ABSTRACT

COGGINS, NICOLE. New High-Throughput Screens and Selections for Synthetic Biology. (Under the direction of Gavin J. Williams).

Synthetic biology aims at creating interchangeable parts for new biological function. However, often the process used to create these parts is slow, impeding the progression of the field. One of the major bottlenecks is the lack of high-throughput screens and selections. Here we focus on developing new high-throughput screens and selections to engineer naturally existing scaffolds for new regulators of gene expression and pharmaceutically relevant peptides.

Transcription factors are traditionally used in synthetic biology to regulate gene expression. However, of the few that are used, many suffer from cross-talk, which hinders more than one being used per cell line. Instead, synthetic biology has turned to engineering RNA-based gene regulators. Of those regulators, riboswitches (cis-acting gene expression regulators) are of particular interest due to the simple mechanism by which they function and their ability to orthogonally and finely tune gene expression. However, few synthetic riboswitches exist due to the lack of ability to select or screen for functional mutants. Here, we are developing new in vitro selections by way of in vitro compartmentalization to engineer naturally occurring riboswitches to have new function. Feasibility of this approach has been demonstrated by showing that both the naturally occurring SAH sensing riboswitch and the synthetic theophylline riboswitch still retain function in emulsions, and the difference between the “on” and the “off” state could easily be detected by a fluorescent activated cell sorter (FACS).
Many pharmacologically relevant natural products come from secondary metabolites. However, due to their complex structure, many of these natural products cannot be synthesized or altered chemically. Instead, synthetic biology has turned to engineering biosynthetic pathways to incorporate new functional groups. Yet, this process is hindered by the lack of high-throughput screens available to alter the substrate specificity. Here this work focuses on developing new high-throughput screens to alter the substrate specificity of non-ribosomal peptide synthetases (NRPSs). Specifically, we have duplicated a previously published work, showing that the adenylation domain (A domain), which is responsible for activating and incorporating each amino acid into the growing peptide chain, can also act as a Coenzyme A (CoA) ligase. This ligation forms a thioester bond, which removes the free thiol from solution. We propose that this loss in thiol concentration can be used as an indicator of protein activity, which could then be quantified by Ellman’s reagen assay in a high-throughput fashion.
New High Throughput Screens and Selections for Synthetic Biology

by

Nicole Coggins

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Master of Science

Chemistry

Raleigh, North Carolina

2012

APPROVED BY:

_________________________  __________________________
Gavin Williams, Ph.D.        Alexander Deiters, Ph.D.
Committee Chair

_________________________  __________________________
Reza Ghiladi, Ph.D.          Robert Kelly, Ph.D.
BIOGRAPHY

Nicole Coggins was born in Scranton, PA on July 23rd, 1986 to Mr. and Mrs. Gerard Coggins. Nicole has one sister, Jennifer and two brothers, Jerry and Jeff. Nicole grew up in Moosic, PA and graduated from Riverside Jr./Sr. High School. Following high school, Nicole went to Wilkes University in Willkes-Barre, PA and then transferred to Marywood University in Scranton, PA. At Marywood University, Nicole was a member of Kappa Mu Epsilon and received a Bachelor’s of Science degree in Biotechnology and minors in Math and Chemistry. Nicole graduated Cum Laude in May 2010 where she received the Award for Excellence in Biotechnology. Immediately after her undergraduate studies, Nicole moved to Raleigh, NC where she received her Masters of Science in Chemistry from North Carolina State University.
# TABLE OF CONTENTS

List of Figures .................................................................................................................. v

List of Tables .................................................................................................................... vii

Chapter 1: Increasing the Synthetic Biology Tool Box ..................................................... 1

1.1 References ................................................................................................................ 4

Chapter 2: New High-Throughput Screens And Selections For Riboswitch Engineering.......................................................................................................................... 6

2.1 Introduction to Regulation of Gene Expression by Riboswitches ....................... 6

2.1.1 Control of Gene Expression ................................................................................. 6

2.1.2 Riboswitches Structure and Function ............................................................... 8

2.1.3 Engineering Riboswitch Function ....................................................................... 15

2.2 Results and Discussion ............................................................................................ 21

2.2.1 Cloning and Characterization of Target Riboswitches ..................................... 21

2.2.2 High-throughput FACS Selection ...................................................................... 36

2.2.2.1 An In Vivo Approach ...................................................................................... 37

2.2.2.2 An In Vitro Approach ..................................................................................... 39

2.2.2.3 In Vitro Selection of Synthetic Riboswitches Using the Streptavidin Biotin Linkage ................................................................................................................. 44

2.3 Conclusions and Future Directions ........................................................................ 47

2.4 Materials and Methods ........................................................................................... 50

2.5 References ............................................................................................................... 57
Chapter 3: Altering the Substrate Specificity of Non-Ribosomal Peptide Synthetases by Directed Evolution

3.1 Introduction ............................................................................................................ 63
3.2 Results and Discussion ......................................................................................... 70
3.3 Conclusion and Future Directions ......................................................................... 73
3.4 Materials and Methods ........................................................................................ 74
3.5 References ............................................................................................................ 77

Appendix ..................................................................................................................... 79

Appendix A: Primers ................................................................................................ 80
Appendix B: Constructs ............................................................................................. 81
LIST OF TABLES

Table 3.1. Mass spectrometry analysis of TycA reaction........................................72

Table A.1. Primer names and associated sequences..............................................80
LIST OF FIGURES

Figure 1.1. Synthetic biology spans many different levels of structure................................2
Figure 2.1. The lac Operon........................................................................................................8
Figure 2.2. Metabolites that regulate the expression of naturally occurring riboswitches.................................................................10
Figure 2.3. The physical structure of a riboswitch.................................................................12
Figure 2.4. Riboswitch Control Mechanisms..........................................................................14
Figure 2.5. Using Systematic Evolution of Ligands by Exponential Enrichment (SELEX) to obtain functional aptamers.........................................................16
Figure 2.6. The process of making and recycling SAH in the cell.................................23
Figure 2.7. The proposed structure of the SAH riboswitch in P. syringae....................24
Figure 2.8. How riboswitches function in an in vitro transcription translation reaction...........................................................................................................26
Figure 2.9. In vitro assay of the ahcY riboswitch.................................................................27
Figure 2.10. On and off states of the TPP riboswitch.........................................................28
Figure 2.11. In vitro assay of the thiM riboswitch.................................................................29
Figure 2.12. The untranslated region of the E. coli lysC gene........................................32
Figure 2.13. Using conserved portions and secondary structure of the Bacillus subtilis lysC aptamer to predict the structure of the E. coli lysC aptamer...............32
Figure 2.14- The predicted structure of the E. coli lysC aptamer........................................33
Figure 2.15. In vivo assay of the E. coli lysC riboswitch.......................................................34
Figure 2.16. Construct that will determine if the *E. coli* riboswitch regulates
transcription or translation…………………………………………………………35
Figure 2.17. lysC controls translation…………………………………………………………36
Figure 2.18. Cells that express gfp can be enriched from populations that do not
using FACS………………………………………………………………………………………………...38
Figure 2.19. The ahcY riboswitch maintains activity in a primary emulsion………41
Figure 2.20. FACS analysis of the on and the off state of the theophylline
riboswitch…………………………………………………………………………………………………………43
Figure 2.21. Streptavidin selection scheme……………………………………………………45
Figure 2.22. His- tagged streptavidin can bind to biotinylated DNA in vitro……….47
Figure 3.1. Core domain and organization of a NRPS……………………………………65
Figure 3.2. Altering the specificity of NRPSs………………………………………………66
Figure 3.3. ATP/pyrophosphate (PPi)-exchange assay…………………………………….67
Figure 3.4. TycA catalysis and quantification………………………………………………69
Figure 3.5. NRPS responsible for tyrocidine production………………………………70
Figure 3.6. SDS-PAGE gel image of purified TycA………………………………………..71
Figure 3.7. RP-HPLC analysis of TycA reactions………………………………………..72
Figure A.1. DNA constructs used to obtain a functional riboswitch…………………..81
Chapter 1: Increasing the Synthetic Biology Tool Box

