Title

Development of in vitro selection strategies for generating new catalytic nucleic acids and peptides

Permalink

https://escholarship.org/uc/item/4hf5p45s

Author

Pobanz, Kelsey

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Development of *in vitro* selection strategies for generating new catalytic nucleic acids and peptides
DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry
with a concentration in Chemical Biology

by

Kelsey Pobanz

Dissertation Committee:
Professor Andrej Lupták, Chair
Professor Gregory Weiss
Professor Thomas Poulos

2015

DEDICATION

To

my family and friends

in recognition of their patience and fortitude

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I would like to thank my chair and advisor, Professor Andrej Lupták for all of his support, ideas and patience over the last few years. I really enjoyed being a part of the lab and sharing all the science and outdoor adventures with the group.

I would like to thank my committee members, Professor Gregory Weiss and Professor Thomas Poulos, for taking the time to be on my committee. I am especially appreciative of Professor Greg Weiss' advice on my future endeavors into science.

I have made great friends in the Lupták lab and I could not have done it without the help and support from Randi Jimenez, especially, Cassandra Burke and Marie Myszka.

And last but not least I could not have made it through these years without the love and support of my parents Kris and Deanna Pobanz and my sister and best friend, Hannah Santiago, thank you for hanging in there with me.

CURRICULUM VITAE

Kelsey Pobanz
LinkedIn: www.linkedin.com/in/kelseypobanz/ E-Mail: knpobanz@gmail.com

Education
Ph.D., Bio-organic Chemistry, University of California, Irvine   **2008-2014**
Advisor: Professor Andrej Lupták

Anticipated Thesis Title: Development of selection strategies for generating new and unnatural catalytic nucleic acids

BS, Chemistry, University of Puget Sound, Tacoma, WA   **May 2008**
Advisor: Professor John Hanson

Thesis Title: Synthesis of two possible substrates for T4 lysozyme

Research Experience
Graduate Assistant, University of California, Irvine
Advisor: Prof. Andrej Lupták   **2009-2014**
Duties:
* Cultured a lab stock of T7 RNA polymerase plasmid and then lysed, and purified the resulting enzyme of polymerase for lab use.
* Rotating laboratory duties included DNA oligomer ordering, making lab buffer stocks and reagent stocks, and FPLC maintenance.

Accomplishments, Related Skills:
* In vitro selection of peptide-DNA conjugations in water-in-oil emulsions
   o Designed and optimized a DNA pool, coding for a random peptide, and flanking regions to increase stability and yield of transcripts in coupled transcription and translation reactions.
   o Incorporated unnatural nucleotides into DNA primers for promotion of bio-orthogonal chemistry during selection.
   o Optimized in vitro coupled transcription and translation reactions in water-in-oil emulsions.
   o Analyzed products and controls with SDS-PAGE, nucleic acid PAGE, MALDI-TOF.
* In vitro selection of a galactosidase DNAzyme
   o Selected for a DNAzyme by using gel-shift assays and immobilized bead incubations.
   o Radiolabeled DNA pool to quantify the amount that was eluted from the beads each round of selection.
   o Amplified, implemented error-prone PCR and purified DNA pool during selection.
   o Cloned and transformed DNA into cells and sequenced individual clones.
   o Collected kinetics data of the selected pool and individual clones with a fluorescent substrate, using a fluorometer and microplate reader.

Undergraduate Researcher, University of Puget Sound
Advisor: Prof. John Hanson   2006-2008
Accomplishments:
- Synthesized two possible T4 lysozyme substrates
  - Used organic synthetic methods to make two small molecules
  - Purified intermediates using standard methods and HPLC
  - Used H-NMR and C-NMR to verify compound structure
- Presented at a meeting:
  - Pobanz, Kelsey N.; Hanson, John E. "Total synthesis and evaluation of T4 lysozyme activity of two N-acetylmuramic acid derivatives" 235th ACS National Meeting & Exposition, New Orleans, LA, April 6-10, 2008 (CHED-510).
  - "Synthesis of two possible substrates for T4 lysozyme." 16th Regional Murdock Conference on Undergraduate Research, November 2-3, 2007. Salem, Oregon.

Teaching Experience
University of California, Irvine   September 2008-present
- Medicinal Chemistry Lab
  - Maintained and passaged cancer cell lines for student use.
  - Wrote an enzyme-ligand computational docking lab protocol.
  - Prepared a scratch-wound assay on cancer cells with a compound library in a 96-well format for a high-throughput screening lab.
- Chemical Biology lab
- Organic Chemistry lab

University of Puget Sound September 2005-May 2008
- Peer Advisor, Academic Advising, August 2007 – May 2008
- Physical Chemistry Teacher Assistant, Department of Chemistry, August 2007-May 2008
- Organic Chemistry teacher assistant, Department of Chemistry, August 2006-May 2007
- Teacher assistant, Academic Access Programs, June 2006 – July 2006
- General Chemistry teacher assistant, Department of Chemistry, August 2005-May 2006

Professional Memberships and Other Activities
- American Chemical Society, June 2007-present
- Philanthropic Education Organization (P.E.O.), December 2009-present
- Iota Sigma Pi Women's Honors Society, May 2009-present
  - Chapter treasurer, June 2012-present—managed the club budget for outreach activities and recruited new members
- Student Affiliates of the American Chemical Society, August 2006-present
  - Education Chair, August 2007-2008
- Habitat for Humanity, August 2004-May 2006

ABSTRACT OF THE DISSERTATION

Development of *in vitro* selection strategies for generating new catalytic nucleic acids and
peptides
By

Kelsey Pobanz

Doctor of Philosophy in Chemistry

with a concentration in Chemical Biology

University of California, Irvine, 2015

Professor Andrej Lupták, Chair

Design of a nucleic acid library is an essential foundation for an *in vitro* selection. Increasing the starting diversity of a pool will increase the sampling of sequence space but other strategies must be considered to access rare secondary structures including varying random region length, modularity of the secondary structure and introducing complexity. Efforts to understand these characteristics are reviewed.

In vitro evolution of peptides and proteins has been enabled by display technologies such as phage and mRNA display, but each has limitations. To bypass these, a new system that utilizes peptide-DNA conjugations was designed by employing *in vitro* compartmentalization, which colocalized genotype and phenotype allowing selections. This system could permit a 10^{14} starting diversity. To optimize DNA and RNA stability in the coupled transcription/translation reaction and the compartmentalized reaction, DNA constructs were designed to code for reporter peptides and altered to increase mRNA stability and translation

yields. Pool construction was then optimized and the flanking fixed sequences as well as four bioorthogonal moieties were installed. A mock selection was optimized using an electrophoretic shift mobility assay (EMSA) leading to the same result in the positive and negative controls, which suspended the project.

In vitro selection of a ssDNA pool was performed in the presence of fluorescein mono-β-D-galactopyranoside (FMG) and fluorescein di(β-D-galactopyranoside (FDG), substrates for a β-galactosidase. Twenty-one rounds of selection were performed with increasing stringency including a decrease of substrate to 2 μM, a decrease of incubation to 10s and addition of 7 M urea and incubation at 95 ℃ after the selection reaction. Fluorescence increases were observed when FDG and FMG were incubated with ssDNA from rounds 19 and 21 as well as with individual DNA clones from round 21. Incubation of individual clones with boronate affinity gel did not result in detectable binding, indicating that galactose had not been transferred from the substrate to the DNA clone. Reselection of the DNA pool was performed by competitively eluting bound ssDNA sequences with 0.1 M ribose but it did not result in enrichment of the DNA pool for active sequences over the six additional rounds of selection.

Chapter 1

The importance of pool design on the outcome of *in vitro* selections

1.1 Introduction

Darwinism is a theory of biological evolution, which states that there is variation among individuals in a population, this variation is heritable and that the survival and reproduction of individuals is dependent on which traits they inherit. *In vitro* selection mimics this idea in a laboratory setting. With the development of tools such as chemical synthesis of oligomers, PCR and RT-PCR, the *in vitro* selection of functional nucleic acids was made possible in the late 1980s. On a practical level, *in vitro* selections of DNA and RNA were not possible on a larger scale until the late 1980s when DNA libraries could be chemically synthesized more cost effectively and efficiently using phosphoramidites and an automated process.[1]

Molecular evolution requires the phenotype be linked to its genotype, and in the case of RNA and DNA, functionality (phenotype) and the encoding strand (genotype) are the same.[2] A nucleic acid *in vitro* selection starts with a population of DNA sequences containing a random region flanked by fixed sequences that are required for amplification and transcription. RNA and DNA *in vitro* selections have been done with random regions ranging from 20- to 200-nucleotides (nts) and up to 10^{16} starting diversity.[3] Selection occurs when the pool of random sequences is required to perform a function, such as reaction catalysis or ligand binding and the active sequences are separated from the majority of inactive sequences. Once these selected sequences are amplified, a new generation of variants is available for continual rounds of selection leading to an evolved RNA or DNA species that is efficient and specific for the desired function.

In vitro selections for functional RNA that bind a ligand, or RNA aptamer, have been extensive. RNA aptamers have been discovered for biological molecules, including ATP (Figure 1.1a), other adenosine derivatives, guanosine derivatives, amino acids, cofactors and antibiotics.[2,4] Selections have also successfully identified RNA aptamers for ions,[5] small molecules,[6] peptides[7] and proteins.[8] The first example of an *in vitro* selected catalytic RNA, or ribozyme, was in 1993 with the discovery of a ligation catalysis reaction involving a oligonucleotide substrate and the novel ribozyme.[9] One subset, the class I ligase (Figure 1.1b), was studied extensively and later generations were converted into a polymerase capable of template-directed extension of RNA.[10] While there are no naturally occurring DNA enzymes or deoxyribozymes, a lead dependent deoxyribozyme capable of cleaving a RNA phosphodiester bond was discovered in 1994.[11] Since then many other deoxyribozymes have been discovered using *in vitro* selection including but not limited to catalysis of DNA and RNA ligation,[12,13] DNA phosphorylation,[14] adenylation,[15] depurination,[16] and a Diels-Alder reaction.[17]
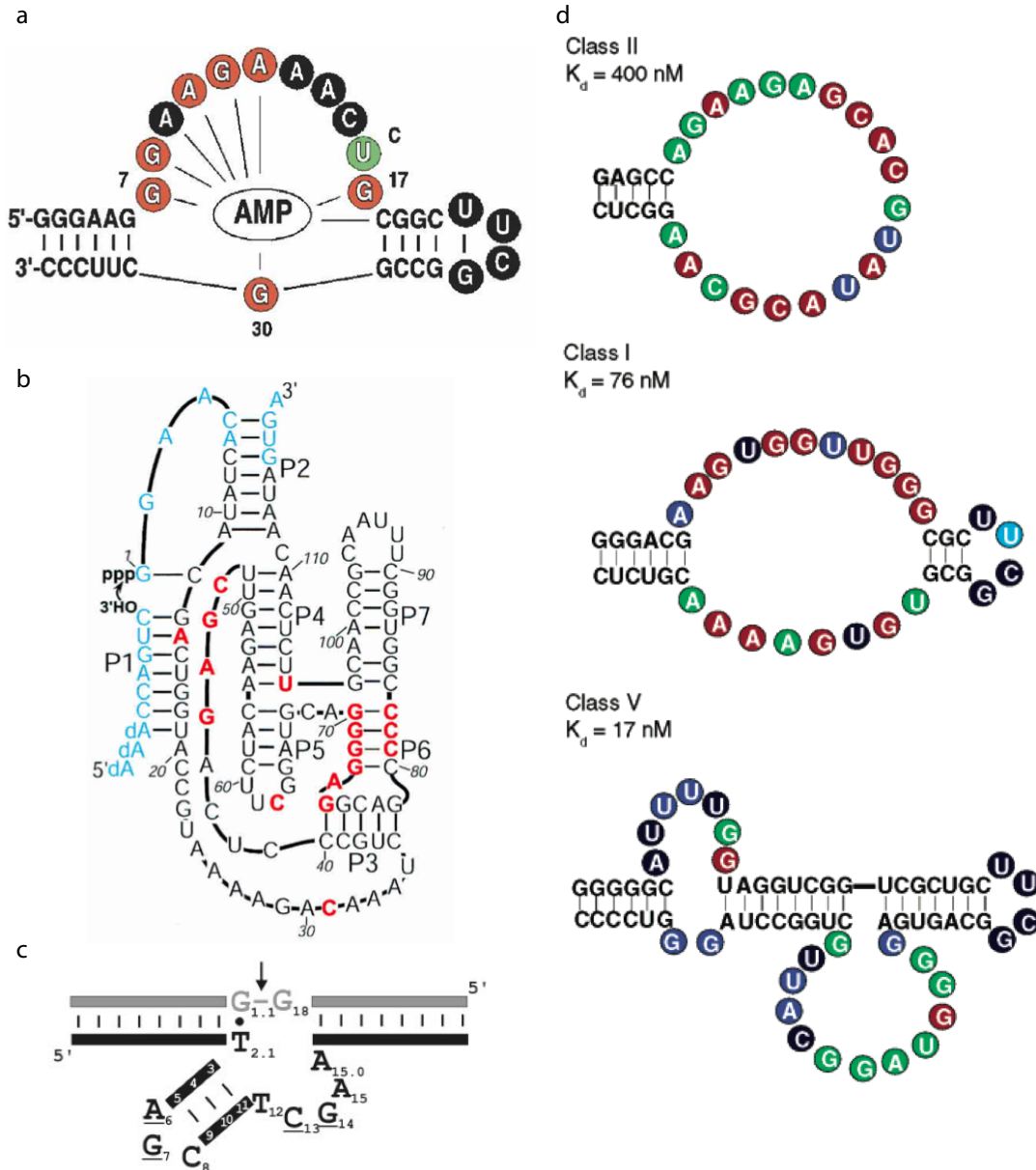
Figure 1.1. Examples of functional nucleic acids discovered with *in vitro* selection. Proposed secondary structures are shown for the a) ATP aptamer,[4] b) class I ligase,[9] c) the 8-17 deoxyribozyme[30] and d) class I, II, and V GTP aptamers with increasing complexity and dissociation constants.[41,44]

In general, an *in vitro* selection experiment requires a selection step, a separation step and an amplification step. Most selections depend on a well-designed separation step as a path to a successful selection of new functional nucleic acids. Some studies have strived to understand the nature of the initial nucleic acid pool, including the effect of length on selection success, how modularity affects structural diversity and strategies to access increased complexity to find rare and new functional nucleic acids. Understanding what can reasonably be expected from a starting library and how the design will affect the outcome of a selection and its ability to isolate new and rare functional nucleic acids may be an important aspect to consider moving forward.

1.2 Effect of random region length on selection success

One dilemma that is considered at the beginning of an *in vitro* selection is the random region length. While shorter lengths will cover all or a large percentage of sequence space, longer regions are thought to allow for more complex structures that may be needed for a functional RNA or DNA. In an effort to start with the largest amount of initial members and access more complex active structures, selections have been successfully carried out with pools of up to 10^{15} members. A large starting diversity may help contribute to the probability of a successful outcome, but the length of pool members is an important aspect that may also contribute to evolving a functional nucleic acid. Successful outcomes have been observed at both extremes with the class I ligase originally evolved from a pool of 220 nts[9] while the small isoleucine aptamer was successfully evolved from a pool with a random region of 22 nts.[18] The question of optimal pool length has been studied in several different ways, computationally and experimentally.

Several labs have also approached the question of random region length

computationally.[19,20] Sabeti derived an equation that estimated the probability of finding a motif in a random sequence which considered the size, modularity and redundancy of the sequence. The probability of finding the hammerhead ribozyme (length=43 nt) in a 220-nt random region versus a 72-nt random region increased 200-fold. This increased the probability of finding the sequence in a small pool of 1.8×10^8 molecules from 0.05 to 0.999, which was significant if motifs rarer than the hammerhead motif were sought in an *in vitro* selection of 1×10^{15} molecules.[21] Knight and Yarus built on this work by eliminating some of the approximations used to estimate the abundance of functional motifs in random pools and were able to consider active sequences and their probability of folding. They used previously discovered functional RNA, an isoleucine aptamer with two modules and the common hammerhead ribozyme that contains three modules and incorporated paired regions that were unspecified sequences but required for structural formation. In both examples, the probability of finding the motifs solely based on length ranged from 1.74×10^{-8} to 7.87×10^{-6} for random regions of 50- to 150-nts, respectively, for the hammerhead motif and 3.46×10^{-9} to 8.94×10^{-8} for the isoleucine aptamer motif for 50- to 150-nts, respectively. However, poorer folding contributed to lower probabilities for each motif at longer lengths which led to maximal probabilities of 4.27×10^{-12} to 8.61×10^{-10} for the hammerhead motif ranging from 50- to 150-nt and 1.88×10^{-10} to 1.06×10^{-9} for the isoleucine motif of the same lengths, indicating that the payoff for using a longer random region decreased due to misfolding. Nevertheless, at both lengths, the probabilities were within the realm of the starting diversity of an *in vitro* selection. This calculation may have been especially useful since they incorporated, not only sequence requirements for the example motifs but also structural requirements, which were not conserved in the primary sequences but in the

secondary sequences.

Sabeti studied the effect of random region length experimentally by ligating arbitrary PCR fragments to class II and class III RNA ligases previously isolated from *in vitro* selections[9,22] and created four new libraries with 10^{12} members. The catalytic activities of the new libraries were compared to the originating libraries and the median effect was a 5-fold decrease in activity, which was inconvenient for selecting smaller, simpler motifs, but worth the cost when accessing rarer, complex structures.[21] Huang and coworkers examined selection success of an RNA-mediated CoA-thioester synthesis using random regions of 30-, 60-, 100-, and 140-nts.[23,24] Only sequences from the 30- and 60-nt random region pools were isolated, indicating that the abundance of smaller and faster replicating sequences will outcompete longer sequences containing the same active core structure. This led to experimental work by Legiewicz who observed the outcomes of six parallel *in vitro* selections for the previously discovered isoleucine aptamer using random region lengths of 16-, 22-, 26-, 50-, 70- and 90-nts. This aptamer selection was chosen because previous work indicated that there was a high probability that the recurrence of the motif in all pool sizes would occur.[25-28] Unexpectedly, the aptamer was 20- to 40-fold more abundant in the 50- and 70-nt sized pools compared to all other lengths.[29]

As a follow up to Sabeti's work with appended excess length, Schlosser examined the effect of excess length on the 8-17 DNA motif in search for a more complex and efficient RNA cleaving deoxyribozyme.[30] The 8-17 motif generally contains 15 nucleotides and is defined by a G-T wobble base pair adjacent to a tri-loop and bulge and is flanked by two helices (Figure 1.1c).[31] Although a longer random region was available for accessing more complex structures, 90% of the resulting active sequences were closely related to the recurring small

motif that was previously discovered. Not only was this structure rediscovered over variable incubation times, it dominated the selection and was immune to any inhibitory effects from the excess sequence of the pool. The compact and stable nature of the motif, large catalytic rate and low sequence requirements may eliminate any inhibitory effect excess sequence contributes.

Isolation of small and more simple motifs from *in vitro* selections generally do not benefit from longer random regions, since they have been shown to have some inhibitory effect which may be caused by misfolding or masking of the active sequence. Also, the abundance of these motifs does not increase significantly in longer random regions. The probability of isolating larger or more complex functional nucleic acids increases with increasing length, up to a point, and usually outweighs possible inhibitory effects.

1.3 Structural diversity

Consideration of random region length of a pool is only one factor that affects the success of an *in vitro* selection. Since nucleic acids only contain four nucleotides, many secondary motifs such as pseudoknots, tetraloops and uridine turns occur often and represent energetically efficient solutions where only a few key nucleotides are defined. For example, the GNRA tetraloop motif binds proteins or forms tertiary contacts within the large RNA structure of *E. coli* 16S rRNA which contains 9 GNRA tetraloops[32] or ligand binding as is the case in the small ATP aptamer where ATP binds to form the complete GNRA-like tetraloop (Figure 1.1a).[33] The majority of the primary sequence can be easily swapped while still maintaining the same secondary structure. This degeneracy leads to many primary sequences that have the ability to fold into very similar structures that may exhibit similar activities as well. Many authors have studied structural diversity by dividing the primary sequence into

three categories: Key conserved positions required for activity, positions required for secondary and tertiary structure and conserved positions. This approach has led to several observations that give insight into the feasibility of a successful *in vitro* selection.

Abundance of RNA folds. A simplistic approach to understanding the difference between a random primary sequence versus a random structure is to consider the probability of each case. In the case of a 200-nt random region, there are over 10^{120} possible sequences and if a starting pool contained 10^{15} molecules, the probability of finding any functional nucleic acid larger than 28 bases would be negligible. However, many *in vitro* selections have successfully found larger ribozymes or aptamers, pointing to only needing a select few conserved bases and structural support. For example, the hammerhead ribozyme only needs approximately 14 specifically placed nucleotides, which leads to a higher probability of isolation. Wedel outlined a simple example of this with a cloverleaf ribozyme of 76 bases containing 16 conserved bases. The probability of this example including helices was 1.4×10^{-20} in a 200-nt pool. This probability went up to 6×10^{-13} if extra nucleotides in the pool are allowed to be interspersed within the structure which led to approximately 600 molecules in a starting pool of 10^{15} that would match the required structure.[34] If the random region was reduced to 100 nucleotides, only one molecule matching the structural requirements would be found in a starting pool of 10^{15} molecules.

Estimating the abundance of functional motifs in random pools relies heavily on considering the modularity of these motifs. Modularity is defined as the number of interacting segments that form the motif.[21] For example, a stem-loop has a modularity of one, the isoleucine aptamer has a modularity of two while the branched internal loop of a hammerhead ribozyme has a modularity of three. Knight utilized combinatorial analysis to

determine the probability of finding a sequence within a longer random region. Utilizing the minimal structures of the isoleucine aptamer[18] and the hammerhead ribozyme[27] as test cases for the robustness of their model, combinatorial analysis was used to determine the probability of finding a sequence within a longer random region. The minimal isoleucine aptamer with m=2 was represented by [4,8] while an average structure as depicted in Figure 1.2a was represented by [6,10] and the minimal hammerhead ribozyme with m=3 was represented by [1,7,4] as depicted in Figure 1.2a and an average structure was represented by [6,10,9]. If considering an average structure, it was calculated that a pool size of 2.1×10^5 was required to find the isoleucine aptamer but once correct folding was considered this would increase to 4.1×10^9, conversely for the average hammerhead ribozyme structure the pool size was determined to be 1.1×10^{10} and increased to 1.6×10^{10} when correct folding was considered. One large conclusion that was drawn that motifs with evenly divided, smaller modules were more abundant than asymmetric and larger ones. Over 1 million unique sequences, considering a 100-nt random sequence, were calculated for a structure represented by [5,5,5,5] while only ~8000 unique sequences existed for a structure of [17,1,1,1].[19,35] Knight expanded on previous calculations of Sabeti and others and determined that the modules that are split evenly will increase the probability of occurrence in a random pool based on a corrected Poisson frequency of occurrence. More broadly, the authors determined that in an *in vitro* selection experiment with 10^{15} molecules, a motif of up to 26-nt with m=1 in a 40-nt random region and a motif with a maximum of 34-nt and modularity of 4 in a 100-nt random region would be possible. An example of modularity beyond probability calculations was the minimized nucleotide synthase ribozyme that was determined to have 5 helices and a required ~40 nucleotides for activity. With a modularity of five, the probability of

isolating the structure from 228-nt random pool increased from 1 in 10^{22} to 1 in 10^{15}.[36]
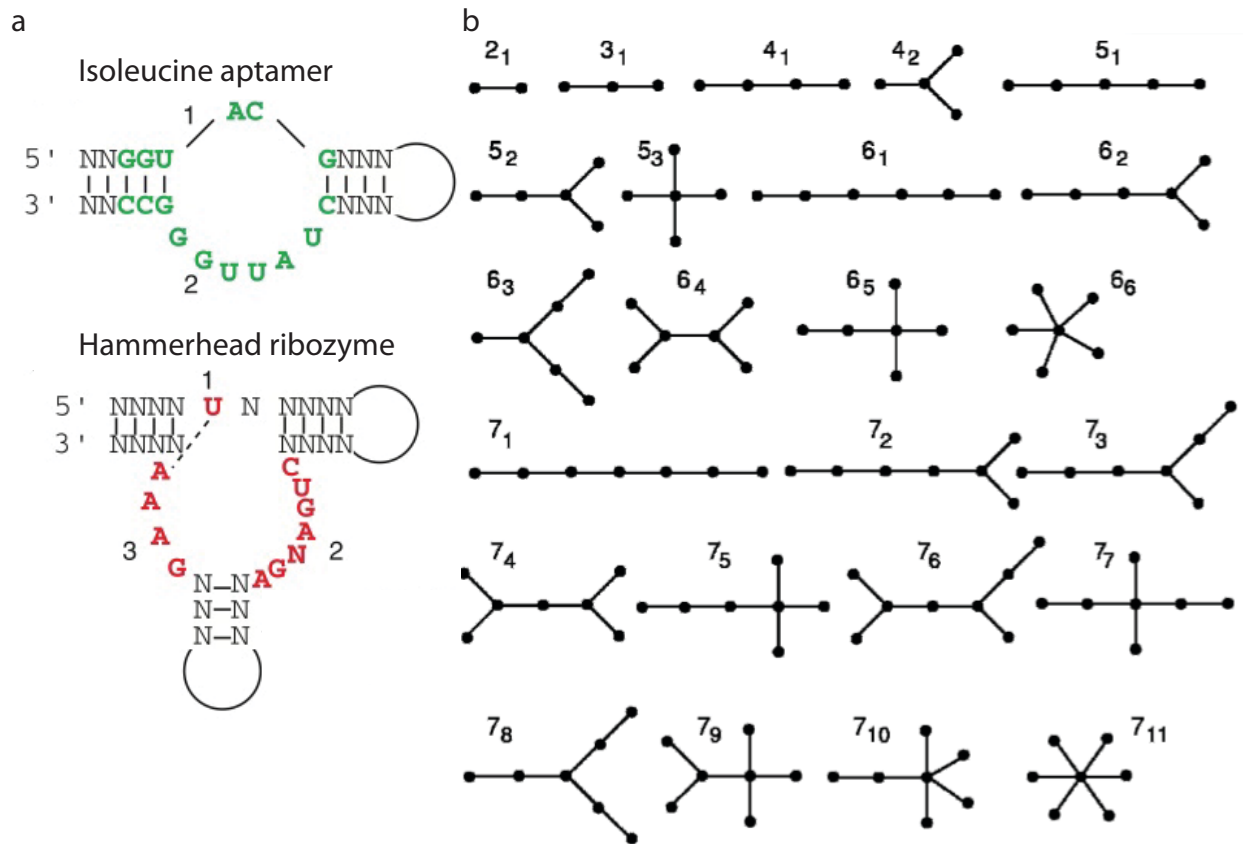
Figure 1.2. Computational efforts to understand structural diversity. Simplification of secondary structures led to a) defining the minimal structures of the isoleucine aptamer and hammerhead ribozyme as modules represented by [6,10] and [1,7,4], respectively, and b) creating a complete set of tree graphs containing two to seven vertices that represent 2D RNA folds defined by the lines as stems and vertices as loops, bulges and junctions.

Another approach to understand the structural abundance was calculated first by Fontana and improved by Gevertz by utilizing graph theory to analyze complete sets of RNA secondary structures for complexity in random pools of 25-, 40-, 60-, 80-, and 100-nts.[37,38] The authors used the Vienna folding package[39] on libraries of 10^4 random RNA sequences and then converted the 2D topologies to tree graphs with bulges, loops and junctions as the vertices and stems as the lines (Figure 1.2b). This approach collapsed a diverse pool of sequences into easily categorized shapes without consideration of details of base-pair information or stem and loop sizes, which allowed analysis of an entire random pool instead of a smaller sample size. More than 90% of folded structures were found to be simple topologies such as stem-loop motifs. Structure complexity as defined by vertex number here went up with length of the pool and a general relation was determined to that a pool of length L would be most abundant with L/20 stems. For example, a 100-nt pool would have the highest abundance of 5-stemmed structures and their results showed that 40% of the 100-nt pool had 6 vertices. While the Vienna folding package is unable to form more complicated tertiary structures such as pseudoknots, these results were insensitive to folding stability parameters as well as the pool size analyzed. An application of this method to the three-stem class V GTP aptamer, represented by 4_1 in graph theory, and was calculated to have the highest abundance in a 60-nt random RNA pool, which corresponded with the aptamer that was discovered, *in vitro*.[40]

Computational efforts to understand the abundance of RNA folds rely on the assumption that the population of more rare and complex RNA secondary structures will exhibit higher or new activities. Carothers experimentally supported this assumption by using information

theory to connect structural diversity to activity by studying a set of 11 RNA aptamers that bound GTP (Figure 1.1d).[40,41] The relationship between sequence, conformation and function is defined as the fitness landscape where each sequence is given a value corresponding to function.[42,43] Each member of the RNA aptamer set was mutagenized up to 21% per position and reselected for GTP-binding activity, which allowed optimization of the variants. The minimal information to define the resulting sequences was used to compute complexity and the dissociation constants were determined by binding assay. Dissociation constants for the set of 11 RNA aptamers ranged from 8 μM to 9 nM were seven of the 11 RNA aptamers had a designed stem-tetraloop from the original in vitro selection[44] and six had dissociation constants >300 nM. The relationship observed between activity and complexity showed that every 10-fold improvement in binding was 1000x less frequent in a random pool. This also held true in the case of two unique RNA ribozyme ligases[45,46] and it was speculated that this relationship would hold as long as additional complexity contributed to stability of the overall RNA fold.

