with DMSO for a total volume of 20 μL. The reactions were vortexed at room temperature for 1-5 days. The sample was evaporated to dryness on speed vacuum, dissolved in 50 μL of water and purified by Princeton column. Reactions of 1-4 days were evaporated to dryness after purification and stored at -20 ºC. Purified oligo products were dissolved in 3.33 μL of streptavidin in buffer (10 mg/mL), diluted to 5 μL with water, and incubated at room temperature for 30 minutes. Loading dye (6X, 1 μL) was added to solutions and directly loaded onto gel. 5% non-denaturing PAGE was run at 150 V for 100 min, stained with EtBr, and UV-imaged.

**Order of addition experiment**

10 pmol of reactant-propro-PEG oligo was evaporated to dryness in speed vacuum from aqueous solution. Addition of the biotinylated reagent was followed by 1 of 2 methods: (a) 10 µL of 40 mM biotinylated aldehyde or ketone in anhydrous MeCN was added to the dried oligo. The solutions were vortexed for 30 minutes and diluted to 0.5 µM with MeCN for a total volume of 20 µL. Or (b) 10 µL of MeCN was added to dry oligo and vortexed for 30 minutes. Biotinylated aldehyde (10 µL, 40 mM) was added to dissolved oligo. Reaction mixtures were vortexed for 5 days at room temperature. Crude products were evaporated to dryness in vacuo, dissolved in 50 µL of water, and purified by Princeton column. Purified oligo products were dissolved in 3.33 µL of streptavidin in buffer (10 mg/mL), diluted to 5 µL with water, and incubated at room temperature for 30 minutes. Loading dye (6X, 1 µL) was added to the solutions and directly loaded onto gel. 5% non-denaturing PAGE was run at 150 V for 100 min, stained with EtBr, and UV-imaged.

**BtTa1N – Synthesis of 3'-biotin-5'-amino 48-mer**




Sequence: 5'-$C_6NH_2$-CGT ACG GTC GAC GCT AGC ATG TCC AGT TAG CAC GTG GAG CTC GGA TCC-Bt-3'

48nt oligomer was synthesized on 3'-protected biotin serinol CPG universal support (1 µmol). Resin and 5'-C6amino modifier were both purchased from Glen Research. The synthesis was performed with the final DMT-on. Resin-bound oligo was transferred to an Eppendorf tube,

suspended in AMA (400 μL) and incubated at 65 °C for 10 min and cooled to room temperature. 22 mg of TRIS base was dissolved in supernatant subsequent to filtration from resin (to protect MMT detritylation). Solution was concentrated *in vacuo* to remove volatile base. Crude product was purified by RP-HPLC (DMT-ON method). Product fractions were collected, frozen, and lyophilized overnight. Detritylation was performed in 20% glacial acetic acid at rt for 1 hr. Product was lyophilized and purified by RP-HPLC (retention time – 4.3 min). Product fractions were collected, frozen, and lyophilized to afford 23.4 nmol (**2.35 % yield**). MS Calcd 15,415.87 found-15,415.9.

## 3.4 REFERENCES


1. Hook, K. D.; Chambers, J. T.; Hili, R., A platform for high-throughput screening of DNA-encoded catalyst libraries in organic solvents. *Chemical Science* **2017,** *8* (10), 7072-7076.

2. Northrup, A. B.; MacMillan, D. W. C., The First Direct and Enantioselective Cross-Aldol Reaction of Aldehydes. *Journal of the American Chemical Society* **2002,** *124* (24), 6798-6799.

3. Groves, J. K., The Friedel–Crafts acylation of alkenes. *Chemical Society Reviews* **1972,** *1* (1), 73-97.

4. Chaiet, L.; Wolf, F. J., The properties of streptavidin, a biotin-binding protein produced by Streptomycetes. *Archives of Biochemistry and Biophysics* **1964,** *106*, 1-5.

5. Green, N. M., Avidin. In *Advances in Protein Chemistry*, Anfinsen, C. B.; Edsall, J. T.; Richards, F. M., Eds. Academic Press: 1975; Vol. 29, pp 85-133.

6. Langer, P. R.; Waldrop, A. A.; Ward, D. C., Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proceedings of the National Academy of Sciences* **1981,** *78* (11), 6633-6637.

7. Leary, J. J.; Brigati, D. J.; Ward, D. C., Rapid and sensitive colorimetric method for visualizing biotin-labeled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: Bio-blots. *Proceedings of the National Academy of Sciences* **1983,** *80* (13), 4045-4049.

8. Davie, E. A. C.; Mennen, S. M.; Xu, Y.; Miller, S. J., Asymmetric Catalysis Mediated by Synthetic Peptides. *Chemical Reviews* **2007,** *107* (12), 5759-5812.

9. Tang, Z.; Marx, A., Proline-Modified DNA as Catalyst of the Aldol Reaction. *Angewandte Chemie International Edition* **2007,** *46* (38), 7297-7300.

10. Rik, P. M.; Gerard, R., Organic co-solvents in aqueous DNA-based asymmetric catalysis. *Organic and Biomolecular Chemistry* **2010,** *8* (6), 1387-1393.

11. Hein, J. E.; Fokin, V. V., Copper-catalyzed azide-alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper(I) acetylides. *Chem Soc Rev* **2010,** *39* (4), 1302-15.

12. Li, X.; Liu, D. R., Stereoselectivity in DNA-Templated Organic Synthesis and Its Origins. *Journal of the American Chemical Society* **2003,** *125* (34), 10188-10189.

13. McKay, Craig S.; Finn, M. G., Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. *Chemistry & Biology* **2014,** *21* (9), 1075-1101.
14. Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G., Labeling Live Cells by Copper-Catalyzed Alkyne−Azide Click Chemistry. *Bioconjugate Chemistry* **2010,** *21* (10), 1912-1916.

15. Zhang, H.; Chao, J.; Pan, D.; Liu, H.; Qiang, Y.; Liu, K.; Cui, C.; Chen, J.; Huang, Q.; Hu, J.; Wang, L.; Huang, W.; Shi, Y.; Fan, C., DNA origami-based shape IDs for single-molecule nanomechanical genotyping. *Nature Communications* **2017,** *8*, 14738.

16. Lowe, A. B., Thiol-ene "click" reactions and recent applications in polymer and materials synthesis. *Polymer Chemistry* **2010,** *1* (1), 17-36.

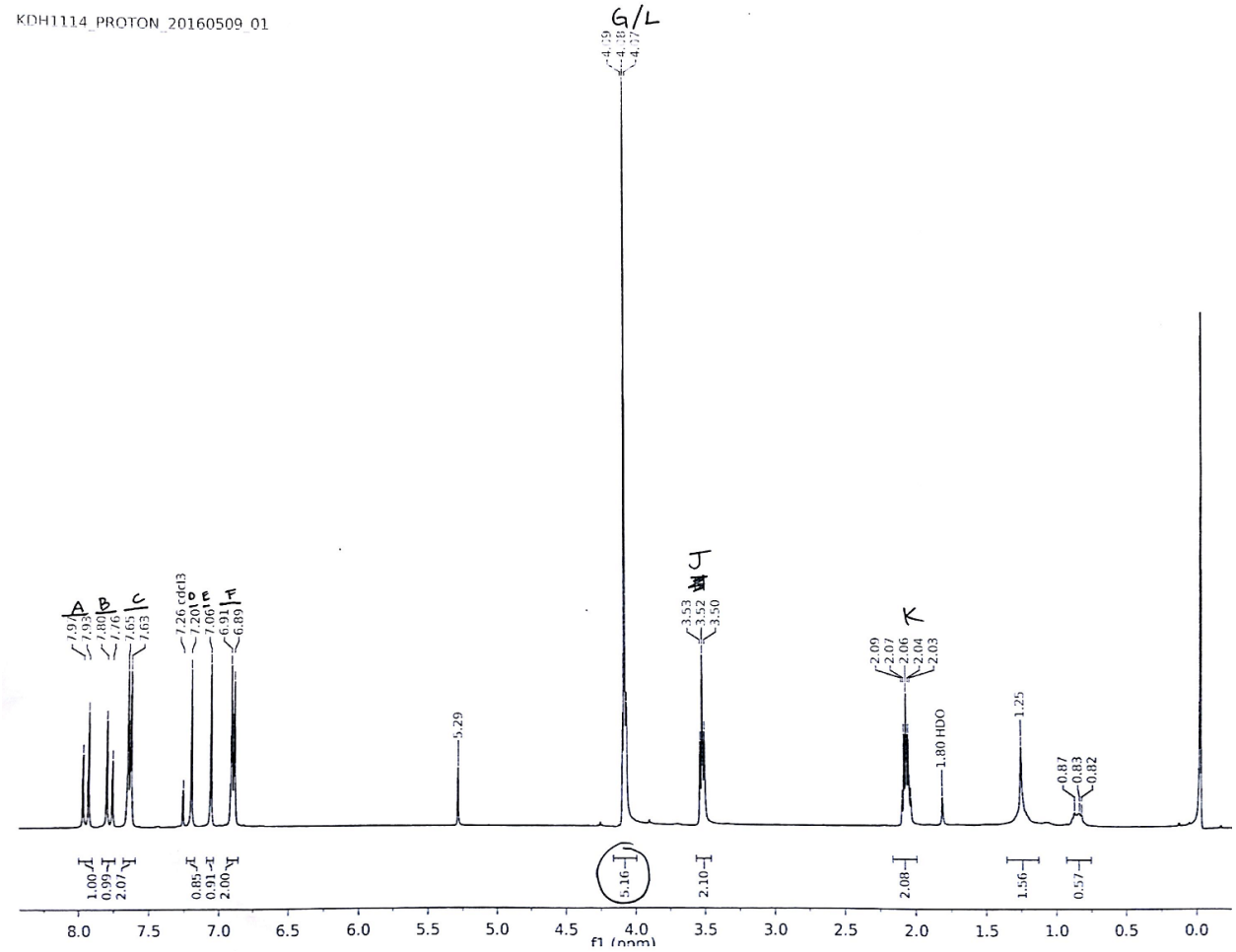
17. Oyo, M.; Masaaki, Y., Preparation of Esters of Carboxylic and Phosphoric Acid via Quaternary Phosphonium Salts. *Bulletin of the Chemical Society of Japan* **1967,** *40* (10), 2380-2382.

## S2. SUPPORTING INFORMATION

### S2.1 Selected Spectra

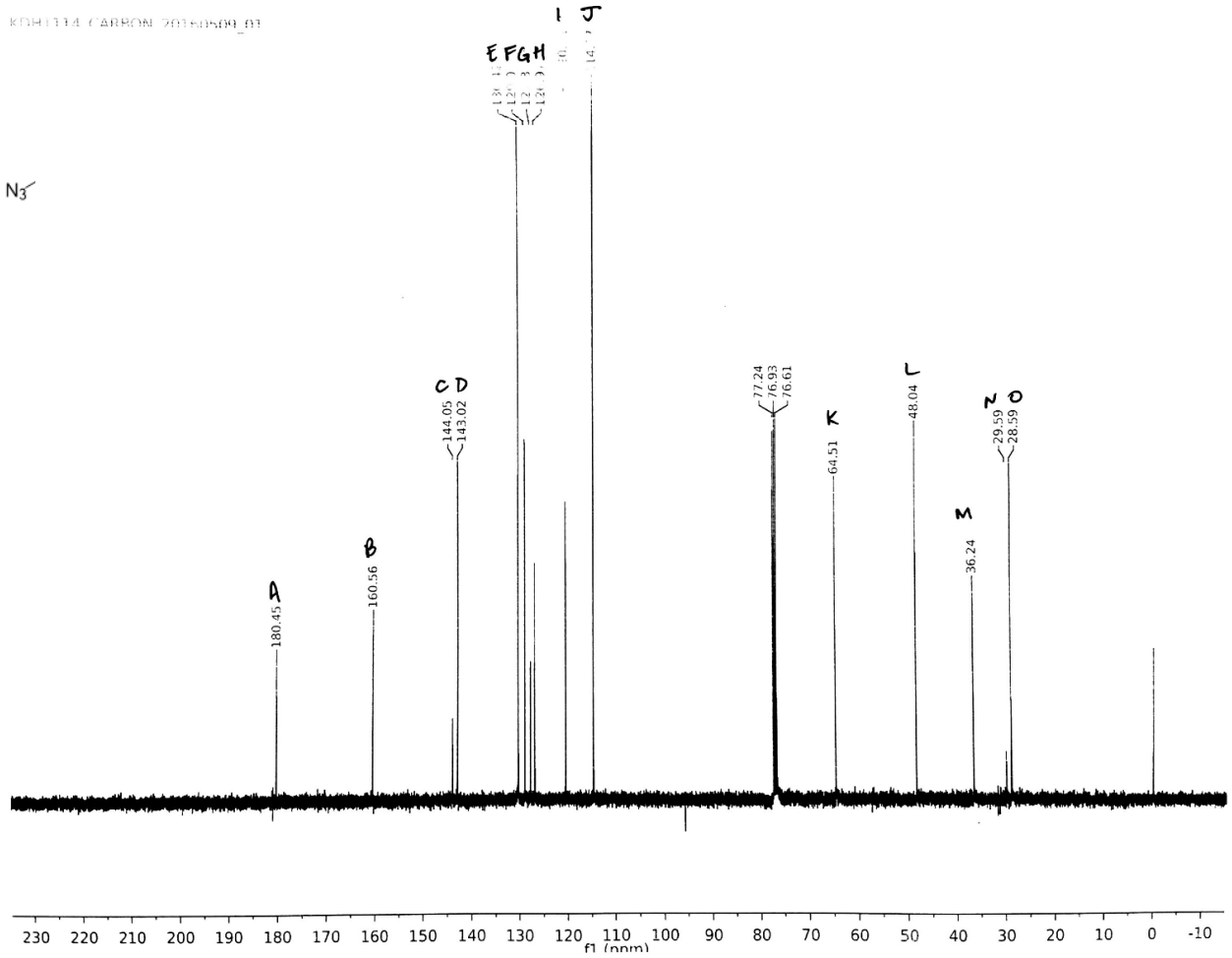
#### S2.1.1 $^{1}$H NMR of 3-(4-(3-azidopropoxy)phenyl)-1-(1-methyl-1H-imidazol-2-yl)prop-2-en-1-one (A3)