1. Introduction

Synthetic biology can be defined as the engineering of new parts, devices, or systems that mimic biology, or the re-engineering of existing parts to exhibit new functions. One component essential of synthetic biology is the creation of interchangeable parts for the rapid construction of new biological functions, ranging from synthetic bacteria that can sense pollutants and degrade them to a microorganism for production of potent cancer therapeutics. Aside from the application-driven aspect of synthetic biology, there is also a significant learning component; if biological parts can be constructed and used to build a new synthetic system, then much can be learned regarding what is necessary for that system to function.

The application of synthetic biology spans many different levels of structure: molecular, pathway/network, organism, and multi-cellular (Figure 1). At the molecular level, synthetic biologists aim to alter the basic biomolecules of life. For example, the genetic code has been expanded via incorporation of non-natural nucleic acid base pairs that are compatible with natural DNA polymerase. Also, the DNA backbone has been shown to be modifiable and still retain ability to be replicated and evolved by engineered polymerases. In addition, unnatural amino acids have been installed into proteins giving proteins new properties and functions.
In addition to creating synthetic genetic codes, synthetic biology aims to provide new tools for gene expression. Protein expression can be regulated in several different ways such as via transcription\textsuperscript{7} or translation\textsuperscript{8}, via protein-independent\textsuperscript{9} or dependent systems\textsuperscript{10}, and via \textit{in cis}\textsuperscript{11} or \textit{in trans}\textsuperscript{12} mechanisms. Systems that regulate gene expression can be functionally described as logic gates, where a defined set of input conditions (e.g. small molecules, protein, or nucleic acid) control the expression of a single gene or a set of genes. One key goal of the field is to be able to independently tune the transcription and translation of multiple genes.\textsuperscript{13} Biosynthetic pathways are also of interest in synthetic biology due to their
potential to be engineered to produce novel biomolecules. One long-term goal for synthetic biology is to be able to engineer a protein, either *de novo* or by modifying an existing scaffold, and to modularize biosynthesis so that new small molecules can be synthesized through efficient combinatorial approaches.$^2$

Microbes that function as multicellular organisms are able to carry out complex functions that no individual cell can carry out and are able to withstand harsher environmental fluctuations. This is accomplished by being able to communicate with each other and divide the labor, both being targets for synthetic biologist that focus on at multi-cellular level.$^3$ Arai *et al.* were able to engineer two populations of *B. subtilis* to secrete two different proteins that are necessary to degrade cellulose, showing that *B. subtilis* could work in concert to perform a function that was previously unsuccessful for an individual population.$^{14}$

Synthetic biologists ultimately aim to mimic nature for a beneficial output. Yet, nature has evolved life to be complex, involving many intricate parts. In order for synthetic biologists to mimic nature, these scientists require many different modular parts at their disposal. However, despite recent advances, the tool box is limited. This work aims a closing this gap by developing new screens and selections that will be used to identify novel nucleic acid-based and protein-based parts from libraries of mutants. These new parts could potentially be used as regulators of gene expression, biosensors for small molecule detection, and enzymes for the synthesis of novel therapeutic analogs.
1.1 References:


Chapter 2: New High-Throughput Screens and Selections for Riboswitch Engineering

2.1 Introduction to Regulation of Gene Expression by Riboswitches

In order to control complex cellular processes, many genes need to be regulated. Currently, gene regulation is achieved by inducible promoters. However, compared to the standard \textit{E. coli} cell which is estimated to have 300 transcription factors, the number of modular inducible promoters for synthetic biology is severely limited (less than 10), making the ability to achieve complex control almost impossible. In addition, many of these inducible promoters suffer from cross talk, and due to the complex interactions necessary for transcription factor function (protein small-molecule interactions and protein DNA interactions) transcription factors can be hard to redesign. Instead, synthetic biologists are now turning to RNA based gene regulators.

2.1.1 Control of Gene Expression

Jacob and Monod were the first to theorize that small molecules were capable of influencing the expression of individual genes. This theory came about when they first noticed that some mutant \textit{E. coli} either constitutively expressed $\beta$-galactosidase, or did not express it at all. Through a series of experiments they were able to isolate and characterize the gene responsible for the controlled expression, experiments which earned them a Nobel Prize in 1965.
Figure 2.1 is the operon identified and characterized by Jacob and Monod. Operons are composed of two different types of genes, structural genes which are sequences that give rise to non-regulatory proteins and operators which are DNA sequences that regulate the expression of the structural genes via specific interactions with repressor proteins. In the lac operon pictured in Figure 2.1, the constitutively expressed Lac repressor protein binds to the operator in the absence of lactose. However, on the occasion where glucose is low and lactose is high, an isomer of lactose (allolactose), can bind to the Lac repressor protein, causing a conformational change in the protein, reducing the binding affinity of the Lac repressor for the lac operator. Removal of the Lac repressor then allows RNA polymerase to bind and transcribe the downstream genes.
Figure 2.1. The lac operon. The repressor protein, LacI binds to the operator sequence in the absence of allolactose (isomer of lactose) preventing transcription of the downstream gene. However, when allolactose is present, LacI binds to the allolactose, decreasing its affinity for the operator 3-fold. This causes LacI to dissociate from the operator and allowing transcription.

2.1.2 Riboswitches Structure and Function

From Jacob and Monod’s isolation of the first operon, it was thought that all inducible genes were regulated by proteins. However, researchers were unable to identify the proteins responsible for regulating the expression of numerous small molecule biosynthetic genes which include thiamine pyrophosphate (TPP), S-adenosyl methionine (SAM), and flavin mononucleotide (FMN). Instead, sequencing data revealed that many of these small molecule biosynthetic genes had conserved 5’ untranslated regions (5’ UTR), and the conservation spanned many different
phylogenetically diverse species indicating they may have ancient origins. Yet, it wasn’t until 2002 that Ronald Breaker’s lab showed that these conserved 5’ UTR’s were actually RNA metabolite sensing domains which were termed riboswitches. These RNAs form complex structures that are able to undergo a conformational change in response to the presence or absence a small molecule metabolite. This conformational change then regulates the expression of the downstream genes without any protein mediation.