Evolving more complex structures. *In vitro* selections tend to find the simplest solutions, which has been supported repeatedly experimentally. Knight calculated that adding one more nucleotide to a minimal motif would make it a factor of 4 more rare in a random pool.[19] More complex structures may exhibit better activities but if they are not magnitudes more efficient than the more abundant and simpler, motif, they may not be discovered in the initial rounds of *in vitro* selection. In general, smaller motifs are observed in higher abundance even if more complex structures are more efficient and consequently all selections must be further optimized by deletion and mutation experiments. Other strategies have been proposed to

increase the abundance of rare secondary structures that may be more efficient as well.

Davis examined a possible strategy designed to isolate more complex structures and thus higher affinity motifs in an *in vitro* selection for GTP aptamers.[44] Many aptamers contain a stem-loop that contributes to the stability of the structure.[4,47-49] A library was designed to contain a stem-tetraloop flanked by random regions and mixed with an unstructured library in a 1:1 ratio. High affinity GTP RNA aptamers were evolved with dissociation constants ranging from 500 to 25 nanomolar and seven unique aptamers were isolated with three of the highest affinity GTP aptamers originating from the structured pool. Out of 13 identified class I GTP aptamers, 12 originated from the structured pool. In this case, it was shown that the advantage of introducing complexity in the form of a stem-loop to a random pool outweighed the possible disadvantage of eliminating possible registers within the random region where the GTP recognition loop could occur.

Many discovered functional nucleic acids have exhibited compositional biases and tend to be purine-biased.[50] This idea was examined as a possibility to increase the initial structural complexity within a random pool computationally. Knight was able to calculate the composition that would lead to the optimal sequence abundance and folding for both the hammerhead ribozyme and isoleucine aptamer.[35] Assuming a random region of 100 nucleotides, the probability of finding the minimal hammerhead ribozyme increased 2.3-fold with an optimal composition of 35% A, 10% C, 25% G, and 30% U over a random region of unbiased composition. Conversely, the probability of finding the minimal isoleucine aptamer within a biased pool of 15% A, 25% C, 35% G, and 25% U increased 3.5-fold over an unbiased random region. While the optimal composition for each example is vastly different, a composition that maximized the probability of finding both motifs was determined to be

20% A, 15% C, 40% G, and 25% U and required 6.23×10^9 molecules for 99% probability of occurrence of both motifs, a factor of 10 smaller than an unbiased pool. Although the sample size was limited, compositional changes in the initial pool of an *in vitro* selection may assist in favoring new and rare functional structures. Gevertz's work with graph theory supported this in a limited way by showing that 20% A, U and 30% G, C increased the proportion of higher vertex structures for a 40-nt pool but structural abundances for a 100-nt pool were not altered significantly.[38]

Increased modularity in a random region has been shown to lead to higher abundances of more complex and rare structures. A modular evolution model has been introduced as a strategy to increase structural complexity.[51] Manrubia used computational simulations of two independently evolving pools of short, random sequences (n=35) to analyze the likelihood of the combination of smaller RNA modules to form a more complex functional nucleic acid. Several advantages of modular evolution were observed including that the selection was equally efficient with 35-nt modules at double the mutation rates of those allowed for 70-nt sequences, the time to select smaller motifs was significantly shorter and smaller pools sizes were required. Applied to an *in vitro* selection, small motifs could be quickly evolved using high mutation rates, exploring most of sequences space, and the resulting modules could be ligated and evolved further for more complex functions. Experimentally, modular evolution has been employed by appending random region sequences to previously discovered ribozymes,[52,53] functional motifs have been joined to form allosteric ribozymes,[54,55] functional RNA has been evolved to exhibit both ligation and RNA cleavage reactions,[56] and has contained separate domains that allow ligand binding and subsequent cleavage.[57]

Accessing rare secondary structures for new activities may be important for isolating new

functional nucleic acids. Increasing the abundance of complex secondary structures in an initial random pool by installing a structured loop or changing the compositional nature are simple measures that could lead to new functional isolates. Independently selected and combination of small modules could be another powerful tool for isolating complex and rare functional nucleic acids containing more than one active domain.

1.4 Conclusion

Design of a nucleic acid library is an essential foundation for a successful *in vitro* selection, especially in search for new and more active functional nucleic acids. Increasing the starting diversity of a pool will increase the sampling of sequence space but most practical considerations limit the number of library members to 10^{15}, so other strategies must be employed to access rare secondary structures that may have interesting activity.

Generally, longer random regions are more prone to misfolding which can be inhibit expected functional activities by masking the key structural motif or not allowing it to form. This effect is most pronounced with smaller, simpler motifs while the probability of finding more complex and rare structures in long random regions increases significantly and negates the possible cost of misfolding.

Since nucleic acids only contain four nucleotides, many secondary motifs such as pseudoknots, tetraloops and uridine turns occur often. The majority of the primary sequence can be easily exchanged while still maintaining the same secondary structure. This degeneracy leads to many primary sequences that have the ability to fold into very similar structures that may exhibit similar activities as well. While sequence space is usually sparsely sampled, the structural diversity is within the practical limits of an *in vitro* selection. Small, evenly sized modular motifs were shown to be the most abundant. Increasing complexity of

secondary structures correlated with increasing vertices in graph theory, pointing to longer random regions for unique structures that might reveal new activities.

In vitro selections tend to find the simplest solutions, which has been supported repeatedly experimentally. More complex structures may exhibit better activities but if they are not magnitudes more efficient than the more abundant and simpler motif, they may not be discovered in the initial rounds of *in vitro* selection. Strategies such as altering nucleotide composition of the random region, introducing a stem-loop into the initial random pool and evolving small motifs that can be combined into more complex functional nucleic acids have been proposed to increase the abundance of rare secondary structures that may be more efficient as well.

Molecular diversity affects the ability of the pool to evolve a new function. Understanding the relationship between increasing complexity of a random pool and complexity of a given phenotype. Studies have sought to understand how the design of the initial nucleic acid pool affects the phenotype. Aspects of this include the effect of length on selection success, how modularity affects structural diversity and strategies to access increased complexity to find rare and new functional nucleic acids. Understanding what can reasonably be expected from a starting library and how the design will affect the outcome of an *in vitro* selection and its ability to isolate new and rare functional nucleic acids may be an important aspect to consider moving forward.

References

(1)   McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.
(2)   Wilson, D. S.; Szostak, J. W. *Ann. Rev. Biochem.* **1999**, *68*, 611.
(3)   Peracchi, A. *ChemBioChem* **2005**, *6*, 1316.
(4)   Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550.
(5)   Ciesiolka, J.; Yarus, M. *RNA* **1996**, *2*, 785.
(6)   Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
(7)   Nieuwlandt, D.; Wecker, M.; Gold, L. *Biochemistry* **1995**, *34*, 5651.
(8)   Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
(9)   Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411.
(10)   Ekland, E. H.; Bartel, D. P. *Nature* **1996**, *382*, 373.
(11)   Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223.
(12)   Purtha, W. E.; Coppins, R. L.; Smalley, M. K.; Silverman, S. K. *J. Am. Chem. Soc.* **2005**, *127*, 13124.
(13)   Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611.
(14)   Wang, W.; Billen, L. P.; Li, Y. *Chem. Biol.* **2002**, *9*, 507.
(15)   Li, Y.; Liu, Y.; Breaker, R. R. *Biochemistry* **2000**, *39*, 3106.
(16)   Sheppard, T. L.; Ordoukhanian, P.; Joyce, G. F. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7802.
(17)   Chandra, M.; Silverman, S. K. *J. Am. Chem. Soc.* **2008**, *130*, 2936.
(18)   Lozupone, C., Changayil, S., Majerfeld, I., and Yarus, M *RNA* **2003**, *9*, 1315.
(19)   Knight, R.; Yarus, M. *RNA* **2003**, *9*, 218.
(20)   Yarus, M.; Knight, R. D. *The Genetic Code and Origin of Life*; Landes Bioscience: Georgetown, TX, 2004.
(21)   Sabeti, P. C.; Unrau, P. J.; Bartel, D. P. *Chem. Biol.* **1997**, *4*, 767.
(22)   Ekland, E. H.; Szostak, J. W.; Bartel, D. P. *Science* **1995**, *268*, 364.
(23)   Huang, Z. J.; Haugland, R. P.; Szalecka, D.; Haugland, R. P. *Biotechniques* **1992**, *13*, 543.
(24)   Coleman, T. M. a. H., F *Chem. Biol.* **2002**, *9*, 1227.
(25)   Nutiu, R. a. L., Y. *Angew. Chem. Int. Ed. Engl* **2005**, *44*, 1061.
(26)   Hanczyc, M. M. a. D., R.L *Mol. Biol. Evol.* **2000**, *17*, 1050.
(27)   Salehi-Ashtiani, K. a. S., J.W. *Nature* **2001**, *414*, 82.
(28)   Cruz, R. P., Withers, J.B., and Li, Y *Chem. Biol.* **2004**, *11*, 57.
(29)   Legiewicz, M.; Lozupone, C.; Knight, R.; Yarus, M. *RNA* **2005**, *11*, 1701.
(30)   Schlosser, K.; Lam, J. C. F.; Li, Y. F. *Nucleic Acids Res.* **2006**, *34*, 2445.
(31)   Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4462.
(32)   Woese, C. R., Winker, S., Gutell, R.R. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8476.
(33)   Jiang F., K. R. A., Jones R.A., Patel D.J. *Nature* **1996**, *382*, 183.
(34)   Wedel, A. B. *Trends Biotechnol.* **1996**, *14*, 459.
(35)   Knight, R.; De Sterck, H.; Markel, R.; Smit, S.; Oshmyansky, A.; Yarus, M. *Nucleic Acids Res.* **2005**, *33*, 5924.
(36)   Chapple, K. E.; Bartel, D. P.; Unrau, P. J. *RNA* **2003**, *9*, 1208.

(37)   Fontana, W.; Konings, D. A. M.; Stadler, P. F.; Schuster, P. *Biopolymers* **1993**, *33*, 1389.

(38)   Gevertz, J.; Gan, H. H.; Schilick, T. *RNA* **2005**, *11*, 853.

(39)   Hofacker, I. L. *Nucleic Acids Res.* **2003**, *31*, 3429.

(40)   Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 5130.

(41)   Carothers, J. M.; Davis, J. H.; Chou, J. J.; Szostak, J. W. *RNA* **2006**, *12*, 567.

(42)   Smith, J. M. *Nature* **1970**, *225*, 563.

(43)   Kauffman, S. A.; Weinberger, E. D. *J. Theor. Biol.* **1989**, *141*, 211.

(44)   Davis, J. H.; Szostak, J. W. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11616.

(45)   Ekland, E. H.; Bartel, D. P. *Nucleic Acids Res.* **1995**, *23*, 3231.

(46)   Schultes, E. A.; Bartel, D. P. *Science* **2000**, *289*, 448.

(47)   Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425.

(48)   Geiger, A.; Burgstaller, P.; von der Eltz, H.; Roeder, A.; Famulok, M. *Nucleic Acids Res.* **1996**, *24*, 1029.

(49)   Fan, P.; Suri, A. K.; Fiala, R.; Live, D.; Patel, D. J. *J. Mol. Biol.* **1996**, *258*, 480.

(50)   Schultes, E.; Hraber, P. T.; LaBean, T. H. *RNA* **1997**, *3*, 792.

(51)   Manrubia, S. C.; Briones, C. *RNA* **2007**, *13*, 97.

(52)   Jaeger, L.; Wright, M. C.; Joyce, G. F. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14712.

(53)   Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. *Science* **2001**, *292*, 1319.

(54)   Tang, J.; Breaker, R. R. *Chem. Biol.* **1997**, *4*, 453.

(55)   Komatsu, Y.; Nobuoka, K.; Karino-Abe, N.; Matsuda, A.; Ohtsuka, E. *Biochemistry* **2002**, *41*, 9090.

(56)   Kumar, R. M.; Joyce, G. F. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9738.

(57)   Romero-Lopez, C.; Barroso-del Jesus, A.; Puerta-Fernandez, E.; Berzal-Herranz, A. *Biol. Chem.* **2005**, *386*, 183.

Chapter 2

Progress towards the selection of a catalytic peptide-DNA conjugate in water-in-oil emulsions

2.1 Introduction

Darwinism is a theory of biological evolution that states that new species arise by natural selection of inherited traits, which enable them to thrive, survive and reproduce more over successive generations. *In vitro* selection mimics this idea in a laboratory setting and was a technique first conceptualized in the 1960s when Speigelman utilized RNA and the mutation rate of Qβ replicase in RNA bacteriophage Qβ to evolve new viral genome variants with shortened replication times in a cell-free environment.[1]

Molecular evolution requires the phenotype be linked to its genotype, permitting selection, amplification and further mutation for several generations. In the case of RNA and DNA, functionality (phenotype) such as binding and catalysis and the encoding strand (genotype) are the same.[1] Molecular evolution of peptides and proteins is less straight-forward because there is no way to amplify the phenotype, but physical linkage of peptides/proteins to their encoding DNA or RNA enables selection and amplification over several generations.[2] Protein and peptide selections became possible in the laboratory setting with the development and implementation of display of fusion proteins on filamentous phage in 1985.[3]

High-throughput *in vitro* evolution of peptides and proteins has been enabled by display technologies[3,4] and *in vitro* compartmentalization (IVC) of the transcription and translation machinery with the DNA library in simple water-in-oil emulsions.[5,6] These methods link the

peptide or protein (phenotype) to its own coding genetic material (genotype) and permit selection and amplification of peptides/proteins that function as enzymes or receptor ligands.[7] Phage display and mRNA display start with large pools, up to 10^{10} and 10^{13}, respectively and IVC can accommodate 10^{10} library members.[4,7] However, the phage display and mRNA display methods each have their own limitations. Phage display requires incorporation of the pool into the phage DNA *via* host bacterium, thus it is limited in scale and cannot contain products that are detrimental to the phage or the bacterium host.[8] The method of mRNA display utilizes *in vitro* translation, therefore it is able to code for potentially toxic or unnatural peptides. However, mRNA display requires purification of the protein-RNA conjugate away from the ribosomes,[9] a cross-linking reaction between the mRNA-linked puromycin and the peptide, and these construction steps of the library can be temperamental.[10] Therefore, a new *in vitro* display system with straightforward construction and high diversity has the potential to be applied broadly and economically to peptide and protein evolution experiments.

In order to bypass the limitations of current peptide and protein display systems, we have chosen to develop a new method that involves the use of *in vitro* compartmentalization (IVC)[5] to evolve a peptide-DNA conjugate (Figure 2.1a). Compartmentalization of the transcription-translation machinery with the DNA library occurs in simple water-in-oil emulsions where surfactants help stabilize the water droplets for at least 24 hours.[5] The benefits of using IVC to house a library for selection are three-fold: each synthesized peptide is colocalized with its coding RNA and DNA;[5,6] a starting diversity of up to 10^{14} library members is within the practical limit (exceeding previous methods); and the library can include members that are incompatible with cellular function.

Selecting for peptides that covalently bond to their coding DNA or a bioorthogonal moiety may allow for the discovery of a peptide motif that can be used for easy construction of the peptide/protein display technology. To find a peptide that is capable of catalyzing conjugation to its own DNA, a pool will be inserted between the known T7 RNA polymerase promoter (T7 P), ribosomal binding site (RBS) and terminator regions (Figure 2.1b). In order for a successful peptide-DNA conjugation to occur, the peptide must specifically bind the 5' DNA termini or a 5' bioorthogonal moiety, form a catalytic center and react. It has been estimated that 1 in 10^{11} library members will have ATP-binding activity[11] and while catalytic centers can be as small as a catalytic triad, a longer peptide sequence gives the opportunity for successful discovery of a catalytic center and nucleotide binding. Twenty-seven amino acids (81 nucleotide random region) is a practical limit for library construction that avoids large-scale enzymatic ligation and preserves high diversity of the starting DNA pool.

Intensive construction of the peptide-DNA conjugation would be avoided by discovery of a catalytic peptide capable of reacting with its own DNA or a bioorthogonal moiety attached to the DNA. A bioorthogonal group has the advantage that it does not interfere with biological processes and reacts quickly and specifically under physiological conditions.[12] This peptide-DNA adduct would permit any pool to be appended to the DNA coding for the active peptide and allow for easy construction of a protein/peptide display system that could tolerate toxic sequences or unnatural amino acids (Figure 2.1c). Successful conjugation would permit larger library diversity for broader application of peptide and protein selections for drug and vaccine discovery involving small molecule binding, antibody design, protein-protein interaction studies and enzymatic reactions.[7]
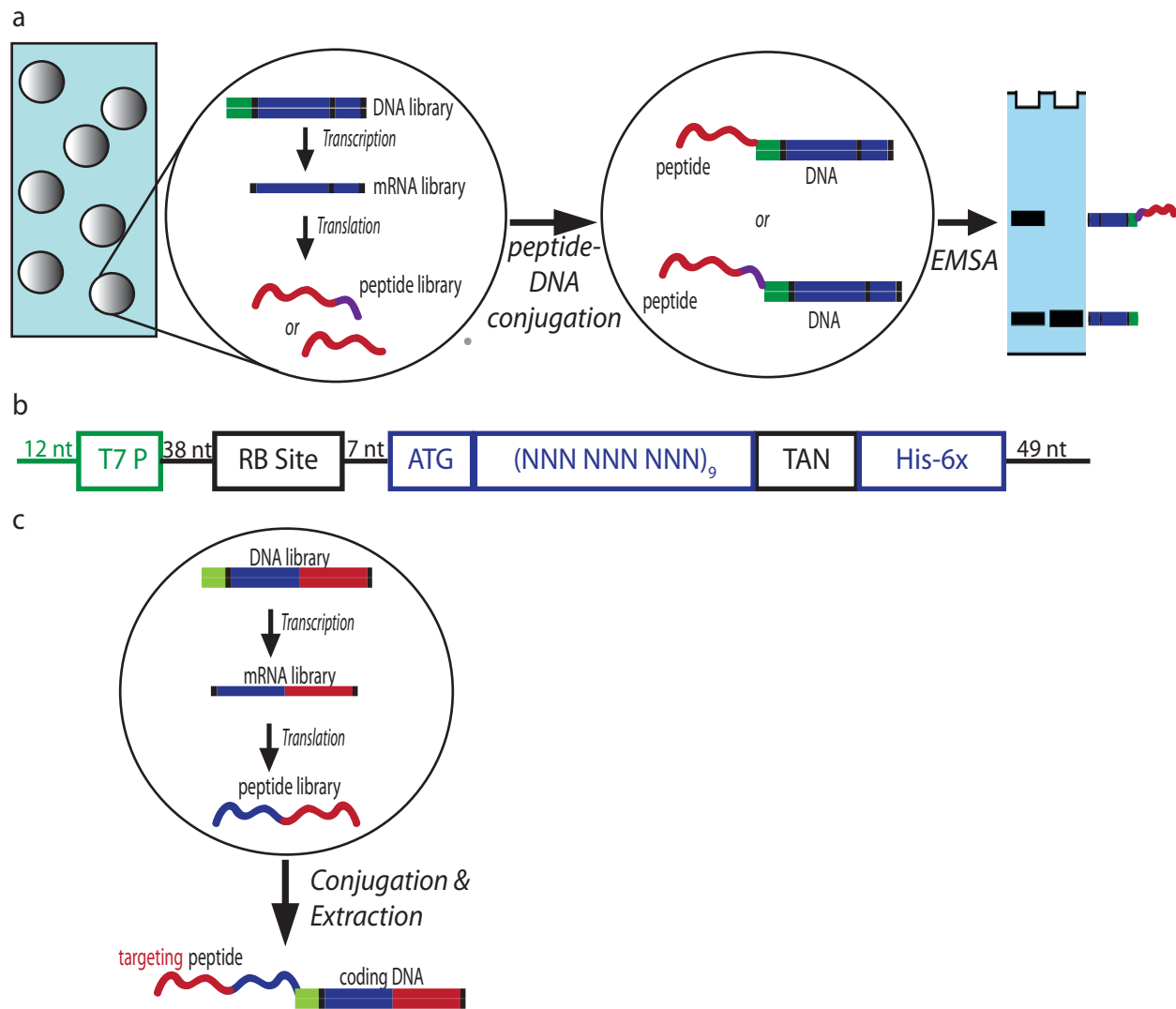
Figure 2.1. Selection overview for a peptide-nucleic acid conjugation. a) DNA, mRNA and components to express a peptide library and select for a catalytic peptide that will form a nucleic-acid conjugation and be separated from inactive sequences by an electrophoretic mobility shift assay. b) The DNA pool design including a T7 RNA polymerase promoter, ribosome-binding site, start codon, 27 random codons and a stop codon. c) Once a peptide-DNA conjugate is selected, the DNA sequence coding for the active peptide could be appended to a DNA library and utilized as a new display system that connects genotype to phenotype.

23

Here we seek to improve upon the mRNA display and IVC method by selecting for a catalytic peptide that will serve as a link between the genotype and phenotype for peptide and protein *in vitro* selections. The prerequisites for developing a new display system are: 1) optimization of the coupled *in vitro* transcription and translation reaction to create a stable initial pool; 2) formulation of an efficient IVC system to ensure colocalization of library transcription and translation; 3) successful addition of the necessary fixed DNA sequences and the bioorthogonal moieties required for efficient transcription and translation; and 4) the ability to select and separate active sequences from the inactive majority.

Results and Discussion

2.2 Construct design and optimization

In order to establish a large starting diversity of the expressed peptide pool, an efficient and coupled transcription-translation platform is required. To accomplish this, a cell free *Escherichia coli* expression system from Invitrogen was chosen. The major components of this system that allow for efficient coupled transcription and translation are an optimized *E. coli* extract,[13] ATP regenerating reaction buffer[14] and feed buffer[15] as well as a T7 enzyme mix optimized for T7-based transcription promotion.[16] To optimize the coupled expression system for the most productive reaction, DNA test constructs were designed to code for a reporter peptide that could be easily purified and quantified. The initial test construct **2a**, in accordance with Invitrogen's manual suggestions, included a T7 RNA polymerase promoter[17] 10 nucleotides from the 5′ end followed by the ribosomal binding site[18] 8 nucleotides upstream of the start codon (AUG),[17,18] open reading frame and stop codon (TAA). All constructs were comprised of codons that are most commonly used by *E. coli*.[19] To maximize translation yields, the intervening bases were chosen to reduce secondary structure as

predicted by Vienna RNA secondary structure server, which predicts the minimum energy secondary structures and base pair probabilities.[20] The codons included in the open reading frame of construct **2a** contained a His_{10} tag used for purification and a tetracysteine motif (CCXXCC) (Figure 2.2a). The tetracysteine tag coordinates to Tsien's Fluorescein Arsenical Hairpin (FlAsH) fluorophore with a dissociation constant of approximately 4 pM, creating a fluorescent signal upon binding,[21] thereby enabling small amounts of peptide to be detected.

Early attempts to detect peptide formation from DNA construct **2a** were unsuccessful. Traditional gel electrophoresis staining and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) spectroscopy were unsuccessful in visualizing and identifying the peptide. However, the positive control plasmid provided by Invitrogen, which expresses calmodulin-like protein 3 (CALML3) with a His_6 tag, was observed by Tris-tricine PAGE and MALDI-TOF with the expected mass of 19.5 kDa. In an effort to visualize the peptide, several reactions were performed with increasing amounts of the DNA construct **2a** (119, 340 and 718 nM), but no peptide was observed.

The lack of the peptide from the DNA construct **2a** could be caused by the initial sequence design or the construct stability. If the T7 RNA polymerase promoter or the ribosome-binding site sequences were not placed in optimal locations, the T7 RNA polymerase or ribosomes would not bind efficiently, preventing productive peptide synthesis. Performing a separate transcription reaction with each DNA construct tested T7 RNA polymerase binding efficiency and prolific transcription was observed in all cases, indicating the T7 RNA polymerase promoter was well placed. Ribosome binding problems are less easily tested and could not be ruled out as the underlying problem.
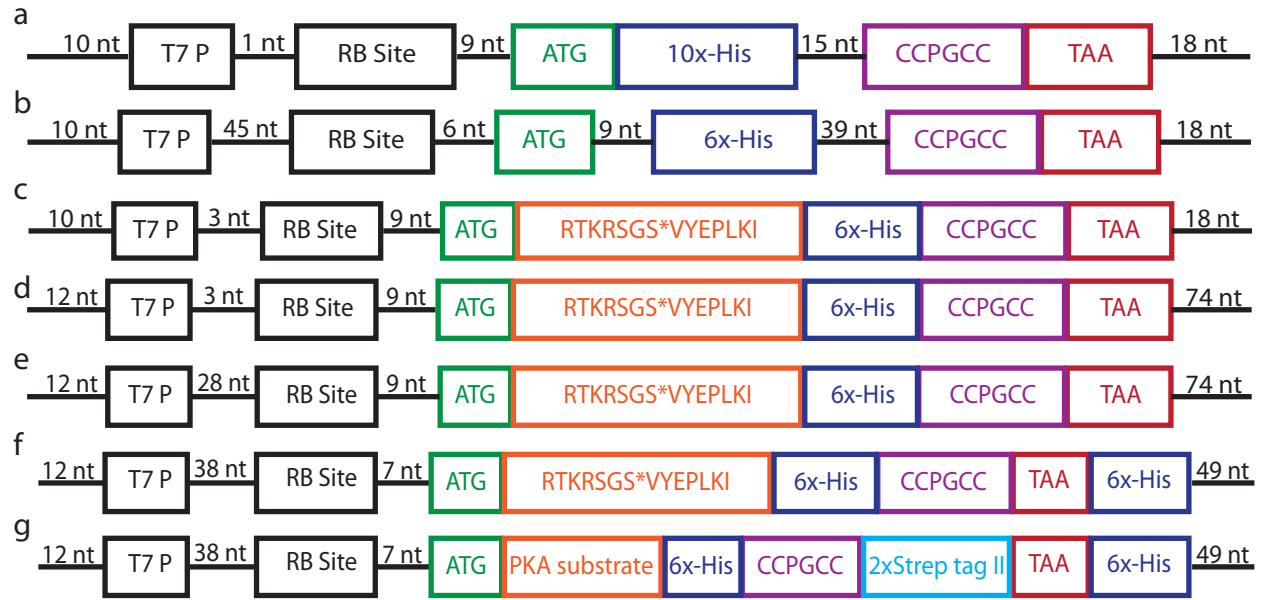
Figure 2.2. DNA construct for peptide translation optimization, including a) T7 RNA promoter, ribosomal binding site, start and stop codon, His$_6$ tag and tetracysteine tag, b) use of a portion of the CALML3 sequence upstream of the start codon, c) addition of a PKA substrate, d) addition of a 3' 15 bp hairpin, e) addition of a 5' 8 bp hairpin, f) incorporation of the His$_6$ tag codons after the stop codon and g) addition of a Twin-Strep-tag®.

If construct instability caused the problem, the DNA construct or transcribed RNA would be degraded by DNase or RNase contamination before transcription and/or translation could produce a detectable amount of peptide. Since the CALML3 control plasmid was successful using several methods, a similar linear PCR product was tested. In order to establish which step of the coupled transcription/translation was not successful, primers were designed to amplify a portion of the productive CALML3 plasmid, containing a His_6 tag (Figure 2.2b). The amplified DNA construct **2b** was included in a synthesis reaction and no product was observed by gel electrophoresis Coomassie staining or MALDI-TOF mass spectrometry. The DNA construct **2b** was transcribed successfully separately, indicating the T7 RNA polymerase promoter location was not problematic. It was assumed that the transcription step during the coupled transcription/translation reaction amplifies the transcript approximately 10-100-fold, so the transcription reaction was purified and 2.5 μM-8.0 μM RNA was added to a synthesis reaction (Table 2-3). The resulting sample was purified using the His_6 tag (MSGSHHHHHHGSSGENLYFQEQVCCPGCC, 3209.4 Da[22]) and analysis by MALDI-TOF mass spectrometry showed a small peak at the expected mass (m/z=3210) and a larger peak at twice the mass (m/z=6419), which may be the peptide dimerized through a disulfide bond. This small success with the addition of RNA suggested that the linear DNA construct **2b** is either degrading or substantial RNA production does not occur in the coupled transcription/translation reaction. Because the *E. coli* extract provided was not DNase treated, residual DNases may be the cause of degradation.[13] Another possibility was that the construct was successfully transcribed in the reaction but the RNA transcript was degraded shortly after by RNase contamination. In this case, the addition of RNase inhibitor could eliminate the issue

and allow an effective concentration of the transcript to be produced.