Solvent: $CDCl_3$

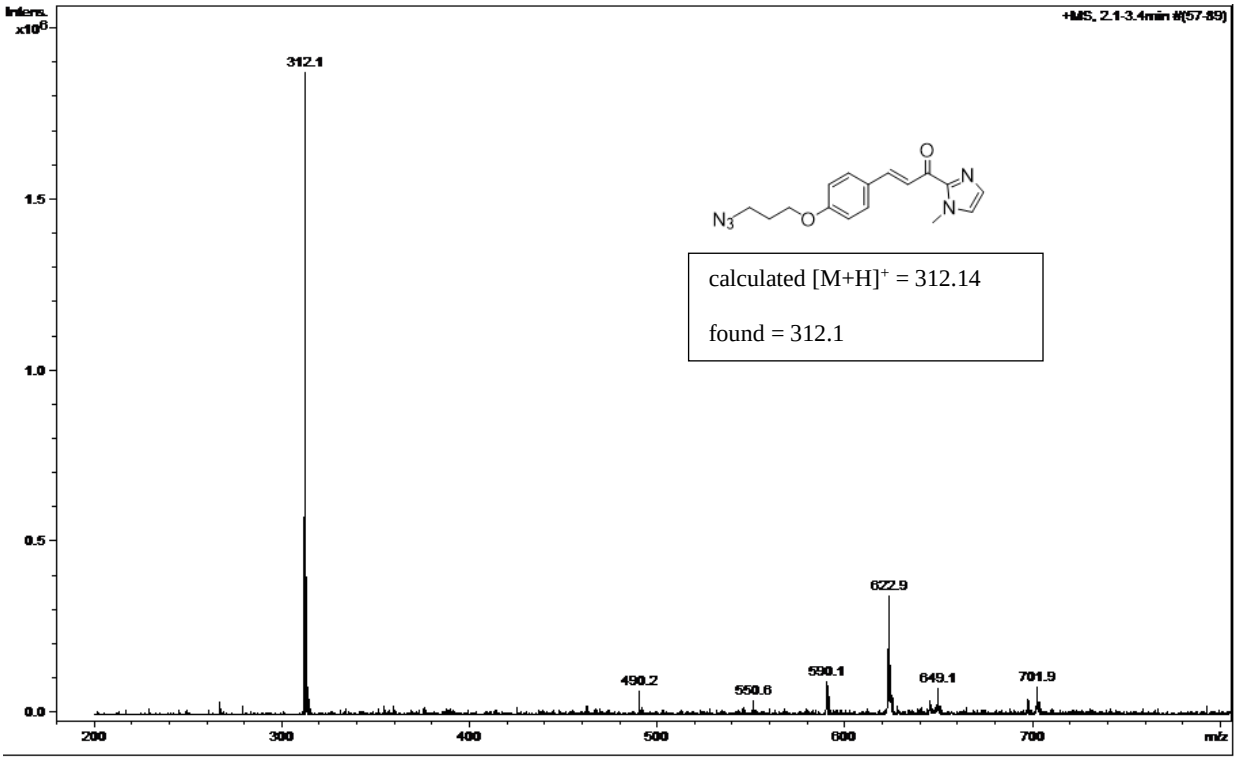


**S2.1.2 $^{13}$C NMR of 3-(4-(3-azidopropoxy)phenyl)-1-(1-methyl-1H-imidazol-2-yl)prop-2-en-1-one (A3)**

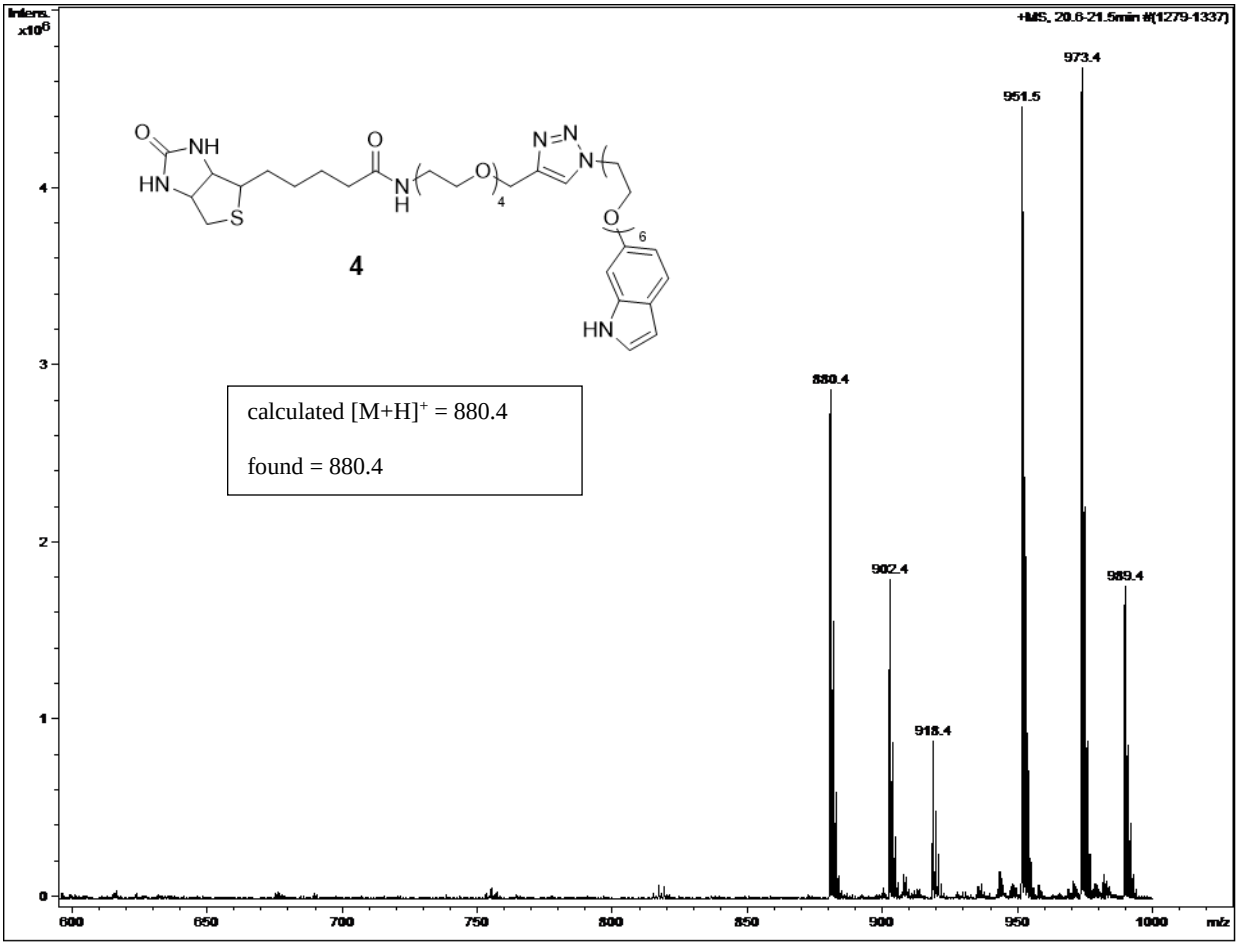
Solvent: $CDCl_3$



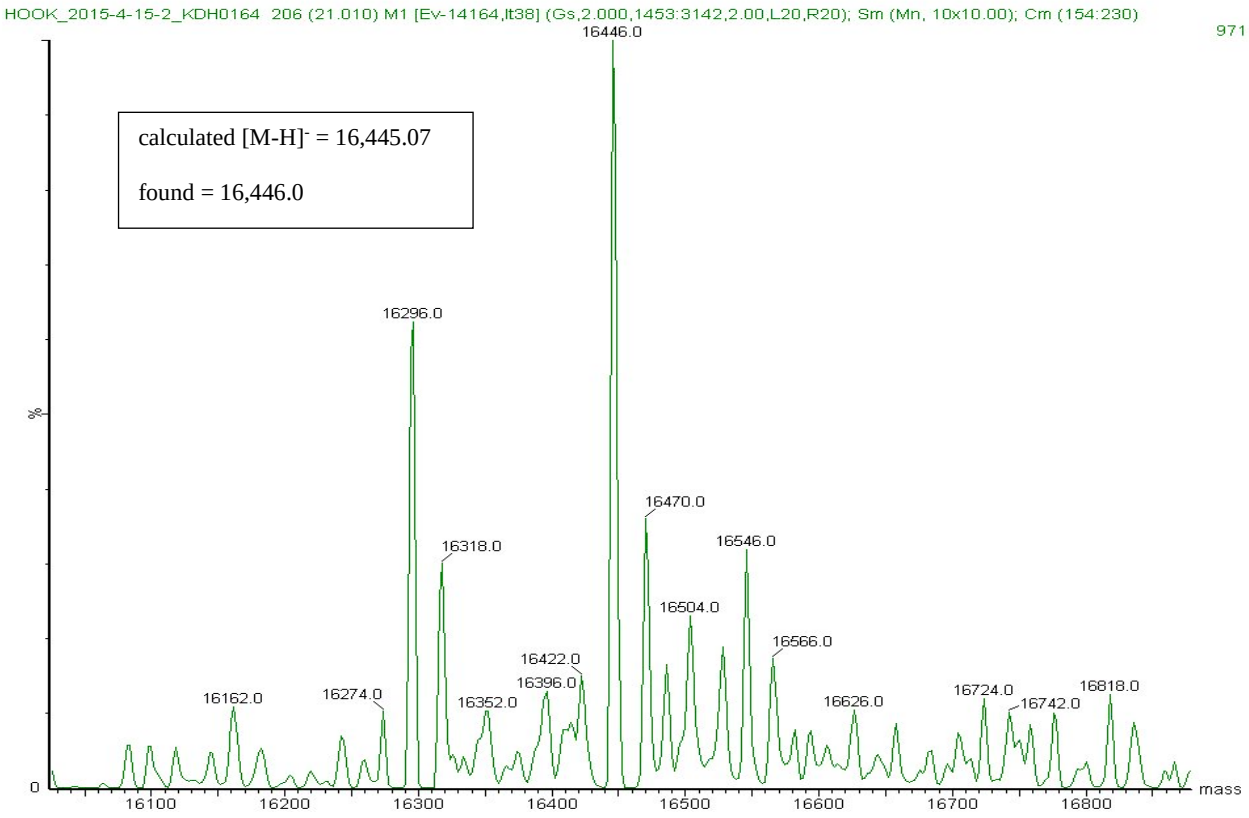
**S2.1.3 (+)ESI MS of 3-(4-(3-azidopropoxy)phenyl)-1-(1-methyl-1H-imidazol-2-yl)prop-2-en-1-one**



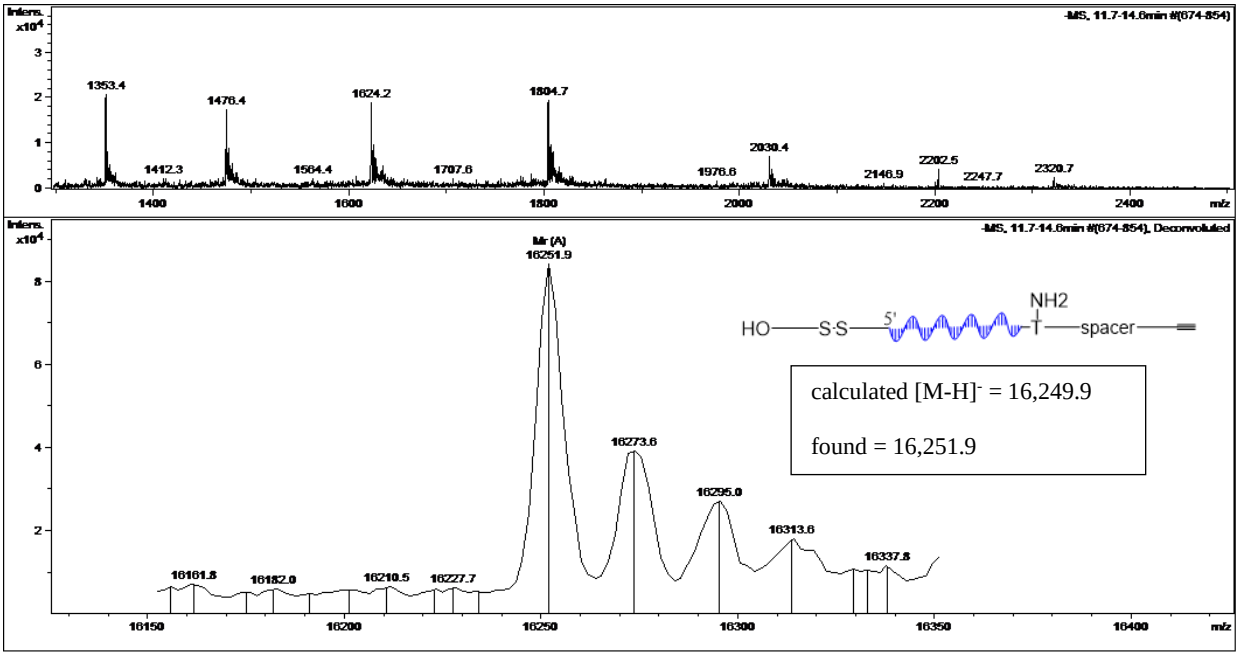
## S2.1.4 (+)ESI MS of biotin-PEG-indole