The diversity of small molecules that could be recognized by riboswitches is impressive. To date, many different naturally occurring riboswitches have been isolate. These naturally occurring riboswitches have the ability to sense a variety of different sized molecules, from those riboswitches that could sense small monoatomic ions (Mg^{2+}) to those that could sense adocobalamin which has a molecular weight of over 1.5 kDa. Also, riboswitches are able to sense ligands that have different charges, from the negatively charged pyrophosphate group on thiamine pyrophosphate to the positively charged sulfonium ion in S-adenosyl methionine. Moreover, there are 24 different structural classes of riboswitches and some times more than one structure could sense the same molecule. In addition, not only are there riboswitches that can sense molecules of varying size and charge, but their specificity for their ligand is proportional to that of an antibody. This can be seen in the SAH sensing riboswitch which discriminates at least 1000-fold against SAM which differs from SAH by only one methyl group. This illustrates that RNA is
capable of functions previously associated only with proteins. From an engineering standpoint, because there are only 4 possible substitutions that could be placed at any nucleotide position versus the 20 different substitutions in proteins, this could make riboswitches an easier target to engineer or create via de novo approaches, which is currently difficult to do with an inducer of transcription.

![Metabolites diagram]

Figure 2.2. Metabolites that regulate the expression of naturally occurring riboswitches.
Currently, most naturally occurring riboswitches are found in prokaryotes, but one has been isolated in eukaryotes\textsuperscript{11}. Even though there are many riboswitches that sense a wide variety of small molecules, riboswitches do share some common structural and functional characteristics (Figure 2.3). Firstly, all known riboswitches regulate some metabolic process. Structurally, prokaryotic riboswitches are organized into 2 distinct domains: an aptamer domain and an expression platform. The aptamer domain is the region that is responsible for binding the cognate ligand with high affinity and specificity. Ligand binding then induces a conformational change in the expression platform that is able to adopt one of two possible conformations which either increases or decreases gene expression.\textsuperscript{5} Another universal feature among bacterial riboswitches is that the aptamer portion is usually located at the 5’ UTR of the mRNA whose expression they control, which gives time for the riboswitch to respond to the metabolite before the full-length mRNA is produced.\textsuperscript{5}
There are several different ways riboswitches can respond to ligand binding, but the most common ways described thus far have been either through translational inhibition or by transcriptional attenuation. Transcriptional attenuation is a process in which transcription is stopped. In Figure 2.4, in the absence of the ligand, a stem-loop structure called the intrinsic terminator forms. This steam loop structure is rich in guanine/cytosine pairs which pauses the ribosome. The ribosome itself does not bind tightly to uracils, so if the ribosome pauses in a region of poly-uracils, (red, Figure 2.4) it will fall off, stopping transcription. However, if the ligand is present,
another stem loop structure is formed, called the anti-terminator. Here again the ribosome is stalled, but just long enough to initiate translation of the leader peptide, causing transcription of the full length transcript.\textsuperscript{12}

Translational inhibition occurs when a ligand binds to the aptamer portion of the riboswitch and inhibits ribosome binding. This is known as an “off switch” It is also worth noting that this process can be reversed; instead of ligand binding inhibiting translation, ligand binding can reveal the RBS and turn on transcription. This riboswitch is known as an “on switch”.

Figure 2.4. Riboswitch control mechanisms.\textsuperscript{13} (A) Riboswitch that controls translation. In the absence of the ligand, the RBS is revealed and the ribosome could bind allowing translation. However, in the presence of the ligand, the riboswitch adopts another conformation which sequesters the ribosome binding site and does not allow transcription to occur. (B) Riboswitch that controls transcription. In the absence of the ligand, a terminator structure forms, this causes the polymerase to stall at a poly-uracil region and truncation of the transcript. In the presence of the ligand, the riboswitch forms an anti-terminator structure, allowing the formation of the full length transcript. (C) Riboswitches acting as ribozymes. The presence of the ligand induces a conformational change and cleavage of the transcript. This cleavage destabilizes the transcript and targets it for degradation by the cell. (Adapted from Gallivan 2007)
2.1.3 Engineering Riboswitch Function

Before natural riboswitches were found, it was already known that aptamers could regulate gene expression when placed in the untranslated region of a gene. In 1998 Werstuck and Green showed that by placing aptamers that senses Hoechst dye in between the 5’ cap and the start codon of a eukaryotic cell, translation could be turned off in the presence of the dye and turned on in the absence by inhibiting ribosome scanning.14

Since 1990, RNA aptamers have been generated in vitro by using a method called systematic evolution of ligands by exponential enrichment (SELEX).15 Here, a linear DNA library is generated where a randomized region is flanked by two constant regions, one of which includes a T7 promoter at the 5’ end. T7 RNA polymerase is then used to transcribe the DNA library in vitro, generating an RNA library. This RNA library is then added to a column where the ligand is bound to the solid phase. RNA that interacts with the ligand will remain bound to the resin, and all loosely bound RNA will be washed off. The bound RNA is then eluted by using a competitor or denaturing agent such as urea, is then reverse transcribed, PCR amplified, and entered back into another round of SELEX. Usually many rounds of SELEX are used before finding the best aptamers.
Figure 2.5. Using Systematic Evolution of Ligands by Exponential Enrichment (SELEX) to obtain functional aptamers.

Many different RNA and DNA aptamers have been generated by this method for a variety of purposes, each binding to their cognate ligand with high affinity and specificity. For instance, an aptamer generated by NeXagen Inc., now Gilead, had a disassociation constant of 100 nM and is 10,000-fold more selective for theophylline than caffeine which differs by a single methyl group. However, a ‘good’ aptamer does not necessarily translate into a ‘good’ riboswitch. In fact, previous studies by the Suess lab showed that only a fraction of tetracycline-binding aptamers generated...
were able to control gene expression in yeast, even though they all had relatively the same affinity for the ligand.\textsuperscript{17} Thus, even though SELEX can select for the best RNA binders out of a library of $10^{15}$ members, an additional screen or selection is required to obtain a functional riboswitch. Yet, current methods available for riboswitch screening/selection are \textit{in vivo} based, which limits the throughput to at most $10^8$ library members. Nevertheless, SELEX has been used in a few cases to generate a variety of RNA regulators in eukaryotes\textsuperscript{18} demonstrating that unlike inducible promoters, RNA regulators can be generated \textit{de novo}.