In order to troubleshoot the low translation yield, construct **2c** was designed, which contains, in addition to the tetracysteine motif, a protein kinase A (PKA) substrate (RTKRSGS*VYEPLKI) (Figure 2.2c-f). This additional motif exhibits a Michealis constant of approximately 15 μM,[23] permitting enzymatic labeling of the substrate serine with γ-^{32}P-adenosine triphosphate. Since a small amount of peptide was observed when RNA was used instead of DNA, the **2c** construct was transcribed, purified and added to the transcription/translation reaction. Initially, 260 pmoles of RNA was used and the gel was imaged with Invitrogen's Lumio™ technology, comprised of FlAsH, to take advantage of the tetracysteine motif and its 1 pmole detection limit (Figure 2.3a), 30-fold more sensitive the Coomassie staining. However, after purification using the His$_6$ tag, a higher fluorescence signal in the flow-through sample was observed relative to the elution, indicating a lot of background binding of the FlAsH reagent. In a second attempt, 480 pmoles of RNA were added to the translation reaction and half of the His$_6$ tag purified sample was labeled with ^{32}P and half was labeled with FlAsH and imaged, in parallel, on a Tris-tricine PAGE (Figure 2.3b, c). Labeling the peptide with ^{32}P, resulted in little signal or a distinct band corresponding to the peptide in the flow-through or elution sample after purification utilizing His$_6$ tag, but there seemed to be more signal in the elution sample. Flow-through and elution samples incubated with the FlAsH reagent did not result in any distinct bands but copious background fluorescence in all lanes. Since a reliable signal was not obtained on a PAGE from the Alexa Fluor® 488 fluorescently labeled protein ladder, the usefulness of detection utilizing the tetracysteine tag was called into question.
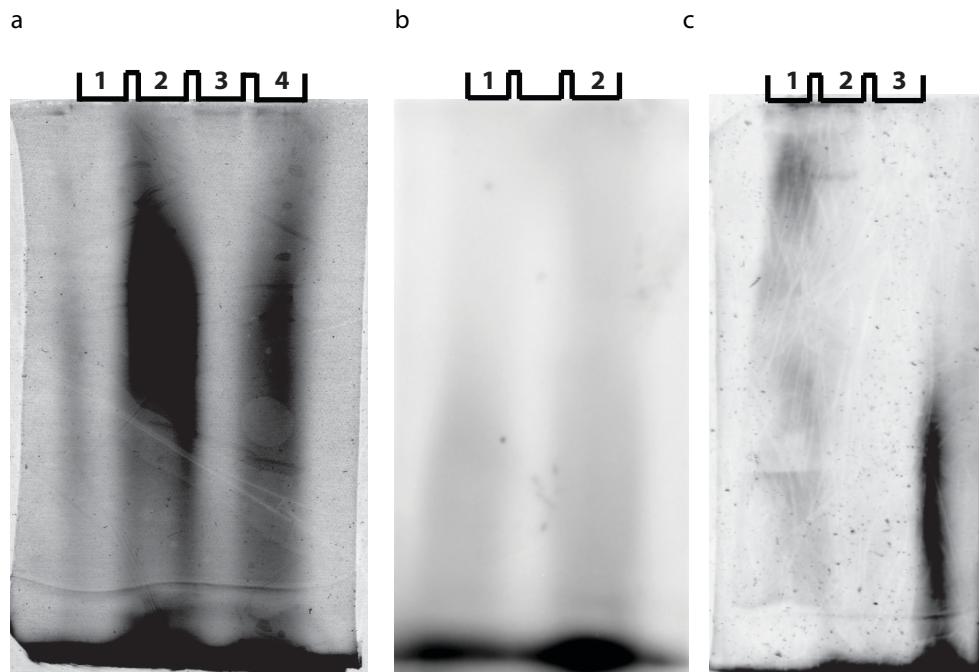
Figure 2.3. **Tris-tricine PAGE using the tetracysteine and PKA substrates for imaging.** An 18% Tris-tricine gel was used to analyze peptide translation success of construct **2b** a) by incubating Lumio™ green detection reagent with His$_6$ tag purified peptide elution (lane 4) and flow-through sample (lane 2) at 70 °C for 10 minutes before loading. The resulting gel was excited at 488 nm and emission was collected in the range of 520-535 nm. An 18% Tris-tricine gel was used to analyze peptide translation success of construct **2c** b) by labeling the His$_6$ tag purified peptide elution (lane 2) and flow-through sample (lane 3) with ^{32}P before loading onto the gel. The gel stored in a phosphor screen overnight and scanned on a Typhoon Trio. c) An 18% Tris-tricine gel was used to analyze peptide translation success of construct **2c** by incubating Lumio™ green detection reagent with His$_6$ tag purified peptide elution (lane 2) and flow-through sample (lane 3) and a ladder (lane 1) at 70 °C for 10 minutes before loading. The resulting gel was excited at 488 nm and emission was collected in the range of 520-535 nm.

Optimization of the test construct and detection of the peptide continued, yielding construct **2d**, which added a 3' 15-basepair hairpin with a tetraloop (UUUG) that favors RNA hairpin formation to increase mRNA stability and expression yields.[24,25] Detection of peptide synthesis for this construct utilized the tetracysteine motif directly in the reaction. The detection reagent FlAsH was added directly to the reaction, with an aliquot taken at T_0 and T_{6} $_{hrs}$. Fluorescence was measured by excitation of the sample at 500 nm and the spectra showed little or no fluorescence signal above the background reading. However, once these reactions were purified, ^{32}P-labeled, fractionated on a Tris-tricine PAGE and imaged, discrete bands were observed. A more intense band was observed with addition of RNase inhibitor to the reaction. These results suggest that the Lumio™ detection reagent binds too many other proteins in the *E. coli* lysate generating background fluorescence that overpowers the detection of such a small amount of peptide, *in situ*. Successful detection of the peptide utilizing the tetracysteine tag, *in situ*, would have allowed easy detection of the peptide without purification as well as easy comparison to a protein ladder standard when imaging on a PAGE. However, because the peptide expressed from construct **2d** could not be easily detected with fluorescence and was successfully visualized *via* labeling with γ-^{32}P-ATP, radiolabeling may have a lower detection limit and is attractive because it avoids the use of other radioactive nuclei (e.g. ^{35}S), significantly reducing waste and material expense.
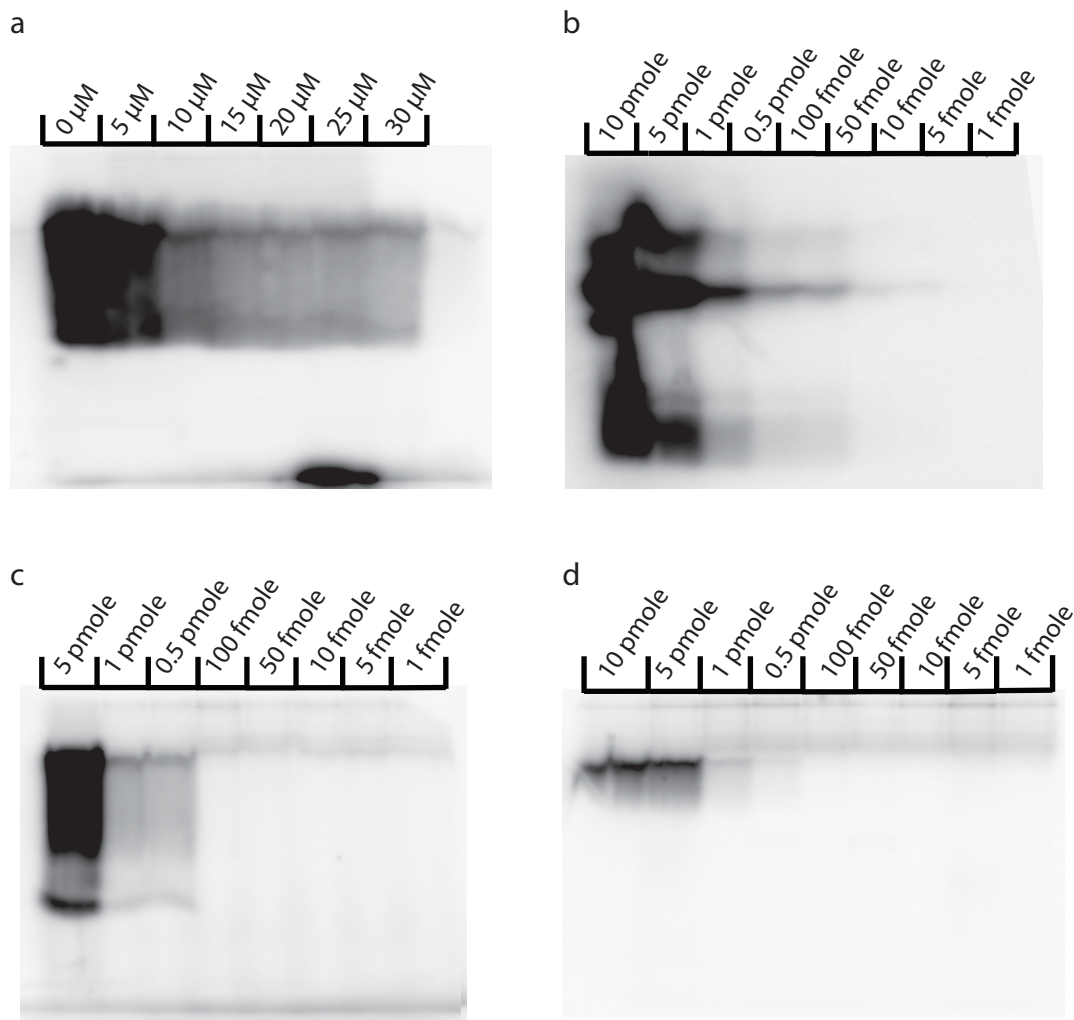
Figure 2.4. PKA substrate labeling optimization. A 18% Tris-tricine gel was used to image peptides (chemically synthesized: RTKRSGSVYEPLKI) labeled by incubating with 2 units of PKA and 0.165 μM γ-^{32}P-ATP for 5 hrs at 37 ℃. Reactions were performed with a) increasing concentrations of unlabeled ATP in the labeling reactions, b) serial dilutions of the peptide after labeling, c) decreasing amounts of peptide in each labeling reaction, and d) decreasing amounts of peptide in each labeling reaction followed by a TCA precipitation. Each gel was stored in a phosphor screen overnight and scanned on a Typhoon Trio.

Once a ^{32}P radiolabeled peptide was successfully imaged on a PAGE, the radiolabeling reaction was optimized using a chemically synthesized malantide (RTKRSGS*VYEPLKI)[26] with titrations of polyethylene glycol to increase the effective concentrations of the reactants and unlabeled ATP to approach the PKA's K_M of ATP (25 μM).[27] However, neither increased the efficiency of the labeling reaction (Figure 2.4a). The detection limit of the storage phosphor screen and Typhoon Trio were tested which showed that a PKA substrate labeled with ^{32}P could be diluted down to 10 fmoles before the signal disappeared (Figure 2.4b). Dilutions of the purchased peptide was also titrated into the ^{32}P labeling reaction to test the lower limit of the PKA enzyme and 1 pmole and 0.5 pmoles were the lowest amount that could be labeled and detected with and without TCA precipitation, respectively (Figure 2.4c, d). This method was at least 2-fold more sensitive than detection of the peptide through fluorescence and exhibited lower amounts of background signal, which was vital for gel imaging of the labeled peptide. Optimization of the labeling reaction with the short malantide (1.6 kDa) and analysis also confirmed that smaller peptides could be successfully resolved on a Tris-tricine PAGE.
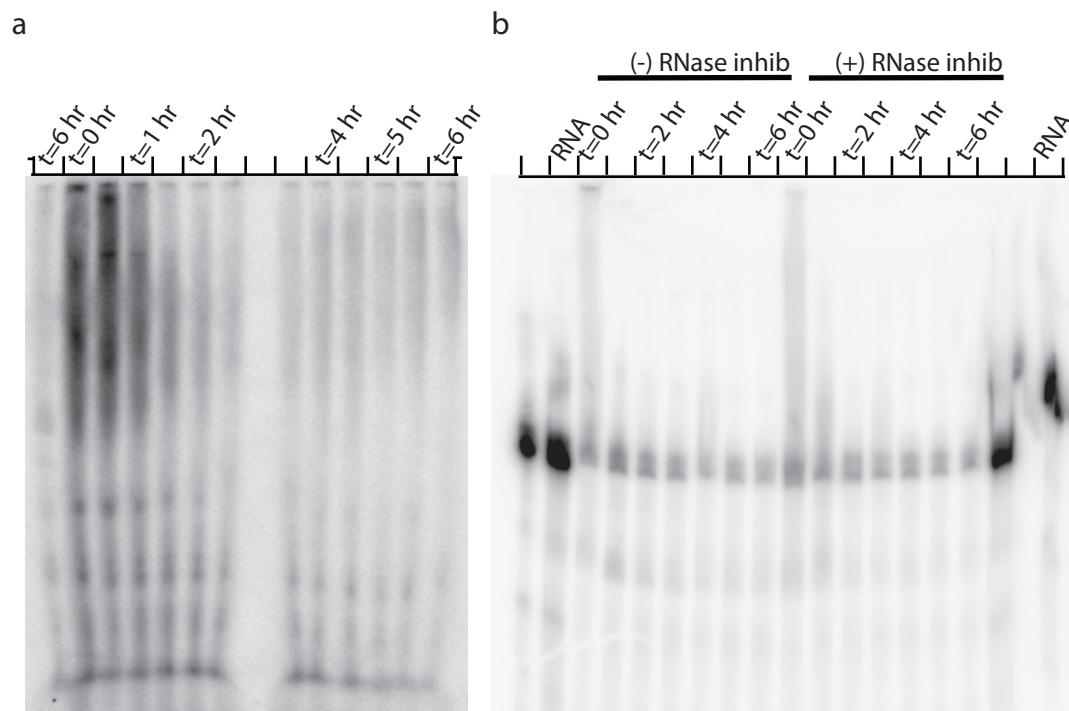
Figure 2.5. RNA content analysis. A 7.5% denaturing polyacrylamide gel was used to image ^{32}P body-labeled RNA aliquoted from a protein synthesis a) every 30 minutes using ~160 pmoles RNA from construct **2d** and b) every 1 hr using ~260 pmoles RNA from construct **2e** and extracting with 20%/80% phenol/chloroform before loading the gel. Each gel was stored in a phosphor screen overnight and scanned on a Typhoon Trio.

In vitro peptide synthesis may have been successful, but only RNA starting material was used. To probe at the question of whether transcription or the resulting RNA stability in the coupled transcription/translation reaction was problematic, a portion of α-^{32}P-labeled RNA was doped into a peptide synthesis reaction. Aliquots were taken and RNA content analysis under denaturing conditions of the translation reaction showed radiolabeled RNA bound to larger proteins and a decrease of the RNA within the first 2 hours (Figure 2.5a) and approximately 50% lost after 6 hrs. Revisiting the modest success with the appended CALML3 construct **2b**, RNAfold showed an 8-basepair hairpin between the T7 RNA polymerase promoter and the ribosomal binding site. This hairpin was appended to construct **2d** to produce **2e** and α-^{32}P-labeled ATP was incorporated into a fraction of the new RNA construct **2e** and content analysis was repeated showing little reduction in RNA over the course of the incubation with and without RNase inhibitor, but there was approximately 50% less RNA than calculated, indicated by the stronger intensity of the RNA band compared to the bands from the aliquots of Figure 2.5b. The samples from this denaturing PAGE were phenol/chloroform extracted beforehand which indicates that a significant amount of RNA was caught up with the ribosomes or other proteins, which could have led to less peptide expression. The stability of the RNA construct **2e** was increased relative to construct **2d** and the presence of RNase inhibitor did not affect RNA concentrations significantly.

Parallel peptide expression reactions with DNA and RNA from construct **2e** were performed. Possible success was seen with the RNA from construct **2e**, but the band was diffuse. There may have been a faint band corresponding to product originating from the DNA construct, but this result could not be reproduced with several attempts.

Up to this point, all portions of the sequence were optimized except the ribosome-

binding site. A program for designing ribosome-binding sites with optimal translation initiation rates based on binding energies of interactions[28] was used to increase initiation and thereby translation. This program was used to modify the ribosome-binding site (ACGACACA TAGATAAGGAGGAACGAGGA) of construct **2e**, resulting in **2f**. Construct **2f** was also slightly modified after the stop codon to reflect the sequence planned for the library (Figure 2.1b). Previously, the radiolabeled malantide was successfully imaged using a Tris-tricine PAGE, however, this method only resolves peptides and proteins above 5 kDa reliably. The small size of the synthesized peptide of all previous constructs made detection more challenging, which led to the design of construct **2g**. Construct **2g** contained all optimized elements of construct **2f** but added 28 amino acids that coded for a streptavidin binding motif[29] and increased the mass of the peptide product from 3.2 kDa to 6.1 kDa. After an initial peptide synthesis success, DNA constructs **2f** and **2g** were added to several parallel reactions in order to confirm peptide synthesis reactions were producing the correct sized product as well as verify the ability of this reaction to occur in water-in-oil emulsions.

2.3 Water-in-oil emulsions

A peptide synthesis reaction was incorporated into water-in-oil emulsions according to Tawfik,[5] optimized, (Figure 2.6a) and imaged after 24 hrs. The diameter distribution of the resulting mixed emulsions was highly dispersed which could be disadvantageous to the selection platform. An ideal sized droplet would have a diameter of 2 μm, which would allow for all components of the *in vitro* expression reaction to be incorporated and if 100 DNA pool members were incorporated into each droplet, the volume of starting aqueous reaction would be approximately 4 mL which is still within practical limits. Small emulsion droplets would not contain all the necessary reactions components and large emulsions may contain

too many copies of the initial pool, eliminating the connection between genotype and phenotype. Due to these reasons, a mini-extruder was acquired along with 1 μm sized membranes, which produced more uniformly sized emulsions close to the ideal 2 μm diameter (Figure 2.6b). Emulsions were reliably uniform using this method, but extraction of the aqueous *in vitro* reaction after incubation was difficult and required more extractions and inclusion of phenol to assist in breaking the stable droplets.
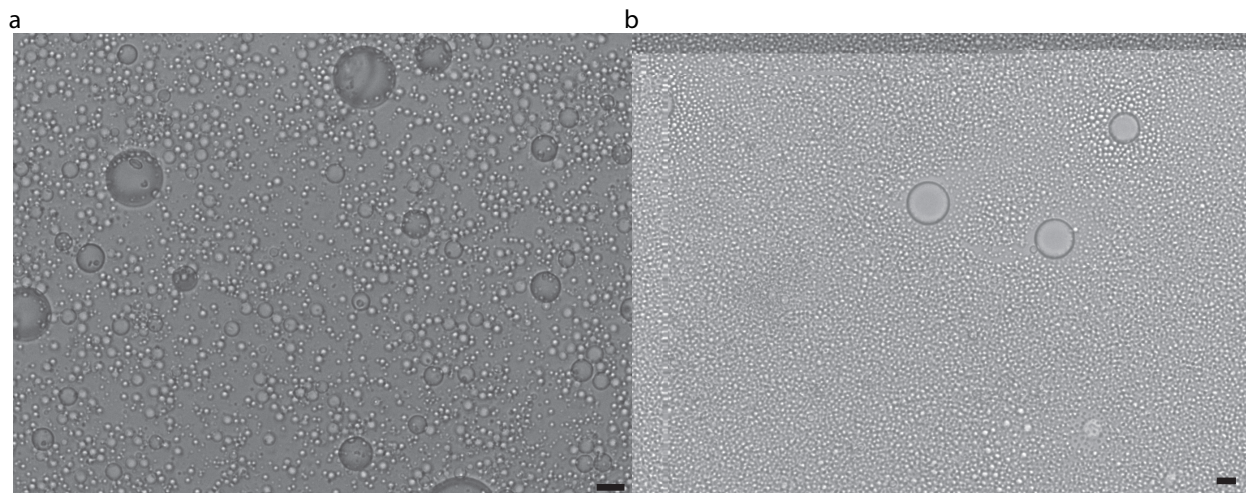
Figure 2.6. Water-in-oil emulsions a) mixed by adding 900 µL light mineral oil with 4.5 % Span® 80 and 0.5 % Tween® 80 to a 100 µL peptide synthesis reaction and stirring at 1150 rpm for 3 minutes and b) extruded by adding the 900 µL mineral oil mixture and the 100 µL peptide synthesis reaction to a mini-extruder and passing it through a 1 µm polycarbonate membrane 11x. The droplets were imaged after 24 hrs at room temperature using an inverted microscope at 40x magnification with a 10 um bar in the right corner.

To examine how the peptide synthesis reaction would perform in emulsions, DNA constructs **2f** and **2g** were tested in parallel. Reactions were performed for both constructs in order to observe the differently sized peptide products, 3.1 kDa and 6.1 kDa, respectively. Four peptide synthesis reactions were performed, including an aqueous synthesis reaction, an aqueous synthesis reaction overlaid with the light mineral oil mixture and two water-in-oil emulsified reactions, one prepared by mixing and the other by extrusion (Figure 2.7). Initially, the distinct major band in each His$_6$ tag purified elution lane was believed to be the desired peptide, but there was no discernable size difference between peptide products from constructs **2f** and **2g**, as expected. If these bands were the desired peptide products, then the coupled transcription/translation reactions were compatible with the mixed emulsions but less efficient when using the extruded emulsions. A lower, less resolved band was most discernable in the elution of the extruded emulsion lane from construct **2g** and if this was the desired peptide, then the reactions are not producing abundant amounts of the peptide but a size difference between constructs was observed. Unfortunately, a similarly sized radiolabeled standard was not used to help verify the expression of the correct sized peptide and verification had to be performed in a more indirect manner. This modest success in the extruded emulsions may be due to the increased effective concentrations of reaction components due to addition of BSA.
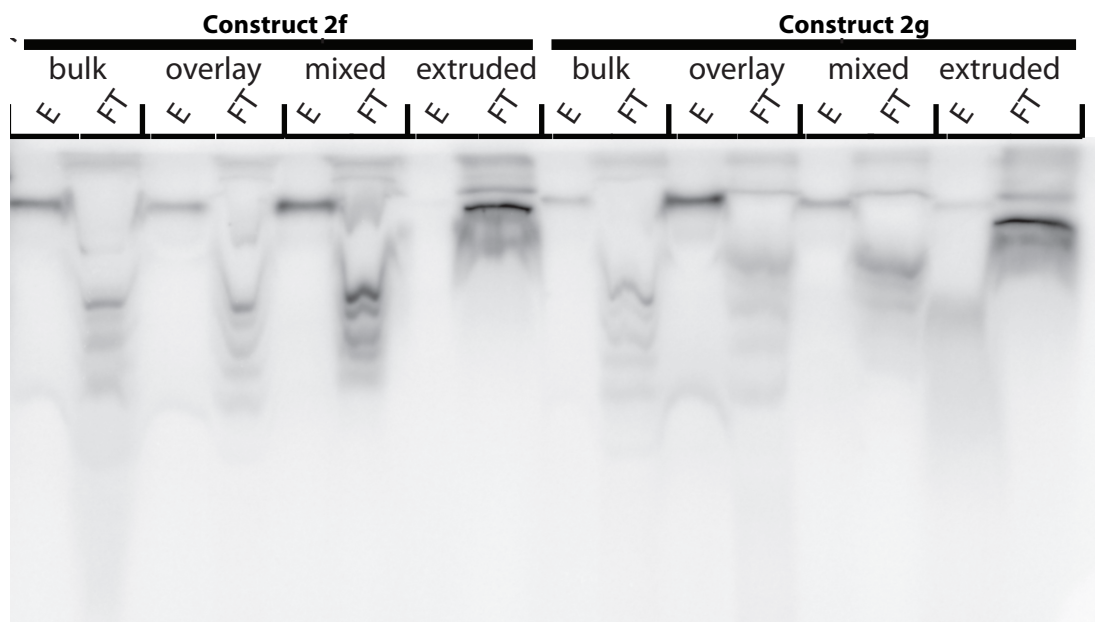
Figure 2.7. Testing construct compatibility with water-in-oil emulsions. DNA construct **2f** and **2g** (13.5 pmoles) were added to a 100 µL aqueous peptide synthesis reaction, an aqueous reaction overlaid with 4.5 % Span® 80 and 0.5 % Tween® 80 in light mineral oil and to a 100 µL peptide synthesis reaction in emulsions prepared by mixing and extrusion by adding 900 µL light mineral oil with 4.5 % Span® 80 and 0.5 % Tween® 80 and stirring or extruding, respectively. The reactions were incubated at 37 °C for 6 hrs and extracted with mineral oil and hexanes. The resulting aqueous layer was purified using a Ni^{2+} charged resin and the elution and flow-through samples were labeled using 2 units PKA and 0.165 γ-^{32}P-ATP and loaded on to an 18 % Tris-tricine PAGE and the gel was stored in a phosphor screen overnight and scanned on a Typhoon Trio.

To investigate which band from the previous experiment corresponded to the peptide product several experiments were performed including ultrafiltration of the elution and flow-through samples, comparison of denaturing versus native purification of the peptide products and a pepsin digest of the radiolabeled peptide product. In general, coupled transcription/translation reactions were extracted from the emulsions using light mineral oil and hexanes, purified utilizing the translated His$_6$ tag, with the elution and flow-through samples both desalted and concentrated using a spin filter and labeled with ^{32}P using the catalytic subunit of PKA and imaged with a Tris-tricine PAGE (Figure 2.8a).

An ultrafiltration experiment was conducted with peptide synthesis reactions of constructs **2f** and **2g** after desalting and concentrating the elution and flow-through samples by using a Centricon with a 10 kDa cut-off (Figure 2.8a). The desired peptide products for both reactions should flow through leaving other larger species above in the upper fraction. This led to seemingly large products in the lower fraction of the elution and a range of sizes in the upper fraction in each case (Figure 2.8c). The sizes in each lane do not make sense from the outset but since the synthesized His$_6$ tagged peptide was purified under native conditions, the peptide could be associated with another protein through non-covalent interactions or a disulfide bond before being denatured and reduced in loading buffer for the Tris-tricine PAGE. The major band seen in the lower fractions was also seen in the upper fractions and may be the PKA enzyme or other contaminant from the enzyme stock that was being labeled with ^{32}P. This indicates that the distinct higher bands seen previously, such as in Figure 2.7, was a ^{32}P labeled contaminant and not the desired peptide and while the malantide was specifically labeled by PKA, it appears that the enzyme was able to successfully several proteins present in the *E. coli* lysate. Due to the native purification conditions, the

major band in the upper fraction of each elution of Figure 2.8c is believed to be the desired peptide even though no size difference is observed between the two peptide products.

Emulsified peptide reactions using constructs **2f**, **2f** with the bioorthogonal azide, alkyne, aldehyde and acrydite moieties incorporated (**2f**-mod) and **2g** were purified with a His$_6$ tag under denaturing conditions and run in parallel on a Tris-tricine PAGE (Figure 2.8b). Again, no resolution occurred between peptides expressed from constructs **2f** and **2g**, but a diffuse band was present in all elution sample lanes. The intensity of this ban increases 230 % in the case of **2f**-mod and 310% for construct **2f** alone. The dramatic difference observed when the peptides were purified under denaturing conditions could possibly be the underlying reason for little coupled transcription/translation success up until this point. The expressed peptides may have been bound to other proteins form the *E. coli* lysate and not available for radiolabeling with the PKA enzyme or the other detection methods that were attempted. Aggregation of expressed peptides and proteins using a cell-free *E. coli* lysate have been reported.[30]

Finally, a pepsin digest of the peptide synthesis reaction using construct **2g** was performed to see if the band thought to be the peptide product would shift lower once specifically cleaved after the proline. While the digestion reaction did not complete, approximately 60% of the synthesized peptide from the eluted sample appeared lower relative to the undigested peptide under denaturing conditions (Figure 2.8d). Results of the digestion after native purification of the peptide had similar results, but the change is less dramatic and close to background signal. This final experiment confirms that the diffuse band that was imaged previously was the successfully transcribed and translated peptide and not a radiolabeled contaminant.
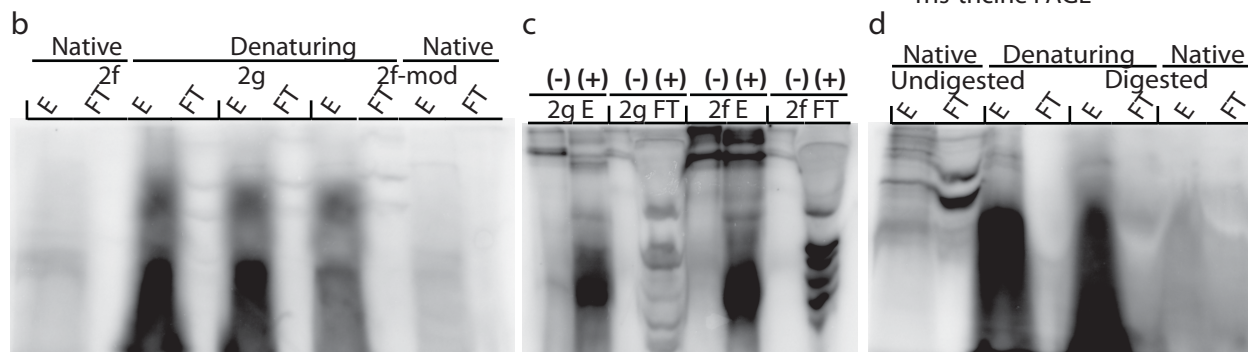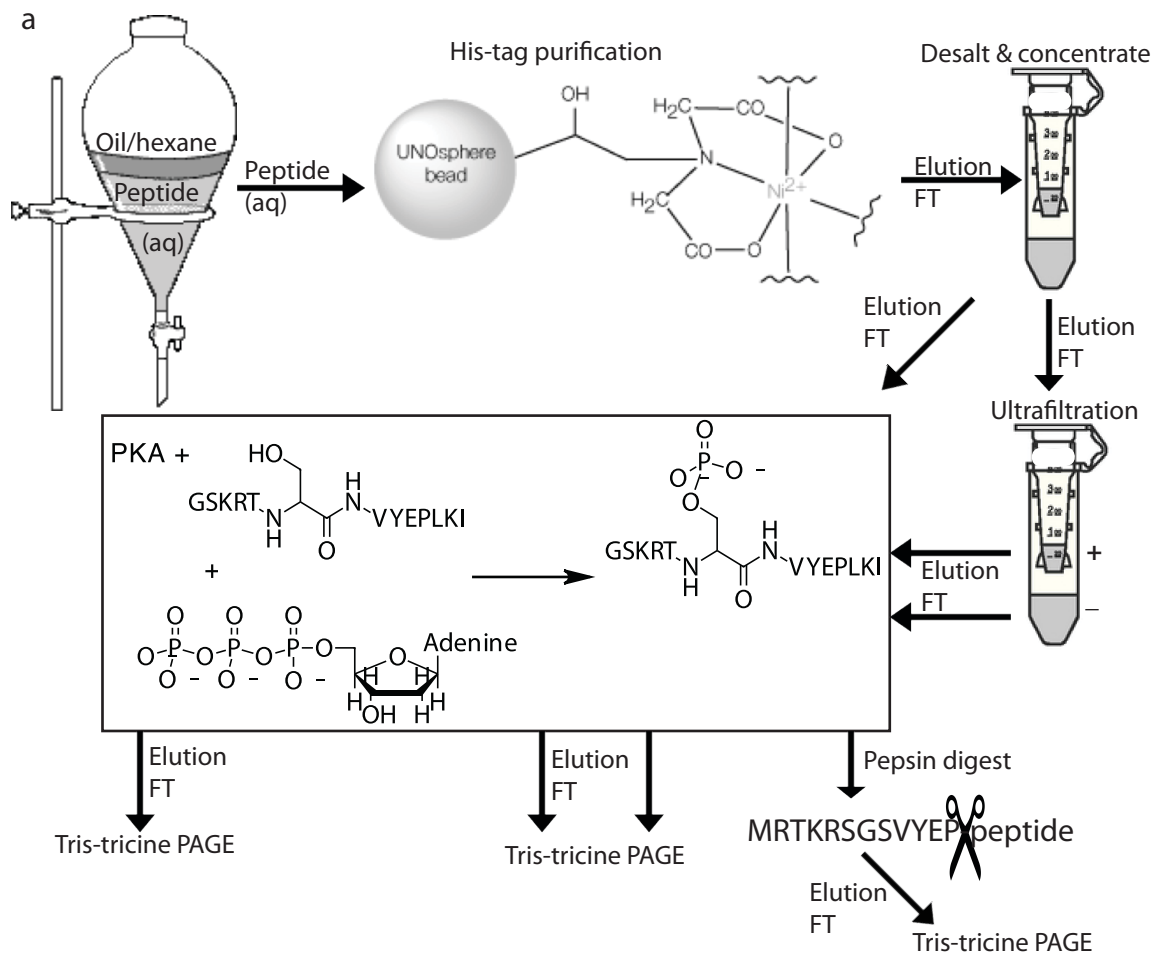
Figure 2.8. Analysis of peptide translation in water-in-oil emulsions. a) In general, a 100 µL peptide synthesis reaction from an emulsion was extracted with mineral oil and hexanes. The resulting aqueous layer was purified using a Ni^{2+} charged resin and the elution and flow-through samples were desalted, concentrated with a spin filter with a MWCO of 3 kDa, and labeled using 2 units PKA and 0.165 γ-^{32}P-ATP and analyzed on a 15 % Tris-tricine PAGE that was stored in a phosphor screen overnight and scanned on a Typhoon Trio. Analysis of peptide reactions were performed by comparing b) native and 6 M urea His_6 tag purification, c) upper and lower fractions from ultrafiltration using a 10 kDa MWCO before labeling, and d) native and 6 M urea His_6 tag purified and labeled peptide before and after incubating with 150 units of pepsin at pH 2.2 for 1 hr at 37 °C.