## S2.1.4 LC-ESI-MS - Alkyne-ProPro-Ta1-Disulfide (DNA-1)

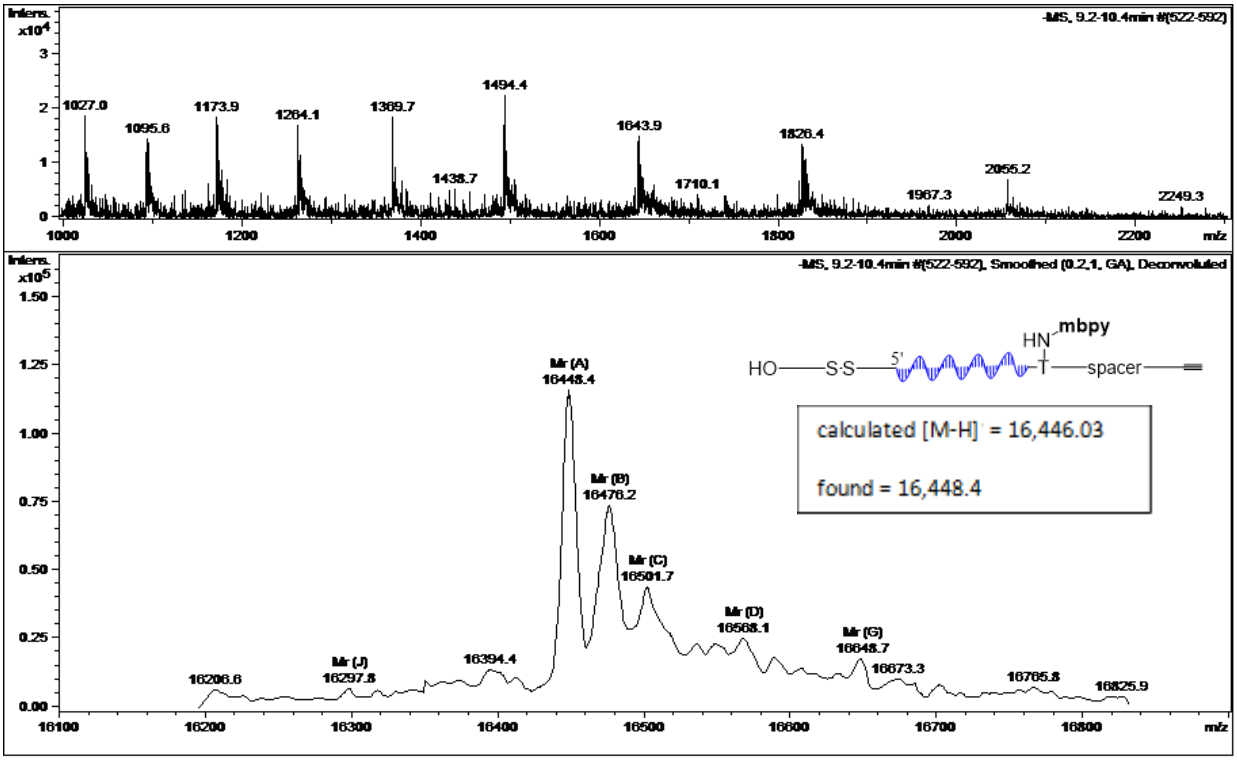


## S2.1.5 (-) ESI MS



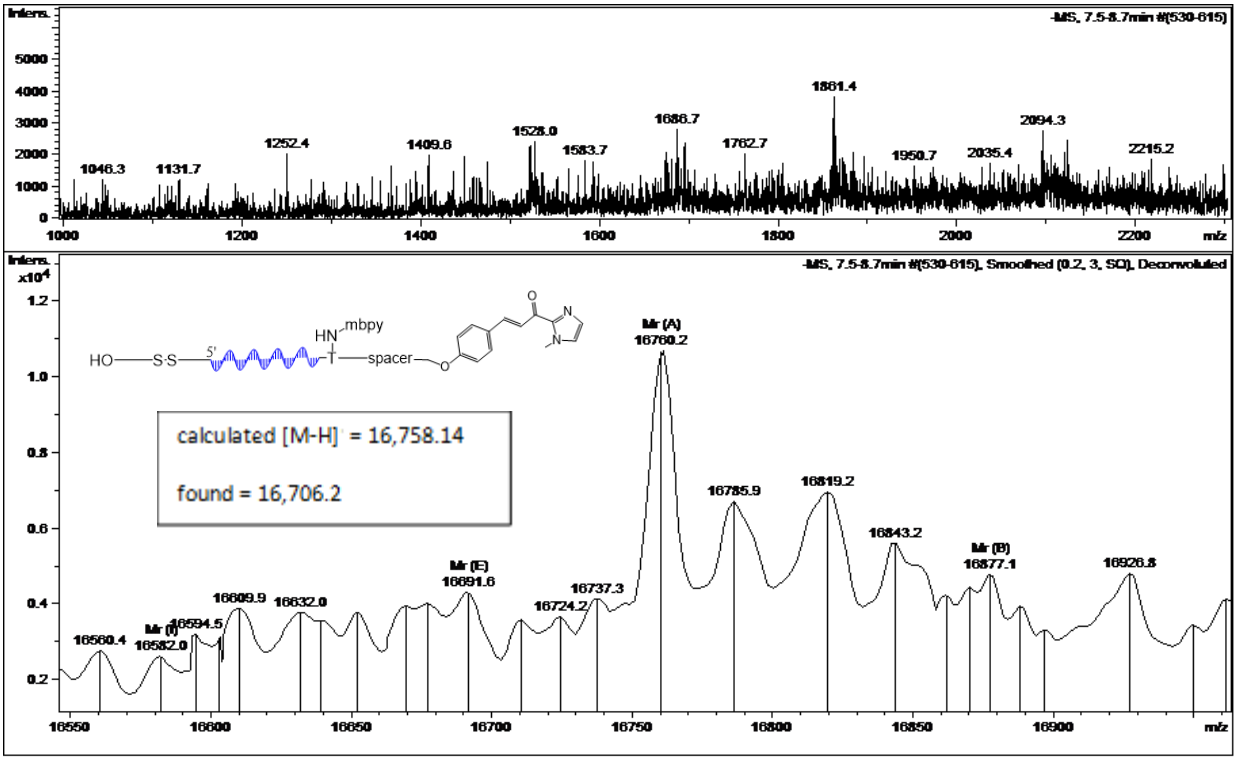

*Architecture A*

## S2.1.6 (-)ESI MS

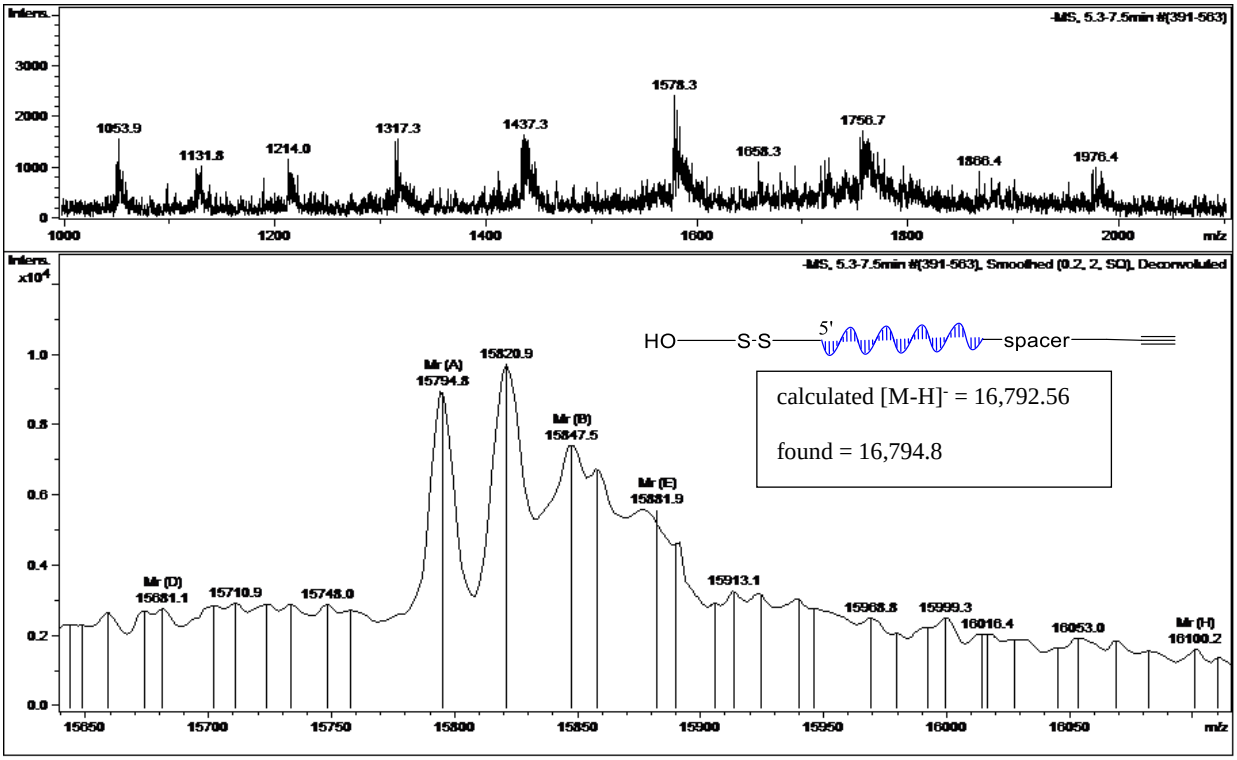



*Architecture A*

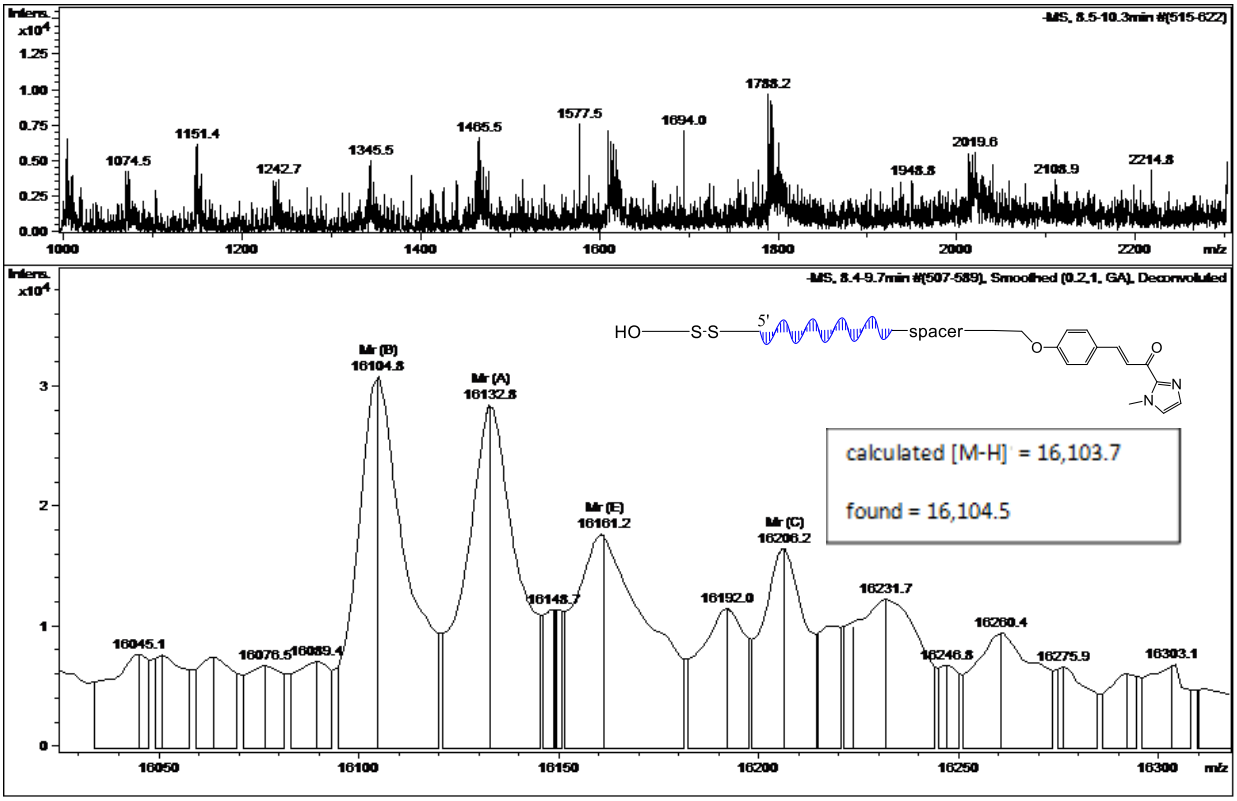
## S2.1.7 (-)ESI MS of acyl imidazole-sp18-bpy-TaEcoRV-disulfide



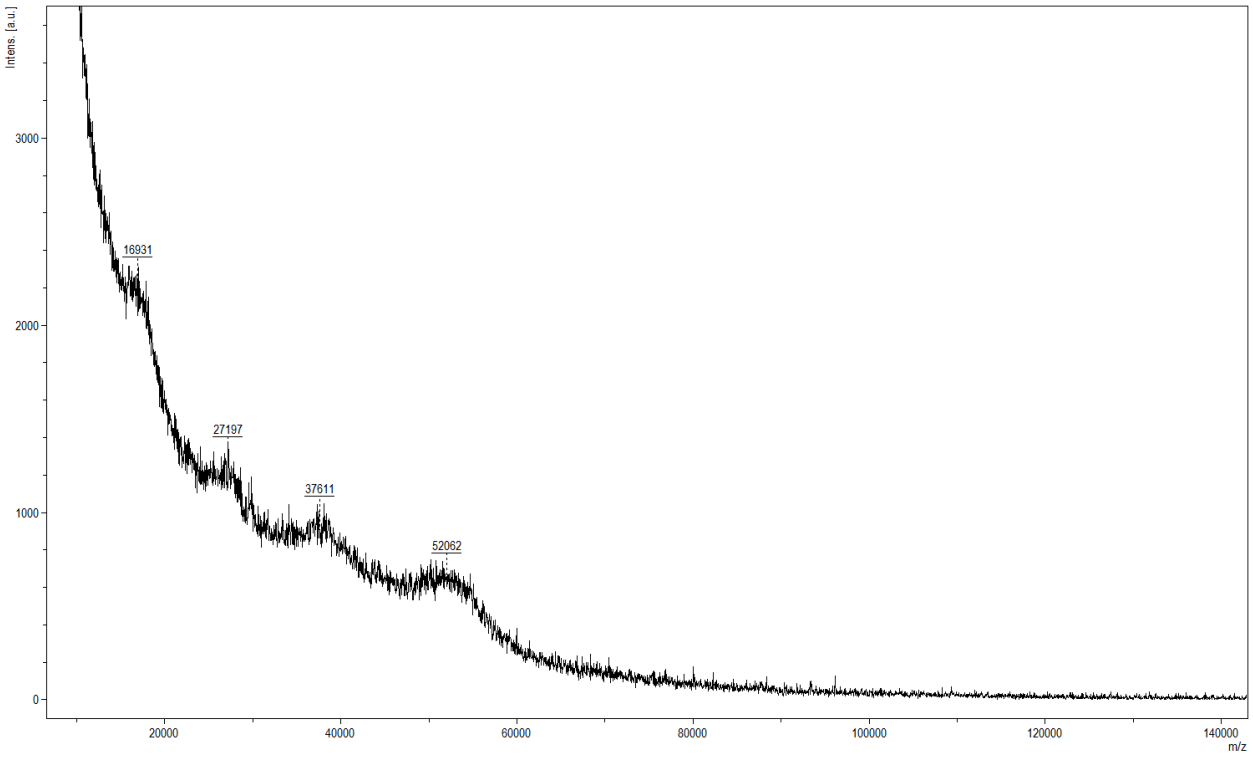
## S2.1.8 (-)ESI MS of alkyne-TaEcoRV-disulfide (no ligand control)



## S2.1.9 (-)ESI MS acyl imidazol-TaEcoRV-disulfide



**S2.1.9 MALDI of KDH1130- alkyne-sp18-bpy-TaEcoRV-PEG40K** (no reactant control)



CHAPTER 4

# MOCK SELECTION OF A PEGYLATED DNA ENCODED CATALYST IN ORGANIC SOLVENT

## 4.1 INTRODUCTION

We next sought to determine if this platform would enable the selective enrichment of active DNA-encoded catalysts from a large library of DNA-encoded molecules. As previously discussed in Chapter 3, there were several issues that might diminish the enrichment of the known aldol catalyst during the selection, including: (*i*) DNA bases non-specifically react to form stable covalent adducts with the biotinylated aldol reactant; (*ii*) DNA catalyzes the aldol reaction and results in non-specific biotinylation of DNA; (*iii*) the catalyst forms stable covalent adducts with the biotinylated aldol reactant; and (*iv*) inter-strand catalysis results in biotinylation of inactive library members by an active member. Issues *i-iii* were addressed with control experiments which demonstrated that catalysis and bond-formation with biotinylated reactant occurs only when the DNA molecule has both the catalyst and the aldol reactant attached (See Chapter 3).

To address issue *iv*, and potentially the efficacy of this catalyst discovery platform, we designed a selection system with a restriction digest-based readout to assess the enrichment of the catalyst. We implemented a selection pressure whereby survival of a library member required its ability to catalyze the aldol reaction. As a model selection, we sought to enrich the diproline positive control sequence from a large library of uncompetitive members. Each member of the

library contained a ketone reactant at the 3'-end. Importantly, the diproline positive control contained an EcoRV restriction digest site within its encoding region, which enabled monitoring of its enrichment after the selection round by restriction digest and PAGE analysis (Figure 4.1).

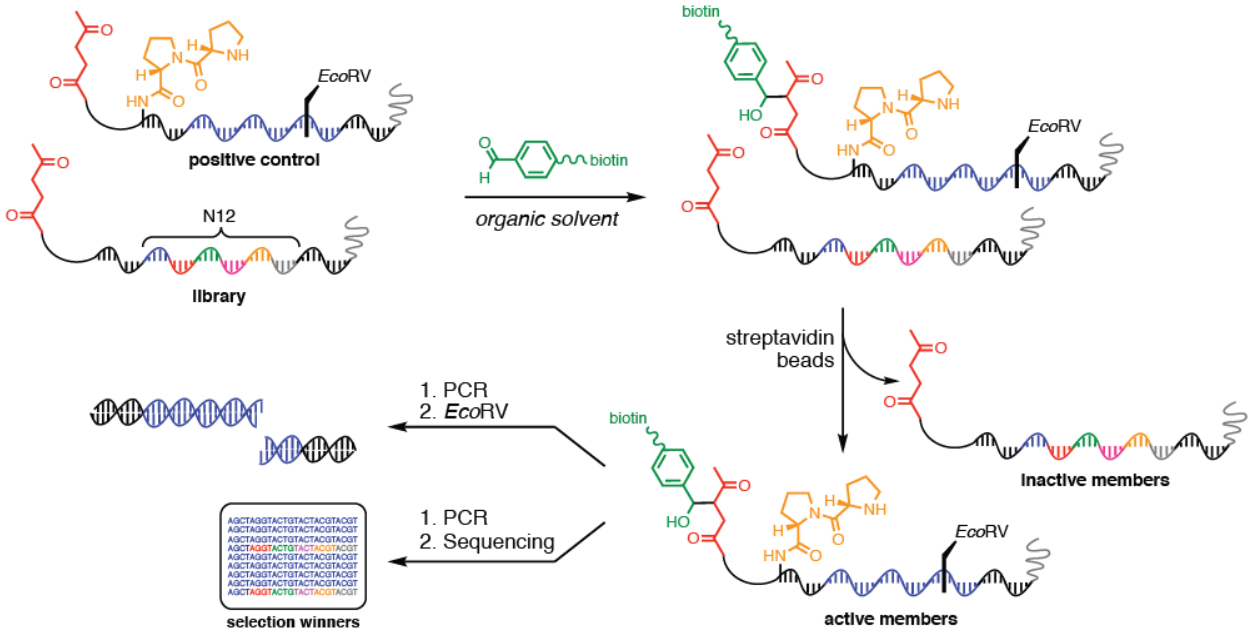



**Figure 4.1:** General scheme for in vitro mock selection of a DNA-encoded aldol catalyst

## 4.2 RESULTS AND DISCUSSION

### 4.2.1 Selection Procedure Overview

The first step of the selection system protocol begins by reacting a mixed PEGylated DNA library with the biotinylated reactant in organic solvent. The catalysis reaction step is performed with the optimized conditions concluded in Chapter 3; however, a 'mixed library' of sequences are present as starting materials in the reaction. The 'mixed library' contains the catalyst-tethered 'positive control' oligonucleotide held at a specific ratio to a 'N12 library' of

PEGylated DNAs that are functionalized with the substrate, *Reactant A*, but do not display a catalyst. It was previously shown that the presence of the diproline catalyst was necessary for catalytic bonds formation in the aldol reaction with the PEGylated DNA catalyst architecture; therefore, the PEGylated ssDNA 'N12 library' are essentially non-competitive with the 'positive control' in the reaction mixture due to the absence of a catalyst tethered to the randomized sequence encoded ssDNA.

The 'positive control' oligonucleotide, **Ta_EcoRV**, which displays the aldol catalyst proximal to the a 3'-end tethered ketone (*Reactant A*), includes a uniquely identifiable 12 nt encoding sequence (Table 4.1). The reactant-only PEGylated ssDNA 'N12 library' sequence, **Ta_N12**, is composed of randomized 12N encoding sequences. 'N' is defined as all four of the canonical DNA nucleotides (N = A,T,G,C); therefore, the Ta_N12 library contains $4^{12}$ or about 16.7 million different sequences in total. Both oligonucleotides contain the same forward and reverse primer binding sequences, flanking the encoding sequence, for later amplification of the selection survivors by polymerase chain reaction (PCR).

**Table 4.1:** DNA Sequences (5'→3') [a]

| | |
|---|---|
| **Ta_EcoRV** (positive control) | *CGTACGGTCGACGCTAGC* **TGGATATCACTG** *CACGTGGAGCTCGGATCC* |
| **Ta_N12** (N12 library) | *CGTACGGTCGACGCTAGC* **NNNNNNNNNNNN** *CACGTGGAGCTCGGATCC* |
| **Ta_XbaI** (biotin selection control) | *CGTACGGTCGACGCTAGC* **TCTAGAAACAAC** *CACGTGGAGCTCGGATCC* |

[a] *Italicized* sequences correspond to the forward and reverse 18 nt primer binding regions. The 12 nt encoding sequence is displayed in **bold** text. The entire sequence is 48 nt in length.

The PEGylated ssDNA 'N12 library' was synthesized with a 3'-end alkyne modification and functionalized with *Reactant A* by CuAAC reaction with the azido-modified substrate. A 5'-end disulfide modification was incorporated into the DNA architecture and post-synthetic modification was performed by thiol-Michael addition of the 40,000 MW PEG NHS-ester (Figure 4.2). Randomized encoding region sequences were synthesized *en masse* by adding an equal mixture of each A, T, C and G nucleotide during each phosphoramidite coupling step.

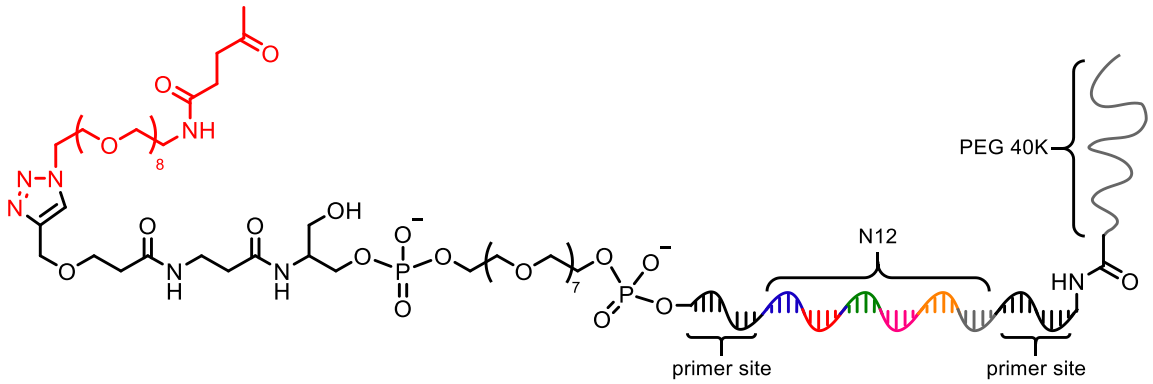



**Figure 4.2**: Molecular architecture of the non-competitive N12 library (Ta_N12) used in the selection protocol. *Reactant A* is shown here as the ketone substrate of the aldol reaction

The 'mixed library' is made by addition of the 'positive control' oligonucleotide to the 'N12 library' at a defined ratio (e.g. 1:100, 1:1,000, or 1:10,000, respectively) in aqueous solution. The 'positive control' oligonucleotide is present in a relatively smaller quantity so that its enrichment can be visualized by gel analysis. When the 'positive control' is present in larger quantities, the difference between the pre-selection library and post-selection library read out is reduced. The 'mixed library' is then evaporated to dryness and redissolved in anhydrous organic solvent. Once dissolved, the biotinylated reactant is added to the reaction mixture and stirred at room temperature. Following completion of the catalysis reaction, the crude reaction mixture is then evaporated to dryness to remove organic solvent and dissolved in DI water for purification with a centrifugal gel filtration column to remove excess biotinylated reactant.