Instead of targeting eukaryotes, others have taken a note from nature and mimicked naturally occurring prokaryotic riboswitches. Furthermore, prokaryotic riboswitches that control translation have been targeted because they operate under the simplest mechanism. Although the mechanism is simple, many parameters still need to be addressed when designing a synthetic riboswitch. One of the most important parameters is the dynamic range or the activation ratio of a riboswitch. The activation ratio is usually reported as a ratio of maximum and minimum gene expression that is achieved by omission or addition of the ligand. Usually, the larger the dynamic range the better the riboswitch. This is because gene expression is proportional to the concentration of the ligand, and therefore with a larger dynamic range many different levels of gene expression can be achieved. For example, using rational design, the Suess lab used the RBS sequestration/exposure mechanism to engineer a theophylline riboswitch into \textit{B. subtilis}.\textsuperscript{19} First, an RNA structure was engineered to hide the ribosome binding site, but upon insertion of one nucleotide,
the ribosome binding site would be in a loop region where the ribosome could bind. Then a theophylline aptamer which was known to undergo a one nucleotide helix slip, was placed at the 5’ end of the RBS. By placing this synthetic riboswitch next to a repressor protein (XylR) which controlled the expression of the lacZ gene, the authors were able to show that galactosidase activity was dependent on the concentration of theophylline present, establishing a maximum activation ratio of 8.8.\textsuperscript{19}

Since then, the Gallivan lab has used various evolutionary techniques to create theophylline riboswitches with high activation ratio’s in \textit{E. coli}. To do this, the Gallivan lab used both genetic screens and selections. Both methods are used to obtain a population with a desired phenotype. With a screen, every mutant is observed and the positives are usually isolated by hand or via robot. Because of this, most libraries are limited to $10^4$ library members. However, with a selection a desired trait is usually linked to cell survival. Hence, the libraries are usually larger but no information is gained for each individual. Desai \textit{et al.} cloned the theophylline aptamer five base pairs upstream of the β-galactosidase reporter gene and subsequently demonstrated that riboswitch activation could be detected in \textit{E. coli}.\textsuperscript{20} The activation of the crude riboswitch was improved by developing a fluorescent activated cell sorting (FACS)-based selection, whereby \textit{E. coli} harboring riboswitch mutants that produced larger fluorescent signals compared to the wild-type riboswitch were collected and used for the next round of evolution.\textsuperscript{21} Still, the library size was limited by their transformation efficiency of \textit{E. coli} which for this
experiment was $10^8$ transformants/µg of DNA. Also, because FACS analyzes individual cells, the fluorescence change could be due to a different copy number of the vector or individual concentration changes of RNA in the cell which could potentially provide false positives and negatives.

Selecting for an aptamer using SELEX and placing it next to an expression platform is one way to obtain a synthetic riboswitch. Others have taken a different approach which involves re-engineering the function of existing, naturally occurring riboswitches. For instance, the Yokobayashi lab reversed the ligand response of the thiamine pyrophosphate (TPP) riboswitch in *E. coli* from an off-switch to an on-switch. This was accomplished by using a dual genetic selection where the TPP riboswitch library controlled the expression of the tetracycline resistance gene, TetA. TetA allows cells to grow in the presence of tetracycline but die in the presence of Ni$^{2+}$. In order to use this selectable marker for a dual genetic selection, the group reasoned that as a positive selection, a functional riboswitch should express TetA in the presence of TPP and hence survive in the presence of tetracycline. Yet, for the negative selection, the positive library members should not express TetA in the absence of TPP and hence grow in the presence of Ni$^{2+}$. This selection was successful and was used to obtain a functional riboswitch out of a library of 75000 members that turned on in the presence of TPP with an 11-fold activation ratio.$^{22}$

Still, as in the methods previously described, this selection can only work *in vivo* which hinders the type and rate of which riboswitches could be developed. Because of working *in vivo*, only ligands that can be used are those that cross the cell
membrane and are non-toxic to the cell. Also, due to transformation efficiency, the maximum library size for a selection is $10^8$ members. In addition, with the selections used, no quantitative information can be obtained about the activation of the switch. Therefore an additional screen needs to be used, which currently only allow for library sizes of up to $10^4$ members.

The ligand specificity of naturally occurring riboswitches has been altered by semi-rational design. Here, the ligand specificity of the adenine riboswitch from the 5'-UTR of the adenine deaminase encoding gene (add) of *Vibrio vulnificus* was changed by creating a small library of mutants, where all of the residues that came into contact with the ligand were randomized. An antibiotic selection was then used to search for activation toward 80 different heterocyclic adenine analogs. However, because an *in vivo* selection was used, the selection was followed by a gfp screen to identify the mutant and adenine analog with the largest activation and orthogonality. In fact, the selected mutant showed no activity toward the original ligand, but showed a 6.5-fold activation with its cognate analog.\(^\text{23}\) This shows that natural riboswitches have the ability to be engineered through semi-rational design to become orthogonal components for synthetic biology. However, this mutant did not function as well as the wild-type, indicating a better functioning riboswitch may have been obtained if a more high through-put selection was used.

Riboswitches are a great target for synthetic biology due to the fairly simple mechanism by which gene expression is controlled, the wide variety of ligands that are able to be recognized, and the ability to be engineered to recognize new ligands.
Also, riboswitches have shown to be modular, controlling the expression of any downstream gene. In addition, unlike transcription factors which act in trans, the regulatory mechanism is directly attached to the gene it encodes, which maintains the genotype phenotype linkage. However, few synthetic riboswitches exist. This is primarily due to the lack of high-throughput methods used to synthetically engineer riboswitches and of those that currently exist, all are *in vivo*. Therefore, there is a need then to develop new higher-throughput methods that allow riboswitches to be engineered to recognize any molecule.

2.2 Results and Discussion

2.2.1 Cloning and Characterization of Target Riboswitches

Two different approaches have been taken to develop synthetic riboswitches: engineer an aptamer to regulate gene expression or alter the function of a naturally occurring riboswitch. Because of the limitations of aptamer based design, naturally occurring scaffolds were used to validate the efficiency of which the developed selections functioned. Of the naturally occurring riboswitches, three different riboswitches that sense s-adenosylhomocysteine (SAH), lysine, and thiamine pyrophosphate (TPP) were selected. More specifically, the SAH sensing riboswitch, ahcY, from *Pseudomonas syringae* and two *E. coli* riboswitches, lysC (lysine) and the thiM (TPP) were chosen because they likely operate translationally. Riboswitches that modulate translation operate under a simple mechanism as compared to riboswitches that regulate transcription which are usually more
complex. Also, some studies have shown that riboswitches that regulate transcription and may be organism specific. Moreover, SAH, lysine, and TPP are small molecules that are typically difficult to regulate \textit{in vivo}, and could therefore represent good targets for development of cell-free screens or selections.

The ahcY riboswitch is an on-switch that turns on translation of the downstream gene in the presence of S-adenosylhomocysteine (SAH). In nature, the SAH riboswitch is responsible for sensing the toxic byproduct SAH which then up-regulates genes responsible for the recycling SAH back to S-adenosylmethionine (SAM)\textsuperscript{10}, an important co-substrate in methyl group transfer reactions. Many genes are believed to have the aptamer portion that could sense SAH, all regulating the expression of SAH recycling. Particularly, the SAH sensing riboswitch from \textit{P. syringae} controls the expression of SAH hydrolase (ahcY). This enzyme catalyzes the breakdown of the thioester bond of SAH to produce adenosine and L-homocysteine. L-homocysteine, through a series of other enzymatic steps, is converted back to SAM (Figure 3.1)\textsuperscript{10}. 