2.4 Pool construction

With confidence that the optimized construct **2g** was successfully expressed, *in vitro*, testing the modified primers and optimization of pool construction were addressed. From commercially available modified bases, four were chosen including a 5′ acrydite phosphoramidite, an alkyne-modified deoxythymidine phosphoramidite, a 5′ aldehyde phosphoramidite and an amine-modified deoxythymidine phosphoramidite, later converted to an azide (Figure 2.9b). The acrydite moiety is commonly used in polymers, is bioorthogonal and supplies a slightly reactive moiety for a possible conjugation.[31] Aldehydes typically react with amine nucleophiles but they favor the carbonyl compound in water over reaction with biological nucleophiles such as amines, thiols and alcohols.[32] Azides and alkynes are non-biological and have been utilized in a 1,3-dipolar cycloaddition originally catalyzed by copper and later in a biocompatible approach without copper and using ring strain.[33,34] The objective of addition of these modified bases to the 5′ end of the dsDNA pool was to facilitate selection for a catalytic peptide that would specifically and bioorthogonally react. To verify the compatibility of the modified primers, several control reactions were performed. Each individually modified primer as well as primers containing all four modifications was subjected to PCR amplification without any loss to amplification efficiency. Transcription was also performed with singly modified DNA as well as fully modified DNA and only transcription yield for the azide-modified construct suffered slightly.
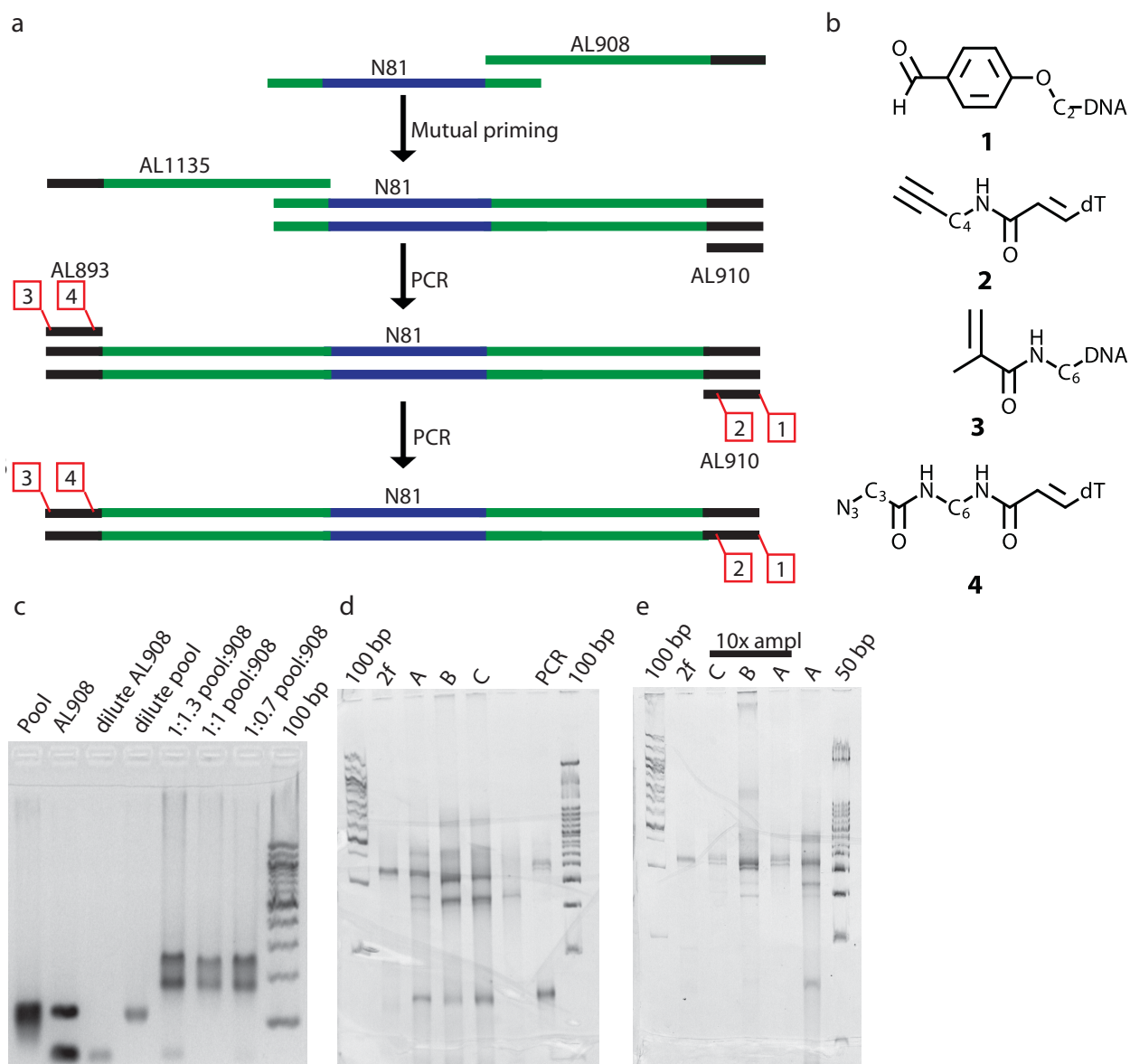
Figure 2.9. Pool construction and modification. a) Overview of assembling the pool and addition of the b) chemically modifed oligomers including an aldehyde **1**, an alkyne **2**, an acrydite **3**, and an azide **4.** c) Mutual priming was optimized by decreasing the ratio of pool to AL908 in a reaction with 1x DreamTaq™ Master Mix and imaged on a 2 % agarose gel. d) Mutual priming and amplification steps with 1x DreamTaq™ Master Mix were combined with reaction A containing 1 µM AL908/AL1135 and 0.5 µM pool, reaction B containing 2 µM AL908/AL1135 and 1 µM pool and reaction C containing 1 µM AL908/pool and 2 µM AL1135/AL910. e) Reactions A, B, and C were amplied 10-fold in 1x DreamTaq™ Master Mix and the samples as well as construct **2f** and product from reaction A were imaged on a 7.5% polyacrylamide gel with 2.75 M urea.

Optimization of pool construction was performed next because if 99 % yield was assumed to occur with each addition of a nucleotide, the overall yield of a 235 nucleotide oligomer would be below 10 % which would be an unacceptable loss of starting diversity. The pool ordered was 127 nucleotides containing 81 random nucleotides flanked by 23 fixed nucleotides and, at the time, smaller than the maximum length allowed. Addition of the fixed sequences (AL908 & AL1135) necessary for *in vitro* coupled transcription and translation and incorporation of the modified bases were required to complete pool construction (Figure 2.9a). However, to preserve the highest diversity of the pool, each step had to be optimized for minimal amplification and losses. Initially, pool construction was attempted stepwise with mutual priming of varied ratios of the 127-nucleotide pool to AL908 (1:1.3, 1:1, 1:0.7) to force 100 % of the pool to be mutually primed and extended to the resulting 174 base pair product (Figure 2.9c). However, the hairpin designed into AL908 led to dimerization and two products and no improvement was observed when an annealing temperature gradient was performed. Next, all pool construction steps were combined into one reaction at varied concentrations, in parallel (Figure 2.9d). Each reaction was designed to amplify the pool 2-fold, with reaction A and B containing 1 μM and 2 μM of AL908 and AL1135 and 0.5 μM and 1 μM pool, respectively. Reaction C was designed to mutually prime and extend 1 μM AL908 with 1 μM pool in the first annealing/extension cycle and succeeding cycles would amplify the pool 2-fold with 2 μM AL1135 and AL910. While none of these reactions converted the 127 nt pool to the full length product efficiently, reaction A produced approximately 40 % of full-length product (235 bp) from starting reactants which was the best outcome observed.

In order to install the modified bases onto 235 base pair full-length pool, PCR

amplification had to be performed with the modified primers and resulted in a final dsDNA pool that had been amplified 20-fold (Figure 2.9e). The resulting modified pool PCR reaction still contained the 127 nt pool oligomer and the 174 bp intermediate, so it required a final purification step before it could be employed. During assembly of the modified pool, approximately 60 % was lost and after a total of 20-fold amplification, only 5 % could be used to insure that the starting pool only had one copy of each member. The goal for the starting diversity of the selection was 10^{14} members, which meant starting with ~166 nmole modified dsDNA and assembly and amplification steps would require ~83 µmole of the 127 nt oligomer. Assembly and amplification of the pool using these methods were not practical for maintaining the starting diversity goal. Today, an oligomer with random bases can be synthesized up to 200 nucleotides and it would be more advantageous to order a pool with the required flanking regions already incorporated than attempt construction through multiple steps and loss of diversity. Elimination of 12 codons from the random region or removal of unnecessary nucleotides after the stop codon could reduce the 235-nucleotide sequence to 200 nucleotides for easy ordering. Then only amplification with the modified primers would be required before a selection could be initiated.
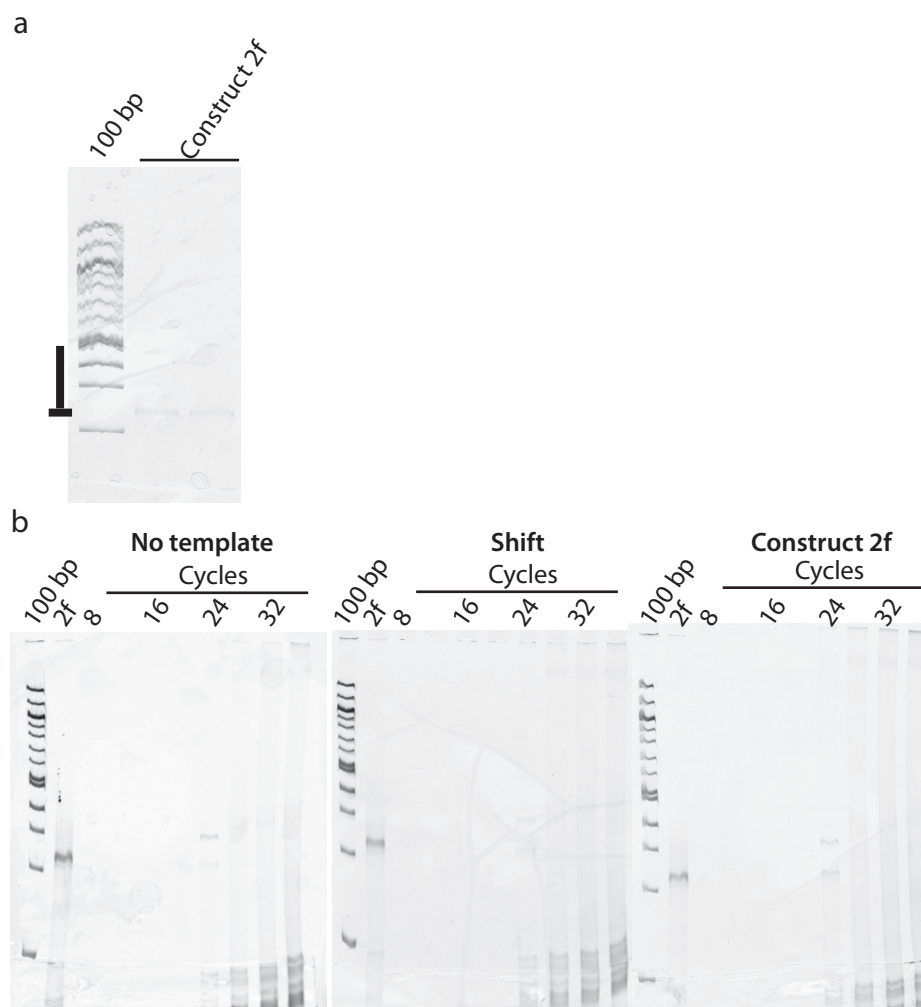
Figure 2.10. Mock selection of construct 2f. Mock selection by EMSA of construct **2f** DNA was performed on a 7.5 % PAGE with 2.75 M urea and a) the 235 bp band and area above it excised, eluted with 10 mM Tris, pH 7.5, concentrated with a spin filter with a 30 kDa MWCO and PCR amplified with 1x DreamTaq™ Master Mix and 2 µM AL893 and AL910 with b) aliquots taken every 4 cycles of a no template control, the shift and the band and imaged on a 7.5 % PAGE with 2.75 M urea.

2.5 Mock selection

In order to select for a peptide-DNA conjugate, an electrophoretic mobility shift assay (EMSA) was planned. A successful reaction during selection could lead to two different outcomes. If an active peptide sequence reacted with one of the reactive moieties, the resulting peptide-DNA conjugate would shift relative to the dsDNA pool on a partially denaturing PAGE and active sequences could be separated away from the inactive ones. If the peptide sequence brought separate bioorthogonal moieties together for a reaction, a DNA-DNA or DNA-RNA conjugate could be formed, resulting in a product mass that could be as much as doubled. For these possible outcomes, a mock selection was performed by fractionating the unmodified, purified construct **2f** on a partially denaturing PAGE and separately excising the band corresponding to the dsDNA, a positive control, and the area above the band up to 500 bp, a negative control (Figure 2.10a). The excised gel was eluted with 10 mM Tris, concentrated, amplified and aliquots were taken and imaged (Figure 2.10b). Unfortunately, there was no difference between the no template control and the amplification reactions from the dsDNA band and shift. Since incorrectly sized products were observed in late cycles of the no template PCR control, these can be attributed to primer-dimers. Similar primer-dimers appear in the late PCR cycles of the shift and construct **2f** and no other PCR product. Since the positive control, construct **2f**, did not amplify, elution of the excised gel with Tris was not effective and had to be optimized. This outcome also did not bode well for PAGE purification of the dsDNA pool that was required after assembly and amplification.

To increase the elution efficiency of the dsDNA from a polyacrylamide gel, elution conditions were optimized by radioactively labeling AL893 to use in amplification of the pool

and construct **2f**. The resulting bands of the labeled pool were eluted with 10 mM Tris, 10 mM Tris/0.1% SDS, 0.3 M NaCl and 0.3 M NaCl/0.1 % SDS, in parallel. Radioactivity was quantified before and after eluting for 4 hours and 48 hours and the fraction eluted was calculated. Eluting with 0.3 M NaCl/0.1 % SDS resulted in the highest fraction with 0.4 at 4 hours and 0.6 at 48 hours. Using 10 mM Tris led to smallest fraction of 0.2 at 4 hours and 0.3 at 48 hours and elutions using 10 mM Tris/0.1% SDS or 0.3 M NaCl led to very similar fractions of 0.3 and 0.4 at 4 and 48 hours, respectively. This indicated that an elution buffer was most effective with higher ionic strength and surfactants present, but only 60 % of the DNA duplex or any selected peptide-DNA conjugate was expected to be recovered by EMSA.

Construct **2f** was amplified with the individually modified primers, resulting in four doubly modified DNA duplexes (alkyne/acrydite, aldehyde/azide, alkyne/azide & aldehyde/acrydite). These modified DNA constructs were subjected to a mock selection on a partially denaturing PAGE with the band corresponding to the dsDNA and the area up to 500 bp excised and eluted with the optimized elution buffer. The resulting eluent was amplified and aliquots were taken from each reaction. While elution was effective, amplification of the band and the shift occurred at the same rate in every case. For a successful selection to occur, the positive control, the DNA duplex, should have amplified at a faster rate relative to the negative control. The hope was that little to no amplification of the excised shift would occur so that when a selection for a DNA-conjugate was performed there would be no background amplification to contend with and only active sequences would be amplified during the electrophoretic mobility shift assay. This outcome was catastrophic and repeated mock selections showed less amplification of the shift but amplification of the band corresponding to construct **2f** suffered as well.
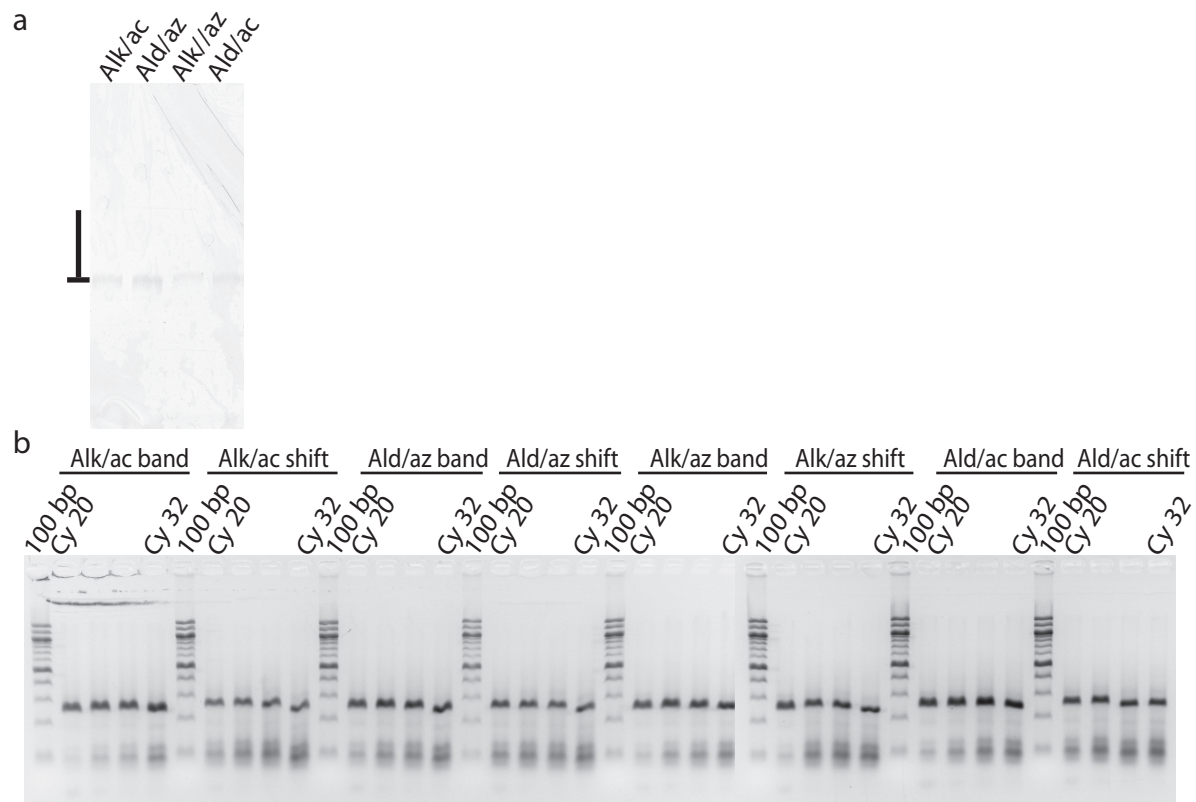
Figure 2.11. Mock selection of construct 2f with modified primers. Mock selection by EMSA of construct **2f** with modified primers was performed on a 7.5 % PAGE gel with 2.75 M urea and a) the 235 bp band and area above it was excised, eluted with 10 mM Tris, pH 7.5, concentrated with a spin filter with a 30 kDa MWCO and PCR amplified with 1x DreamTaq™ Master Mix and 2 μM AL893 and AL910 with b) aliquots taken every 4 cycles of the shift and the band and imaged on a 7.5 % PAGE with 2.75 M urea.

2.6 Conclusions

The development of a system that could successfully transcribe and translate a colocalized DNA library into its corresponding peptide library, *in vitro*, was partially completed. A completely optimized system could allow for a selection of a peptide that could specifically react with a bioorthogonal moiety of its coding DNA and lead to a DNA-peptide conjugate that could be utilized in a new display system that could surpass other peptide/protein display methods in ease of construction and starting diversity. Each component of the system required optimization, including the coupled *in vitro* transcription and translation, peptide expression in water-in-oil emulsions, construction of the pool and the selection protocol for separating out and amplifying active sequences away from the inactive majority.

A linear PCR product was optimized for stability in an *in vitro* coupled transcription and translation reaction and included the T7 promoter, optimized ribosomal binding site, start and stop codons and a 3' 15 bp hairpin and an 8 bp hairpin between the T7 promoter and the ribosomal binding site for mRNA stability. Codons for a His_6 tag and PKA substrate were determined to be optimal for purification and imaging, respectively. The radiolabeled peptide product was successfully imaged and verified on a Tris-tricine PAGE. Purification of the peptide product under denaturing conditions indicated that the product easily bound to other proteins in the *E. coli* lysate. Successful expression of the peptide in water-in-oil emulsions was observed and the product was verified to be the anticipated peptide but never quantified. An ideal selection for a catalytic peptide would have started with 10^{14} DNA molecules and each member would have expressed more than one corresponding peptide. This would have led to at least 165 nmoles of expressed peptide, which should have been

easily detectable with methods less sensitive than radiolabeling.

Construction of the pool was successful but with an overall amplification of 20-fold. However, the multistep assembly led to 60 percent loss, making the goal of a starting diversity of 10^{14} members difficult to obtain on a practical level. Incorporation of the bioorthogonal primers was successful and all testing indicated that they were not disruptive to the *in vitro* expression reaction. By eliminating unnecessary portions of the pool or ligating the fixed sequences to the random portion, construction of the final bioorthogonally-modified pool would be easier and would not diminish the desired starting diversity.

Failure of the negative control in the EMSA mock selection coupled with the fact that the fully modified dsDNA pool required PAGE purification before selection, led to the end of this project. Mock selection of the dsDNA never displayed a higher rate of amplification of the positive control (DNA duplex) relative to the negative control (shift). Instead of trying to maintain the DNA duplex on a partially denaturing PAGE, a fully denaturing PAGE may have allowed separation of inactive sequences from any reacted peptide-DNA conjugates while eliminating stray dsDNA that may have ran slower and been caught up in the shifted area of the PAGE.

Use of a more established display method such as mRNA display or phage display to initially isolate a catalytic peptide that reacts with an immobilized bioorthogonal moiety would be a useful intermediate step before DNA was introduced to form the potential peptide-DNA conjugate. Peptide sequences unable to bind or react with the moieties would be washed away and active sequences would be available for amplification and repeated rounds of selection. Once a bioorthoganol reaction pair is identified, the DNA-peptide conjugate can be used to optimize a more robust selection system. Bioorthogonal probes

have become extremely valuable for studying many cellular macromolecules such as glycans, proteins and lipids and discovery of a catalytic peptide that could react specifically with a bioorthogonal moiety would be another useful tool. Implementation of a catalytic peptide to connect genotype to its phenotype in an efficient way could enable efficient construction of peptide and protein libraries of up to 10^{14} members without requiring purification as seen in mRNA display and containing unnatural or toxic peptides or proteins that would otherwise not be expressed in phage display.

Experimental

Construct design. Initially, constructs were designed to contain a T7 RNA polymerase promoter (TAATACGACTCACTATA) 38 bases upstream of the translation initiation site, a Shine-Dalgarno sequence (AGGAGGT) 16 bases upstream of the translation initiation site, and a start codon (ATG) and a stop codon (TAA), sandwiching 22 to 27 codons (Figure 2.1). Codons included in test constructs included a His tag (HHHHHH), a tetracysteine tag (CCPGCC), and later a PKA substrate (RTKRSGSVYEPLKI). Sequences were inputted into RNAfold to ensure that secondary structure was minimized.

Purification of oligonucleotides. Oligonucleotides purchased from Invitrogen or Integrated DNA technologies (Table 2.1) were resuspended in Millipore filtered water to 100 μM. Oligonucleotides larger than forty nucleotides were gel purified by a 7.5% denaturing polyacrylamide gel containing 7 M urea in 0.5x TBE buffer (0.5x Tris, boric acid, EDTA, pH 8.4 buffer), 0.1% APS and 0.01% TEMED. Samples were diluted 2-fold in 8 M urea loading buffer containing xylene cyanol and bromophenol blue and run at 20 watts with 0.5x TBE buffer for up to 1.5 hours, as necessary. The gel was imaged by UV shadowing and the desired band was excised, eluted with 300 mM KCl, suspended to a final concentration of 70% ethanol and precipitated. The resulting pellets were resuspended in 50 μL water and the oligonucleotide concentration was quantified by UV spectroscopy using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific) at 260 nm.

Synthesis of an azidobutyramide-labeled oligonucleotide. To incorporate the azide group onto the C6-aminoalkyl thymidine of the oligonucleotide AL910, 0.2 μM of the oligonucleotide in 500 μL of 0.1 M Na_2CO_3/$NaHCO_3$, pH 9 was incubated for 2 hours at room

temperature with 45 nmoles of succinimidyl-4-azidobutyrate in 60 μL acetonitrile. Product formation was verified by FPLC using a Mono Q 5/50 ion exchange column (GE Healthcare Life Sciences).

Assembly of test constructs. Forward strands were mutually primed and extended with their corresponding reverse strand, containing approximately twenty complementary bases. Reactions containing a final concentration of 3 μM oligonucleotide forward and reverse strand and 1x of DreamTaq™ 2x Master Mix were denatured at 95°C for 1 minute, annealed at 50 °C for 30 seconds and elongated at 68°C for 2 minutes twice using a BioRad MJMini Personal Thermal Cycler. In preparation for a peptide synthesis reaction, the extended forward and reverse primers (3 μM final concentration) containing the T7 promoter region and additional 3′ terminal hairpin were added by polymerase chain reaction (PCR). The final construct was amplified by PCR containing a final concentration of 20 nM DNA template, 1x Taq Master Mix and 3 μM of the forward and reverse primers. The PCR conditions denatured the reaction at 95 °C for 2 minutes, followed by at least 12 cycles (as necessary) of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds, and completed with a final elongation at 72 °C for 2 minutes. Once an adequate amount of DNA was synthesized, it was desalted and concentrated using a Microcon YM-30 and the final concentration was determined by UV spectroscopy.

Primers were designed (AL525, 526) that would complement and amplify a portion of Invitrogen's provided sequence of the CALML3 plasmid and PCR was performed as described above.

Table 2.1. Oligonucleotide sequences from Invitrogen, Inc or Integrated DNA Technologies.