Following the catalysis reaction with the 'mixed library', the successfully catalyzed products are selectively removed from the mixture using a streptavidin bead affinity pulldown procedure. The affinity pulldown step is performed using commercially available streptavidin magnetic beads for capture and isolation of biotinylated molecules.[1] Magnetic bead separation is a widely used technique for isolation of target biomolecules.[2-5] Streptavidin magnetic beads are

uniform and superparamagnetic beads typically 2-3 μm in diameter, with a monolayer of recombinant streptavidin covalently coupled to the surface. The majority of biotin binding sites (~2-3) on the protein are sterically available for binding of biotinylated ligands/targets. The beads are hydrophilic, negatively charged and have shown rapid liquid-phase reaction kinetics for efficient capture, separation, and downstream handling of target biomolecules. Additionally, the size of the biotinylated molecule affects the binding capacity of the beads. The capacity also depends on steric availability and charge interaction between bead and molecule and between molecules. Typically, a mg of streptavidin beads can bind approximately 200 pmol of biotinylated ssDNA.

Successfully catalyzed reaction products are tagged with the biotinylated aldehyde, *Reactant B,* in solution by means of carbon-carbon bond formation. Theoretically, only the 'positive control' oligonucleotides displaying the active catalyst will produce the desired biotinylated reaction product. This 'selection pressure' enables isolation of active catalysts from non-catalytically active library members. The affinity pulldown consists of incubation of the semi-crude reaction product (after removal of excess biotinylated *Reactant B*) with the streptavidin magnetic beads for binding of the biotinylated PEG-DNAs. Elution of the non-biotinylated, or unreacted, DNA is performed by multiple washings of the beads in buffer solution. Any molecules that are not bound to the streptavidin on the magnetic beads will be removed in solution with each wash. The beads are magnetized to the side of the vial by placement on a magnetic rack when performing each wash and ssDNA bound to the magnetic beads is selectively retained. A notable occurrence that potentially lowers the efficiency of this method is nonspecific protein-nucleic acid interactions. Pertinent attractant forces with respect to protein-nucleic acid interactions include electrostatic interactions, hydrogen bonding,

hydrophobic interactions and Van der Waals forces. Non-specific binding can be controlled by altering the stringency of washing conditions including increased number of washes, salt concentration in the buffer, reduced temperature during binding incubation and washes, and 'blocking' of the beads with nonspecific inhibitor ssDNA (e.g. Salmon sperm DNA) to occupy nonspecific binding sites.[6]

Following affinity purification via streptavidin magnetic bead pull down, the beads are diluted to a concentration suitable for PCR and the streptavidin-bound ssDNAs are directly amplified on bead in PCR solution. The aqueous PCR solution contains DNA polymerase, forward and reverse primer sequences, deoxyribonucleotide triphosphates (dNTPs), and buffer salts. PCR amplification is performed to generate exponential copies of the DNA. Direct evaluation of the DNA selectively isolated from the affinity pull down is not performed due to the small quantity of selection survivors in the system, which cannot easily be characterized by common chemical biology techniques. Therefore, PCR is used to generate exponential dsDNA copies of the selection survivors consisting of the original ssDNA sequence template hybridized to its reverse complement strand for evaluation of the selection outcome.

Additionally, the ability to differentiate the DNA sequences of the selection outcome is necessary to quantify the enrichment fold of the 'positive control' DNA sequence. This is performed by a restriction enzyme digest-based assay. Restriction enzymes are found in bacteria (and other prokaryotes). They recognize and bind to specific sequences of DNA, called restriction sites. Each restriction enzyme recognizes just one or a few restriction sites. When it finds its target sequence, a restriction enzyme makes a double-stranded cut in the DNA molecule. Typically, the cut is at or near the restriction site.[7] The 'positive control' DNA was designed to contain a restriction enzyme recognition site within its encoding sequence Restriction enzyme

digestion of the PCR amplified selection product enables distinction of the 'positive control' selection survivors by cutting the dsDNA at the restriction enzyme recognition site, resulting in two dsDNA fragments. PAGE is used to separate the 'positive control' sequence fragments and full-length N12 library by shift in mobility of the gel due to the difference in mass of the digested and non-digested sequences.

Enrichment fold of the 'positive control' DNA determined by comparing the amount of full-length and digested DNA from the library before and after selection. The digested dsDNA migrates at a faster rate than full-length PCR product due to decreased mass. The amount of digested and non-digested DNA can be quantified by the UV-exposure intensity of the EtBr stained bands using standard gel imaging software. Enrichment fold is calculated by dividing the ratio of non-digested to digested DNA of the post-selection outcome (P) by the pre-selection library (I).

$$Enrichment\ Fold = (N12_I/digest_I)/(N12_P/digest_P)$$

### 4.2.2 Selection Optimization with Biotinylated 'Selection Control' PEG Libraries

Prior to performing selection experiments with the catalysis reaction products, the selection procedure was optimized. Control selections were performed with a biotinylated 'PEG-DNA to serve as the 'positive selection control' in a mixed library containing ketone-N12-PEG DNA as non-competitive library members (Figure 4.3). The control selection mixed library represents what the aldol reaction catalysis output would be if all of the 'positive control' oligonucleotides were successfully catalyzed by the tethered diproline catalyst.

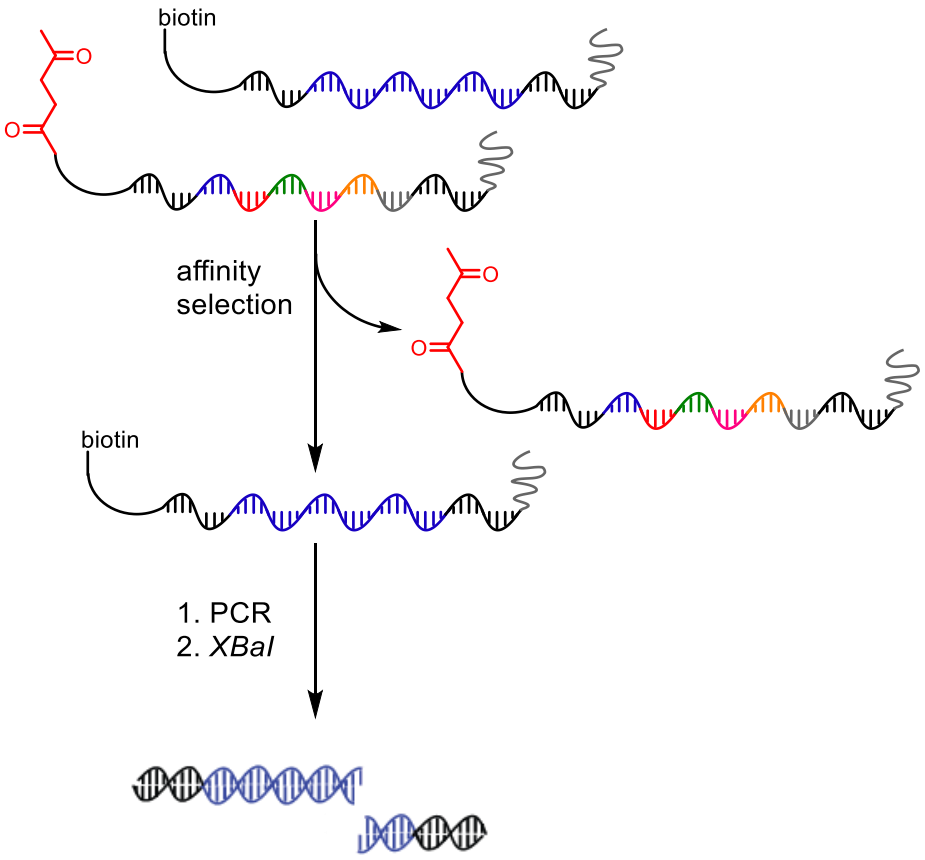



**Figure 4.3**: General selection scheme of positive selection control

**Optimization of Affinity Pulldown with Biotinylated Selection Control Mixed Libraries**

Multiple procedures are available for biotin-streptavidin affinity pulldown. We compared the fold enrichment of the selection control, biotinylated Ta_XbaI from a library of non-biotinylated N12 randomized encoding sequences with multiple affinity pull down procedures to determine the efficiency of each method with our system. Affinity pull-down experiments were performed with two separate brands of commercially available streptavidin magnetic beads. Additionally, gel purification of the biotinylated sequences from the selection control libraries was performed. Biotinylated selection control DNAs were purified from a mixed control library, consisting of N12 library members in 10 or 100 molar excess, by excision of the shifted

migration bands on a polyacrylamide gel following incubation with streptavidin in solution. While gel migration-based methods for purification of active selection products are executed successfully in SELEX-based protocols,[8] whereby multiple rounds of selection are performed, an obvious limitation is the very small quantity of DNAs that are being purified. For our selection platform this amounts to quantities of DNA less than 1 pmol. This amount of DNA is unable to be detected by blue light exposure of the gel following staining with a SYBR safe DNA intercalating stain. Therefore, the position of the bands must be estimated based on a standard migration control, electrophoresed concurrently with the selection outcome products.

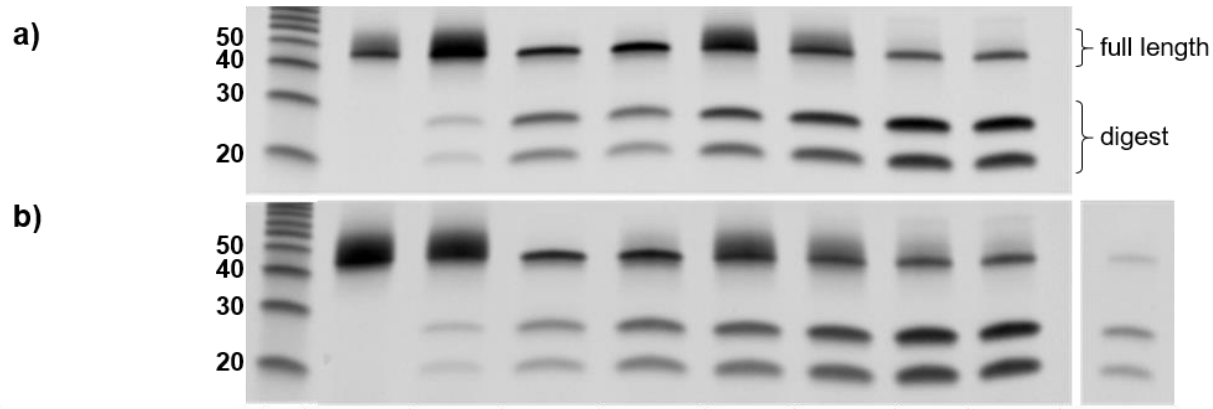
Mixtures of 1:10 and 1:100 biotinylated DNA to ketone-bound TA_N12, respectively, were used to determine the optimal method for selection of biotinylated PEG-DNA from non-biotinylated PEG-DNA. The libraries of oligonucleotides were prepared with biotin-PEG-DNA (biotin_TaXbaI) and N12 PEGylated DNA library (Ta_N12) in 1:10 and 1:100 ratios. Selection of the control libraries was performed by either gel purification or streptavidin magnetic bead separation. Gel purification was performed by binding the selection control libraries with streptavidin in solution for thirty minutes, followed by gel electrophoresis on a non-denaturing 5% acrylamide gel. The gel was stained with SYBR safe and visualized by blue light exposure. Although the streptavidin-bound DNA could not be visualized, bands were excised based on the migration of the positive biotinylated-PEG DNA standard. The DNA was extracted into solution from the gel excisions by shaking overnight in an 0.3 M NaCl aqueous solution. The extracted DNA solution was desalted using a centrifugal gel purification column prior to amplification by PCR.

Selection by streptavidin magnetic beads (NEB or Dynabeads$^{TM}$) was performed by incubating the 10 pmol of selection control library with the magnetic beads in the manufacturer

recommended buffer for thirty minutes. The beads were then magnetized to a magnetic rack and the eluent was removed. Buffer washes of the beads consisted of adding fresh buffer to the beads, pipet mixing, magnetization to the magnetic rack and removal of the eluent. A total of three washes were performed with each brand of magnetic beads. The beads were then suspended in DI water to their original concentration and equivalent amounts were removed for amplification by PCR.

Following affinity selection, either by streptavidin bead or gel purification, the exact same conditions and starting amounts of DNA were used for PCR and restriction enzyme digestion. PCR and RE digestion are responsible for amplifying the selection output to a manageable size and distinction of the control sequence from the N12 library, respectively.

PCR amplification of each selection control experiment was performed, followed by purification by a commercially available PCR purification kit. The purified dsDNA PCR products were subjected to restriction enzyme digestion and visualized by non-denaturing PAGE (Figure 4.4). The band percentage of each lane was calculated by the imaging software based on intensity of UV-exposure in the image. Ratios of digested to full length selection product, PCR amplified and RE digested, and enrichment fold of the biotinylated control sequence were calculated (Table 4.2). These experiments show that the NEB streptavidin magnetic beads provide the highest enrichment potential over the other compared methods. Therefore, this method was determined to be most effective and selected for use and optimization.




| Lane | L | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | DC |
|---|---|---|---|---|---|---|---|---|---|---|
| Selection | | – | – | + | + | + | + | + | + | |
| Gel Cut | | – | – | + | + | – | – | – | – | |
| Dynabeads ™ | | – | – | – | – | + | + | – | – | |
| NEB Beads | | – | – | – | – | – | – | + | + | |
| XbaI | | – | + | + | + | + | + | + | + | + |
| Bead Mass (µg) | | – | – | – | – | 25 | 12.5 | 40 | 20 | |

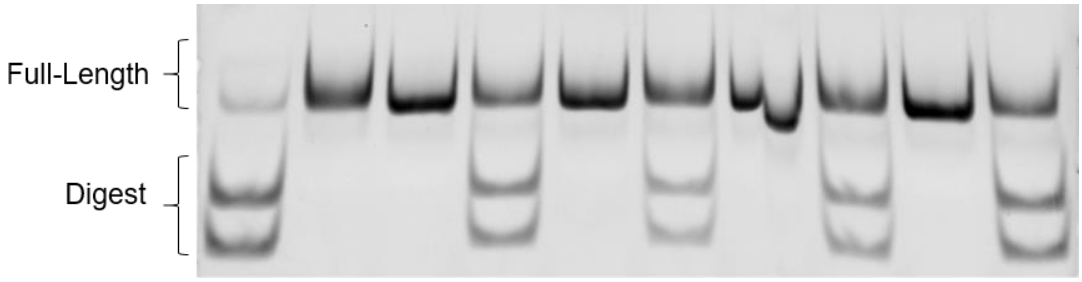
**Figure 4.4**: Affinity pulldown selection method experiment– digest product comparison. a) 1:10 Selection Library. b) 1: 100 Selection Library. Selection libraries 1:10 or 1:100 – biotin selection control (Ta_XbaI) vs. N12 library (Ta_N12). L: 10 bp DNA ladder, DC: Digest control of biotinylated Ta_XbaI, Restriction Enzyme: XbaI. Magnetic Bead DNA Binding Capacity: Invitrogen Dynabeads™ (650 pmol/mg), NEB Hydrophilic Streptavidin Beads (400 pmol/mg).

**Table 4.2**: Selection Method Experiment Results

| Lane | Digest (%) | Full-Length (%) | Enrichment Factor |
|---|---|---|---|
| 1:10 | | | |
| Pre-Selection | 7.9 | 92.1 | - |
| 3 | 51.3 | 48.7 | 12.28 |
| 4 | 40.8 | 59.2 | 8.04 |
| 5 | 39.2 | 60.8 | 7.52 |
| 6 | 52.1 | 47.9 | 12.68 |
| 7 | 75.7 | 24.3 | **36.32** |
| 8 | 76.2 | 23.8 | **37.33** |
| 1:100 | | | |
| Pre-Selection | 6.6 | 93.4 | - |
| 3 | 39.2 | 60.8 | 9.12 |
| 4 | 43.0 | 57.0 | 10.67 |
| 5 | 38.2 | 61.8 | 8.75 |
| 6 | 49.7 | 50.3 | 13.98 |
| 7 | 66.3 | 33.7 | **27.84** |
| 8 | 70.6 | 29.4 | **33.98** |

## Temperature and Solvent Effects of Streptavidin Magnetic Bead Selection Incubation

We next devised an experiment to determine whether the addition of heat and/or DMSO to the streptavidin magnetic bead incubation step could increase the binding affinity and, therefore, enrichment of the biotinylated sequence in a 1:100 mixture of biotinylated control sequence to N12 sequence library. Previous literature shows increased capture efficiency of the streptavidin-coated beads to biotinylated, polymer-bound ssDNA when heated to 40 ºC. We recreated this experiment to with a 100:1 biotinylated positive selection control mixed library to determine if these alterations would increase the efficiency of the streptavidin magnetic bead affinity pulldown (Figure 4.5).