Figure 2.6. The process of making and recycling SAH in the cell.\textsuperscript{10}

Figure 2.7 shows the predicted structure of the ahcY riboswitch from \textit{P. syringae}. Here the riboswitch is believed to function by a simple helix slipping mechanism, where the two nucleotides that are base paired with the ribosome binding site or Shine-Dalgarno sequence (SD) are believed to alternatively base pair with the two other boxed nucleotides in the P4 stem, revealing the ribosome binding site and allowing the ribosome to bind and initiate translation. In order to prove that this untranslated region is in fact a riboswitch, the ahcY 5’-UTR was fused with the β-galactosidase reporter gene and placed back into \textit{P. syringe}. The cells were then grown in minimal media containing small molecules that are known to increase the cellular level of SAH, resulting in a five to six fold increase in gene expression.\textsuperscript{10}
Figure 2.7. The proposed structure of the SAH riboswitch in *P. syringae*⁷⁰

The ahcY riboswitch is a naturally occurring riboswitch that is a good target for our studies for several reasons. Firstly, the Breaker lab has shown, this riboswitch appears to operate by a fairly simple helix slip mechanism. Second, the structure only contains 151 nucleotides; many naturally occurring riboswitches are larger. In addition, most naturally occurring riboswitches turn off gene expression in
the presence of the ligand. Yet, this switch is different than most in that it is an on switch, and it is easier to quantitate an increase in gene expression rather than a decrease. However, given SAH does not cross the cell membrane, this is more difficult to use as a target for engineering using currently described in vivo screens and selections. In order to evolve this riboswitch, an in vitro technique is likely required.

Given the activity of the ahcY riboswitch has not been determined in vitro, we first set out to prepare a suitable reporter construct for use in commercial in vitro transcription/translation (IVTT) kits. The untranslated region of the P. syringae ahcY gene was built by assembly PCR. In order to determine if the ahcY riboswitch could even function in vitro, the ahcY riboswitch was fused to the 5’ end of the superfolder green fluorescent protein gene (sfgfp) affording the vector pET-17b/ahcY/gfp. The RTS 100 E. coli HY kit from 5 Prime was used for IVTT assays. The HY kit is an S30 E. coli lysate which potentially better mimics a natural cellular environment compared to the purified PURExpress kits from New England Biolabs. In a bulk IVTT reaction, only the linearized riboswitch and reporter is expressed to eliminate all of the unnecessary genes present on the vector (Figure 2.8) The DNA is then transcribed using T7 RNA polymerase. If functional, the riboswitch can bind the cognate ligand, which is added to the assay mixture when required, reporter expression is induced. However, in the absence of ligand, no binding should occur, and the ribosome is unable to translate the reporter gene.
This simple in vitro assay was used to determine the activation ratio of the ahcY riboswitch. Gratifyingly, the ahcY riboswitch was activated 2.5-fold in the response to SAH but had little effect on gene expression in the absence of the riboswitch (Figure 2.9). Interestingly, this activation ratio is ~2 fold less than that previously reported via an in vivo assay. Similar activation ratios were obtained when other IVTT kits were used in this assay (data not shown). One possible explanation for this slight discrepancy is that the E. coli based IVTT kit is not optimal, compared to the endogenous environment provided by P syringae. Additionally, the SD sequence of this riboswitch is a weak ribosome binding site for E. coli, which could lower the rate of translation. Lastly, in order to be compatible with most IVTT kits, which are T7-dependent, the endogenous promoter for the ahcY riboswitch was
not used. Some authors believe that there is a link between promoter and riboswitch function.\textsuperscript{5, 20}

\begin{figure}[h]
\centering
\includegraphics[width=0.6\textwidth]{figure2_9.png}
\caption{\textit{In vitro} assay of the ahcY riboswitch. The pET-17b/ahcY/gfp construct was assayed in the presence or absence of 100 µM SAH. Fluorescence values were normalized to the average fluorescence of the on state.}
\end{figure}

Of all of the naturally occurring riboswitches found to date, the riboswitch that senses the coenzyme thiamine pyrophosphate(TPP) is the only riboswitch found in all three domains of life.\textsuperscript{26} Like all riboswitches, the aptamer domain, called the T-box, is highly conserved, but the expression platform can vary greatly especially in the way it regulates gene expression. In fact, depending on the organism it is found in, it has the ability to regulate transcription, translation, and can even regulate mRNA splicing in higher order organisms.\textsuperscript{22} Also, others have shown that it has the ability to be engineered, to evolve it from being an off switch to an on switch.\textsuperscript{22}
Because of this, the TPP riboswitch has become a target for synthetic biology. In fact, if the TPP aptamer can be modified to accept a non-endogenous and cell permeable small molecule, this synthetic riboswitch may be able to control gene expression in all 3 domains of life. However, like the SAH riboswitch, TPP is not cell permeable. Accordingly, thiamine, a cell permeable TPP precursor, is usually added to bacterial cultures for assaying the TTP riboswitch.\textsuperscript{27} Thus, controlling the intracellular concentration of TPP is difficult and inconsistent. A cell-free \textit{in vitro} screen or selection could therefore offer a more efficient approach for engineering the TPP riboswitch because the TPP concentration could be more easily controlled.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{tpp_riboswitch.png}
\caption{On and off states of the TPP riboswitch\textsuperscript{22}}
\end{figure}

Once again, we first set out to build a T7-based reporter construct that could be used in commercial IVTT kits. The region -215 to +23 relative to the thiM gene
start was cloned out of *E. coli*. This gene was then subcloned into the pET-17b, affording the plasmid pET-17b/thiM/gfp. Subsequent *in vitro* assay revealed poor activation (data not shown). Consequently, the thiM/gfp fusion was subcloned into a pUC19 vector which is under the control of the *lac* promoter affording the plasmid pUC19/thiM/gfp. Under this control, very low level of activation was observed. Literature reports describe an activation between 2.5 and 18.\textsuperscript{22, 27-28}

![Figure 2.11: *In vitro* assay of the thiM riboswitch. *E. coli* Top10 cells harboring the pUC19/thiM/gfp vector were incubated for 4 hours in the presence or absence of 100 µM thiamine. The values of each colony assayed were then normalized by the optical density and highest fluorescence value. All normalized values were then averaged.](image)

The mechanisms of even well-characterized riboswitches are still unclear. Several riboswitches, particularly those that are transcriptionally regulated, are thought to be under kinetic control, which means that the switching mechanism is
mainly determined by the rate limiting step,\textsuperscript{29} which is likely ligand binding. Because of this, the effective concentration for riboswitch function has to be far greater than the $K_D$ in order for binding to beat transcription. Therefore, with a fast promoter like the T7 promoter, a large concentration of ligand may be needed for a significant change in gene expression. This can also be true for riboswitches that control translation. To further validate this, the Breaker lab has shown using isothermal titration calorimetry (ITC) experiments that the $K_D$ for the \textit{B. subtilis} lysine aptamer is 1 mM. However, when the lysine aptamer is attached to its expression platform, the $K_D$ goes up to 10 mM.\textsuperscript{30} This shows that the aptamer has to adopt the ligand bound conformation before the expression platform is expressed, meaning that the rate of transcription has to be slow enough to allow this binding. In the case of the thiM riboswitch, it may be that a slower promoter is needed for the riboswitch to be functional. In fact, as I have previously shown, there have been many different activation ratios that have been published for this same riboswitch, and all of which used a different promoter. The Breaker lab, who obtained the largest activation at 18-fold, used the endogenous thiM promoter. Hence, the endogenous promoter may be the best to use to obtain the largest dynamic range.