Oligo	Sequence	L
AL447	GGGAGAAGGAGGTAAAAAAAAAAT	24
AL448	TTTTTTCATTTTATTCATTTAGCAGC	26
AL449	CCCGCGAAATTAATACGACTCACTATAGGGAGAAGGAGGTAAAAAAAAAAT	51
AL450	GGGAGAAGGAGGTAAAAAAAAAATGCACCATCACCATCACCATCACCATCACCATGGCGCGAGCG	65
AL451	TTTTTTCATTTTATTCATTTAGCAGCAGCCCGGACAGCATTTGCCGCTCGCGCCATGGTGATGGT	65
AL455	GGGAGAAGGAGGTAAAAAAAAAATGCGCACCAAACGTAGCGGCAGCGTGTATGAACCGCTGAAAA	65
AL456	TTTTTTCATTTTATTCATTTAGCAGCAGCCCGGACAGCAGTTAATTTTCAGCGGTTCATACACGC	65
AL504	GGGAGAAGGAGGTAAAAAAAAAATGCGCACCAAACGTAGCGGCAGCGTGTATGAACCGCTGAAAATTAACCAC	74
AL505	TTTTTTCATTTTATTCATTTAGCAGCAGCCCGGACAGCAATGGTGATGGTGATGGTGGTTAATTTTCAGCGGTTC	75
AL525	TTTATTCATTTAGCAGCAGCCCGGACAGCAGACCTGCTCCTGAAAATACAGG	52
AL526	TTTATTCATTTAGCAGCAGCCC	22
AL690	TTCCCGCGAAATTAATACGACTCACTATAGGGAGAAGGAGGTAAAAAAAAA	51
AL692	CAAAAAACCCCCCCCCT	17
AL693	TTCCCGCGAAATTAATACGA	20
AL697	CAAAAAACCCCCCCCCTCGGCAAACCGAGGGGGGGGGGGTTATGCTAGTTAGCTTTTTTTCATTTTATTCATTTAGCAGC	78
AL757	CAAAAAACCCCCCCCCTCGGCAAGCCGAGGGGGGGGGGGTTATGCTAGTTAGCTTTTTTTCATTTTATTCATTTAGCAGC	78
AL758	TTCCCGCGAAATTAATACGACTCACTATAGGGAGA CCTTCGGGTTTCCCTC GGGAGAAGGAGGTAAAAAAAAAAT	75
AL893	TTCCCGCGAAATTAATACGAC	21
AL898	CAAAACCCCCCCCCTCGGCGAACCGAGGGGGGGGGGGTTTTATTTATTCAAATGGTGATGGTGATGGTG	67
AL900	GACAGCAATGGTGATGGTGATGGTGGTTAATTTTCAGCGGTTCATACACGCTGCCGCTACGTTTGGTGCGCATTCCTCGTTCCTCCTTATCTA	93
AL908	CTTAACGCGCGCCCTCGGCGAACCGAGGGCGCGCGTTTTATTTATTCAAATGGTGATGGTGATGGTG	67
AL910	CTTAACGCGCGCCCTC	16
AL1118	GTGAGACCACGCGGAACCGCCGCTACCGCCACCAGAGCCACCGCCTTTTTCAAACTGCGGGTGGCTCCAGCAGCAGCCCGGACAGCAATGGTGATGGTG	99
AL1133	CTTAACGCGCGCCCTCGGCGAACCGAGGGCGCGCGTTTTATTTATTCAAATGGTGATGGTGATGGTGTCATTTTTCGAACTGCGGGTGAGACCACGCGGA	100
AL1135	TTCCCGCGAAATTAATACGACTCACTATAGGGTGCCTTCGGGCACCCCCTTTTACGACACATAGATAAGGAGGAACGAGGAATG	84

AL1883	TAGATAAGGAGGAACGAGGAATG N_{81} TGN CA CCA TCA CCA TCA CCA TTT	127
AL893ac/ az	Acrydite-TTCCCGCGAAA/C6-azide-T/TAATACGAC	21
AL893az	TTCCCGCGAAA/C6-azide-T/TAATACGAC	21
AL893ac	Acrydite-TTCCCGCGAAATTAATACGAC	21
AL910ald	Aldehyde-CTTAACGCGCGCCCTC	16
AL910ald /alk	Aldehyde-CTTAACGCGCGCCC/C8-alkyne-T/C	16
AL910alk	CTTAACGCGCGCCC/C8-alkyne-T/C	16

Construct 2a (136 nts):

CCCGCGAAATTAATACGACTCACTATAGGGAGAAGGAGGTAAAAAAAAAATGCACCATCACCATC

ACCATCACCATCACCATGGCGCGAGCGGCAAATGCTGTCCGGGCTGCTGCTAAATGAATAAAATGA

AAAAA

Forward primer: AL447

Reverse primer: AL449

Forward strand: AL449

Reverse strand: AL450

Expected peptide sequence (mass): MHHHHHHHHHHHGASGKCCPGCC (2488.)

Construct 2b (178 nts):

TAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGA

AGGAGATATACCATGTCTGGTTCTCATCATCATCATCATCATGGTAGCAGCGGCGAAAACCTGTATT

TTCAGGAGCAGGTCTGCTGTCCGGGCTGCTGCTAAATGAATAAA

Forward primer: AL525

Reverse primer: AL526

Template: CALML3 plasmid

Expected peptide sequence (mass): MSGSHHHHHHGSSGENLYFQEQVCCPGCC (3210.)

Construct 2c (154 nts):

CCCGCGAAATTAATACGACTCACTATAGGGAGAAGGAGGTAAAAAAAAAATGCGCACCAAACGTA

GCGGCAGCGTGTATGAACCGCTGAAAATTAACCACCATCACCATCACCATTGCTGTCCGGGCTGCT

GCTAAATGAATAAAATGAAAAAA Forward primer: AL447

Forward primer: AL449

Reverse primer: AL448

Forward strand: AL504

Reverse strand: AL505

Expected peptide sequence (mass): MRTKRSGSVYEPLKINHHHHHHCCPGCC (3269.)

Construct 2d (208 nts):

TTCCCGCGAAATTAATACGACTCACTATAGGGAGAAGGAGGTAAAAAAAAAATGCGCACCAAACG

TAGCGGCAGCGTGTATGAACCGCTGAAAATTAACCACCATCACCATCACCATTGCTGTCCGGGCTG

CTGCTAAATGAATAAAATGAAAAAAAGCTAACTAGCATAACCCCCCCCCTCGGTTTGCCGAGGGGG

GGGGTTTTTTG

Forward primer: AL693

Reverse primer: AL692

Extended forward primer: AL690

Extended reverse primer: AL697

Forward strand: AL504

Reverse strand: AL505

Expected peptide sequence (mass): MRTKRSGSVYEPLKINHHHHHHCCPGCC (3269.)

Construct 2e (230 nts):

TTCCCGCGAAATTAATACGACTCACTATAGGGAGACCTTCGGGTTTCCCTCGGGAGAAGGAGGTAA

AAAAAAAATGCGCACCAAACGTAGCGGCAGCGTGTATGAACCGCTGAAAATTAACCACCATCACC

ATCACCATTGCTGTCCGGGCTGCTGCTAAATGAATAAAATGAAAAAAAGCTAACTAGCATAACCCC

CCCCCTCGGTTCGCCGAGGGGGGGGGGTTTTTTG

Forward primer: AL693

Reverse primer: AL692

Extended forward primer: AL757

Extended reverse primer: AL758

Forward strand: AL504

Reverse strand: AL505

Expected peptide sequence (mass): MRTKRSGSVYEPLKINHHHHHHCCPGCC (3269.)

Construct 2f (197 nts):

TTCCCGCGAAATTAATACGACTCACTATAGGGTGCCTTCGGGCACCCCCTTTTACGACACATAGATA

AGGAGGAACGAGGAATGCGCACCAAACGTAGCGGCAGCGTGTATGAACCGCTGAAAATTAACCA

CCATCACCATCACCATTGCTGTCAATAAAACGCGCGCCCTCGGTTCGCCGAGGGCGCGCGTTAAG

Forward primer: AL893

Reverse primer: AL910

Forward strand: AL1135

Reverse strand: AL900

Extended reverse primer: AL908

Expected peptide sequence (mass): MRTKRSGSVYEPLKINHHHHHHCCQ (3037.)

Construct 2g (319 nts):

TTCCCGCGAAATTAATACGACTCACTATAGGGTGCCTTCGGGCACCCCCTTTTACGACACATAGATA

AGGAGGAACGAGGAATGCGCACCAAACGTAGCGGCAGCGTGTATGAACCGCTGAAAATTAACCA

CCATCACCATCACCATTGCTGTCCGGGCTGCTGCTGGAGCCACCCGCAGTTTGAAAAAGGCGGTGG

CTCTGGTGGCGGTAGCGGCGGTTCCGCGTGGTCTCACCCGCAGTTCGAAAAATGACACCATCACCA

TCACCATTTGAATAAATAAAACGCGCGCCCTCGGTTCGCCGAGGGCGCGCGTTAAG

Forward primer: AL893

Reverse primer: AL910

Forward strand: AL1135

Reverse strand: AL900

Extended reverse primer 1: AL1118

Extended reverse primer 2: AL1133

Expected peptide sequence (mass): MRTKRSGSVYEPLKINHHHHHHCCPGCCWSHPQFEKGGGS

GGGSGGSAWSHPQFEK (6138.)

Pool (235 nts):

TTCCCGCGAAATTAATACGACTCACTATAGGGTGCCTTCGGGCACCCCCTTTTACGACACATAGATA

AGGAGGAACGAGGAATGN$_{81}$TGNCACCATCACCATCACCATTTGAATAAATAAAACGCGCGCCCTC

GGTTCGCCGAGGGCGCGCGTTAAG

Forward primer: AL893

Reverse primer: AL910

Forward strand: AL1883

Reverse strand: AL908

Extended forward primer: AL1135

Transcription of test constructs. If the RNA construct was used in the protein synthesis reaction, a final concentration of ~0.5 μM DNA construct was used in a transcription reaction, containing 40 mM Tris, 0.01% Triton X-100, 20 mM DTT, 2 mM spermidine, 20 mM rNTPs, 25 mM MgCl$_2$, 10% DMSO and 2.5% T7 RNA polymerase and incubated for 3 hours at 37 °C. The reaction was stopped with a final concentration of 25 mM EDTA and gel purified, eluted, and precipitated as described in oligonucleotide purification.

Peptide synthesis. Depending on the reaction, the DNA or RNA construct was added (Table 2.2) to a reaction including 20 μL *E. coli* extract from Expressway Cell-Free *E. coli* Expression System (Invitrogen), 20 μL 2.5x *in vitro* protein synthesis (IVPS) buffer, a final concentration of 2.5 mM amino acids (-Met) and 3 mM methionine, 1 μL T7 enzyme mix, if initiated with DNA, and water up to 50 μL. The reaction was agitated while incubating at 30 °C for 30 minutes before adding solution including 12.5 μL 2x IVPS feed buffer, a final concentration of 1.25 mM amino acids (-Met) and 1.5 mM methionine and increasing the temperature to 37 °C.

RNA quantification. A modified transcription reaction was performed to radiolabel construct 2e with [32]P. The reaction contained a final concentration of ~0.5 μM **2e** DNA construct, 40 mM Tris, 0.01% Triton X-100, 20 mM DTT, 2 mM spermidine, 20 mM rGTP, rCTP, and rTTP, 4 mM rATP, 25 mM MgCl$_2$, 10% DMSO, 2.5% T7 RNA polymerase, and ~75 nM α-^{32}P-ATP and incubated for 3 hours at 37 °C. The reaction was gel purified, eluted, and precipitated as described in oligonucleotide purification. In two parallel peptide synthesis reactions, as described above, 40 pmoles of labeled RNA supplemented 240 pmoles of **2e** unlabeled RNA, including and excluding 100 units of RNase inhibitor (New England Biolabs). Aliquots were

taken on an hourly basis and extracted with 20%/80% phenol/chloroform. RNA extract

samples were diluted 2-fold with 8 M urea loading buffer containing xylene cyanol and

bromophenol blue and loaded onto a 7.5% denaturing polyacrylamide gel containing 7 M

urea in 0.5x TBE buffer (0.5x Tris, boric acid, EDTA, pH 8.4 buffer), 0.1% APS and 0.01% TEMED.

The gel was exposed on a scanning storage phosphor screens (Molecular Dynamics) for 1 day

and scanned at 200 μm resolution on a Typhoon Trio (GE Healthcare).

Table 2.2. Peptide synthesis conditions and visualization methods utilized.

Construct	Nucleic acid	Amount (pmoles)	Analysis methods
2a	DNA	11.9	Brilliant Blue R
	DNA	34.0	MALDI/ Brilliant Blue R
	DNA	71.8	Brilliant Blue R /MALDI
(+) Control	DNA	7.78	Brilliant Blue R /MALDI
2b	DNA	12.1	Brilliant Blue R /MALDI
	RNA	375	MALDI
2c	RNA	260	Lumio™/ Brilliant Blue R
	RNA	483	Lumio™/^{32}P labeling
	RNA	805	Lumio™
	RNA	347	Lumio™/^{32}P labeling
2d	RNA	387	Lumio™/^{32}P labeling
	RNA	245	Lumio™/^{32}P labeling
	RNA	163	^{32}P labeling
2e	RNA	279	^{32}P labeling
	RNA	312	^{32}P labeling
	DNA	10.5	^{32}P labeling
	RNA	257	^{32}P labeling
	DNA	9.3	^{32}P labeling
	DNA	19.7	^{32}P labeling
	DNA	13.1	^{32}P labeling
	DNA	13.1	^{32}P labeling
	DNA	15.0	^{32}P labeling
	DNA	15.0	^{32}P labeling
	DNA	15.0	^{32}P labeling
	DNA	13.1	^{32}P labeling
2f	DNA	18.3	^{32}P labeling
	DNA	13.5	^{32}P labeling
	DNA	10.5	^{32}P labeling
	DNA	10.9	^{32}P labeling

	DNA	21.8	^{32}P labeling/ pepsin digest
	DNA	20.5	^{32}P labeling/ pepsin digest
2g	DNA	13.5	^{32}P labeling
	DNA	10.5	^{32}P labeling
	DNA	10.9	^{32}P labeling
	DNA	21.8	^{32}P labeling/ pepsin digest
	DNA	20.5	^{32}P labeling/ pepsin digest

Peptide synthesis in water-in-oil emulsions. A DNA or RNA construct was added (Table 2.2) to an aqueous 100 μL reaction containing 20 μL *E. coli* extract, 20 μL 2.5x IVPS buffer, 25 μL 2x IVPS feed buffer, 5 μL 50 mM amino acids (-Met), 4 μL 75 mM methionine, 2.5 μL T7 RNA polymerase, 100 units murine RNase inhibitor (New England Biolabs), 9 μL 20 mg/mL BSA (New England Biolabs) and 10 μL 45% Span® 80 with 5% Tween® 80 (aq).

Separately, an oil mixture was prepared by adding 450 μL Span® 80 and 50 μL Tween® 80 to 9.5 mL light mineral oil to produce a 4.5% and 0.5% surfactant concentration, respectively. Emulsions were made by two different methods. Initially, the 100 μL peptide synthesis reactions were stirred into 900 μL of the cold oil mixture at 1150 rpm over 1 minute and then stirred for an additional 2 minutes. Later, emulsions were made by combining 100 μL peptide synthesis reaction and 900 μL cold oil mixture and passing the mixture through a mini-extruder (Avanti Polar Lipids, Inc) with a 1.0 μm polycarbonate membrane (Avanti). The emulsified reaction was incubated at 30 °C for 30 minutes to initiate the reaction and then the temperature was increased to 37 °C. Aliquots of the mixed and extruded emulsions were spotted on acid-cleaned slides, recorded on a confocal microscope and the size was compared to magnetic beads (Pierce) with a mean diameter of 3.16 μm.

Peptide purification. After 6 hours of incubation time, the expression reactions were spun down and the surfactants were extracted with 3x1000 μL light mineral oil and the

mineral oil was extracted with 3x1000 μL hexanes. A binding buffer with a final concentration of 50 mM NaH$_2$PO$_4$, 300 mM NaCl, pH 8.0 was added to the resulting aqueous mixture and was incubated for 5 minutes with 400 μL 50% IMAC resin charged with Ni^{2+} (Bio-Rad), pre-equilibrated in 5 column volumes of binding buffer. The column was washed with 500 μL of binding buffer and eluted with 500 μL 50 mM NaH$_2$PO$_4$, 300 mM NaCl, 300 mM imidazole, pH 8.0. The first flow-through and peptide elution were precipitated in 80% acetone and centrifuged to obtain a pellet used later in analysis.

Tris-tricine polyacrylamide gel electrophoresis. Peptide products were analyzed by several methods as summarized in Table 2.2. For gel loading, peptide samples were resuspended in 15 μL tricine sample buffer (50 mM Tris buffer, 12% glycerol, 8% SDS, 100 mM DTT, 0.01% Brilliant Blue R dye) and loaded onto a Tris-tricine SDS/PAGE gel, which was run at 50 V for 30 minutes through a 5% polyacrylamide stacking gel and at 175 V for 2 hours through the 18% polyacrylamide resolving portion of the gel, using 200 mM Tris, pH 8.9, as the anode buffer and 100 mM Tris, 100 mM tricine, 0.1% SDS, pH 8.4, as the cathode buffer. The gel was stained for three hours using a 10 % acetic acid, 50 % methanol and 0.5% Brilliant Blue R dye solution and destained using a 10 % acetic acid, 50 % methanol solution and imaged.

Analysis by MALDI-TOF. His$_6$ tag purified peptides were desalted, concentrated and resuspended in α-cyano-4-hydroxy-cinamic acid using OMIX pipette C18 tips (Varian). The sample was eluted using 75% acetonitrile in water with 0.1% triflouroacetic acid onto a plate for analysis by MALDI-TOF (Perceptive Biosystems DE STR system).

Analysis by Lumio™ Reagent. Two microliters of fluorescein arsenical hairpin binding

(FlAsH) reagent (Life Technologies) was added to tetracysteine tagged peptides (CCPGCC) and an aliquot of 5 μL diluted up to 50 μL was analyzed in a quartz cuvette with a Jasco FP-6300 Spectrofluorometer exciting at 500 nm and scanning the emission from 510-600 nm.

Purified peptide pellets containing the tetracysteine tag were resuspended in 4x Lumio™ sample loading buffer (Life Technologies) and water up to 20 μL and incubated with 0.2 μL Lumio™ green detection reagent (Life Technologies) at 70 °C for 10 minutes and adding 1μL detection enhancer reagent before loading on to a gel as described above. The gel was imaged on a Typhoon Trio (GE Healthcare) by exciting at 488 nm and collecting the image data in the range of 520-535 nm.

PKA substrate positive control optimization. To optimize the PKA substrate labeling reaction, a chemically synthesized peptide, RTKRSGSVYEPLKI or malantide (Abbiotec), was tested. Each 10 μL optimization reaction was performed with 0.5 μL 3.3 μM γ-^{32}P-ATP and 2 units (1 μL) of the PKA catalytic subunit in a reaction buffer of 50 mM Tris, 10 mM $MgCl_2$, pH 7.5. Separately, titrations of polyethylene glycol (0%, 7.5%, 15%, 22.5%, 30%, 37.5%) and unlabeled ATP (0 μM, 5 μM, 10 μM, 15 μM, 20 μM, 25 μM, 30 μM) were carried out with a constant 1 μM peptide. Peptide samples were precipitated by adding a final concentration of 20% trichloroacetic acid to each sample and incubating at 4 °C for 15 minutes, spun twice at 15,000 rpm for 5 minutes and washed twice with 60 μL cold acetone. Pelleted peptide samples were heated at 95 °C for 2 minutes and resuspended in loading buffer and run on a Tris-tricine polyacrylamide gel, as described above.

To determine the detection limit of the Typhoon trio and labeling limit of the enzyme, peptide titrations (1 μM, 0.5 μM, 0.1 μM, 0.5 μM, 10 nM, 5 nM, 1 nM, 0.5 nM, 0.1 nM) and

dilution of a 1 μM reaction (5x, 10x, 50x, 100x, 500x, 1000x, 5000x, 10,000x) were also carried out and precipitated with 20% trichloroacetic acid. Pelleted peptide samples were heated at 95 °C for 2 minutes and resuspended in loading buffer and run on a Tris-tricine polyacrylamide gel, as described above.

Analysis by PKA substrate labeling. Subsequent His_6 tag-purified peptides containing the protein kinase A substrate (RTKRSGS*VYEPLKI) and their corresponding flow-through sample were labeled with ^{32}P by incubating 0.5 μL 3.3 μM γ-^{32}P-ATP and 2 units (1 μL) of the PKA catalytic subunit in 10 μL of 50 mM Tris, 10 mM $MgCl_2$, pH 7.5. The reaction was incubated for 5 hours at 37 °C, desalted using Sephadex® G-10, diluted using tricine sample buffer and loaded on to a Tris-tricine PAGE, as described above. The gel was exposed on a scanning storage phosphor screens (Molecular Dynamics) for 1 day and scanned at 100 μm resolution on a Typhoon Trio.

Size verification of the peptides synthesized were performed by resuspending the His_6 tag-purified peptide elution and flow-through samples in 100 μL H_2O and centrifuging each sample through a Microcon® YM-10 (Millipore) at 14,000 x g for 20 minutes. The sample that was retained on the filter as well as the filtrate was concentrated with a Microcon® YM-3 (Millipore) by centrifuging at 14,000 x g for 1 hour and labeled with ^{32}P and analyzed on a Tris-tricine PAGE, as described above.

A pepsin digest of the expressed peptide was performed. Synthesized peptides were His_6 tag-purified under denaturing conditions by adding a final concentration of 6 M urea to the binding and elution buffers, as described in peptide purification. The flow-through and elution samples were desalted using Microcon® YM-3 and labeled with ^{32}P. The resulting

reactions were incubated with 15 or 150 units of pepsin for 1 hour at 37 °C and pH 2.2. The digested reactions were then analyzed on a gel, as described in Tris-tricine polyacrylamide gel electrophoresis.

Pool construction. To add the flanking sequences to chemically synthesized pool (AL1883), 1 μM final concentration of AL1135 and AL908 were added to a reaction containing a final concentration of 0.5 μM AL1883 and 1x DreamTaq™ PCR Master Mix (Thermo Scientific). The reaction was denatured at 95°C for 1 minute, annealed at 50 °C for 30 seconds and elongated at 72 °C for 2 minutes four times using a BioRad MJMini Personal Thermal Cycler. The elongated pool (235 bp) was concentrated ~4x using a Microcon® YM-30 by centrifuging at 14,000 x g for 8 minutes and added to 2x loading buffer containing SYBR gold (Life Technologies) and loaded onto a 10% semi-denaturing polyacrylamide gel containing 2.75 M urea in 0.5x TBE buffer, 0.1% APS and 0.01% TEMED. The gel was run, as described in oligonucleotide purification. The band corresponding to 235 bp was excised and eluted with 10 mM Tris, pH 7.5. The resulting dsDNA pool was then amplified with the modified primers (AL910ald/alk, AL893az/ac), as described in assembly of test constructs, with 4 cycles of PCR and an annealing temperature of 58 °C.

Mock selection of pool by an electrophoretic mobility shift assay. The purified, unmodified dsDNA pool was added to 2x loading buffer containing SYBR gold (Life Technologies) and loaded onto a 7.5% semi-denaturing polyacrylamide gel containing 2.75 M urea in 0.5x TBE buffer, 0.1% APS and 0.01% TEMED. The gel was run as described in oligonucleotide purification. The DNA duplex band and the area above it up to 500 bp, based

on a 100 bp ladder (New England Biolabs), was excised and eluted with 10 mM Tris, pH 7.5 and concentrated with a Microcon® YM-30 by centrifuging at 14,000 x g for 8 minutes.

To quantify elution efficiency, each eluted sample was diluted to 100 μL with water and PCR amplified. One microliter of each sample was amplified, as well as a no template control by a 100 μL PCR reaction containing a final concentration 200 μM dNTPS, 2 μM forward and reverse primers (AL893, AL910) and 1 μL PFU in a buffer of 10 mM Tris•HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3. The PCR conditions denatured the reaction at 95 ℃ for 2 minutes, followed by 36 cycles of 95 ℃ for 30 seconds, 58 ℃ for 30 seconds and 72 ℃ for 30 seconds, and completed with a final elongation at 72 ℃ for 2 minutes. Aliquots were taken every 4 cycles after the 8th cycle and added to 2x loading buffer and loaded onto a 7.5% polyacrylamide containing 2.75 M urea, as described above. A similar procedure was also performed with the modified dsDNA pool, except the excised bands were eluted with 300 mM NaCl and 0.1% SDS.

Elution efficiency of the double-stranded pool was optimized by first labeling an oligonucleotide with [32]P. Ninety-five pmoles of an oligonucleotide (AL893) was incubated for 1 hour at 37 ℃ with 20 units of T4 polynucleotide kinase (New England Biolabs) with ~9 pmoles γ-^{32}P-ATP in a buffer of 70 mM Tris, 10mM $MgCl_2$, 5 mM DTT, pH 7.6. The labeled oligonucleotide was used in PCR amplification of the purified pool, added to 2x loading buffer and loaded onto a 7.5% polyacrylamide containing 2.75 M urea, as described above. The resulting gel was exposed on a scanning storage phosphor screens (Molecular Dynamics) for 1 hour and scanned at 200 μm resolution on a Typhoon Trio. The bands containing the amplified dsDNA were excised and eluted with for different buffers: 300 mM NaCl, 300 mM

NaCl and 0.1% SDS, 10 mM Tris, 10 mM Tris and 0.1% SDS. Radiation (CPM) was recorded for

the gel and supernatant using a Geiger counter at 4, 24 and 48 hours.

References

(1)   Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611.
(2)   Aharoni, A.; Griffiths, A. D.; Tawfik, D. S. *Curr. Opin. Chem. Biol.* **2005**, *9*, 210.
(3)   Smith, G. P. *Science* **1985**, *228*, 1315.
(4)   Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297.
(5)   Tawfik, D. S.; Griffiths, A. D. *Nat. Biotechnol.* **1998**, *16*, 652.
(6)   Griffiths, A. D.; Tawfik, D. S. *EMBO J.* **2003**, *22*, 24.
(7)   Smith, G. P.; Petrenko, V. A. *Chem. Rev.* **1997**, *97*, 391.
(8)   Amstutz, P.; Forrer, P.; Zahnd, C.; Pluckthun, A. *Curr. Opin. Biotechnol.* **2001**, *12*, 400.
(9)   Lipovsek, D.; Pluckthun, A. *J. Immunol. Methods* **2004**, *290*, 51.
(10)   Wilson, D. S.; Keefe, A. D.; Szostak, J. W. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3750.
(11)   Keefe, A. D.; Szostak, J. W. *Nature* **2001**, *410*, 715.
(12)   Sletten, E. M.; Bertozzi, C. R. *Angew. Chem.-Int. Edit.* **2009**, *48*, 6974.
(13)   Zubay, G. *Annu. Rev. Genet.* **1973**, *7*, 267.
(14)   Lesley, S. A.; Brow, M. A. D.; Burgess, R. R. *J. Biol. Chem.* **1991**, *266*, 2632.
(15)   Kim, D.-M.; Swartz, J. R. *Biotechnol. Bioeng.* **1999**, *66*, 180.
(16)   Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. *Methods Enzymol.* **1990**, *185*, 60.
(17)   Rosa, M. D. *J. Mol. Biol.* **1981**, *147*, 199.
(18)   Barrick, D.; Villanueba, K.; Childs, J.; Kalil, R.; Schneider, T. D.; Lawrence, C. E.; Gold, L.; Stormo, G. D. *Nucleic Acids Res.* **1994**, *22*, 1287.
(19)   Wada, K.; Wada, Y.; Ishibashi, F.; Gojobori, T.; Ikemura, T. *Nucleic Acids Res.* **1992**, *20*, 2111.
(20)   Hofacker, I. L. *Nucleic Acids Res.* **2003**, *31*, 3429.
(21)   Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. *J. Am Chem. Soc.* **2002**, *124*, 6063.
(22)   Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. *The Proteomics Protocols Handbook*; Humana Press, 2005.
(23)   Murray, K. J.; England, P. J.; Lynham, J. A.; Mills, D.; Schmitz-Peiffer, C.; Reeves, M. L. *Biochemistry Journal* **1990**, *267*, 703.
(24)   Ahn, J.; Chu, H.; Kim, T.; Oh, I.; Choi, C.; Hahn, G.; Park, G.; Kim, D. *Biochem. Biophys. Res. Comm.* **2005**, *338*, 1346.
(25)   Antao, V. P.; Lai, S. Y.; Tinoco, I. *Nucleic Acids Res.* **1991**, *19*, 5901.
(26)   Malencik, D. A.; Anderson, S. R. *Anal. Biochem.* **1983**, *132*, 34.
(27)   Armstrong, R. N.; Kondo, H.; Kaiser, E. T. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 722.
(28)   Salis, H. M.; Mirsky, E. A.; Voigt, C. A. *Nat. Biotech.* **2009**, *27*, 946.
(29)   Schmidt, T. G. M.; Batz, L.; Bonet, L.; Carl, U.; Holzapfel, G.; Kiem, K.; Matulewicz, K.; Niermeier, D.; Schuchardt, I.; Stanar, K. *Protein Expression Purif.* **2013**, *92*, 54.
(30)   Spirin, A. S. *Trends Biotechnol.* **2004**, *22*, 538.
(31)   Mitra, R. D.; Church, G. M. *Nucleic Acids Res.* **1999**, *27*.
(32)   Jencks, W. P. *J. Am Chem. Soc.* **1959**, *81*, 475.

(33)   V. V. Rostovtsev, L. G. G., V. V. Fokin, K. B. Sharpless *Angew. Chem.* **2002**, *114*, 2708.

(34)   Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am Chem. Soc.* **2004**, *126*, 15046.

Chapter 3

Progress towards the selection of a deoxyribozyme with galactosidase activity

3.1 Introduction

Darwinism is a theory of biological evolution that states that new species arise by natural selection of inherited traits that are better able to thrive, survive and produce more over successive generations. *In vitro* selection mimics this idea in a laboratory setting and was a technique first conceptualized in the 1960s when Speigelman utilized RNA and the mutation rate of Qβ replicase in RNA bacteriophage Qβ to evolve new viral genome variants with shortened replication times in a cell-free environment.[1] On a practical level, *in vitro* selections of DNA and RNA were not possible on a larger scale until the late 1980s when DNA libraries could be chemically synthesized more cost effectively and efficiently using phosphoramidites and an automated process.[2]

Molecular evolution requires the phenotype be linked to its genotype, and in the case of RNA and DNA, functionality (phenotype) and the encoding strand (genotype) are the same.[1] A RNA or DNA *in vitro* selection starts with a population of DNA sequences containing a random region flanked by fixed sequences that are required for amplification and transcription. RNA and DNA *in vitro* selections have been done with random regions ranging from 20 to 200 nucleotides and up to 10^{16} starting diversity.[3] Selection occurs when the pool of random sequences is required to perform a function, such as reaction catalysis or ligand binding and the active sequences are separated from the majority of inactive sequences. Once these selected sequences are amplified, a new generation of variants is available for continual rounds of selection leading to an evolved RNA or DNA species that is efficient and specific for the desired function.