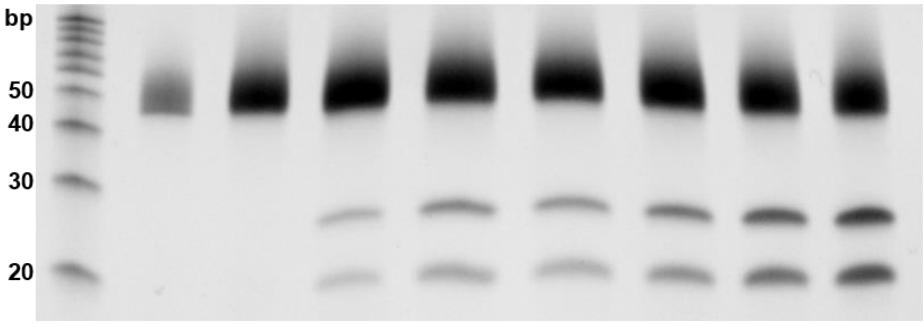

| Lane | 1(POS) | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Selection | | - | + | + | + | + | + | + | + | + |
| Heat (40°C) | | | - | - | + | + | - | - | + | + |
| DMSO | | | - | - | - | - | + | + | + | + |
| Digest | + | + | - | + | - | + | - | + | - | + |
| Ratio (FL:D) | 90% D | 100:1 | | 1.1:1 | | 2.2:1 | | 1.7:1 | | 1:1 |
| Enrichment Fold | | | | 91 | | 46 | | 59 | | 100 |

**Figure 4.5:** Optimization of streptavidin affinity pull-down incubation conditions

The gel image shows positive enrichment for every 1:100 positive selection control library mixture (lanes 4/6/8/10). The largest enrichment resulted from the samples incubated with the streptavidin magnetic beads is in 10% DMSO at 40 ºC (Lane 10). However, the increase in enrichment is negligible compared to the incubation under normal room temperature conditions without DMSO. Interestingly, the incubation with only heat resulted in the least amount of enrichment, contrary to the published findings of Mirkin[12] when studying biotinylated DNA capture by the SA mag beads. From these results it was concluded that addition of heat and/or DMSO during streptavidin bead incubation does not provide a significant change in affinity separation efficiency to be explored in further selection experiments.

**Wash Efficiency**

A wash efficiency experiment was next conducted to determine the optimal number of washes needed to produce the largest enrichment. A ratio of 1:10 positive catalyst to N12 library, respectively was chosen based on the positive enrichment results seen from previous selection experiment. After incubation with the streptavidin coated magnetic beads, they were washed twice, 500 μL each, once with a tween-buffer and once with a no tween-buffer. Following each additional wash, 1.0 μL was removed for PCR after resuspension to the of the beads to their original volume after the previous wash. PCR products were restriction digested and visualized by non-denaturing PAGE (Figure 4.6). The results of this experiment demonstrate gradually increasing fold enrichment of the catalyst DNA with each bead-wash performed. Increased enrichment from each bead-wash step was hypothesized to originate from removal of non-specific binding of streptavidin and the PEGylated DNA library.

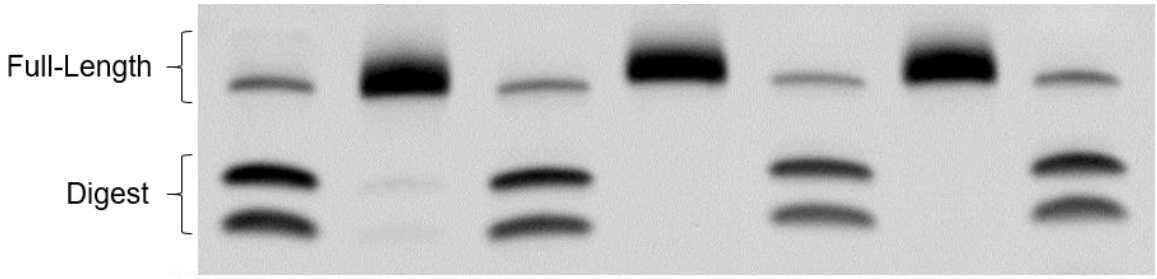



| Lane | L | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Selection | | – | + | + | + | + | + | + | + |
| Wash Number | | – | 3 | 3 | 4 | 5 | 6 | 7 | 8 |
| XBaI | | – | – | + | + | + | + | + | + |

**Figure 4.6:** Wash efficiency of streptavidin magnetic beads. L) 10 base pair (bp) DNA Ladder. Restriction Enzyme: XbaI (NEB). 15% acrylamide non-denaturing PAGE.

**Positive Control Selection with Optimized Conditions– Biotinylated PEG-DNA vs. N12 PEG DNA Library**

Greater stringency of washing was employed in the positive selection control protocol following the wash efficiency experiment. Using the pre-biotinylated positive control sequence, mixtures of a random, non-biotinylated sequence library were prepared in 10:1, 100:1, and 1000:1 (N12 library: positive selection control). These selection libraries were subjected to ten, 1 mL washes with a combination of tween and no-tween, high salt concentration buffers. This was followed by PCR amplification and restriction enzyme digestion of the affinity pulldown products. The ratio of the starting library, after PCR and digestion, was compared to the selection enrichment product by non-denaturing PAGE (Figure 4.7).




| Lane | 1(POS) | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Selection | - | - | + | - | + | - | + |
| Ratio (FL:D) | 80% D | 10:1 | 1:4.5 | 100:1 | 1:4.9 | 1000:1 | 1:4 |
| Enrichment Fold | | | 45 | | 490 | | 4000 |

**Figure 4.7:** Positive selection control with increased washes. Non-Denaturing PAGE – 15 %

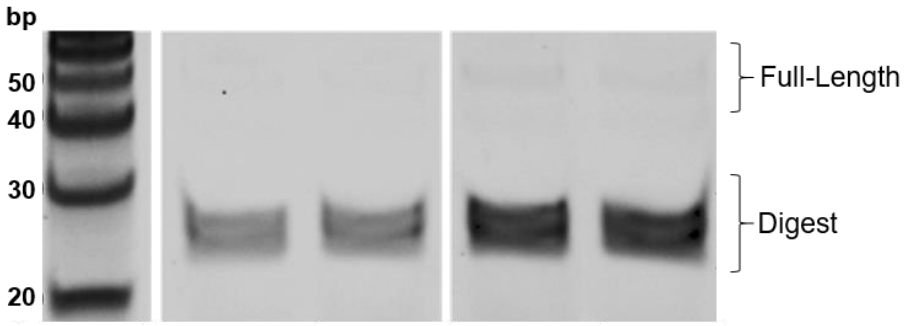
First, it is important to note that the biotinylated template PEG-DNA, used to serve as a digest positive control, was only 80% digested (Figure 4.7, Lane 1). This means that the

enrichment fold calculated for each selection experiment could possibly be higher, due to incomplete digestion of the target sequences. Interestingly, these results show that all of the selection products are 80% digested. The enrichment fold increases by factor of approximately 10 when the quantity of the target sequence in the starting library is decreased the same amount. Therefore, it was concluded that this selection procedure was efficient in isolating biotinylated DNA from a mixture of non-biotinylated PEG-DNAs. Overall, these results indicate that the procedure used for the affinity separation and PCR purification are well within acceptable ranges of what should be seen with a successful selection round.

### 4.2.3 Mock Selection of PEGylated DNA-Encoded Aldol Reaction Catalyst

**Restriction Enzyme Validation- EcoRV**

The observation of incomplete digestion of the biotinylated control sequence in previous experiments prompted the use of an alternate sequence for the mock aldol catalysis selection DNA. The efficiency of the new restriction enzyme, EcoRV, was tested with the new sequence. The non-modified template sequence ssDNA was purchased from Integrated DNA Technologies, Inc. (IDT). Various amounts of PCR amplified template were digested with 20 units of EcoRV various times. This amount of restriction enzyme was used based on the manufacturer's recommended protocol. The restriction enzyme is a 'quick-digest' enzyme, purchased from NEB, who claim complete digestion of the target dsDNA sequence can be achieved in 15 minutes. The results of the RE digest experiment show that the enzyme is very efficient and was able to digest greater than 98% of the template in 30 minutes or an hour. Based on these results, this sequence was used in the synthesis of the diproline-tethered 'positive control' for forthcoming selection experiments of the DNA-encoded aldol reaction catalyst.



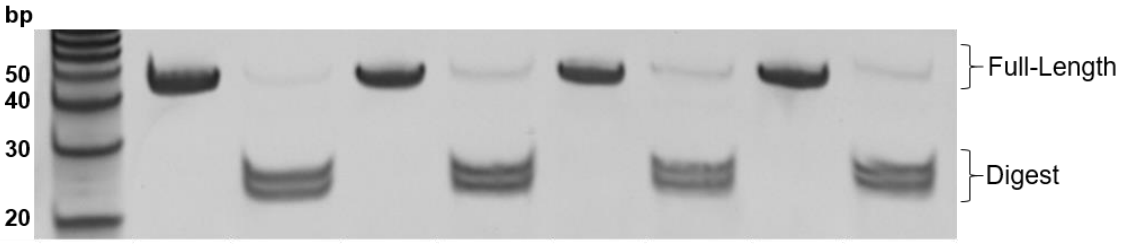

| Lane | L | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| EcoRV (µg) | | 0.05 | 0.05 | 0.10 | 0.10 |
| Digest (min) | | 30 | 60 | 30 | 60 |
| Digest (%) | | 97 | 99 | 97 | 99 |

**Figure 4.8:** Restriction enzyme digestion of oligonucleotide sequence Ta_EcoRV – purchased standard without modifications

**PCR Fidelity of Modified PEG-DNA with Ta_EcoRV Encoding Sequence**

To ensure the sequence fidelity of the synthesized oligonucleotides was retained during PCR, the PEGylated oligonucleotides were amplified and restriction digested with the restriction enzyme EcoRV. The fidelity of a DNA polymerase is the result of accurate replication of a desired template. Specifically, this involves multiple steps, including the ability to read a template strand, select the appropriate nucleoside triphosphate and insert the correct nucleotide at the 3´primer terminus, such that Watson-Crick base pairing is maintained. Additionally, if depurination or degradation of the DNA bases occurred in our synthesized oligos due to the exposure of various reaction conditions, the fidelity of the amplification may also be lowered. In turn, this could cause a false negative when determining digest efficiency, as the digest specific sequence would not be present in all of the amplified DNA.

The oligonucleotides tested include the diproline catalyst PEG- DNA with 3' display of an aldehyde, ketone, or alkyne (Figure 4.9). The PAGE results show that 95% or greater of the PCR amplified product was digested by the restriction enzyme in 30 minutes. These results provide evidence that the RE digest readout of the selection protocol is able to show at least 95% of the target sequence and provide more accurate results than the Ta_XbaI sequence when determining fold enrichment.



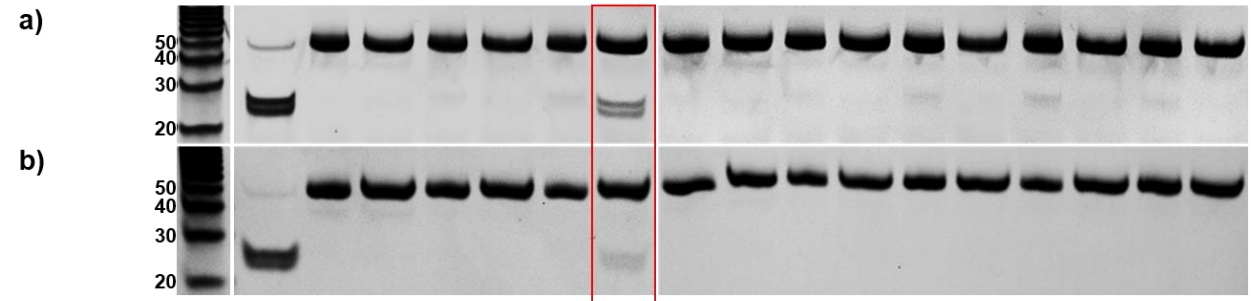

| Lane | L | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 3' End | | POS | POS | Y | Y | K | K | A | A |
| Digest | | - | + | - | + | - | + | - | + |
| Digest (%) | | | 98 | | 96 | | 94 | | 95 |

**Figure 4.9:** PCR fidelity of synthetic oligonucleotides used in selection experiments. Restriction enzyme-based readout of PCR amplified 'positive control' DNA.

## Aldol Catalysis *In Vitro* Selection

Catalysis experiments were performed with the 'positive control' and 'N12 library' PEGylated DNA as shown in Figure 4.1. Catalysis of the aldol reaction 'mixed library', combined with the biotinylated reactant in 1,2-dichloroethane, was shaken for 2 days at room temperature. Both 3'-aldehyde and ketone reactants were studied. Affinity pull-down was performed with streptavidin coated magnetic beads using the optimized wash protocol. Selection

survivors were PCR amplified directly from the magnetic beads, digested with EcoRV, and the results were visualized by non-denaturing PAGE (Figure 4.10).




| Lane | L | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pro-Proline DNA | | | + | + | + | + | + | + | + | + | + | + | + | + | - | - | - | - |
| 3' Reactant | | | + | + | + | + | + | + | - | - | - | - | - | - | - | - | - | - |
| Bt-Reactant | | | - | - | - | - | + | + | - | - | - | - | + | + | - | - | + | + |
| SA beads | | | - | - | + | + | + | + | - | - | + | + | + | + | + | + | + | + |
| Digest | | POS | - | + | - | + | - | + | - | + | - | + | - | + | - | + | - | + |

**Figure 4.10:** Aldol reaction catalyst selection with controls. a) 1:100 **ketone**-diproline-Ta_EcoRV-PEG DNA : ketone-N12-PEG DNA. b) 1:100 **aldehyde**-diproline-Ta_EcoRV-PEG DNA : ketone-N12-PEG DNA. POS: digest control (Ta_EcoRV sequence).

Lanes 7 are the outcome of the key selection experiment when the library is reacted with the biotinylated ketone, the selection mixture is affinity purified by streptavidin magnetic bead pull down, PCR amplified, and RE digested. These results show the positive control DNA-encoded catalyst was enriched. Furthermore, these results demonstrate that inter-strand catalysis resulting in biotinylation of inactive library members by an active member does not occur.

Additional selection experiments were conducted to control for various scenarios that could negate any enrichment outcome of the positive control sequence in the mock selection experiment. Lanes 3 show the initial 1:100 selection library that was not subjected to the aldol

reaction conditions or pulldown conditions to show the ratio of the 'positive control' DNA found in the starting 'mixed library'. Lanes 5 display the initial selection library outcome from being subjected to the affinity pulldown conditions. Both ketone and aldehyde experiments show that the positive control oligonucleotide is not enriched when by the pulldown conditions alone.

When the aldehyde is absent from the 'positive control' architecture, whereby only the diproline catalyst is displayed, the target sequence is not enriched (Lane 13, Figure 4.10 b). These results further exemplify that the catalyst does not form stable covalent adducts with the biotinylated aldol reactant. This scenario can only be examined when biotinylated ketone is in solution because enamine formation can only occur on the ketone, as the biotinylated benzaldehyde reactant does not contain the reactive α-hydrogen.

A control experiment was undertaken to determine if enrichment of the target sequence from the 16.7 million-membered N12 library was possible by subjecting the N12 library, lacking *Reactant A*, to the selection experiment conditions and digest readout of the PCR amplified products. This scenario would only occur if target DNA sequence itself was acting as a DNA catalyst[9] and catalyzing the formation of the biotinylated aldol product. However, it is extremely unlikely due to essentially equal probability of each of the 16.7 million sequences in the library to act as a DNA catalyst. Additionally, enrichment of the target sequence would only occur if it happened to be the most active DNA catalyst compared to every other library sequence. The results of this experiment show that this does not occur (Lanes 17).

Additionally, to determine if positive enrichment could occur from the selection procedures themselves, affinity pull-down without incubation with the biotinylated reactant was also performed with each library mixture. This scenario would occur if any secondary structure

of the target sequence itself was selectively binding the magnetic beads. Analysis of the gel shows that this does not occur (Lanes 5, 11, and 15).

**Diproline Catalyzed Aldol Reaction Purification Optimization**

Subsequent to the aldol reaction, the crude reaction mixture must be purified to remove the remaining excess biotinylated reagent. Typically, this is done by using silica-based size exclusion, or Princeton Separations Centri-Sep column.The efficiency of this method was compared to an alternative method for removing biotinylated reactant. Since the positive selection control normally used, a biotinylated PEG oligo, is not subjected to the actual reaction conditions it cannot be used to test the purification efficiency of the crude reaction product. Therefore, 0.3 pmol of an aldol reaction catalysis experiment, previously found to have a biotinylated product yield of 8.8%, was evaporated to dryness, resuspended in water, and 0.1 pmol was aliquoted into each of two vials. 10 pmol of ketone-N12-library was then added to each vial to produce a 1:100 library of crude aldol reaction catalysis product vs. negative library. One vial was purified by the typical Centri-Sep protocol and the other was purified by the QIAquick Nucleotide Removal Kit. The optimized selection protocol, as explained above, was performed following the purification including streptavidin magnetic bead pull-down, PCR, RE digest, and non-denaturing PAGE (Figure 4.11).