The final riboswitch selected is an amino acid sensing riboswitch. Given amino acids and their analogues are widely available and are readily cell permeable, riboswitches offer exciting possibilities for creating libraries of orthogonal riboswitches. So far, there are two different riboswitch classes that have been identified that interact with amino acids, the glycine and lysine riboswitch.

The lysine
riboswitch has a conserved aptamer portion called the L-box and is probably one of the most well studied riboswitches because it can be targeted by antibiotics. However, only the lysine riboswitches that regulate transcriptional attenuation have been studied. In fact, little is known about the lysine riboswitch in gram negative bacteria except it turns off gene expression, and it is proposed that it regulates translation due to the fact that no transcriptional attenuator is present. Therefore, before the lysC riboswitch could be used as a target for engineering, it first needs to be shown that this riboswitch regulates translation and does not operate by a more complex mechanism.

Figure 3.8 shows the untranslated region of the *E. coli* lysC gene. Because it is unclear where the riboswitch ends (often some of the gene coding sequence is a part of the riboswitch structure), two reverse primers were engineered to give two different constructs: -309 and +3 of the lysC gene and -309 and +42 of the lysC gene. These primers were then used to amplify the lysC riboswitch from *E. coli* genomic DNA. Also, in order to help understand where the riboswitch sequence may end, the proposed *E. coli* lysC riboswitch sequence was compared to the *B. subtilis* lysC riboswitch. It is known that the aptamer portion of the riboswitch is well conserved even among the most diverse organisms. As you can see from Figure 2.14, even though these are fairly distantly related bacteria, much of the secondary structure is still similar. Also, it is known that the lysine riboswitch that regulates transcriptional attenuation makes a loop loop interaction at the ends of the P2 and P3 stem. This also seems to be the case for the *E. coli* lysine riboswitch (Figure 3.9).
Although this is the first time this riboswitch structure has been fully proposed, how the RBS is sequestered still unknown.

Figure 2.12. The untranslated region of the *E. coli* *lysC* gene. The *E. coli* promoter elements are highlighted in yellow, and the transcription start site is highlighted in green.

Figure 2.13. Using conserved portions and secondary structure of the *Bacillus subtilis* *lysC* aptamer to predict the structure of the *E. coli* *lysC* aptamer. The known structure of the Bacillus subtilis (B. sub) *lysC* aptamer was compared to the 5’ untranslated region of the *E. coli* *lysC* gene. The regions highlighted in red are conserved regions of all lysine aptamers. The predicted secondary structure base pairing (P) are individually colored. Portions of this structure were previously predicted.
Figure 2.14. The predicted structure of the *E. coli* lysC aptamer. The untranslated region of the *E. coli* lysC gene (-288 to -68) was entered into M-fold RNA Folding Form with the known secondary structure constraints from Figure 2.13. It was also known that the B. sub lysC riboswitch makes an important loop loop interaction at the end of P2 and P3, and it appears as though this interaction may also be present in the *E. coli* lysC riboswitch. At the 3' end of P1', another stem loop was included that was seen in some M-fold predicted structures. This stem loop appears to be important in the switching mechanism of the switch; the alternative base-pairing is highlighted in red.

In order to obtain the activation of the lysC riboswitch, the DNA was subcloned into pUC19 (via pET-17b) to form the plasmid pUC19/ lysC/gfp. Because the activation of the lysC riboswitch has not been previously reported, determining the optimal assay conditions was by trial and error. *E. coli* readily decarboxylates lysine as a part of its acid metabolism. Thus, *in vivo* lysine concentration is dependent on the pH and the lower the pH, the less free lysine, and since lysC is
an off switch, the higher the gfp expression. In fact, the highest activations were achieved when the cells were incubated for long periods of time, but this result was hard to duplicate. Subsequently, reproducible activation ratios of ~7 were if the cells were diluted in a 1:1000 ratio (Figure 2.17). Presumably, the low concentration of the cells would not be able to change the extra cellular lysine concentration much during the course of the assay. Also, both constructs (-309 and +3 or -309 and +42 relative to the lysC gene) gave the same activation, which leads me to believe that the riboswitch terminates close to the start codon.

Figure 2.15. In vivo assay of the E. coli lysC riboswitch. A stock culture of cells harboring the +3 lysC/gfp fusion was diluted 1000 fold and incubated 4 hours in the presence of either 0 or 100 mM lysine. After re-suspending each culture in PBS, fluorescence measurements were obtained and normalized by the optical density and the highest fluorescence value.

Next, the construct shown in Figure 2.16 was designed to provide evidence that the lysC riboswitch controls translation. If the E. coli lysC riboswitch controls
translation, then gfp should be expressed in both the presence and absence of lysine. However, if the riboswitch operates by transcriptional attenuation or through an unknown ribozyme activity, then the number of full length mRNA transcript will still be dependent on the presence and absence of lysine and a higher GFP fluorescence should be seen in the absence of lysine.

![Diagram](image)

**Figure 2.16. Construct that will determine if the *E. coli* riboswitch regulates transcription or translation** gfp Δ is a truncated form of gfp; it contains only the first 75nt. If the riboswitch regulates translation then since a second RBS is located down-stream with the full length gfp gene, gfp expression should not change in the presence or absence of lysine. However, if lysC regulates transcription, the generation of the full length transcript will still be dependent on the lysine concentration.

When this construct was assayed in the presence or absence of lysine, no significant difference in GFP expression was observed (Figure 3.13). This is the first initial quantitative evidence that the lysine riboswitch controls translation. To further validate this result, methods have been developed to transcribe the DNA *in vitro* in the presence and absence of the ligand. If the lysC riboswitch controls translation, the length of RNA construct should not change in the presence or absence of ligand.
The construct from Figure 3.12 was express in vivo to determine if lysC controls translation. If the lysC riboswitch controlled translation by hiding or revealing the RBS on the expression platform of the switch then the downstream gfp gene should be expressed constitutively. However, if lysC controls transcription, then in the presence of lysine the transcript will be truncated and no GFP will be made. As you can see from this experiment, no significant change in gfp expression is observed in the presence or absence of lysine. This leads me to believe that the lysC riboswitch controls translation.