In vitro selections for functional RNA that bind a ligand, or RNA aptamer, have been extensive. One early example of *in vitro* selection for the biomolecule ATP was performed in 1993.[4] The first example of an *in vitro* selected catalytic RNA or ribozyme was in 1993 with the discovery of a catalyzed ligation reaction.[5] This reaction ligated an oligonucleotide substrate to the triphosphorylated 5' end of the ribozyme and three separate ribozyme classes were identified with unique conserved secondary structures. The class I ligase was studied extensively and later generations of this class were converted into a polymerase capable of template-directed extension of RNA.[6]

It was speculated that deoxyribozymes would not have as much activity as their ribozyme counterparts due the lack of the 2' hydroxyl. While there are no known naturally occurring DNA enzymes or deoxyribozymes, the first deoxyribozyme to be selected was a Pb^{2+} dependent deoxyribozyme capable of cleaving a RNA phosphodiester bond in 1994.[7] Since then many other deoxyribozymes have been discovered using *in vitro* selection including catalysis of DNA and RNA ligation,[8,9] DNA phosphorylation,[10] adenylation,[11] depurination,[12] and a Diels-Alder reaction.[13]

Although functional DNA has not been found naturally, there are many *in vitro* applications for selected deoxyribozymes. They have the advantage of being chemically stable, easy to synthesize, biocompatible, and rationally engineered and may present a good alternative to their less-stable RNA counterparts or related protein enzymes, which can only perform catalysis under a narrow range of conditions.[14] Deoxyribozymes may be used as possible chemotherapeutics by disrupting pathogenic mRNAs,[15] could be used in the analysis of nucleic acids,[16] have been used to detect metal ions[17] and been shown to work in

molecular-scale computational devices[18] as well as biosensing and chip-based catalysis and nanoreactors.[19-21]

One category of protein enzymes that have been utilized in biosensing are galactosidases which are able to hydrolyze galactose containing substrates. Fluorescein-di(β-D-galactopyranoside) (FDG) (Fig. 3.1b) was synthesized and was used early to detect a single β-galactosidase enzyme in a crude *E. coli* extract within 10 hours in 1961.[22,23] Utilizing the fluorescence ratio of 105 of fluorescein to FDG, Rotman was able to determine the Michaelis constant of the β-galactosidase extract as well as estimate the molecular mass of the enzyme and the fraction that was active. It was reported that FDG had 0.13% of the fluorescence of fluorescein, making it a sensitive and easily detectable substrate for sensing in the presence of β-galactosidase.[22] Since this enzyme and fluorogenic substrate pair was discovered, it along with fluorescein-mono(β-D-galactopyranoside) (FMG), the hydrolyzed intermediate, has been utilized in many applications.

The β-galactosidase is controlled by the inducible *lac* operon[24] and can be utilized in conjunction with FDG to monitor successful expression in real time. FDG and β-galactosidase have utilized for testing stopped-flow kinetics[19] and new methods that improve on the technique as well as chip-based enzymatic monitoring[20] and microreactors.[25] Enzyme immunoassays have utilized the pair to successfully visualize lymphocyte surface receptors by linking β-galactosidase to sheep anti-rabbit immunoglobulin[26] and this could easily be applied to other enzyme-linked immunosorbent assays (ELISA).

These fluorescein-containing galactose derivatives are advantageous because they allow for easy monitoring of galactose hydrolysis due to the large increase in fluorescence signal with loss of the sugar moiety. These substrates are ideal for selection of a novel

deoxyribozyme with galactosidase activity, because any kinetic activity generated by selection can be easily monitored and maximum fluorescence of fluorescein occurs at pH 8.[27,28] Substitution of β-galactosidase for its deoxyribozyme version would likely extend the range of conditions when galactose or its substrate analogs could be applied. Substitution of the protein enzyme with a deoxyribozyme would likely expand the applications of this bioanalytical method.

Here, we aim to select for a galactosidase DNAzyme that could expand the applications of the enzyme-substrate pair. In order for a deoxyribozyme with galactosidase activity to be selected, a single-stranded DNA sequence must be able to specifically bind the FMG or FDG substrate and cleave the fluorescein-galactose bond (Figure 3.1a). One way that this cleavage could occur is through transferring the sugar to the deoxyribozyme, or in the case of FDG, the FMG intermediate product could also be transferred. This outcome would allow for the unique cis-diol of the glycosylated DNA product to be separated away from inactive sequences (which contain no cis-diols) by a boronate-labeled resin, first used as an affinity chromatography technique in 1970.[29] Boronate forms a cyclic diester with several cis-diols such as mannitol, NADH and D-fructose diester with dissociation constants of 3.3, 5.9 and 0.129 mM, respectively.[30-32] Thus a boronate column can be used to retain self-glycosylated sequences and allow for amplification and selection of deoxyribozyme after repeated rounds of selection.
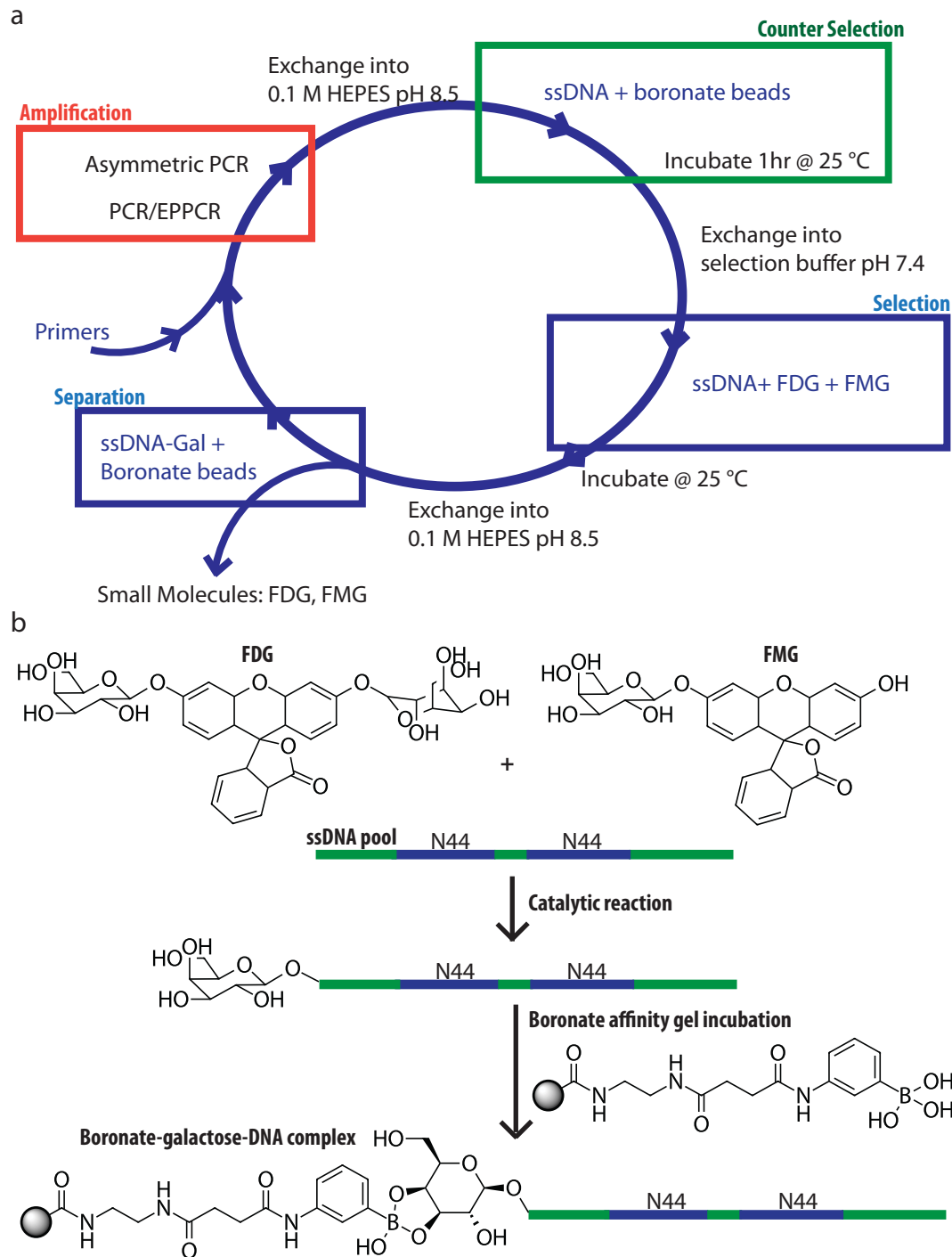
Figure 3.1. *In vitro* selection of self-glycosylating deoxyribozymes. a) Overview of the selection designed to subtract possible Affi-gel aptamers, select for a galactose cleavage reaction, separate active sequences utilizing boronate Affi-gel and amplify sequences for another round of selection. b) A possible reaction catalyzed by a deoxyribozyme would transfer the galactose of fluorescein-mono(β-D-galactopyranoside) (FMG) or fluorescein-di(β-D-galactopyranoside) (FDG) to itself and allow active sequences to be separated by the boronate Affi-gel.

Results and Discussion

3.2 Galactosidase selection

Dr. Cassandra Burke designed the ssDNA pool that was used for this selection. The construct contains two 45-nucleotide random regions flanking a 10-nucleotide tetraloop (TTTC) to favor secondary structure formation of the ssDNA or RNA pool. This 100-nucleotide sequence was flanked by constant primer-binding regions, which totaled a 160-nucleotide pool length.

Before selection for a deoxyribozyme could occur, the boronate Affi-gel binding efficiency needed to be evaluated. The manufacturer (Bio-Rad) claims that the Affi-gel could bind sorbitol with 130 μmol/ml gel. A typical ribonucleotide purification protocol used the Affi-gel with 0.1 M HEPES buffer at pH 8.5 as the application buffer and eluted with 0.1 M NaH_2PO_4, pH 6.0 or 0.05 M NaOAc.[33] Due to this the boronate Affi-gel was equilibrated in 0.1 M HEPES, pH 8.5 buffer and applied samples were exchanged into the same buffer. To evaluate the binding capacity of Affi-gel in HEPES buffer, α-^{32}P-rATP was incubated with pre-equilibrated boronate Affi-gel and washed with repeated amounts of 0.1 M HEPES, pH 8.5 buffer and eluted with 0.1 M KH_2PO_4, pH 5.0 and 0.05 M NaOAc. After four washes with high-pH buffer, approximately 35 % of the total labeled α-^{32}P-rATP was still bound to the gel and elutions with potassium phosphate and sodium acetate were able to remove 60 % and 50 % of bound α-^{32}P-rATP, respectively.

To determine whether the boronate Affi-gel had background affinity for DNA, which contains a deoxyribose but no cis-diols, an oligonucleotide was radioactively labeled on the 5′ terminus with ^{32}P using T4 polynucleotide kinase. The labeled oligomer was incubated with

pre-equilibrated boronate Affi-gel and after washing with 0.1 M HEPES, pH 8.5, approximately 0.4 % total radioactivity remained on the Affi-gel. This was a small amount of the total DNA bound but when considering a DNA pool, background binding could easily be amplified and propagated into a larger percentage of the selected pool after several rounds. Due to this consideration, it was determined that a counter-selection step was required to subtract ssDNA sequences that non-specifically bind to the Affi-gel and prevent selection of an Affi-gel aptamer.

Selection of a deoxyribozyme started with the incubation of 20 μM FMG and FDG with 3.1 nmoles of ssDNA pool, corresponding to a starting diversity of 1.9×10^{15} molecules. Because the initial pool only sampled a fraction of the sequence space, a counter-selection step was not performed until the 2nd round of selection after amplification produced approximately 100 copies of each sequence for the subsequent selection reaction. The 3 initial rounds of selection only required 10 cycles of PCR to regenerate the DNA pool and approximately 100 copies of each library member was used in the subsequent round of selection, indicating that the first 3 rounds of selection were not very stringent or selective and only reduced the diversity approximately 100-fold. Increasing the amount of the boronate Affi-gel in the counter-selection step and reducing it in the selection step introduced more stringency in rounds 4-6. Starting at round 9, the selection reaction time was reduced to promote faster binding or reacting sequences to outcompete their slower counterparts.

In round 12 the selection reactions were split into two parallel selections so that the selectivity towards each substrate would increase. Stringency was increased again by lowering the substrate concentration as well as denaturing and heating the selection reaction

before incubating the selected pool on the boronate Affi-gel in round 13 and was reflected in the amount of amplification that was required to regenerate the pool for the next round of selection. Up until round 13, all steps were performed under native conditions and favored selection for a possible FMG or FDG aptamer because it was assumed that the DNA catalysis being selected for would require some binding affinity for the substrate. Denaturing the selection population would likely eliminate any ssDNA sequences that were only binding, but not reacting with FMG or FDG, and more strongly select for sequences that were transferring the galactose moiety to themselves. The initial selection finished at round 21, with a selection reaction containing 0.5 μM FMG or FDG and a reaction time of 10s before denaturation.

3.3 Fluorescence testing of the selected DNA

While the selection was being performed, fluorescence activity of the selected pool was monitored, trying several different detection methods in search for an optimal assay. To be able to compare results from several methods, control reactions were performed in parallel. To control for variability of the fluorophore over time due to possible hydrolysis or photobleaching, incubation reactions with selection buffer and FMG or FDG were performed. To account for the possible changes in fluorescence due to the presence of the hydrophobic bases of DNA, a random DNA pool was also incubated with FMG or FDG in selection buffer. From reaction to reaction, variability in fluorescence was observed, therefore collected data were processed to be viewed relative to t=0. To establish the change in fluorescence expected from each fluorophore, 20 μM FMG and FDG were incubated with β(1-4)-galactosidase, resulting in a 1.54-fold increase in FMG fluorescence and a 45.6-fold increase in FDG fluorescence after 80 minutes. FMG fluorescence may have saturated the detector,

79

therefore a reaction with 1 μM FMG and β(1-4)-galactosidase was performed resulting in a 2.55-fold increase in fluorescence.

Initially, fluorescence was monitored by taking time points of reactions and controls of FMG or FDG incubated with ssDNA from round 3, 5 and 7 using a NanoDrop™ fluorescence spectrophotometer. The FAM setting of this instrument excites at 490 nm and emission data are collected at 515 nm. While continuous fluorescence over time could not be collected using this instrument, it had the advantage of requiring only 2 μL of sample for each reading thus reducing the amount of fluorophore and DNA that needed to be prepared. Preliminary results with round 7 ssDNA showed a 2.7-fold increase in 20 μM FMG fluorescence after 2 hours versus 1.2-fold increase with the 20 μM FMG fluorescence control. Incubation of the 20 μM FDG fluorescence control and the round 7 ssDNA with 20 μM FDG showed little change in fluorescence. When more time points were collected from incubation reactions with round 3 and 5 ssDNA, no fluorescence increase was observed with the presence of FDG but FMG with R3 ssDNA displayed a 2.2-fold increase versus a 1.6-fold increase with FMG alone after 1 hour. However, earlier time points of the FMG control varied up to 91 % relative to t=0, calling into question the accuracy of pipetting such small volumes for kinetic analysis. Uniform dilution of each aliquot may have alleviated some of the error associated with pipetting such a small volume, but this instrument would be more useful comparing endpoint fluorescence of the reactions instead of time point monitoring.

Looking ahead to the possibility of testing individual clones for fluorescent activity, the Mini Opticon RT-PCR system was tested for fluorophore sensitivity using its FAM channel. The fluorescence of 2-fold serial dilutions of 20 μM FDG and FMG were measured and showed

that the detection limit of unreacted FMG was between 1.25 and 0.625 μM and the detection limit of unreacted FDG was about 5 μM. A baseline fluorescence signal was collected before the fluorophores were added and once FMG or FDG was added, the instrument took up to 6 reads (6 minutes) before accurately reporting the signal. Although this instrument may have been sensitive enough for activity testing, the lag or automatic gain control would not allow for initial reaction kinetics data to be collected and would only be ideal collecting end point data. Until this instrument was replaced with a newer machine that could automatically adjust the gain during a program, this high-throughput method could not be utilized for monitoring initial reaction kinetics that would be required for analyzing active clones.

In order to monitor fluorescence increases continuously over time as well as initial reaction kinetics, settings for an analytical fluorospectrometer were optimized. Since the selection was initially performed using 20 μM FDG and FMG, fluorescence testing was performed at these levels. Fluorescence data was collected every 10 s for 1 hour for each round of selection tested (Figure 3.2a). Even though a master mix was prepared containing the fluorophore and the selection buffer to avoid pipetting discrepancies, raw fluorescence data varied from sample to sample. Due to this result, all fluorescence was viewed relative to an average of the first 3 data points of each sample (t=0). Initially, a change in fluorescence over time was observed compared to the FDG or FMG and random pool DNA controls. Incubation of FMG with DNA from round 11 resulted in a relative fluorescence gain of 1.53 versus 1.16 for the FMG control or 0.89 for the random pool incubated with FMG after 1 hr. Incubation of FDG with DNA from round 11 resulted in a relative fluorescence gain of 1.35 versus 1.21 for the FDG control or 1.1 for the random pool incubated with FDG. Observation

of initial rate kinetics was also attempted with these reactions, but the reactions containing 20 μM FMG were unreliable in the first 10 minutes due to changes in fluorescence because of mixing and/or insolubility problems at such a high concentration in water. FDG fluorescence testing was static over the early rounds of selection but deoxyribozyme activity would not be expected until after the denaturing steps were introduced in later rounds. Also, assuming that the pool was only partially enriched for active sequences and only 1% of the pool was active and able to cleave the fluorescein substrate, incubation with the fluorophore would lead to a 0.026-fold increase in fluorescence using FMG and 0.46-fold increase with FDG based on earlier positive control results. While this change is significant for the FDG fluorescence testing, observing a significant change using 20 μM FMG would be difficult because of the large background fluorescence. By lowering the substrate concentration, a larger relative change might be observed in fluorescence and solubility of FMG would no longer be an issue, especially if the stock was stored in DMSO instead of water. Due to this, the pool from different selection rounds was tested at lower concentrations corresponding to the lowered concentrations used in later rounds of selection, 2 μM and 0.5 μM FDG or FMG.
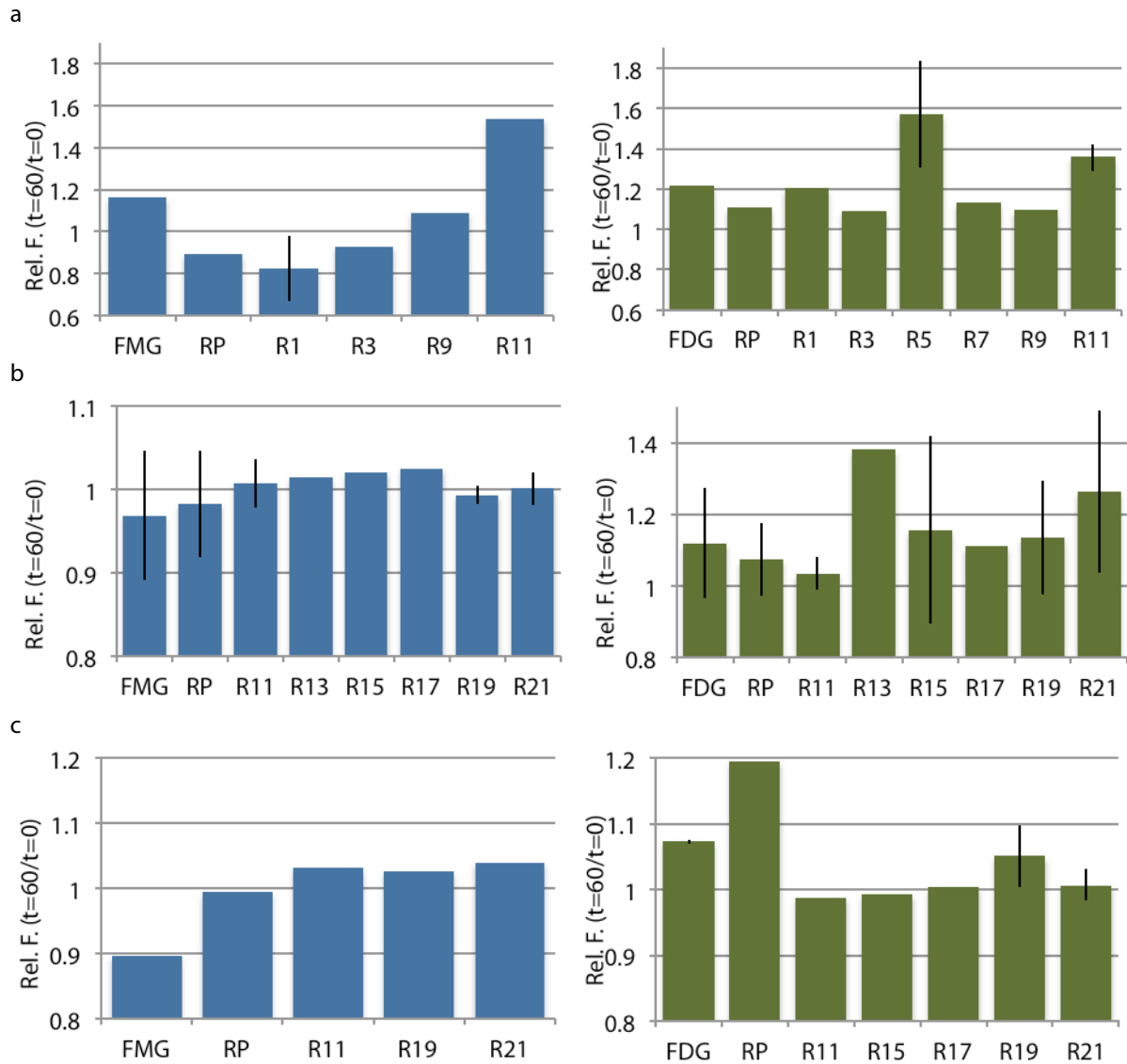
Figure 3.2. Fluorescence monitoring of fluorescein/FMG release by selected DNA pools. Fluorescence after 60 minutes relative to the average of the first 3 time points of fluorescein/FMG release at a) 20 µM FMG (blue) or FDG (green), b) 2 µM FMG (blue) or FDG (green), and c) 0.5 µM FMG (blue) or FDG (green), incubated by itself, with ssDNA from a random pool or ssDNA of selected rounds at room temperature and measured using a spectrofluorometer by exciting at 480 nm and measuring emission at 516 nm for FMG samples and exciting at 490 nm and measuring emission at 512 nm for FDG samples.

Testing with 2 µM FDG and FMG was initiated with controls of FDG and FMG and the random DNA pool. The FMG control varied greatly and in many samples fluorescence decreased over the 60-minute period, resulting in an average relative fluorescence of 0.97 of the FMG control and 0.98 for the random DNA pool control. The selection rounds incubated with FMG showed little change in fluorescence after 60 minutes with round 19 at 0.99 and round 21 at 1.0 relative fluorescence, but the fluorescence did not decrease as much as the controls, indicating that presence of the selected round was either preventing photobleaching of FMG or cleaving the substrate (Figure 3.2b). The FDG substrate control, in general, did not decrease over time with an average relative fluorescence of 1.1, indicating that FDG was not prone to photobleaching. Testing of the ssDNA from the selection rounds resulted in overall increase of the average fluorescence from rounds 19 (1.13) and 21 (1.26) when compared to FDG alone (Figure 3.2b). Measuring fluorescence of the selection rounds with 0.5 µM FMG decreased the background signal and showed increased average relative fluorescence with round 19 (1.026) and 21 (1.039) compared to the average relative fluorescence of FMG (0.90) and FMG incubated with random pool DNA (0.99) controls (Figure 3.2c). Measuring the fluorescence of the selection rounds with 0.5 µM FDG did not maintain the same trend observed when tested with 2 µM FDG, indicating that the FDG selected DNA, 1.05 and 1.01 average relative fluorescence for rounds 19 and 21, respectively, was less active at the lowest concentration used in the selection (Figure 3.2c).
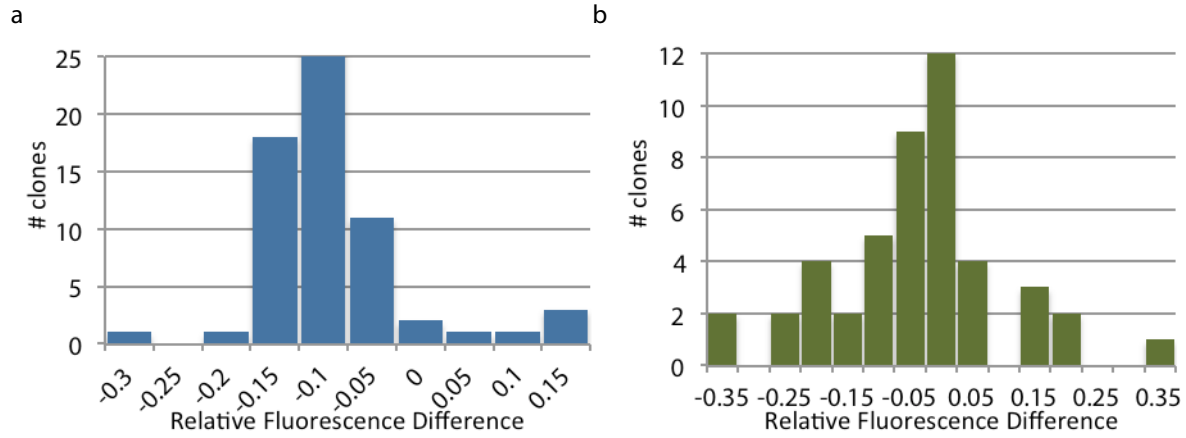
Figure 3.3. Relative fluorescence difference of clones. a) Relative fluorescence difference after 60 minutes of clones incubated with 2 μM FDG; the average relative fluorescence difference of 2 μM FMG was -0.15. b) Relative fluorescence difference after 16 hours of clones incubated with 2 μM FMG; the average relative fluorescence difference of 2 μM FDG was -0.014.

The increase in fluorescence observed in later rounds prompted insertion of the DNA pool into a vector and the individual clones were amplified and tested for fluorogenic activity. Initially, fluorescence of the FDG selected clones was tested using the fluorometer by using a 10 s-interval time course measurement for 1 hour. The fluorescence difference relative to t=0 for 48 clones was plotted (Figure 3.3a). The average fluorescence difference for the FDG control was -0.014, which left 22 clones that showed higher activity after 1 hour. Later, similar testing was performed with the FMG selected clones. However, fluorescent activity was compared after 16 hours (Figure 3.3b) and the average FMG control fluorescence change was -0.154. The change indicated that photobleaching occurred in every sample, however, 43 clones out of the 75 tested showed a higher change in relative fluorescence over 16 hours, indicating presence of the selected DNA clones were preventing photobleaching by binding the fluorophore or cleaving the galactose.

Reagent	1	2	3	4
FDG/FMG	2 μM	0.5 μM	2 μM	2 μM
β-gal	-	4 units/25 μL	4 units/25 μL	-
RP DNA	-	-	-	1.8 μM
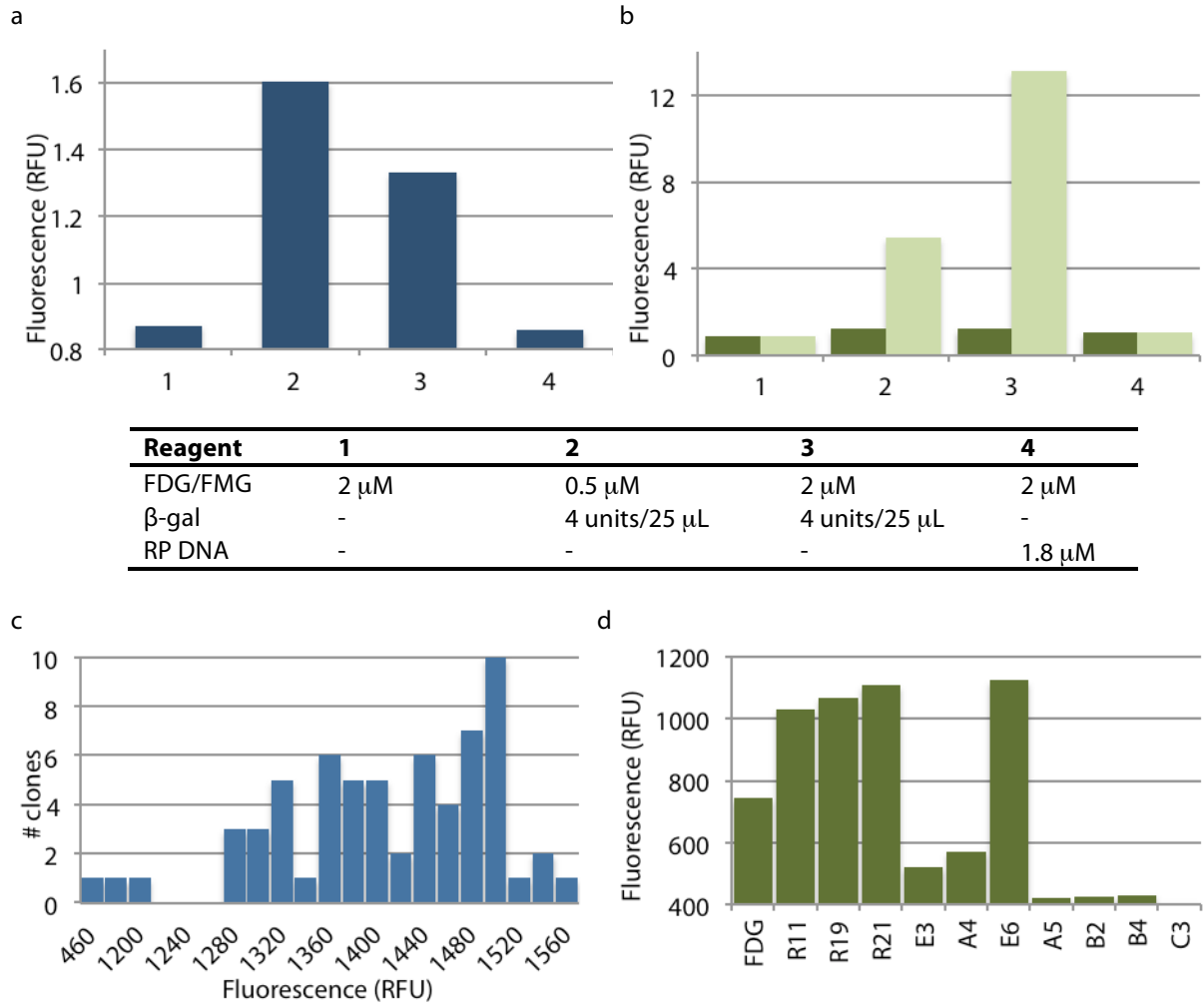
Figure 3.4. Fluorescence control reactions and clone testing using a microplate reader. a) Fluorescence relative to t=0 of 2 and 0.5 μM FMG incubated with β-gal or random DNA pool after 72 hours. b) Fluorescence relative to t=0 of 2 and 0.5 μM FDG incubated with β-gal or random pool DNA after 1 hour and 72 hours. c) Number of clones displaying increased fluorescence incubated with 2 μM FMG after 90 minutes and a 2 μM FMG control with a fluorescence of 460 RFU. d) Fluorescence of clones incubated with 2 μM FDG after 90 minutes using a Biotek Synergy microplate reader by exciting all samples at 480 nm and measuring emission at 528 nm.