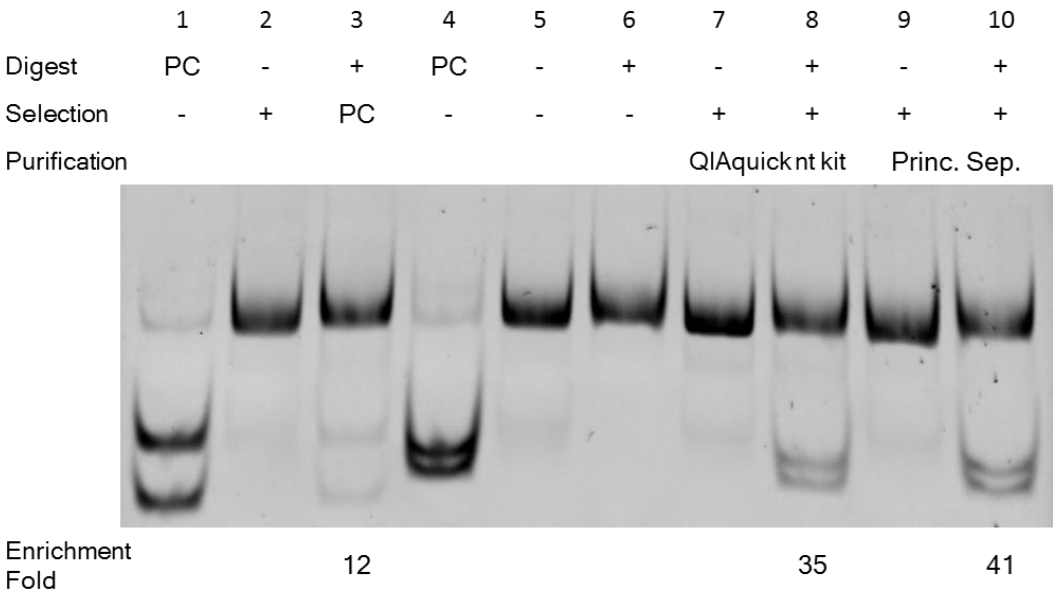



**Figure 4.11:** Aldol reaction purification comparison of QIAquick nucleotide removal kit and Princeton Separations centri-sep column. 100:1 mixture of negative sequence library to catalysis product. 15% native PAGE. (PC = positive control)

This gel image shows that both selection products have positive enrichment of the aldol reaction catalysis product PEG-DNA sequence (Ta_EcoRV). The Centri-sep column afforded a slightly increased enrichment (41-fold) compared to the QIAquick Nucleotide Removal Kit (35-fold). The positive control selection enrichment of the biotinylated DNA sequence was lower than expected based on previous experimental results. The enrichment fold was calculated by dividing the ratio of digested DNA to full length by the initial starting ratio, which is 100:1 randomized sequence to catalyst-PEG-oligo, respectively.

However, it is important to note that the yield of the aldol catalysis biotinylated product was already determined to be 8.8%. If we assume 0.088 pmol (8.8% of 0.1 pmol) of biotinylated DNA was in the selection library mixture (10.1 pmol total), the ratio of biotinylated DNA to non-biotinylated DNA is instead ~1:1000, respectively. If we consider this initial library ratio, the

resultant enrichment folds of this experiment would be approximately an order of magnitude greater than those calculated (340 fold and 410 fold). If this were to be assumed correct, the difference in enrichment between the aldol catalysis reaction purification methods is more significant. Therefore, it was concluded to use the Princeton Separation Technologies, Inc. 'Centri-Sep' gel purification column for subsequent catalysis reaction product purifications.

**Time Course Experiment**

A range of reaction times were tested to determine reaction efficiency and kinetics (Figure 4.12). Time scale reactions show the enrichment increasing from 10 minutes to one hour and then decreasing from three hours to 24 hours. However, the amount of amplified PCR product is substantially less than the longer time scales. This is expected, as there will be less overall reaction product; however, the imaging software calculation of band ratios is with relatively greater error.

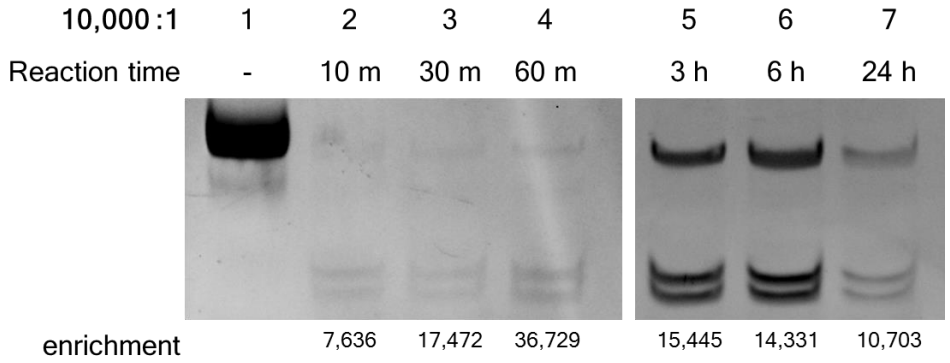



**Figure 4.12:** Time course experiment of aldol reaction catalysis selection – ketone on DNA. Reactions performed in DCE.

Time scale reactions show the enrichment increasing from 10 minutes to an hour and then decreasing from 3 hours to 24 hours. However, the amount of amplified PCR product from the selections performed from the reaction products of 10-60 minutes is substantially lower than that of the increased reaction time selection products. This is understandable as there will be less overall reaction product when a shorter reaction time is implemented; however, the imaging software calculation of band ratios has relatively greater error when less product is visible. A faint band on a gel image displays less contrast from the background intensity of the gel.

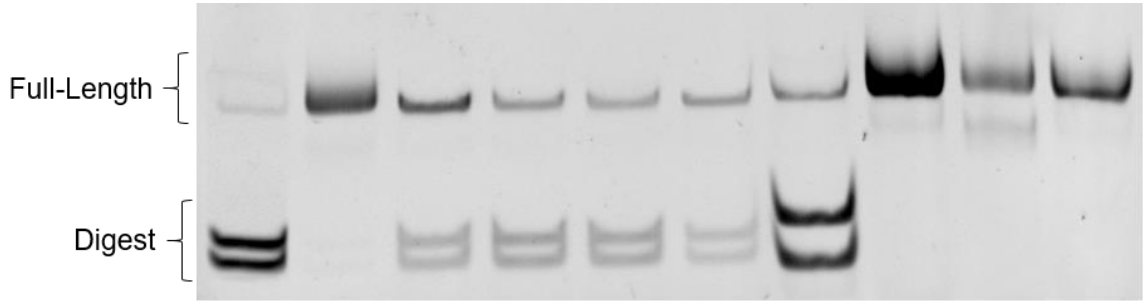
It is important to note that this mock selection exemplifies the ability of only one DNA-scaffolded catalyst to be enriched from a library of DNA with no catalysts attached. Since the selection capability of the system is now able to enrich from a small initial input, this makes it complicated to compare conditions within the system without some degree of error. Essentially, any successful reaction product can be selectively removed from those sequences that did not yield product and therefore only accurate measurement of general reaction success can be observed.

**Post-Selection Streptavidin Affinity Pulldown Bead Dilution**

The aldol reaction selection was performed without modifications to the previously optimized reaction conditions. The only modification made was introduction of the serial dilution of the streptavidin beads after bead binding and washes. Initially, a 1:1,000 starting ratio of ketone-propro-PEG to ketone-N12-PEG was performed (Figure 4.13).

These results show that a substantial increase in observed enrichment is seen with the added serial dilutions ranging from about 700- to 2000-fold. Previously, the greatest enrichment observed for this reaction was 30-fold, which is 20 times less than the lowest enrichment seen here. Additionally, a negative control experiment with ketone-Ta1-PEG vs. ketone-N12-PEG,

subjected to identical conditions, confirms that the observed enrichment can be attributed to the proline-proline catalysis.




| 1,000 : 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Pro-Proline | - | + | + | + | + | + | - | - | - | - |
| Reactant A | DC | A | A | A | K | K | DC | K | K | K |
| Reactant B | - | - | K | K | A | A | - | - | A | A |
| SA Bead Dilution | - | - | 500 | 5000 | 500 | 5000 | - | - | 500 | 5000 |
| Enrichment Fold | | | 680 | 1680 | 2200 | 1045 | | | | |

**Figure 4.13:** 1,000:1 Aldol reaction catalyst selection – post selection streptavidin magnetic bead dilution. Reaction performed in DCE.

**Final Selection with Optimized Conditions**

The positive control was diluted 500-fold into a library of DNA sequences that lacked a catalyst. The library mixture was incubated with the biotinylated aldehyde reactant in DCE for three hours followed by binding to streptavidin-coated magnetic beads. After extensive washing of the beads, on-bead PCR was performed to amplify the selection survivors. Enrichment analysis was determined by restriction enzyme digestion of the PCR product, followed by non-denaturing PAGE analysis. After one round of selection for catalytic activity, the positive control

was enriched 100-fold (Figure 6b). Importantly, when the aldol reactants were exchanged, such that the aldehyde was on the DNA template, and the ketone was biotinylated in solution, similar enrichment values (70-fold) were observed (Figure 4.14).

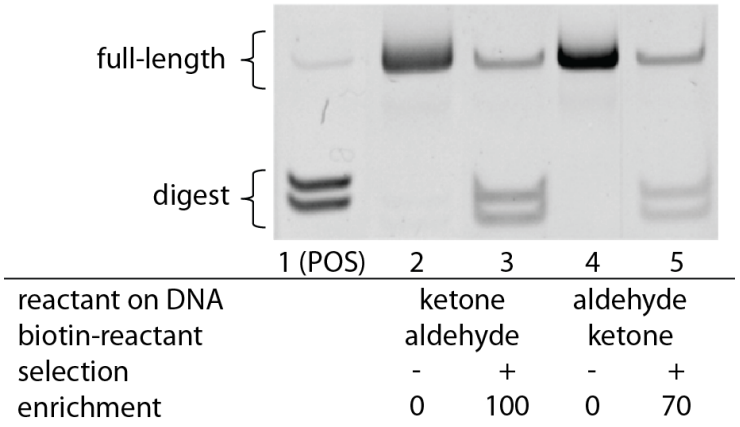



**Figure 4.14:** Gel analysis of mock aldol selection resulting from a 500-fold dilution of DNA-encoded aldol catalyst

### 4.2.3 High-Throughput Sequencing Analysis of Mock Aldol Selection

Satisfied with the outcome of the preliminary mock selections, high-throughput DNA sequencing was performed to quantitatively determine the fold enrichment of the aldol reaction selection under more dilute selection conditions. Compared to the EMSA characterization, which only allows characterization of one specific sequence, DNA sequencing permits characterization of all the sequences in a library allowing for a more in-depth analysis of the selection outcome.

PCR products of initial and selection libraries were PCR amplified with sequencing primers to add unique experiment ID to each sample. 200 attomole of PCR product was amplified with Q5 high-fidelity master mix for 15 cycles, purified with Sigma PCR purification

kit, and gel purified from 5% non-denaturing PAGE. Figure 4.15 shows an example of the gel after excision of the full-length product band (172 bp).

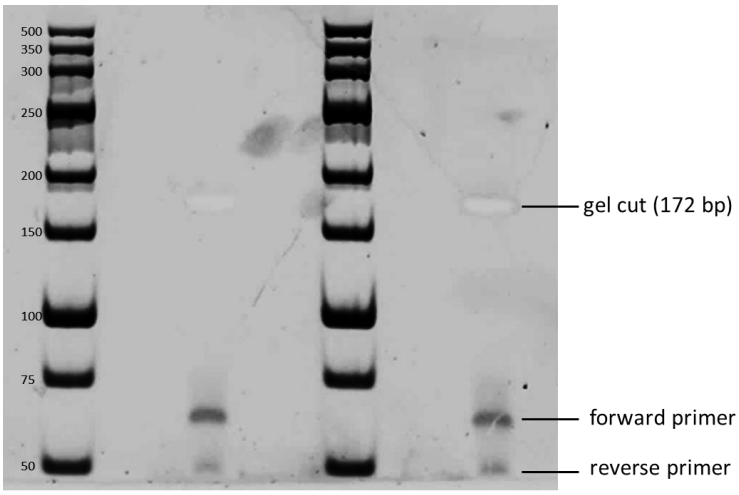



**Figure 4.15:** Non-denaturing PAGE extraction of sequencing PCR samples

Aldol selection was performed as described above with the positive control diluted 2,000-fold into a library of 16.7 million (N12) library members; however, instead of restriction digest as a readout, Illumina barcoded adapters were added to the template sequences by PCR amplification and Illumina Mi-Seq paired-end sequencing was performed.

Post-sequencing analysis involved merging of paired-end reads and trimming off sequencing adapters to yield readouts of the survivors of the aldol selection. By comparing the sequence frequencies of the starting library with those of the post-selection library, enrichment levels could be readily calculated for each sequence (Fig. 4.15). Sequencing analysis revealed that the positive control diproline catalyst was strongly enriched by 1,200-fold. This level of

enrichment suggests that this method could support the de novo discovery of small-molecule catalysts for bond-forming reactions.

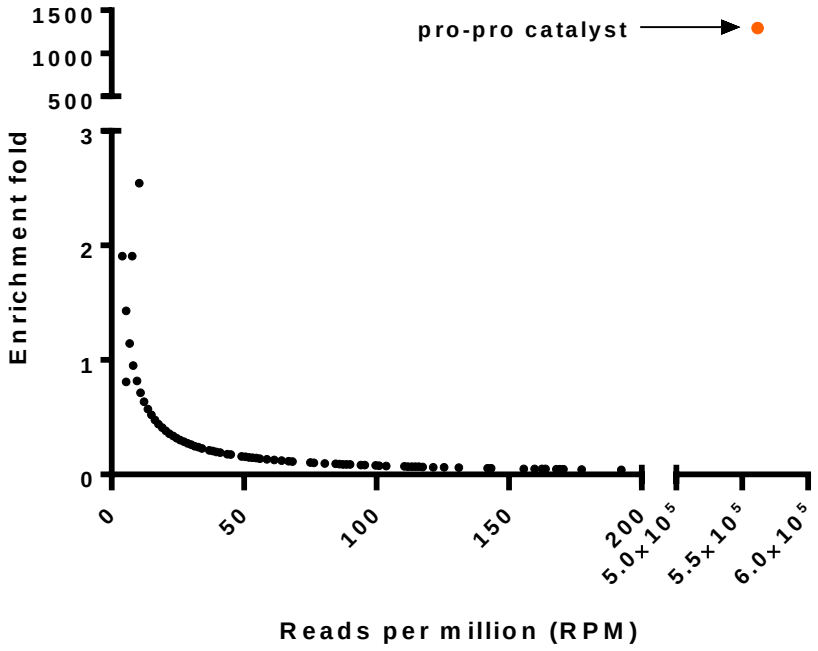



**Figure 4.16:** High-throughput sequencing analysis of mock aldol selection resulting from a 2000-fold dilution of DNA-encoded aldol catalyst

## 4.3 CONCLUSIONS

In summary, we have developed a catalyst selection system based upon the use of DNA-encoded libraries in organic solvents. Survival of the selection requires DNA-encoded catalysts to engage in catalytic bond formation between an in-cis reactant and an in-trans biotinylated reactant. Affinity pull-down and readout by high-throughput DNA sequencing enables the rapid identification of active catalysts. Using the amine-catalyzed aldol reaction as a model, we demonstrated that this approach can be implemented in various organic and aqueous solvents and can enrich a known aldol catalyst by 1,200-fold. This platform has the potential to greatly accelerate the discovery of catalysts by increasing the throughput of catalyst screening efforts and expanding the chemical space explored during conventional catalyst screenings.

## 4.4 EXPERIMENTAL METHODS

### DNA Synthesis Sequences of DNA Used in Selection Experiments

'Positive control' PEG DNA with 5'-40,000 MW PEG, diproline catalyst, and 3'-Reactant A:

Ta_EcoRV: CGT ACG GTC GAC GCT AGC TGG ATA TCA CTG CAC GTG GAG CTC GGA TCC/iAmMC6T//iSp18//3AlkSerCPG/

Biotinylated 'selection control' PEG-DNA with 5'-40,000 MW PEG and 3'-biotin:

Ta_XbaI: /5ThioMC6-D/ CGT ACG GTC GAC GCT AGC TCT AGA AAC AAC CAC GTG GAG CTC GGA TCC/3'BtSerCPG

Non-competitive 'N12 library' with 5'-40,000 MW PEG and 3'-Reactant A:

Ta_N12: /5ThioMC6-D/ CGT ACG GTC GAC GCT AGC NNN NNN NNN NNN CAC GTG GAG CTC GGA TCC/iSp18//3AlkSerCPG/

/Sp18/ 3' Spacer 18

/AmMC6C/ Amino Modifier C6 dC

/ThioMC6-D/ Thiol Modifier C6 S-S (Disulfide)

/3'AlkSer/ 3'-Alkyne Serinol CPG

/3'BtSerCPG/ 3'-Protected Biotin Serinol CPG

Note: Post-synthetic modification of the DNA was performed as explained in Chapter 3.