2.2.2 High-throughput FACS Selections

Currently, there are very few examples of riboswitches being evolved to recognized new ligands. This is probably due to that fact that using current existing methods, ligand choice is limited to those that are non-toxic and cell permeable. Also, most of the screens and selections that have been developed do not allow for large enough libraries. Yet, having a riboswitch ligand that is able to cross the cell membrane could be beneficial, especially for engineering new inducible promoters. This lead me to develop both an in vivo and an in vitro fluorescence-activated cell
sorting (FACS) screen that could be used to evolve the naturally occurring riboswitches into recognize a new ligand.

### 2.2.2.1 An In Vivo Approach

FACS is a special type of flow cytometry that uses a combination of fluorescence and light scattering to sort cell or cell like mixtures into two or more populations. Currently, a high-through-put FACS selection has been used to identify mutant theophylline riboswitches with improved activation ratios. A similar screen may also be used to evolve the ligand specificity of the lysC riboswitch, which is the focus of this subsection. A preliminary experiment was designed to show that *E. coli* cells that poorly express GFP (typical to riboswitch-controlled expression) could be isolated by FACS from a population that did not express GFP. This was done using varying ratios of GFP and non-GFP expressing cells (Figure 3.14). Gratifyingly, when these populations were mixed in a 1:1 ratio, the positive population was enriched 90%, showing that if this was a riboswitch selection, a 90% enrichment should be achieved.
Figure 2.18. Cells that express gfp can be enriched from populations that do not using FACS. A 1:1 mixture of cells that either express gfp or vector were sorted by FACS into two separate tubes. Each sorted population was then reanalyzed by FACS to show enrichment of both populations.

Although, this selection holds promise for evolving a riboswitch that recognizes a cell permeable ligand like lysine, this selection cannot easily be used to evolve a riboswitch which uses ligands that cannot cross the cell membrane (e.g. SAH riboswitch). Instead of using cells to maintain the genotype phenotype link, an alternative strategy called ‘in vitro compartmentalization’ (IVC) will be developed and is the subject of the next section.
2.2.2.2 An In Vitro Approach

IVC encapsulates an aqueous phase, usually containing the IVTT reaction, in an oil phase, forming a water-in-oil (w/o) or primary emulsion. The size of the resulting droplet can differ depending on the method used, but typically for an IVTT reaction they are about 1-10 µm in size. This volume is the smallest average volume that still maintains gene expression. On average, each droplet should contain a single member of DNA library that is transcribed and translated to give multiple copies of a protein it encodes, while still maintaining that genotype phenotype linkage. This makes each droplet one discrete reaction, and typically there are over $10^{10}$ droplets per milliliter reaction.

Depending on the strategy used, this primary emulsion can be re-emulsified to form a water-in-oil-in-water emulsion or a secondary emulsion, which better mimics a cellular compartment. If the output of each reaction is fluorescence, each droplet can be analyzed and sorted by FACS. Here droplets that are over an arbitrary threshold of fluorescence can be sorted at a maximum rate of 50,000 droplets per second and those that are collected can be entered into another round of evolution.

In order to use IVC to evolve riboswitches, we first need to illustrate that riboswitches can function in droplets. In Figure 2.19, an IVTT reaction containing the linear T7/ahcY/gfp fusion was emulsified in a primary emulsion also containing the DsRed protein. DsRed is added to the emulsion to identify droplets that include an internal aqueous phase. Gratifyingly, primary emulsion droplets that contained
pET17b/ahcY/gfp were bright green-orange after transcription/translation was allowed to proceed in the presence of SAH (Figure 2.19). Notably, when SAH was not added to the droplets, only red fluorescence was observed, confirming the absence of GFP expression. A positive control using droplets that contained the plasmid pET17b/gfp also produced green/orange fluorescence, while a negative control that lacked the gfp gene entirely produced only red fluorescent droplets. This result was duplicated yielding the same results, showing that riboswitches do function in a primary emulsion. In a recent literature report, the well-known theophylline riboswitch was also shown to function in emulsions.36
Figure 2.19. The ahcY riboswitch maintains activity in a primary emulsion. A confocal micrograph image of the positive control (A), negative control (B), and the riboswitch in the absence (C) or presence (D) of SAH.

Given cell sorters are not able to sort primary emulsions, these primary emulsions require conversion to a secondary w/o/w emulsion. As a proof of principle, the theophylline riboswitch\textsuperscript{35} (obtained from the Gallivan lab) was placed between
the T7 promoter and the gfp gene in the pET-17b vector. The resulting construct was linearized by PCR and emulsified in a secondary emulsion in the presence or absence of ligand (1mM theophylline). Subsequently, each emulsion was analyzed by flow cytometry, yielding an average fluorescence of 810 (A.U) and 366 (A.U) for the droplets with and without ligand, respectively (Figure 2.20). Previously, this riboswitch was reported to have over a 80-fold activation ratio under the IS10 promoter. However, under the T7 promoter the activation was less than 4-fold in bulk IVTT reactions (Steve Walsh, unpublished work). Regardless, although small, this difference is easily distinguished by FACS analysis and was reproducible (data not shown).
Figure 2.20. FACS analysis of the on and the off state of the theophylline riboswitch. (a) FACS analysis of droplets containing the theophylline controlled gfp gene expressed in the presence of 1mM theophylline. The arbitrary mean fluorescence for this population was ~810. (b) FACS analysis of droplets the theophylline controlled gfp gene expressed in the absence of theophylline. The arbitrary mean fluorescence for this population was ~366, showing a 2.2 fold decrease in fluorescence intensity.
2.2.2.3 In Vitro Selection of Synthetic Riboswitches Using the Streptavidin Biotin Linkage

Using the FACS IVC selection described in section 2.2, each round of selection could contain as many as $10^{10}$ mutants, which at a sort rate of 50,000 droplets per second, it would take over two days to sort the entire population. Yet, this is a lot faster than traditional selections (it would take 100 days to go through a library of similar size the most efficient selection). A genuine selection strategy that could identify functional or improved riboswitches from very large libraries ($10^{12}$) in a single step could offer a potentially efficient approach to engineer riboswitches. Inspired by other selection schemes that utilize the tight binding properties of the streptavidin:biotin interaction\textsuperscript{37} to maintain a genotype-phenotype linkage, a similar strategy could be adapted to pull-down functional riboswitches from completely randomized or partially randomized nucleic acid pools (Figure 2.21). The streptavidin and biotin dissociation constant is $\sim 10^{-14}$ M making it one of the strongest non-covalent interactions known in biology and has been successfully used in IVC-based strategies to evolve streptavidin variants.\textsuperscript{37a} Here, mutant riboswitches will be tested for the ability to control the expression of His-tagged streptavidin within droplets of a primary emulsion. Because the template DNA is biotinylated, expressed streptavidin captures those gene sequences which encoded functional riboswitch sequences.
Figure 2.21. Streptavidin selection scheme: In this selection, only a primary emulsion is made. Here, linear biotinylated DNA containing the riboswitch library attached to a 5’ His-tagged streptavidin is emulsified in a primary emulsion. After allowing time for expression, the emulsion is broken and only those riboswitches that are allowing expression in the presence of the ligand are purified by His-tag affinity purification. The library containing only the functional riboswitches can then be PCR amplified and the selection could be repeated. Also, a negative selection could be done to validate that the riboswitches only turn on gene expression in the presence of the ligand. Here, the biotinylated DNA would be emulsified without the ligand and all DNA that does not bind after the purification will be kept and entered into the next round of selection.