After individually testing the FDG selected clones using a fluorometer, a microplate reader was used to test FMG clones in the hope of increasing the rate of screening. Figure 3.4a and 3.4b shows several β-galactosidase control reactions performed on a Biotek Synergy microplate reader using the 460/40 nm excitation and 528/20 nm emission filter set. Relative fluorescence decreased over 3 days for the 2 µM FMG and random DNA pool controls and relative fluorescence increased when 2 µM FMG or 0.5 µM FMG was incubated with of 8 units of β-galactosidase, 1.6- and 1.3-fold, respectively. Relative fluorescence of 2 µM FDG and random DNA pool controls were static over 3 days and increased approximately 5.4- and 13-fold when incubated with 8 units of β-galactosidase and 0.5 and 2 µM FDG, respectively. A higher relative fluorescence increase was observed when 2 µM FDG was reacted with the enzyme, indicating that the lower 0.5 µM FDG was below the optimal working concentration for the β-galactosidase.

Fluorescence data were collected for all FMG selected clones and some previously quantified FDG clones in the order to test reproducibility of results from instrument to instrument. Fluorescence was collected without fluorophore for 5 minutes and then 2 µM FMG or FDG was added and fluorescence was collected every 5 minutes for 90 minutes. The resulting data did not show increasing fluorescence over time, as expected, but increased dramatically when FMG or FDG was added and then stayed at a static level. Figure 3.4c shows the number of FMG selected clones within each 20 RFU interval of fluorescence after 90 minutes with the FMG control fluorescence data plotted at 460 RFU. Fluorescence data collected from the FDG control and selected samples were inconsistent and dependent on which row of the 96-well plate it came from, indicating that the multi-channel pipette was

inaccurate from dispensation to dispensation and that the small amount of background fluorescence of FDG translated to large observed changes while using the microplate reader (Figure 3.4d). No dependence on row location was observed for the fluorescence data from the DNA clones incubated with FMG, but the apparent jump in fluorescent signal followed by little change was suspect and an analytical fluorometer had to be used. Testing clones required minimization of the reaction volume so that the amount of fluorophore and individual DNA amplification reactions could be minimized as well. In a 96-well format, 50 µL was the smallest volume that could be used to avoid inconsistent plate reader data collection.

Unsure of the data collected from the plate reader due to the pipetting inaccuracies and the insensitivity to fluorescent changes over time, clones were tested further using an analytical fluorometer. The apparent fluorescence jump observed with FMG selected clones led to testing the initial rate of the fluorescence reaction with a select group of clones from each selection. Initial rates were measured by adding 1.5x concentrated FMG or FDG in selection buffer to the cuvette and collecting fluorescence data for two minutes followed by addition, mixing of the 3x concentrated ssDNA clone, resulting in a final concentration of 0.5 µM FMG or 2 µM FDG and ~1.8 µM ssDNA, and an additional 8 minutes of fluorescence monitoring. While fluorescence increases were observed over the 10-minute interval with these reactions, monitoring the initial rate by this method was inconsistent. The mixing rate of a DNA clone with FDG or FMG was not fast enough and required opening the instrument mid-measurement to add the DNA, which led to spikes in the data due to ambient light. A stopped-flow system that could monitor fluorescence directly after mixing would have been ideal for reliable initial rate information.
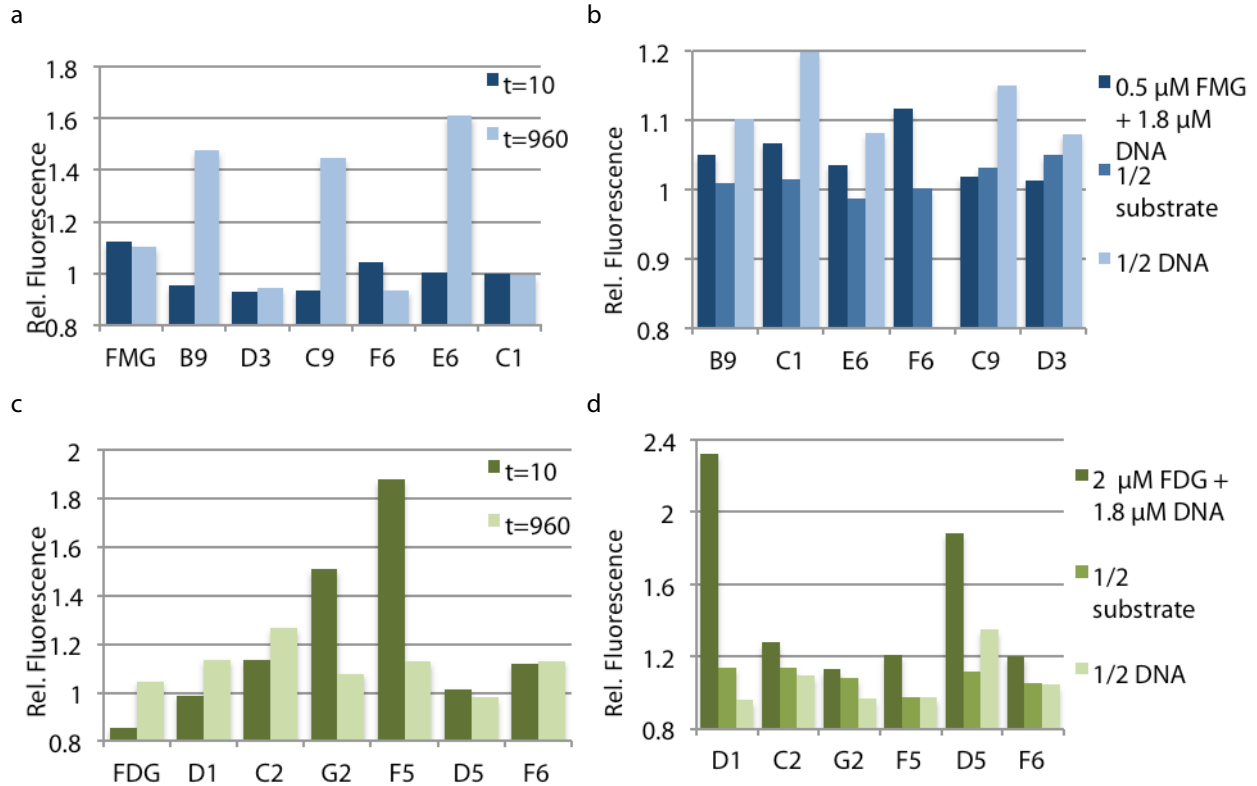
Figure 3.5. Relative fluorescence of individual clones with variable substrate and DNA concentrations. a) Relative fluorescence of test clones incubated with 0.5 μM FMG divided by test clones incubated with 0.25 μM FMG at 10 minutes and 16 hours. b) Relative fluorescence of FMG selected clones after 16 hours at 0.5 μM FMG, 0.25 μM FMG and ~0.9 μM DNA clone. c) Relative fluorescence of test clones incubated with 2 μM FDG divided by test clones incubated with 1 μM FDG at 10 minutes and 16 hours. d) Relative fluorescence of FDG selected clones after 10 minutes at 2 μM FDG, 1 μM FDG and ~0.9 μM DNA clone. All samples were incubated at room temperature and measured using a spectrofluorometer by exciting at 480 nm and measuring emission at 516 nm for FMG samples and exciting at 490 nm and measuring emission at 512 nm for FDG samples.

These experiments may not have resulted in useful initial rate measurements, but they did provide insight into activity of individual clones. Fluorogenic activity was compared for individual clones by halving the amount of fluorophore, 0.25 µM FMG or 1 µM FDG, to determine the dependence of the reaction on substrate concentration (Figure 3.5a, c). Figure 3.5a shows relative fluorescence change of 0.5 µM FMG reactions divided by relative fluorescence of 0.25 µM FMG reactions after 10 minutes and 16 hours. By dividing data of FMG relative fluorescence at 0.5 µM by 0.25 µM, the control increased to 1.1 and was static over 16 hours. Relative fluorescence data at 10 minutes for all clones decreased with more substrate but after 16 hours, clones B9, C9 and E6 showed higher fluorescence activity, over 1.45-fold more, relative to 0.25 µM FMG reactions, indicative of substrate concentration dependence. Clones B9, C1 and E6 exhibit a decrease in relative fluorescence after 16 hours when half the substrate was used, however, no decrease in relative fluorescence was observed when half the DNA was used in the incubation reactions. If the DNA clone was catalyzing a reaction that would lead to fluorescence of the cleaved substrate, a lower DNA concentration would be expected to lead to lower fluorescence activity.

The relative fluorescence data from incubation of 2 µM FDG with each DNA clone was divided by relative fluorescence of the 1 µM FDG reactions after 10 minutes and 16 hours (Figure 3.5c). By dividing data of FDG relative fluorescence at 2 µM by 1 µM, the control decreased to 0.85 and was increased to 1.05 after 16 hours. Relative fluorescence data at 10 minutes for clones C2, G2, F5 and F6 increased with more substrate but after 16 hours, only clone C2 showed higher and increasing fluorescence activity relative to the 1 µM FDG reaction, indicative of substrate concentration dependence. Fluorescence activity of clones

G2 and F5 was promising based on data collected at 10 minutes, a 1.5- and 1.9-fold increase, respectively, however, the decrease of relative fluorescence after 16 hours was troubling and could not be reproduced. All the FDG selected clone reactions exhibited a relative fluorescence decrease with half the substrate and half the DNA after 10 minutes, indicating that fluorescent activity was dependent on both substrate and DNA concentration and that these clones may be involved in the desired hydrolysis reaction. The most fluorescently active clones from both selections were sequenced but the alignment of the resulting sequences did not lead to any insight about activity and there was little similarity observed.

3.4 Boronate binding with clones and the DNA pool

To investigate the most fluorogenic sequences, a boronate Affi-gel binding experiment was designed, mimicking the selection assay (Figure 3.6a). The selection was dependent on separation of active sequences through the galactose-boronate complex; therefore binding and competitively eluting the active sequences should verify that the galactose was indeed transferred during the selection reaction. Previously, no elution step had been performed and boronate-bound sequences were amplified directly from the Affi-gel, which assumed that the counter selection would efficiently subtract sequences prone to bind the boronate Affi-gel. DNA from round 11 was utilized in a reaction with 20 µM FMG or FDG, bound to the boronate Affi-gel, and the flow-through, washes and elution were collected and the fractions were analyzed by real-time PCR. Competitive elutions with 0.1 M ribose, in 0.1 M HEPES, pH 8.5 (E1, 2, 3 of Figure 3.6b and E4 of 3.6c) were more effective than lowering the pH with elutions of 0.05 M NaOAc, pH 5.5 (E4 of Figure 3.6b and E1, 2, 3 of 3.6c), so this was used in subsequent elutions steps after clone and pool binding to the boronate Affi-gel.
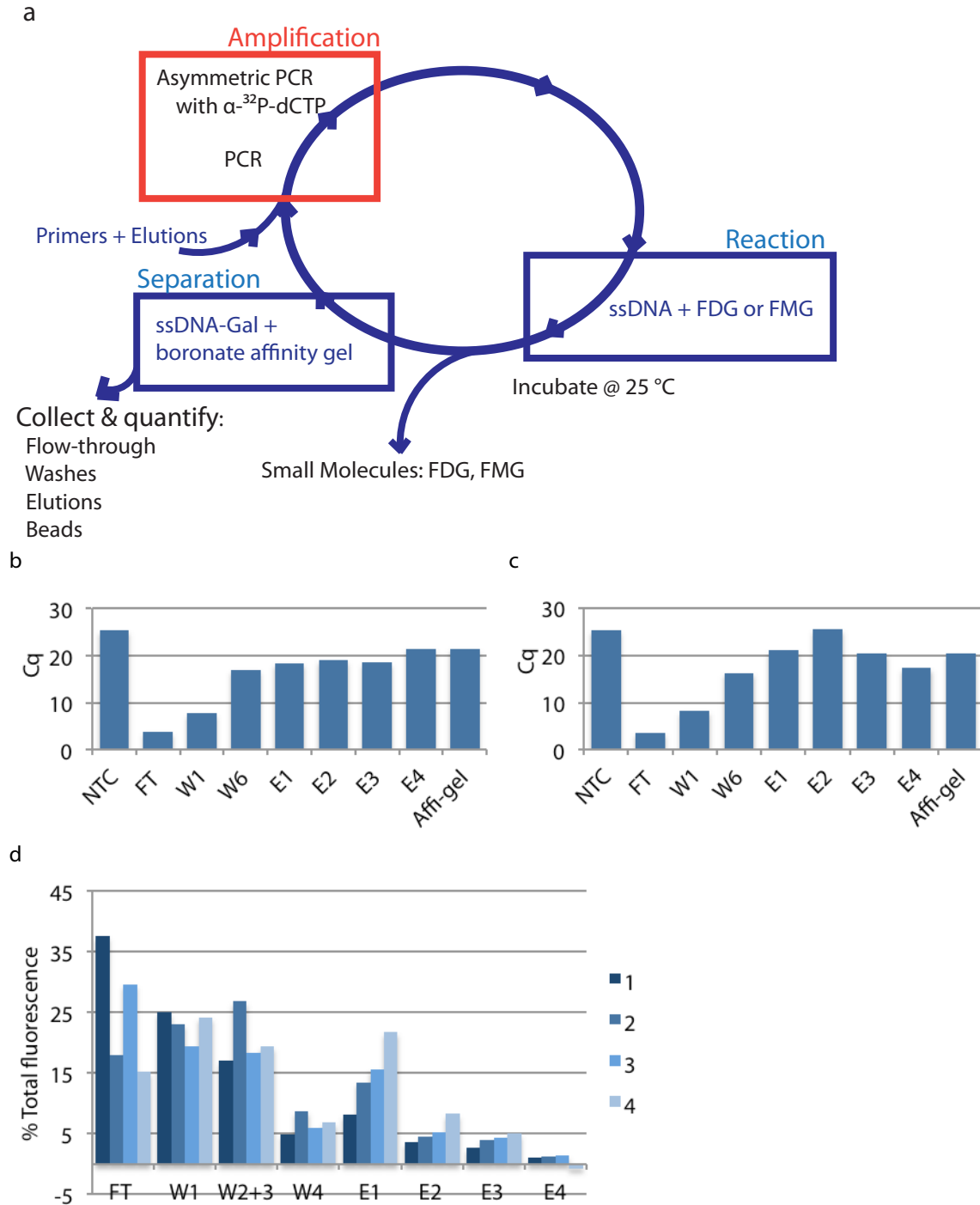
Figure 3.6. Boronate Affi-gel testing with selected DNA and FMG. a) Overview of a boronate binding experiment with preparation of ssDNA, reaction with FDG or FMG, incubation with boronate Affi-gel and collection quantification of the flow-through, washes, elutions and Affi-gel. R11 DNA was incubated on boronate Affi-gel, washed, and eluted with b) 3x 0.1 M ribose in 0.1 M HEPES, pH 8.5 and 1x 0.05 M NaOAc, pH 5.5 and c) 3x 0.05 M NaOAc, pH 5.5 and 1x 0.1 M ribose in 0.1 M HEPES, pH 8.5. d) Optimization of boronate Affi-gel binding with 20 μM FMG using 1) 7.5 mg boronate Affi-gel and 0.1 M HEPES, pH 8.5, 2) 15 mg boronate Affi-gel and 0.1 M HEPES, pH 8.5, 3) 7.5 mg boronate Affi-ge and 0.4 M HEPES, pH 8.5, 4) 15 mg boronate Affi-gel and 0.4 M HEPES, pH 8.5. Fluorescence was quantified using the FAM channel of a real-time PCR detection system.

A fluorogenically active clone from the FMG selection (B9) and the FDG selection (F4) were asymmetrically amplified and radiolabeled with α-^{32}P-dCTP, reacted with 2 μM FDG or FMG, and incubated with the boronate Affi-gel. Flow-through, washes, elutions and the Affi-gel were collected, their radioactivity measured (cpm), and the fraction of total counts was plotted (Figure 3.7a). Unfortunately, the negative controls (B9- and F4-), which did not contain the fluorophore during incubation, were not appreciably lower compared to the reactions containing FDG or FMG. All elution samples did not display binding above 0.5 % of the total counts which was similar to the results observed when a radioactive oligomer was tested with the boronate Affi-gel. Another indication that the tested clones did not contain a transferred galactose was the fact that the percentage of total counts decreased from the last wash to the first elution, which would not be expected if the bound DNA was being competitively eluted by the ribose.
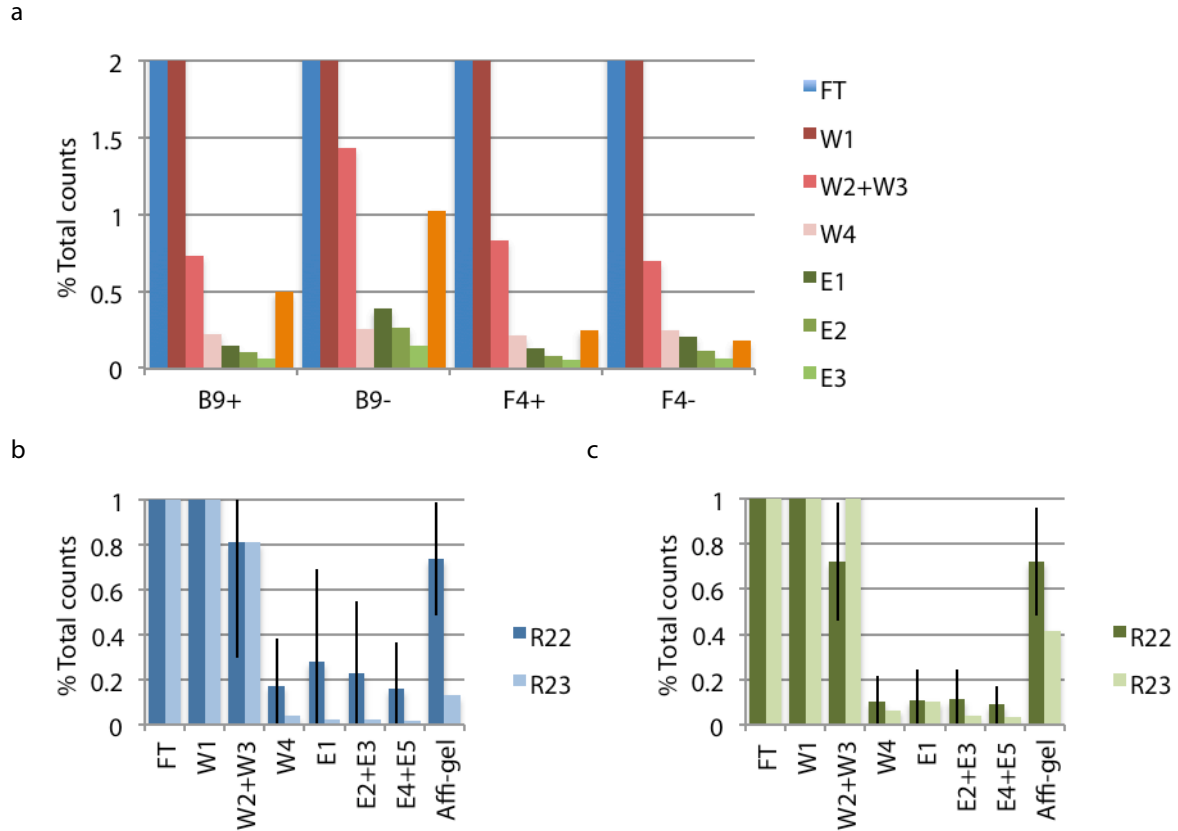
Figure 3.7. **Selected clones and DNA pool boronate Affi-gel binding and elution experiment.** a) Two fluorescently active clones B9 of FMG selection and F4 of the FDG selection were body labeled with ^{32}P and reacted with (B9+, F4+) and without (B9-, F4-) 2 μM FMG/FDG at 25 °C for 1 minute and bound to boronate Affi-gel and washed with 4x 0.1 M HEPES, pH 8.5, eluted with 4x 0.1 M ribose in 0.1 M HEPES, pH 8.5 and the resulting samples were quantified by scintillation counter. An analogous boronate Affi-gel binding and elution procedure was performed with ssDNA from round 21 DNA and reacted with b) 2 μM FDG, c) 2 μM FMG and the R22 elutions were concentrated and amplified for repeated binding and elution, resulting in R23.

Since no specific binding or elutions were observed with individual clones, it was thought the selected pool, comprised of many fluorogenically active sequences, would also exhibit galactose-boronate binding and so similar testing was performed with DNA from round 21 of the FDG and FMG selections (Figure 3.7b, c). After initially binding the round 21 DNA to the boronate Affi-gel, it was determined that combining the elution fractions, amplifying and re-incubating them with the fluorophore might enrich the pool for sequences that specifically formed the galactose-boronate complex. While the first elution increased relative to the last wash in round 22, elution fractions above 1 % were not observed in either the FMG or FDG selection and the elution percentages did not increase in the next round. In addition to observing no enrichment of the elution fractions, almost 0.8% of the radioactively labeled DNA was bound strongly to the boronate Affi-gel, which meant that the counter-selection step of the selection was not completely effective.

3.5 Reselection of the DNA pool

Due to the problems with binding and eluting DNA from the boronate Affi-gel, the selection was redesigned (Figure 3.8a) to include elutions of the radiolabeled, boronate-bound DNA with 0.1 M ribose in 0.1 M HEPES, pH 8.5 and the resulting flow-through, wash and elution samples were quantified by radioactivity (CPM). This new design would leave behind DNA sequences that were prone to bind the boronate Affi-gel that were not subtracted by the counter-selection step, where before amplification of active sequences occurred directly from the boronate Affi-gel. After performing new rounds 8, 9 and 10 with the modified selection method, amplification of the resulting round 10 elutions did not regenerate DNA for the next round and the elution percentage did not increase above 0.4 %, even after several attempts. Previous to the selection redesign, DNA from the previous round

was asymmetrically amplified and used directly in the counter-selection of the next round, which meant that approximately 20% of the DNA was double stranded. It was assumed that this portion was inert and would be washed away in the boronate-binding step, however, to accurately quantify radiolabeled ssDNA, it first needed to be purified away from the DNA duplex. This was performed, initially, using a kit that could purify the ssDNA away from the DNA duplex quickly using spin columns, but with about 50 % lost. When the combined elution fractions were not successfully amplified after round 10, the radiolabeled DNA from the previous round was purified by using a partially denaturing PAGE, excising the area corresponding to the ssDNA and eluting the samples in 0.3 M KCl. While this procedure took longer, an increased fraction bound and eluted from the boronate and the was successfully amplified afterward, revealing purification of the ssDNA pool using a PAGE was more effective than the spin columns.

At this point, the binding efficiency of the boronate Affi-gel was revisited by incubating the pre-equilibrated gel with 20 μM FMG. This incubation was performed under varied conditions (Figure 3.6d), including equilibration and washes with a more concentrated buffer, 0.4 M HEPES, pH 8.5, and increased Affi-gel, 15 mg, and both conditions combined. Using 0.4 M HEPES, pH 8.5 and double the boronate Affi-gel led to about 32 % FMG being retained on the beads and eluted for quantification, 10 % more than just increasing ionic strength. A large percentage did not bind the beads, indicating that a fraction of the FMG was hydrolyzed and unable to bind the boronate Affi-gel or that the binding capacity of the gel was lower than reported by the manufacturer. If only 32 % of a positive control was able to bind and elute

from the boronate Affi-gel under optimal conditions, successful binding and elution of a small fraction of active sequences from a pool would be less likely.
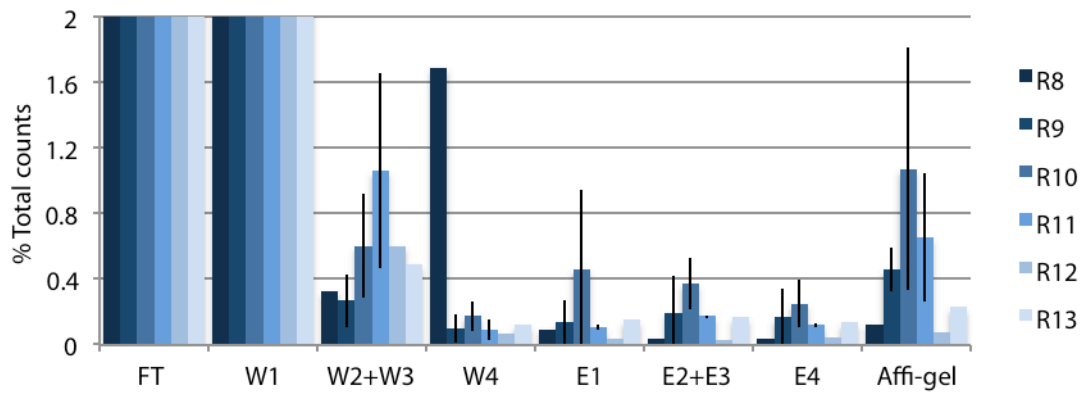
Selection rounds were continued, however, until round 13 (Figure 3.8b, c) and selection of round 22 and 23 (Figure 3.8d, e) were repeated using 0.4 M HEPES buffer for equilibration of the boronate Affi-gel and washes. Elutions from these rounds were not enriched for boronate binding, demonstrating that sequences that were amplified in each round did not contain the catalytic activity that was desired. One exception seemed to be round 13 from the FDG selection, however, the total counts from the radioactively labeled DNA was initially low and the counts from the flow-through sample was suspiciously low, 73 % compared to 85 % from the FMG selection, which skewed the rest of data. In addition to this, DNA from rounds 8 through 13 were not subjected to any denaturing steps because enrichment of the elution fractions of the pool were desired before increased stringency was applied to the selection, so, at best, a galactose or fluorescein aptamer may have been amplified. Reselection of round 22 and 23 was performed in the hopes that active sequences that previously showed fluorescent activity could be further selected but no elution enrichment was observed with less than 0.1% of the total counts being eluted from the boronate Affi-gel, so this outcome is unlikely. Since the pool in the selection and reselection seemed to have little affinity for the boronate Affi-gel, selection for a galactosidase was not successfully carried out.

a

Amplification
Asymmetric PCR with α-³²P-dCTP
PCR

Exchange into 0.1 M HEPES pH 8.5

Counter Selection
ssDNA + boronate affinity gel
Incubate 1hr @ 25 °C

Exchange into selection buffer pH 7.4

Selection
ssDNA + FDG + FMG

Primers + Elutions

Separation
ssDNA-Gal + boronate affinity gel

Incubate @ 25 °C

Collect & quantify:
Flow-through
Washes
Elutions
Beads

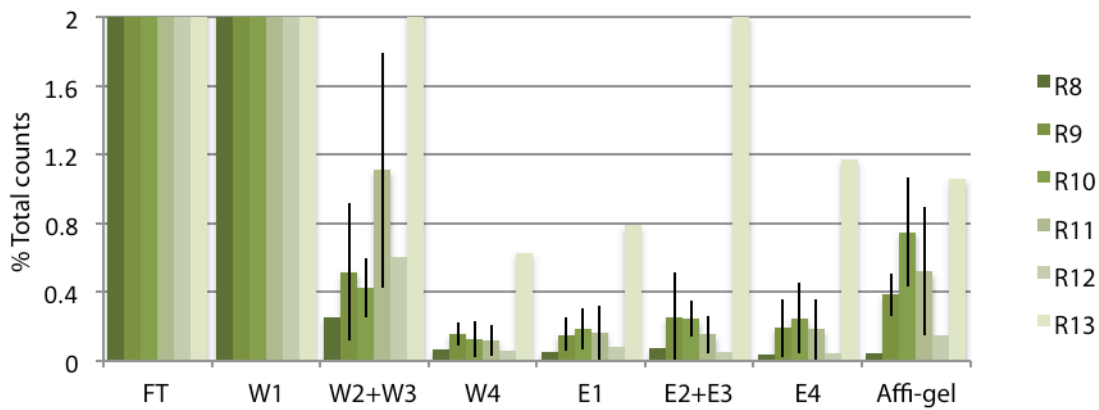Exchange into 0.1 M HEPES pH 8.5
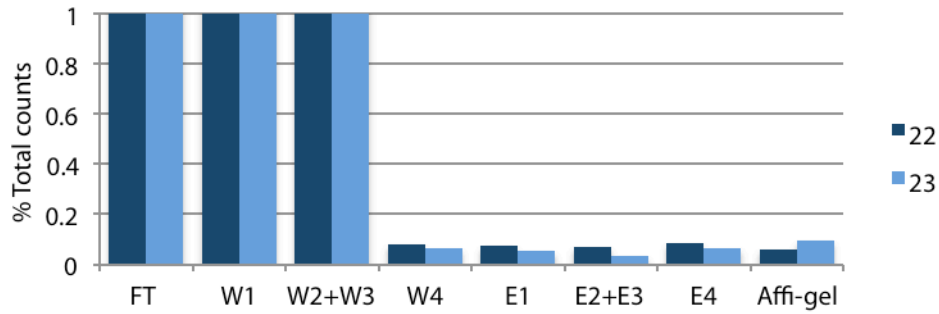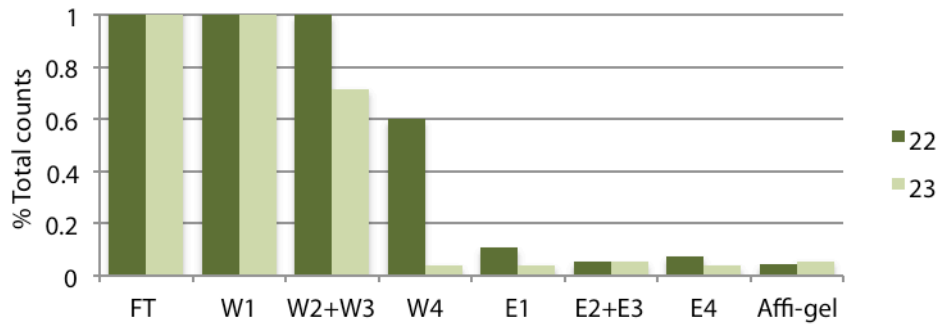
Small Molecules: FDG, FMG

b

c

d

e

Figure 3.8. Reselection overview and results. a) Reselection overview, which includes asymmetric amplification of the previous round with α-^{32}P-dCTP, counter selection with the boronate Affi-gel, a 2 hr selection reaction with 2 µM FMG/FDG, separation of active sequences by incubation with boronate Affi-gel and flow-though, wash, elution and Affi-gel samples quantified by scintillation counter, resulting in the combined elutions being re-amplified for another round of selection. The percent of total radioactivity counts was plotted for all flow-through, wash, elution and Affi-gel samples from the rounds of re-selection performed with averages reported for rounds that had to be repeated using ssDNA from the previous round and b) 2 µM FDG or c) 2 µM FMG by washing the boronate Affi-gel with 4x 0.1 M HEPES, pH 8.5 and eluting with 4x 0.1 M ribose in 0.1 M HEPES, pH 8.5. Continued selections with ssDNA of R21 were performed with d) 2 µM FMG or e) 2 µM FDG by washing the boronate Affi-gel with 4x 0.4 M HEPES, pH 8.5 and eluting with 4x 0.1 M ribose in 0.1 M HEPES, pH 8.5.