### Catalysis

Libraries of 100:1 reactant-neglib-PEG DNA to propro-reactant-Ta_EcoRV-PEG were prepared for each reactant-bound (aldehyde, ketone, or alkyne/negative control) catalyst DNA to their respective aldehyde or ketone bound negative libraries (alkyne was mixed with both). 10 pmol of each mixture was evaporated to dryness and resuspended in 10 μL of 1,2-dichloroethane (DCE).

The solutions were vortexed overnight to allow maximum solubility. Biotinylated reactant (10 μL, 20 mM) was as added to each oligo mixture (0.5 μM) and incubated at 25 ºC for 48 hrs. Crude mixtures were evaporated to dryness, dissolved in water and excess biotinylated reactant was removed by a size-exclusion Princeton column. The purified product was evaporated to dryness and dissolved in 4 μL water.

**Affinity Pull-Down**

For each wash, the buffer was kept cold and tubes were rotated at room temperature for 5 minutes. Streptavidin-coated magnetic beads (2 μL, Invitrogen) were aliquoted into an Eppendorf tube. The buffer was removed with the beads on the magnetic rack and the beads were washed three times with 500 μL of 1X Tween buffer (5 mM Tris HCl pH 8, 0.5 mM EDTA, 1M NaCl, 0.05% tween). The buffer was removed and 4 μL (twice the volume of initial beads) of 2X No Tween Buffer (10 mM Tris HCl pH 8, 1 mM EDTA, 2M NaCl) was added. Oligo sample (4 μL) was added to the beads and incubated for 30 min with rotation at RT. The buffer was removed and the beads were washed eight times with 500 μL of 1X Tween buffer and twice with 500 ml of 1X No Tween Buffer. The buffer was removed and the beads were resuspended in 2 μL water. 1 μL was removed and used for PCR.

**Serial Dilution of Selection Beads**

Streptavidin beads were incubated with 10 eq. biotin- ketone or biotin-aldehyde overnight in binding buffer (10% DMSO), and washed 3x with DCE. Oligo library (5 pmol) was incubated with beads overnight. 8 washes with tween buffer and 2 washes with no-tween buffer were performed. The products were 10-fold serially diluted in DI water, PCR amplified, and restriction enzyme digested.

## Positive Control Experiments

### Streptavidin Magnetic Bead – Affinity Pulldown

Two separate libraries of oligos were prepared with biotin-PEG-DNA to negative library-PEG-DNA in 1:10 and 1:100 libraries. Positive control consisted of entirely biotin-PEG-DNA and the negative control was used consisting of only library-PEG-DNA. 10 pmol of each oligo was evaporated to dryness and dissolved in 40 μL of no-tween buffer. Oligo buffer solutions were added to 10 μL of NEB streptavidin beads that were pre-washed 2X with tween buffer and 1X with no-tween buffer, and incubated for 30 minutes at room temperature with rotation. The binding buffer was removed and the beads were washed in the same procedure as stated above without changing any steps. – NEB magnetic streptavidin beads (10 μL) was used for each experiment, buffer removed, and washed with tween buffer (500 μL; 1 M NaCl, 0.5 mM EDTA, 5 mM Tris-HCl pH 8, 0.05 % Tween 20) two times and no-tween buffer (500 μL; 1 M NaCl, 0.5 mM EDTA, 5 mM Tris-HCl pH 8) one time with 5 min rotation between each wash. Evaporated oligo samples (10 pmol) were dissolved in 40 μL (4X bead volume) of no-tween buffer and transferred to magnetic beads, rotated at room temperature for 30 min, and buffer removed on magnetic rack. The magnetic beads were washed 3X with tween buffer and 2X with no-tween buffer, resuspended in 10 μL of water, and 1 μL was removed for PCR.

### Streptavidin Binding Gel Purification – Affinity Pulldown

10 pmol of each of the samples listed above was combined with 4 eq. of streptavidin binding solution and incubated for 30 min and room temperature. The samples were diluted to 10 μL with water and combined with 6X loading dye (2 μL). The samples were loaded onto a 5% TBE non-denaturing gel and run at 150 V for 95 minutes. The gel was dyed with SYBR safe for 30 min at room temperature while covered in foil. Gel was washed with water and imaged on blue

light imaging plate to detect bands of interest. Streptavidin bound bands were excised from the gel with plastic gel cutting tool and transferred to an Eppendorf tube. The extracted gel was crushed with a pipet tip, centrifuged down, and suspended in equal volume of 0.3 M NaCl solution overnight. Sample was centrifuges at 12,000 rpm for 1 minute and liquid transferred by pipet to new vial. Products were filtered through Centri-Sep, evaporated to dryness, and dissolved in 10 μL of water.

**Affinity Pull-Down**

NEB magnetic streptavidin beads (5 μL) were used for each experiment, buffer removed, and washed with tween buffer (500 μL; 1 M NaCl, 0.5 mM EDTA, 5 mM Tris-HCl pH 8, 0.05 % Tween 20) two times and no-tween buffer (500 μL; 1 M NaCl, 0.5 mM EDTA, 5 mM Tris-HCl pH 8) one time with 5 min rotation between each wash. Evaporated oligo samples (10 pmol) were dissolved in 20 μL (4X bead volume) of no-tween buffer and transferred to magnetic beads, rotated at room temperature for 30 min, and buffer removed on magnetic rack. The magnetic beads were washed 8X with tween buffer and 2X with no-tween buffer (1.0 mL each), resuspended in 5 μL of water, serially diluted to 1:1000.

**PCR**

1.0 μL of streptavidin beads from selection was combined with forward primer (PFa, 1 μL, 20 μM), reverse primer (PRa, 1 μL, 20 μM) Hot Start Taq (2X, 10 μL) and water (7 μL). Reactions were amplified according to normal PCR protocol with an annealing temperature of 58 ºC and extension temperature of 72 ºC for 20 cycles. PCR products were immediately put on ice and purified by MinElute PCR purification kit, solution evaporated to dryness, and concentration determined by absorbance at 260 nm on Nano-drop.

**Wash Efficiency Experiment**

Remaining 9 μL streptavidin beads from the 1:10 selection (ketone-propro-PEG: ketone-neglib-PEG) were systematically washed to determine if separation efficiency could be increased. Buffer was removed from beads on magnetic platform. Beads were washed with tween and no-tween buffer once (500 μL), buffer removed beads suspended in 9 μL water (starting volume of beads), and 1 μL was removed for PCR. This procedure was repeated 5 X. Each aliquot was PCR amplified following the same procedure as stated above for 25 cycles, purified with nucleotide removal kit, evaporated to dryness, dissolved in water, and absorbance measured. RE digest was performed according to the same procedure as above with 0.9 μg of dsDNA. Digests were purified by Centri-Sep, combined with 6X gel loading buffer, and run on 15% TBE non-denaturing gel for 50 min.

**Restriction Enzyme Digest**

Purified PCR product DNA (0.1 μg) was combined with 10 X CutSmart Buffer (NEB) (5 μL), restriction enzyme XbaI (1.0 μL, 20,000 U/mL) and diluted to 50 μL with water. Digestion reactions were incubated overnight (approx. 18 hr.) at 37 ºC, filtered through Centri-sep, evaporated to dryness, and dissolved in 5 μL of water. 6X loading dye (1 μL) was added to each sample. Samples were loaded onto 15% TBE non-denaturing gel, run at 150 V for 50 min, stained with EtBr, and imaged. CutSmart Buffer® (NEB): 10 mM TRIS-HCl, 20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9 (25ºC).

**Aldol Reaction Purification Method Optimization**

0.1 pmol of aldehyde-propro-PEG oligo aldol reaction product from biotinylated reactant addition to dry oligo was removed and evaporated to dryness. 10 pmol of aldehyde negative library was added to oligo and vortexed. Solution was purified by Princeton column, evaporated

to dryness, and dissolved in 10 μL of water. 5 μL of NEB beads were washed with Tween buffer 3X (500 μL) and resuspended in 10 μL of 2X binding buffer. Magnetic beads were added to the aqueous oligo solutions and incubated for 1 hr. at room temperature with rotation. The beads were magnetized, and the buffer was removed and discarded. The beads were washed with the following protocol:

500 μL of cold, 1X Tween buffer added to beads in Eppendorf tube. Tubes are briefly and mildly vortexed (1 second) and rotated for 5 min a room temperature. Tubes are very briefly spun down to remove liquid from top of cap and placed on magnetic rack for 1 min. Buffer is removed and discarded. Brief centrifuge of tubes again, place on magnetic rack, and remove any remaining buffer. Repeat process 7 X. Repeat process 2 X with 1X binding buffer after Tween buffer washes. Resuspend beads with water to original volume.

1 μL of beads was removed from tube and added to 19 μL polymerase mix (1X Hot start Taq polymerase, forward primer (PFa, 1 μM), reverse primer (PRa, 1 μM)). PCR was performed for 20 cycles and immediately removed and put on ice. Crude PCR products purified by MinElute PCR Purification Kit. Samples evaporated to dryness, suspended in 10 μL of water, and vortexed. Absorbance was measured on Nanodrop UV-Vis spectrophotometer.

0.1 μg of purified PCR product was added to PCR tube followed by water for a total volume of 10 μL. CutSmart buffer (10X, 2μL) and 7 μL of water added to oligo. Restriction enzyme (1 μL, 20 U/μL) was added last. Tubes were hand mixed and incubated at 37º C for 90 min. 4 μL of 6X loading dye was added to each digest reaction solution to stop the enzyme activity. 12 μL of digest reaction was loaded onto 15% non-denaturing gel and run at 150 V for 1 hr. Gel was removed from apparatus, stained with EtBr, and imaged under UV light.

**Selection Positive Control Incubation Optimization**

0.1 pmol of biotin-Ta_XbaI-PEG and 10 pmol of ketone-PEG negative library were combined to make a 1:100 mixture. 4 identical samples were prepared. The oligos were evaporated to dryness on speed vacuum. Two of the samples were dissolved in 10 µL of water and the other two were dissolved in 8 µL of water and 2 µL of DMSO. 5 µL of prewashed magnetic beads in 10 µL of 2X Binding buffer were added to the oligos. One aqueous sample and one 10% DMSO sample were rotated at room temperature for 1 hr. The other two samples were incubated at 40 ºC with periodic, mild vortexing for 1 hr. Magnetic beads were washed, PCR amplified, digested, and gel loaded as described in above section.

**DNA Sequencing**

**Illumina Primers**

**iTurS_i7_D707C** CAA GCA GAA GAC GGC ATA CGA GAT CTG AAG CT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGG ATC CGA GCT CCA CGT G

**iTurS_i7_D708C** CAA GCA GAA GAC GGC ATA CGA GAT TAA TGC GC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATCT GG ATC CGA GCT CCA CGT G

**PRIMEC_D** AAT GAT ACG GCG ACC ACC GAG ATC T ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CGT ACG GTC GAC GCT AGC

**Post-Sequencing Analysis**

Pre-processing of the sequencing data was performed by merging the raw Illumina paired-end read .fastq files using the PEAR software[10]. The sequences were trimmed to 48nt by removing the sequencing primer codes using a Python script. All post processing was performed using the FASTAptamer software[11]. The sequences we clustered using a Levenshtein edit distance of 3.

Enrichment was determined by comparing normalized reads per million (RPM) values the starting library sequences and the post-selection sequences.

## 4.4 REERENCES


1. Xu, T.-H.; Yan, Y.; Harikumar, K. G.; Miller, L. J.; Melcher, K.; Xu, H. E., Streptavidin Bead Pulldown Assay to Determine Protein Homooligomerization. *Bio-protocol* **2017,** *7* (22), e2901.

2. Leunissen, M. E.; Dreyfus, R.; Cheong, F. C.; Grier, D. G.; Sha, R.; Seeman, N. C.; Chaikin, P. M., Switchable self-protected attractions in DNA-functionalized colloids. *Nature Materials* **2009,** *8*, 590.

3. Lindblad-Toh, K.; Winchester, E.; Daly, M. J.; Wang, D. G.; Hirschhorn, J. N.; Laviolette, J.-P.; Ardlie, K.; Reich, D. E.; Robinson, E.; Sklar, P.; Shah, N.; Thomas, D.; Fan, J.-B.; Gingeras, T.; Warrington, J.; Patil, N.; Hudson, T. J.; Lander, E. S., Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics* **2000,** *24*, 381.

4. Persson, H.; Kvist, A.; Vallon-Christersson, J.; Medstrand, P.; Borg, Å.; Rovira, C., The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nature Cell Biology* **2009,** *11*, 1268.

5. Wang, F.-B.; Rong, Y.; Fang, M.; Yuan, J.-P.; Peng, C.-W.; Liu, S.-P.; Li, Y., Recognition and capture of metastatic hepatocellular carcinoma cells using aptamer-conjugated quantum dots and magnetic particles. *Biomaterials* **2013,** *34* (15), 3816-3827.

6. Jutras, B. L.; Verma, A.; Stevenson, B., Identification of novel DNA binding proteins using DNA affinity chromatography-pulldown. *Current protocols in microbiology* **2012,** *CHAPTER*, Unit1F.1-Unit1F.1.

7. Boffey, S. A., Restriction Endonuclease Digestion and Agarose Gel Electrophoresis of DNA. In *Experiments in Molecular Biology*, Slater, R. J., Ed. Humana Press: Totowa, NJ, 1986; pp 17-27.

8. F., J. G., Forty Years of In Vitro Evolution. *Angewandte Chemie International Edition* **2007,** *46* (34), 6420-6436.

9. Silverman, S. K., Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends in Biochemical Sciences* **2016,** *41* (7), 595-609.

10. Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A., PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **2014,** *30* (5), 614-620.

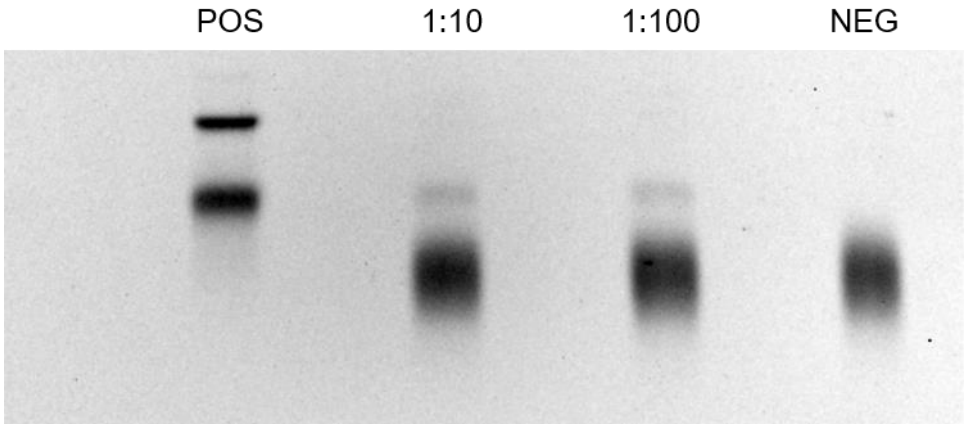
11. Alam, K. K.; Chang, J. L.; Burke, D. H., FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Molecular Therapy. Nucleic Acids* **2015,** *4* (3), e230.

12. Nam, J.-M.; Park, S.-J.; Mirkin, C. A., Bio-Barcodes Based on Oligonucleotide-Modified Nanoparticles. *Journal of the American Chemical Society* **2002,** *124* (15), 3820-3821.