3.6 Conclusions

Identification of an active deoxyribozyme could lead to expanded biosensing applications of the galactosidase and fluorescent substrate pair, and since DNA is easily and cheaply chemically synthesized and can exhibit activity outside of physiological conditions, it would be useful for nonbiological applications such as chip-based catalysis and nanoreactors. Progress towards the selection of a deoxyribozyme with galactosidase activity was accomplished. Initially, 21 rounds of selection were performed with increasing stringency by lowering substrate concentrations, lowering reaction times as well as denaturing the reaction before the separation and amplification of active sequences.

Several instruments were used to measure fluorescence of the FMG or FDG with the selected pool and clones and all would be useful to monitor fluorescence if only interested in endpoint data. However, if initial rate kinetics data was required, a different method such as a stopped-flow method should be employed since mixing the fluorophore and DNA was not instant. Cuvettes with tubing to allow for stopped-flow fluorescence monitoring exist and would allow for initial rate kinetics data to be collected in the absence of ambient light. Increased relative fluorescence was observed with DNA from round 19 and 21 of the selection when incubated with 2 µM FDG or FMG after 60 minutes when compared to their corresponding controls. Photobleaching was observed with the FMG substrate and in the future an antioxidant such as ascorbic acid should be added for more reliable fluorescent data.

After cloning DNA from round 21 of the FMG and FDG selections, 22 clones from the FDG selection and 40 clones from the FMG selection exhibited higher relative fluorescence than their corresponding controls when incubated with 2 µM substrate after 60 minutes for FDG or

16 hours for FMG. Possible substrate concentration dependence was observed for four clones from the FMG selection, but dependence on DNA concentration was observed. Six clones from the FDG selection seemed to exhibit substrate concentration dependence and as well as DNA concentration dependence. Fluorescently active clones and the selected DNA pool that were bound to boronate Affi-gel did not exhibit any selective binding or elution that would indicate covalent attachment of the galactose from the substrate to the possible deoxyribozyme. Due to this outcome, the selection was redesigned to include competitive elution of boronate-bound DNA sequences but this did not lead to a selected pool that was enriched for galactosidase activity.

Since the boronate-galactose complex is the most direct way to separate out DNA sequences that have successfully transferred the galactose to itself, further optimization of the system may be worth the time. Increased binding may be observed with inclusion of Mg^{2+} in the boronate-binding buffer. Modification of the selection reaction by using higher concentrations of Mg^{2+} or addition of other metals, such as Pb^{2+} or Mn^{2+}, which have been included in other *in vitro* selections for deoxyribozymes, may allow for easier identification of active sequences. The boronate Affi-gel may exhibit too much background binding to selectively isolate the boronate-galactose-DNA complex and other methods of separation should be considered. Utilization of the galactose moiety for recognition and binding by an immobilized protein such as GAL3 from *S. cerevisiae*[34] or galectin,[35] which specifically binds galactose, could be another strategy to efficiently separate active DNA sequences that have acquired the galactose from the fluorescent substrate. Introduction of a denaturing step earlier in the selection may also help select for the desired catalytic reaction instead of just an aptamer. Development of a reliable high-throughput format fluorescence assay would allow

screening of clones for fluorogenic activity as well as fluorescence dependence on substrate or DNA concentration and would assist in identifying a deoxyribozyme with galactosidase activity. Increasing the efficiency of the separation step of the *in vitro* selection and better characterization of the resulting clones would allow isolation of a deoxyribozyme that would be a key component of a more versatile biosensing pair that could operate under physiological and non-physiological conditions.

Experimental

Pool design. A pool was ordered from Keck Biotechnology Resource Laboratory with two 45 nucleotide random regions flanking a 10-nucleotide tetraloop (TTTC). This 100-nucleotide sequence was flanked with constant regions for primer binding sites and a reverse complement T7 promoter for RNA selections. Forward and reverse primers were ordered from Integrated DNA Technologies, shown in Table 3.1.

Table 3.1. Pool and oligonucleotides.

Oligo	Sequence	L
AL228	CTGAGCTTGACGCATTG	17
AL1954	GATCTGTAATACGACTCACTATAGGG	26
AL1859	CTGAGCTTGACGCATTG N45 GCGTTTCCGC N45 GTAGTGAGGCACGTCTGCCC TATAGTGAGTCGTATTACAGATC	160
AL1296	GTAAAACGACGGCCAGT	17
AL1297	CAGGAAACAGCTATGACC	18

Boronate Affi-gel binding efficiency. A spin column of 15 mg Affi-gel boronate affinity gel (Bio-Rad) was equilibrated with 500 μL 1 M HEPES, pH 8.5 followed by 500 μL 0.1 M HEPES, pH 8.5. The Affi-gel was incubated with ~16 nM ^{32}P-α-ATP in 0.1 M HEPES, pH 8.5 at room temperature for 20 minutes. The Affi-gel column was spun for 1 minute at 4000 x g and washed four times with 0.1 M HEPES, pH 8.5 and eluted twice with 600 μL 0.1 M KH$_2$PO$_4$, pH 5 and eluted once with 600 μL 3 M NaOAc, pH 5.4. The radioactivity counts of each flow-through sample and the Affi-gel after each step was quantified using a Geiger counter.

Galactosidase selection. In round 1 of the selection, 3.1 nmoles (final [5 μM]) of AL1859 was incubated with 20 μM FDG and 20 μM FMG in selection buffer (140 mM KCl, 10 mM NaCl, 10 mM MES, 1mM MgCl$_2$, pH 7.5) for 2 hours at room temperature. The selection reaction was

exchanged into 0.1 M HEPES, pH 8.5 using a spin column of Sephadex® G-25. At the same time, a spin column of 37.7 mg boronate Affi-gel that was pre-equilibrated by washing with 2x 500 μL 1 M HEPES, pH 8.5, followed by 2x 500 μL 0.1 M HEPES, pH 8.5 and centrifuging after each wash at 4000 x g for 1 minute. The flow-though from the Sephadex® G-25 column was added to the boronate Affi-gel column and was shaken on a vortex for 20 minutes. The column was centrifuged at 4000 x g for 2 minutes and washed four times with 0.1 M HEPES, pH 8.5. The boronate Affi-gel was resuspended in the minimum amount of water, ~225 μL.

Each 50 μL PCR amplification reaction contained a final concentration 200 μM dNTPs, 2 μM primers (AL228 and AL 1954) and 1 μL PFU in a buffer of 10 mM Tris•HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3. The rest of volume, 37.07 μL, was made up of boronate Affi-gel resuspended in water. The PCR conditions denatured the reaction at 95 °C for 2 minutes, followed by 8 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds, and completed with a final elongation at 72 °C for 2 minutes. 5 μL of the amplified Affi-gel reaction was added to a new 50 μL PCR amplification reaction with the same conditions and amplified for an additional 22 cycles. 5 μL aliquots were taken every 2 cycles and added to 2x loading buffer containing 8 M urea, xylene cyanol and bromophenol blue. Samples were loaded onto a 2% agarose gel containing 0.002% ethidium bromide in 0.5x TBE buffer (0.5x Tris, boric acid, EDTA, pH 8.4) and run at 200 V for 15 minutes. The resulting gel was analyzed under UV light.

Once the optimal cycles of PCR were determined and carried out, an asymmetric PCR reaction was performed. Five microliters (final ~[0.2 nM]) of the previous PCR reaction was added to each 50 μL asymmetric PCR amplification reaction containing a final concentration

200 μM dNTPs, 2 μM forward primer (AL228), 0.4 μM reverse primer (AL 1954) and 1 μL PFU in a buffer of 10 mM Tris, 50 mM KCl, 1.5 mM MgCl$_2$, pH 8.3. The asymmetric PCR conditions denatured the reaction at 95 °C for 2 minutes, followed by 10 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds, and completed with a final elongation at 72 °C for 2 minutes. The ssDNA generated was concentrated ~5-fold and desalted using an Amicon® centrifugal filter (Millipore) with a 10 kDa cut-off and used in the next round of selection.

All rounds following the first round of selection were subjected to a counter-selection step where DNA pool was exchanged into 0.1 M HEPES, pH 8.5 buffer using an equilibrated Sephadex G-25 spin column. The exchanged DNA solution was then incubated on pre-equilibrated boronate Affi-gel (Bio-Rad) for 1 hour before continuing with another round of selection, as described above. A total of 21 rounds of selection were performed similarly and Table 3.2 summarizes the approximate amount of DNA used, amount of boronate Affi-gel used in counter-selection (CS), selection (S), selection reaction time and the PCR cycles required to regenerate the selected DNA for each round of selection.

The selection was split into two parallel selections after round 11, which selected for catalysis with FMG and FDG individually. The incubated selection reactions, starting at round 13, were diluted 3-fold with 8 M urea and heated at 95 °C for 3 minutes. Additionally, rounds 11, 14, 16, 19 and 21 were purified on a 2% agarose gel by excising the band corresponding to the dsDNA pool and eluting it with 0.3 M KCl and concentrating the eluent with a Amicon® centrifugal filter (Millipore) with a 50 kDa cut-off. The resulting DNA was re-amplified using the conditions above and then asymmetrically amplified.

Table 3.2. Summary of the galactosidase selection.

R	DNA (pmoles)	CS (mg)	Reaction time (min)	[FMG/FDG]	S (mg)	PCR cycles
1	3100	0	120	20	38	10
2	1300	30	120	20	15	10
3	1200	66	120	20	20	10
4	40	41	120	20	14	16
5	40	70	120	20	15	14
6	100	68	120	20	11	16
7	40	25	120	20	10	16
8	40	25	120	20	8	20
9	40	28	40	20	8	20
10	40	30	10	20	8	14
11	40	30	1	20	8	14
12	40	28	1	20	12	0
12	40	24	1	20	8/8	14D/16M
13	40	20	1	2	8/7	14D/M
14	40	22/25	1	2	8/8	28D/M
15	40	19/19	1	2	5/5	20 D/M
16	40	22/22	1	2	7/7	20 D/M
17	40	20/20	1	0.5	6/6	16 D/M
18	40	16/16	1	0.5	6/6	16 D/M
19	40	16/16	1	0.5	6/6	14 D/M
20	40	12/12	10 s	0.5	5/5	18D/20M
21	40	12/12	10 s	0.5	9/9	16D/18M

Pool fluorescence testing. Several methods to test fluorescence activity of the selected pool were attempted. Generally, the ssDNA produced by the asymmetric PCR reaction from several different rounds were exchanged into selection buffer using a spin column of Sephadex® G-25 equilibrated in selection buffer. The resulting ssDNA solution (final ~[1.8μM]) was added to a reaction of FMG or FDG, ranging from a final concentration of 20 μM to 0.5 μM, and selection buffer and incubated at room temperature for up to 2 hours.

To test fluorescence of the selected DNA using a 3300 NanoDrop fluorospectrometer (Thermo Scientific), aliquots were collected and frozen at t=0, 1hr and 2hr. The incubated

reactions contained 20 μM FDG/FMG and ssDNA from round 7, round 1, or no DNA. Using the fluorescein setting, emission spectra were collected for each sample and maximum emission (515 nm) was plotted.

To test fluorescence of the selected DNA using a Mini Opticon RT-PCR system (Bio-Rad), dilutions of FMG and FDG in selection buffer were tested first (20, 10, 5, 2.5, 1.25, 0.63, 0.31, and 0.16 μM) using the FAM channel. The fluorophores were incubated at 25 °C and the emission signal was collected every minute for 1 hour. In parallel reactions, a final concentration of 20 μM FMG/FDG was added to selection buffer and the resulting signal was collected every 1 minute for 5 minutes, followed by addition of selection buffer-exchanged ssDNA of round 3 (final ~[1.8 μM]) or selection buffer. The reactions were incubated at 25 °C and the fluorescent signal was collected every minute for 1 hour.

To test fluorescence of the selected DNA using a fluorescence spectrometer (Jasco FP-6300), settings for each fluorophore were optimized. The optimal signal for 2 μM FMG was obtained using an excitation bandwidth of 10 nm, an emission bandwidth of 5 nm, an excitation wavelength of 480 nm and emission wavelength of 516 nm and a PMT voltage of 400 V. The optimal signal for 2 μM FDG was obtained using an excitation bandwidth of 2.5 nm, an emission bandwidth of 5 nm, an excitation wavelength of 490 nm and emission wavelength of 512 nm and a PMT voltage of 1000 V. Varied FMG/FDG concentrations (20, 2, and 0.5 μM) were mixed with selection buffer and added to a quartz cuvette. Selection buffer-exchanged ssDNA was added (final ~[1.8 μM]) and the fluorescence emission was collected every 10s for 1 hour.

Cloning the pool. Purified dsDNA from round 21 was PCR amplified and 2 μL of the reaction was added to a TOPO® cloning reaction (Invitrogen) containing 0.5 μL TOPO® vector and 0.5 μL salt solution (1.2 M NaCl, 60 mM $MgCl_2$) and incubated for 5 minutes at room temperature. The 3 μL reaction was added to 50 μL of One Shot® Top 10 Chemically Competent *E. coli* cells and incubated on ice for 5 minutes. The cells were heat-shocked for 45 seconds at 42 °C and put on ice for 2 minutes. Four-hundred microliters of S.O.C. medium (Invitrogen) was added and the tube was rotated horizontally at 37 °C for 50 minutes. 100 μL of the resulting transformations were spread on pre-warmed LB agar plates with ampicillin, IPTG, and X-Gal and incubated at 37 °C for 16 hours. Successfully transformed colonies were diluted into a 50 μL solution of 50% glycerol and PBS buffer.

Individual colonies were amplified by adding 1 μL of the diluted clone to a reaction containing a final concentration of 200 μM dNTPs, 2 μM primers (AL228 and AL 1954) and 0.25 μL PFU in a buffer of 10 mM Tris•HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3. The PCR conditions denatured the reaction at 95 °C for 2 minutes, followed by 20 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds, and completed with a final elongation at 72 °C for 2 minutes. Each clone was asymmetrically amplified as described above in the galactosidase selection and exchanged into selection buffer for analysis.

Clone fluorescence testing. Using the fluorescence spectrometer (Jasco FP-6300), 48 clones from the FDG selection were tested. Forty-five microliters of the selection buffer exchanged DNA was added to 5 μL of 20 μM FDG in selection buffer and mixed. Fluorescence data was collected using the settings described above, every 10s for 1 hour. Similar tests on the 75 clones from the FMG selection were performed collecting data for 10 minutes or a final

concentration of 0.5 μM FMG. A fluorescence reading for each clone was performed approximately 16 hours after initiation of the reaction.

Asymmetrically amplified DNA of some of the most active clones were concentrated ~6-fold using an Amicon® centrifugal filter (Millipore) with a 30 kDa cut-off. 25 μL FDG or FMG (6 μM or 1.5 μM) in 25 μL selection buffer was added to a quartz cuvette and fluorescence data was collected every 10 s for 2 minutes followed by the addition of 25 μL concentrated DNA and fluorescence data collection for another 8 minutes.

Time course data was also collected on a Synergy HT microplate reader (Bio-Tek) by using a 485 nm excitation filter with a 20 nm bandwidth and a 528 nm emission filter with a 20 nm bandwidth. Forty-five microliters of the exchanged DNA clones from both FDG and FMG selections were added to individual wells of a black 96-well plate and 5 μL of 20 μM FDG or FMG was added and fluorescence readings were collected every 5 minutes for 90 minutes.

Sequencing active clones. Active clones were PCR amplified from the vector as described in cloning the pool with a final concentration of 2 μM primers (AL1296 and AL1297). The amplified 364 bp product was purified by excising the band from a 2% agarose gel and adding 300 μL solubilization buffer (Buffer QG, Qiagen) and heating the sample at 50 °C for 10 minutes. One-hundred microliters of isopropanol was added to the solution and incubated on a QIAquick spin column for 1 minute and spun at 16,000 x g for 1 minute. The spin column was washed and spun with 500 μL buffer QG and 750 μL PE and the flow-through buffer was discarded. The column was spun at 16000 x g to dryness and incubated with 50 μL H_2O for 2 minutes and spun at 16000 x g. 10 μL of each sample was sent to GENEWIZ, Inc. for sequencing. Sequences were aligned using ClustalW2 software (EMBL-EBI).

Boronate Affi-gel binding and elution testing. Initially, to determine the optimal elution buffer, 40 μL ~2 μM selection buffer-exchanged DNA from round 11 was incubated with a final concentration of 20 μM FDG or FMG in selection buffer for 1 minute at room temperature. The reaction was diluted 2-fold with 7 M urea and exchanged into 0.1 M HEPES, pH 8.5 using a spin column of Sephadex® G-25. The resulting solution was incubated on pre-equilibrated 15 mg of boronate Affi-gel, as described in galactosidase selection, for 1 hour at room temperature. Each boronate Affi-gel column was spun at 4000 x g for 1 minute and washed with 6x100 μL 0.1 M HEPES, 8.5. The FMG reaction was eluted with 3x100 μL 0.1 M ribose in 0.1 M HEPES, pH 8.5, followed by 1x100 μL 0.05 M NaOAc. In parallel, the FDG reaction was eluted with 3x100 μL 0.05 M NaOAc, followed by 1x100 μL 0.1 M ribose in 0.1 M HEPES, pH 8.5. After each wash and elution, the columns were spun at 4000 x g for 1 minute and each flow-through was collected and diluted 100-fold. One microliter of each sample was added to a 10 μL reaction containing a final concentration of 1x iQ™ SYBR® Green Supermix (Bio-Rad) and 300 nM of forward and reverse primers (AL228 and AL1954). The samples amplified by denaturing at 95 °C for 2.5 minutes, followed by 40 cycles of 95 °C for 5 seconds and 60 °C for 30 seconds and fluorescence data collected using a CFX Connect™ Real-Time PCR Detection System (Bio-Rad).

To determine the binding and eluting efficiency of the boronate Affi-gel, a standard curve of FMG (600, 60, 6 and 0.6 pmoles) was generated using the FAM channel of the CFX Connect™ Real-Time PCR Detection System (Bio-Rad). Four parallel boronate Affi-gel binding experiments were performed by adding 8.1 mg boronate Affi-gel to two of columns and 16.0 mg to the other two columns. Each column was equilibrated by washing with 2x250μL of 1 M

HEPES, pH 8.5 and then 2 of the columns were washed with 1x250 μL 0.1 M HEPES, pH 8.5 and

the other 2 were washed with 1x250 μL 0.4 M HEPES, pH 8.5. 30 μL of 20 μM FMG was added

to each column and incubated for 1hr. Each column was spun at 4000 x g for 1 minute and

washed 4x250 μL 0.1 M or 0.4 M HEPES, pH 8.5, followed by eluting with 3x250 μL 0.1 M

ribose in 0.1 M or 0.4 M HEPES, pH 8.5 and then 250 μL 3 M NaOAc, pH 5.4. All flow-through

samples were collected and fraction was added to a 48-well white plate for quantification on

the CFX Connect™ Real-Time PCR Detection System. The resulting data was corrected for

volume fraction and plotted.

Column-binding testing of clones and pool. Clones F4 (FDG selection) and B9 (FMG

selection) were asymmetrically amplified and radioactively labeled by adding 5 μL (final ~[0.2

nM]) of the clone PCR reaction to a 50 μL reaction containing a final concentration 200 μM

dATP, dGTP, dTTP, 50 μM dCTP, ~66 nM ^{32}P-α-dCTP, 2 μM forward primer (AL228), 0.4 μM

reverse primer (AL 1954) and 1 μL PFU in a buffer of 10 mM Tris•HCl, 50 mM KCl, 1.5 mM

MgCl$_2$, pH 8.3. The asymmetric PCR conditions denatured the reaction at 95 °C for 2 minutes,

followed by 10 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 30 seconds,

and completed with a final elongation at 72 °C for 2 minutes.

Each reaction was purified by diluting it 1.5-fold with 7 M urea loading buffer containing

bromophenol blue and xylene cyanol and loaded onto a 7.5% denaturing polyacrylamide gel

containing 2 M urea in 0.5x TBE buffer (0.5x Tris, boric acid, EDTA, pH 8.4 buffer), 0.1% APS and

0.01% TEMED. The gel was electrophoresed for 1 hour at 20 W, exposed on a scanning

storage phosphor screens (Molecular Dynamics) for 40 minutes and scanned at 200 μm

resolution on a Typhoon Trio. The bands corresponding to the ssDNA were excised and

eluted in 0.3 M KCl and precipitated in a final concentration of 70% ethanol. An alternative purification method was performed by using an ssDNA/RNA Clean & Concentrator™ kit (Zymo Research) resulting in a final 27 μL H$_2$O elution.

The DNA pellets were resuspended in 20 μL of selection buffer or 3 μL 10x concentrated selection buffer was added to the 27 μL H$_2$O elutions and a final concentration of 2 μM FMG/FDG was added to each reaction and incubated for 1 minute at room temperature. The reaction and the no fluorophore control were exchanged into 0.1 M HEPES, pH 8.5 using a spin column of Sephadex® G-25. The solution was incubated on pre-equilibrated 7.5 mg boronate Affi-gel, as described in galactosidase selection, for 30 minutes. The column was spun at 4000 x g for 1 minute, washed with 4x150 μL 0.1 M HEPES, pH 8.5 and eluted with 4x150 μL 0.1 M ribose in 0.1 M HEPES, pH 8.5. The flow-through after each wash and elution and the boronate Affi-gel were collected and the radioactivity (CPM) was quantified using a liquid scintillation counter (Beckman). The procedure was repeated with asymmetrically amplified and radioactively labeled DNA from round 21 of the FMG and FDG selections.

Continued selection and re-selection for a galactosidase. DNA from round 7 or round 21 were asymmetrically amplified and purified, as described in column binding testing. The DNA pellets or DNA elutions were added to a buffer with the final concentration of 0.1 M HEPES, pH. The solution was counter-selected by incubating the DNA with pre-equilibrated boronate Affi-gel for 1 hour, as described in the galactosidase selection. The boronate Affi-gel column was spun at 4000 x g for 1 minute and the flow-through was exchanged into selection buffer (140 mM KCl, 10 mM NaCl, 10 mM MES, 10 mM MgCl$_2$, pH 7.5) using a spin column of Sephadex® G-25. The solution was incubated with 2 μM FMG/FDG for 2 hours at

room temperature. The reaction was exchanged into 0.1 M HEPES, pH 8.5 using a spin column

of Sephadex® G-25 and incubated on pre-equilibrated boronate Affi-gel, as described in

galactosidase selection, for 1 hour. The column was spun at 4000 x g for 1 minute, washed

with 4x250 µL 0.1 M HEPES, pH 8.5 and eluted with 4x250 µL 0.1 M ribose in 0.1 M HEPES, pH

8.5 and 1x250 µL 3 M NaOAc. The flow-through after each wash and elution and the boronate

Affi-gel were collected and the radioactivity was counted using a liquid scintillation counter

(Beckman). All elutions were combined, concentrated and exchanged into a buffer of 10 mM

Tris, 0.1% SDS, pH 7.5 using an Amicon® centrifugal filter (Millipore) with a 30 kDa cut-off. The

concentrated sample was PCR amplified for the next round of selection, as described in the

galactosidase selection.

For rounds 12, 22 and 23, 0.4 M HEPES, pH 8.5 replaced the previously used 0.1 M HEPES,

pH 8.5 in all incubation and washing steps. Table 3.3 summarizes the amount of boronate

Affi-gel used and PCR cycles required to regenerate the pool.

Table 3.3. Summary of the galactosidase re-selection.

R	DNA (pmoles)	CS (mg)	Reaction time (min)	[FMG/FDG]	S (mg)	PCR cycles
8	~100	17	120	2	7.5	24D/24M
9	~100	16.5	120	2	7	20D/20M
10	~100	16	120	2	8	20D/20M
11	~100	15	120	2	7	20D/20M
12	~100	9.5	120	2	15	20D/20M
22	~100	15	120	2	15	20D/20M
23	~100	9.5	120	2	15	20D/20M

References

(1)   Wilson, D. S.; Szostak, J. W. *Ann. Rev. Biochem.* **1999**, *68*, 611.
(2)   McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.
(3)   Peracchi, A. *ChemBioChem* **2005**, *6*, 1316.
(4)   Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550.
(5)   Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411.
(6)   Ekland, E. H.; Bartel, D. P. *Nature* **1996**, *382*, 373.
(7)   Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223.
(8)   Purtha, W. E.; Coppins, R. L.; Smalley, M. K.; Silverman, S. K. *J. Am. Chem. Soc.* **2005**, *127*, 13124.
(9)   Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611.
(10)   Wang, W.; Billen, L. P.; Li, Y. *Chem. Biol.* **2002**, *9*, 507.
(11)   Li, Y.; Liu, Y.; Breaker, R. R. *Biochemistry* **2000**, *39*, 3106.
(12)   Sheppard, T. L.; Ordoukhanian, P.; Joyce, G. F. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7802.
(13)   Chandra, M.; Silverman, S. K. *J. Am. Chem. Soc.* **2008**, *130*, 2936.
(14)   Baum, D. A.; Silverman, S. K. *Cell. Mol. Life Sci.* **2008**, *65*, 2156.
(15)   Dass, C. R. *Trends in Pharmacol. Sci.* **2004**, *25*, 395.
(16)   Sando, S.; Sasaki, T.; Kanatani, K.; Aoyama, Y. *J. Am. Chem. Soc.* **2003**, *125*, 15720.
(17)   Liu, J.; Lu, Y. *J. Am. Chem. Soc.* **2003**, *125*, 6642.
(18)   Stojanovic, M. N.; Stefanovic, D. *Nat. Biotech.* **2003**, *21*, 1069.
(19)   Burke, B. J.; Regnier, F. E. *Anal. Chem.* **2003**, *75*, 1786.
(20)   Xu, H. W.; Ewing, A. G. *Anal. Bioanal. Chem.* **2004**, *378*, 1710.
(21)   Okhonin, V.; Liu, X.; Krylov, S. N. *Anal. Chem.* **2005**, *77*, 5925.
(22)   Rotman, B.; Edelstein, M.; Zderic, J. A. *Proc. Natl. Acad. Sci. USA* **1963**, *50*, 1.
(23)   Rotman, B. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1981.
(24)   Revel, H. R.; Luria, S. E.; Rotman, B. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1956.
(25)   Kumemura, M.; Collard, D.; Yoshizawa, S.; Wee, B.; Takeuchi, S.; Fujita, H. *ChemPhysChem* **2012**, *13*, 3308.
(26)   Cameron, D. J., Erlanger, B.F. *J. Immunol.* **1976**, *116*, 1313.
(27)   Diehl, H.; Markuszewski, R. *Talanta* **1989**, *36*, 416.
(28)   Hofmann, J.; Sernetz, M. *Anal. Biochem.* **1983**, *131*, 180.
(29)   Weith, H. L.; Wiebers, J. L.; Gilham, P. T. *Biochemistry* **1970**, *9*, 4396.
(30)   Zittle, C. A. *Advances in Enzymology and Related Subjects of Biochemistry* **1951**, *12*, 493.
(31)   Fulton, S. *Amicon Corp., Danvers, MA* **1981**.
(32)   Soundararajan, S., Badawi, M. *Anal. Biochem.* **1989**, *178*, 125.
(33)   Hageman, J. H., Kuehn G.D. *Anal. Biochem.* **1977**, *80*, 547.
(34)   Yano, K.; Fukasawa, T. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 1721.
(35)   Andre, S.; Wang, G. N.; Gabius, H. J.; Murphy, P. V. *Carbohydr. Res.* **2014**, *389*, 25.