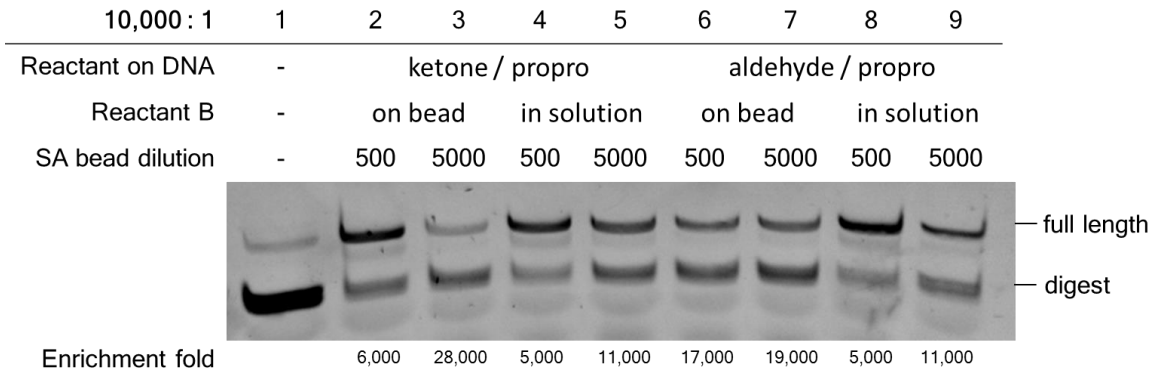
## S3 SUPPORTING INFORMATION

### S3.1 Selected Gel Images

#### S3.1.1 Gel Extraction PAGE of Biotinylated Selection Control Libraries

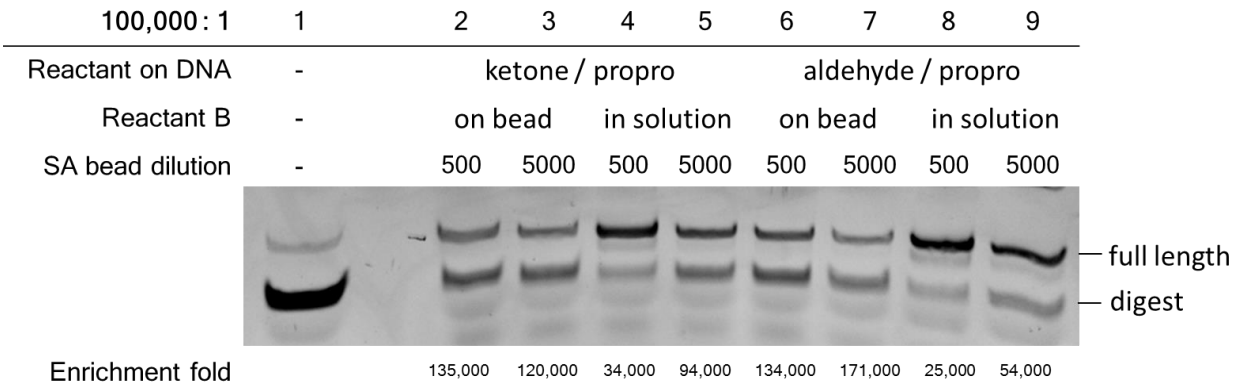


### S3.1.2 10,000:1 Aldol Reaction Selection




Reaction solvent: DCE

### S3.1.3 100,000:1 Aldol Reaction Selection




Reaction solvent: DCE

**S3.2: Top 100 reads of selection readout and initial library.**

Sequence reads normalized to reads per million (RPM).

| Rank | Sequence | Initial (RPM) | Selection (RPM) |
|---|---|---|---|
| 1 | CGTACGGTCGACGCTAGCTGGATATCACTGCACGTGGAGCTCGGATCC | 569.58 | 590795.45 |
| 2 | CGTACGGTCGACGCTAGCGAAGTTGGTTTACACGTGGAGCTCGGATCC | 2.60 | 38.18 |
| 3 | CGTACGGTCGACGCTAGCCTAGCTCATTCTCACGTGGAGCTCGGATCC | 2.60 | 79.11 |
| 4 | CGTACGGTCGACGCTAGCGCTATATCTTATCACGTGGAGCTCGGATCC | 2.60 | 81.83 |
| 5 | CGTACGGTCGACGCTAGCTACTTCTGTTTACACGTGGAGCTCGGATCC | 2.60 | 80.46 |
| 6 | CGTACGGTCGACGCTAGCATCATTTTTCACCACGTGGAGCTCGGATCC | 2.60 | 105.02 |
| 7 | CGTACGGTCGACGCTAGCTTGACAAGCCTTCACGTGGAGCTCGGATCC | 2.60 | 45.00 |
| 8 | CGTACGGTCGACGCTAGCGGGCTATCGTGTCACGTGGAGCTCGGATCC | 2.60 | 60.01 |
| 9 | CGTACGGTCGACGCTAGCTGCTGTGTCCGCCACGTGGAGCTCGGATCC | 2.60 | 27.29 |
| 10 | CGTACGGTCGACGCTAGCATGACCCCCGGCCACGTGGAGCTCGGATCC | 2.60 | 24.55 |
| 11 | CGTACGGTCGACGCTAGCACAATTTCATCTCACGTGGAGCTCGGATCC | 2.60 | 87.29 |
| 12 | CGTACGGTCGACGCTAGCACCCAATAACCACACGTGGAGCTCGGATCC | 2.60 | 20.46 |
| 13 | CGTACGGTCGACGCTAGCCATTTACCTATCCACGTGGAGCTCGGATCC | 2.60 | 50.46 |
| 14 | CGTACGGTCGACGCTAGCTTTGAGTATAGCCACGTGGAGCTCGGATCC | 2.60 | 32.73 |
| 15 | CGTACGGTCGACGCTAGCCTAACTACTCCACACGTGGAGCTCGGATCC | 2.60 | 36.82 |
| 16 | CGTACGGTCGACGCTAGCTCAATAAAGGGGCACGTGGAGCTCGGATCC | 2.60 | 34.10 |
| 17 | CGTACGGTCGACGCTAGCTAGTCCGTCATGCACGTGGAGCTCGGATCC | 2.60 | 46.37 |
| 18 | CGTACGGTCGACGCTAGCCGTTATAGTAAACACGTGGAGCTCGGATCC | 2.60 | 51.83 |
| 19 | CGTACGGTCGACGCTAGCCCGTCTCCTGGACACGTGGAGCTCGGATCC | 2.60 | 34.10 |
| 20 | CGTACGGTCGACGCTAGCGCTGCTTTATTCCACGTGGAGCTCGGATCC | 2.60 | 38.19 |
| 21 | CGTACGGTCGACGCTAGCCTGGTCAACTATCACGTGGAGCTCGGATCC | 2.60 | 32.73 |
| 22 | CGTACGGTCGACGCTAGCTTCAGTTCTTTCCACGTGGAGCTCGGATCC | 2.60 | 75.02 |
| 23 | CGTACGGTCGACGCTAGCAAATTATCAATCCACGTGGAGCTCGGATCC | 2.60 | 72.29 |
| 24 | CGTACGGTCGACGCTAGCATGGGTGATCTTCACGTGGAGCTCGGATCC | 2.60 | 16.37 |
| 25 | CGTACGGTCGACGCTAGCTGGCATACTTGTCACGTGGAGCTCGGATCC | 2.60 | 58.65 |
| 26 | CGTACGGTCGACGCTAGCCTGCTTCTGTCACACGTGGAGCTCGGATCC | 2.60 | 47.73 |
| 27 | CGTACGGTCGACGCTAGCTGACAACGATAGCACGTGGAGCTCGGATCC | 2.60 | 20.46 |
| 28 | CGTACGGTCGACGCTAGCAACTTCACTGTTCACGTGGAGCTCGGATCC | 2.60 | 42.28 |
| 29 | CGTACGGTCGACGCTAGCTCCCTTATCTGGCACGTGGAGCTCGGATCC | 2.60 | 32.73 |
| 30 | CGTACGGTCGACGCTAGCTTCACTCTCTACCACGTGGAGCTCGGATCC | 2.60 | 61.37 |
| 31 | CGTACGGTCGACGCTAGCCAATCCGACACACACGTGGAGCTCGGATCC | 2.60 | 50.46 |
| 32 | CGTACGGTCGACGCTAGCTCGGTGTTGACGCACGTGGAGCTCGGATCC | 2.60 | 65.47 |
| 33 | CGTACGGTCGACGCTAGCCTGAGATGTCCCCACGTGGAGCTCGGATCC | 2.60 | 21.83 |
| 34 | CGTACGGTCGACGCTAGCGGTAAATAACATCACGTGGAGCTCGGATCC | 2.60 | 38.19 |
| 35 | CGTACGGTCGACGCTAGCATGATAGTCTTACACGTGGAGCTCGGATCC | 2.60 | 36.82 |
| 36 | CGTACGGTCGACGCTAGCACTCATTCCGGTCACGTGGAGCTCGGATCC | 2.60 | 45.00 |
| 37 | CGTACGGTCGACGCTAGCAGTATGGAGAAGCACGTGGAGCTCGGATCC | 2.60 | 24.55 |

| | | | |
|---|---|---|---|
| 38 | CGTACGGTCGACGCTAGCATTCATCCCCGGCACGTGGAGCTCGGATCC | 2.60 | 24.55 |
| 39 | CGTACGGTCGACGCTAGCTACTGTGCATTCCACGTGGAGCTCGGATCC | 2.60 | 68.17 |
| 40 | CGTACGGTCGACGCTAGCAGCTCTCAGCCCCACGTGGAGCTCGGATCC | 2.60 | 38.19 |
| 41 | CGTACGGTCGACGCTAGCTGTGCCTCAGAACACGTGGAGCTCGGATCC | 2.60 | 40.91 |
| 42 | CGTACGGTCGACGCTAGCTAGAACTGTTAGCACGTGGAGCTCGGATCC | 2.60 | 28.64 |
| 43 | CGTACGGTCGACGCTAGCTCCTCCATACGGCACGTGGAGCTCGGATCC | 2.60 | 27.27 |
| 44 | CGTACGGTCGACGCTAGCACAGTAGACATGCACGTGGAGCTCGGATCC | 2.60 | 25.92 |
| 45 | CGTACGGTCGACGCTAGCATAATGCTACACCACGTGGAGCTCGGATCC | 2.60 | 51.83 |
| 46 | CGTACGGTCGACGCTAGCAATCGGTTCTCCCACGTGGAGCTCGGATCC | 2.60 | 25.92 |
| 47 | CGTACGGTCGACGCTAGCGGTTATGTTGGGCACGTGGAGCTCGGATCC | 2.60 | 19.09 |
| 48 | CGTACGGTCGACGCTAGCGCCACGCCGCGGCACGTGGAGCTCGGATCC | 2.60 | 23.18 |
| 49 | CGTACGGTCGACGCTAGCTGCATATACGAACACGTGGAGCTCGGATCC | 2.60 | 66.82 |
| 50 | CGTACGGTCGACGCTAGCCCGCTGAGCCCTCACGTGGAGCTCGGATCC | 2.60 | 28.64 |
| 51 | CGTACGGTCGACGCTAGCTTGCACATTCACCACGTGGAGCTCGGATCC | 2.60 | 47.73 |
| 52 | CGTACGGTCGACGCTAGCTATATTGCTCTCCACGTGGAGCTCGGATCC | 2.60 | 64.09 |
| 53 | CGTACGGTCGACGCTAGCCCGTAATCGTAGCACGTGGAGCTCGGATCC | 2.60 | 40.91 |
| 54 | CGTACGGTCGACGCTAGCTGGGTTTATCTCCACGTGGAGCTCGGATCC | 2.60 | 31.37 |
| 55 | CGTACGGTCGACGCTAGCTGAATGTCTTAACACGTGGAGCTCGGATCC | 2.60 | 42.28 |
| 56 | CGTACGGTCGACGCTAGCCGGATAAGTTCACACGTGGAGCTCGGATCC | 2.60 | 27.27 |
| 57 | CGTACGGTCGACGCTAGCGATTCATCGAGGCACGTGGAGCTCGGATCC | 2.60 | 32.73 |
| 58 | CGTACGGTCGACGCTAGCATGAATAGATCTCACGTGGAGCTCGGATCC | 2.60 | 73.63 |
| 59 | CGTACGGTCGACGCTAGCTTATGTTATTTACACGTGGAGCTCGGATCC | 2.60 | 47.72 |
| 60 | CGTACGGTCGACGCTAGCGACTACATTAAGCACGTGGAGCTCGGATCC | 2.60 | 36.82 |
| 61 | CGTACGGTCGACGCTAGCTCGAATGGGAGACACGTGGAGCTCGGATCC | 2.60 | 28.64 |
| 62 | CGTACGGTCGACGCTAGCGGTGCTGCAGTGCACGTGGAGCTCGGATCC | 2.60 | 25.92 |
| 63 | CGTACGGTCGACGCTAGCCGTGGATTTCTGCACGTGGAGCTCGGATCC | 2.60 | 39.54 |
| 64 | CGTACGGTCGACGCTAGCATGATCTATAGGCACGTGGAGCTCGGATCC | 2.60 | 32.73 |
| 65 | CGTACGGTCGACGCTAGCCCAAAAGGCAGTCACGTGGAGCTCGGATCC | 2.60 | 27.27 |
| 66 | CGTACGGTCGACGCTAGCATCAAACCTAGTCACGTGGAGCTCGGATCC | 2.60 | 58.64 |
| 67 | CGTACGGTCGACGCTAGCCTTGCATCAGCGCACGTGGAGCTCGGATCC | 2.60 | 34.10 |
| 68 | CGTACGGTCGACGCTAGCTAGGTTGGAAAGCACGTGGAGCTCGGATCC | 2.60 | 19.09 |
| 69 | CGTACGGTCGACGCTAGCTAACGGTTTGTCCACGTGGAGCTCGGATCC | 2.60 | 42.28 |
| 70 | CGTACGGTCGACGCTAGCAATTCTAACAAACACGTGGAGCTCGGATCC | 2.60 | 23.18 |
| 71 | CGTACGGTCGACGCTAGCTTTACATATTCTCACGTGGAGCTCGGATCC | 26.00 | 98.19 |
| 72 | CGTACGGTCGACGCTAGCTAACCCGTTCTCCACGTGGAGCTCGGATCC | 2.60 | 46.37 |
| 73 | CGTACGGTCGACGCTAGCTTGGAAGTACCACACGTGGAGCTCGGATCC | 2.60 | 70.91 |
| 74 | CGTACGGTCGACGCTAGCCTACCAACGTAACACGTGGAGCTCGGATCC | 2.60 | 24.55 |
| 75 | CGTACGGTCGACGCTAGCTGCATTTCAGGTCACGTGGAGCTCGGATCC | 2.60 | 35.45 |
| 76 | CGTACGGTCGACGCTAGCTGAGAACAATCCCACGTGGAGCTCGGATCC | 2.60 | 36.82 |
| 77 | CGTACGGTCGACGCTAGCCAATCTTACAGCCACGTGGAGCTCGGATCC | 2.60 | 45.02 |
| 78 | CGTACGGTCGACGCTAGCTTTCTCCGTTTTCACGTGGAGCTCGGATCC | 2.60 | 68.19 |
| 79 | CGTACGGTCGACGCTAGCGCGGCTTCTGGTCACGTGGAGCTCGGATCC | 2.60 | 19.09 |
| 80 | CGTACGGTCGACGCTAGCACTGGAACTTATCACGTGGAGCTCGGATCC | 2.60 | 19.09 |
| 81 | CGTACGGTCGACGCTAGCGGCTAGAGGGGTCACGTGGAGCTCGGATCC | 2.60 | 10.91 |
| 82 | CGTACGGTCGACGCTAGCACGTTGCGATTTCACGTGGAGCTCGGATCC | 2.60 | 27.27 |

| | | | |
|---|---|---|---|
| 83 | CGTACGGTCGACGCTAGCACGAGCACCTCCCACGTGGAGCTCGGATCC | 2.60 | 39.54 |
| 84 | CGTACGGTCGACGCTAGCCGCCATCAACACCACGTGGAGCTCGGATCC | 2.60 | 38.19 |
| 85 | CGTACGGTCGACGCTAGCAGTCTGAATTGGCACGTGGAGCTCGGATCC | 2.60 | 31.36 |
| 86 | CGTACGGTCGACGCTAGCTATTGATGTTAGCACGTGGAGCTCGGATCC | 2.60 | 42.28 |
| 87 | CGTACGGTCGACGCTAGCTCACATCGCAATCACGTGGAGCTCGGATCC | 2.60 | 73.65 |
| 88 | CGTACGGTCGACGCTAGCGTGGGAAACTAACACGTGGAGCTCGGATCC | 2.60 | 19.09 |
| 89 | CGTACGGTCGACGCTAGCCAGGCCTATCGGCACGTGGAGCTCGGATCC | 2.60 | 36.82 |
| 90 | CGTACGGTCGACGCTAGCCTGCTGCAAAGCCACGTGGAGCTCGGATCC | 2.60 | 27.27 |
| 91 | CGTACGGTCGACGCTAGCTCCCGATCTTAGCACGTGGAGCTCGGATCC | 2.60 | 28.64 |
| 92 | CGTACGGTCGACGCTAGCGGTGTTCTGATGCACGTGGAGCTCGGATCC | 2.60 | 34.10 |
| 93 | CGTACGGTCGACGCTAGCTATAGACACCGGCACGTGGAGCTCGGATCC | 2.60 | 42.28 |
| 94 | CGTACGGTCGACGCTAGCGTATCAGCCGCCCACGTGGAGCTCGGATCC | 2.60 | 24.55 |
| 95 | CGTACGGTCGACGCTAGCTGCACTTCCCTGCACGTGGAGCTCGGATCC | 2.60 | 36.82 |
| 96 | CGTACGGTCGACGCTAGCGTAAGTTGTACACACGTGGAGCTCGGATCC | 2.60 | 31.36 |
| 97 | CGTACGGTCGACGCTAGCAGCTGCGCGAGGCACGTGGAGCTCGGATCC | 2.60 | 23.18 |
| 98 | CGTACGGTCGACGCTAGCTGCCTCTCTCTCCACGTGGAGCTCGGATCC | 2.60 | 84.54 |
| 99 | CGTACGGTCGACGCTAGCAAGATATACAGTCACGTGGAGCTCGGATCC | 2.60 | 15.00 |
| 100 | CGTACGGTCGACGCTAGCGACTATGTACCACACGTGGAGCTCGGATCC | 2.60 | 53.18 |

**ABBREVIATIONS**

| | |
|---|---|
| AMA | 1:1 ammonium hydroxide: methylamine |
| Cq | quantification cycle |
| DMT | dimethoxy trityl |
| DNA | deoxyribonucleic acid |
| dsDNA | double stranded DNA |
| EMSA | electrophoretic mobility shift assay |
| EtBr | ethidium bromide |
| HPLC | high performance liquid chromatography |
| MALDI | matrix-assisted laser desorption/ionization |
| MW | molecular weight |
| NEB | New England Biolabs, Inc. |
| NHS | N-hydroxysuccinimide |
| nt | nucleotide |
| PAGE | polyacrylamide gel electrophoresis |
| PEG | polyethylene glycol |
| PS | polystyrene (solid support) |
| RE | restriction enzyme |
| SELEX | systematic evolution of ligands by exponential enrichment |
| ssDNA | single-stranded DNA |
| TBE | TRIS/borate/EDTA |
| TEAA | triethylammonium acetate |

| | |
|---|---|
| TEAB | triethylammonium bicarbonate |
| TRIS | tris(hydroxymethyl)aminomethane |