The coding sequence of streptavidin was based off of previously published construct\textsuperscript{37a} with an engineered N-terminal His\textsubscript{6}-tag and was synthesized by
Genscript. The construct was then placed in a plasmid under T7 control and linearized using primers with and without a biotin tag. In order to demonstrate whether the designed streptavidin construct binds its own biotinylated DNA, a DNA capture assay was performed. Here, non-emulsified ‘bulk’ control reactions were prepared, in which either biotinylated or non-biotinylated linear streptavidin DNA was expressed \textit{in vitro} and the resulting His$_6$-tagged streptavidin was pulled-down using magnetic Ni-NTA resin following several wash steps. Subsequently, the presence of streptavidin gene was detected by PCR, which showed that the streptavidin gene was specifically eluted from the resin prepared from biotinylated DNA (Figure 2.22). As expected, streptavidin gene was present in the wash steps from both biotinylated and non-biotinylated DNA. This data provides some evidence that \textit{in vitro} expressed streptavidin can bind its own biotinylated DNA, and could now form the basis of a novel IVC-based selection for functional riboswitches. Mock selections now need to be performed in primary emulsions that contain known ratios of functional and non-functional riboswitches or known ratios of other streptavidin positive and negative controls.
2.3 Conclusions and Future Directions

One of the major goals in synthetic biology is to be able to independently and reliably tune the expression of multiple genes, whether for controlling the expression of enzymes responsible for making a small molecule biosynthetic product or for the construction of complex gene circuits.\textsuperscript{1, 38} Traditionally, genes are regulated by inducible promoters. Yet, only a limited number ~10 of these inducible promoters could be used in \textit{E. coli}\textsuperscript{1} and to compound this problem, many inducible promoters cannot be used in conjunction with each other because they suffer from cross-talk.\textsuperscript{39} Instead, much interest is being focused on RNA-based regulators due to their ability to be re-engineered to perform new functions and control gene expression or to be a part of elaborate gene networks.\textsuperscript{40} However, we are limited by the techniques used to engineer these RNA based regulators. Currently, although aptamer portions of riboswitches can be identified \textit{in vitro}, only a small fraction of aptamers make good
riboswitches. Subsequently, functional riboswitches ultimately need to be identified by high-throughput screens or selections, which to date have entirely been performed in vivo, severely restricting both the throughput and scope of small molecule ligands used for the search.

Probably the most powerful method to date to engineer riboswitches is in vivo FACS selections. Yet, this method has only been used to engineer the dynamic range of a natural riboswitch, and not the ligand specificity. In this work, I have opened the door to using a previously uncharacterized riboswitch (lysC) as a possible template to evolve new amino acid sensing riboswitches. Due to the cell permeability of most amino acids, these evolved riboswitches could act as new orthogonal regulators of gene expression and biosensors. Also, lysine has been photocaged with a (2-nitro-valeryl)oxymethyl group on its ε amino end. Due to the photocaged lysine’s ability to enter the cell and dephotocage (be restored to L-lysine) with non-phototoxic effects, a NAND logic gate could be engineered in vivo where only in the present of both light and the photocaged small molecule, will gene expression be turned off.

Currently, no high-through-put method exists to evolve naturally occurring riboswitches in vitro. Because of this, researchers are limited by the type of ligand a riboswitch could be engineered to accept. This is a severe limitation especially for evolving riboswitches to sense a ligand like an important metabolite inside the cell that does not cross the cell membrane. In this chapter, I have established a
functioning riboswitch that has the ability to detect SAH concentrations in vitro. The activation ratio obtained was lower than what was previously reported, but this was expected because the in vitro assay was performed using an E. coli lysate, whereas previously the endogenous host (P. syringae) was used. In fact, it is believed that the dynamic range could easily be increased, by targeting the weak ribosome binding site and the nucleotides that interact with it. By increasing the dynamic range, this riboswitch could potentially be used as a novel SAH biosensor which then could be used to engineer new methyltransferases, which produce SAH as a byproduct. However, because its natural ligand cannot cross the cell membrane, the riboswitch is a model switch to engineer in vitro. Currently, two promising high-through-put methods to evolve these riboswitches were investigated, both of which involve in vitro compartmentalization. With the FACS selection, I have shown that even a small change in activation (2.5 fold) is quantifiable. For the purposes of using this as a screen to evolve the SAH riboswitch, selecting for the droplets that have the highest fluorescence in the presence of the ligand and lowest fluorescence in the absence of the ligand should increase the dynamic range. However, I have not yet been able to show a quantifiable fluorescence difference for the SAH riboswitch using FACS. Because the droplets have a different forward and side scattering than what was observed with the theophylline riboswitch, I believe the SAH may be interacting with the droplets, and if this is the case, this problem may be easily remedied by altering the concentration of SAH.
With the streptavidin-based selection in vitro expressed His$_6$-tagged streptavidin was shown to bond its own biotinylated gene, forming the basis of a proposed emulsion based selection for identifying functional riboswitches. I have also built a construct where the RBS of streptavidin was deleteriously mutated to a restriction site, causing it to mimic a non-functioning riboswitch. Mixing this construct with the active streptavidin construct in varying ratios in droplets, I should be able to show enrichment of the active DNA, further validating this selection.

2.4: Materials and Methods

Cloning of ahcY riboswitch

The ahcY genetic construct$^{10}$ was made by assembly PCR.$^{25}$ Briefly, six DNA oligonucleotides that contained overlapping ends with a similar melting temperature were extended by PCR. The resulting PCR reaction contained the complete gene sequence along with other DNA fragments. In order to make the full length gene the predominate product, a second PCR reaction was done where a small aliquot from the 1$^{st}$ reaction was used as a template for the second. Also during this reaction, only one forward and one reverse DNA oligo was added which would only amplify the full length construct. After the second PCR reaction, the full length ahcY construct was gel purified, digested using the engineered XbaI and BamHI restriction sites, and ligated into pET-28a to make the construct ahcY/pET-28a. The ahcY riboswitch was then PCR amplified out of the pET-28a vector using primers that contained the XbaI and HindIII restriction sites and ligated into the double digested
pET-17b vector giving the ahcY/pET-17b construct. The GFP reporter, SuperFolder gfp (sfgfp) gene, was PCR amplified from the SuperFolder GFP Plasmid (Theranostech, Inc.) and ligated to the N-terminus of the riboswitch using the BamHI and HindIII restriction sites of the ahcY/pET-17b vector to give the ahcY/sfgfp/pET-17b vector.

**General riboswitch cloning**

The resulting ahcY/gfp/pET-17b was used for cloning of all other riboswitches using the same restriction sites.

**In vitro transcription and translation reactions**

The resulting ahcY/sfgfp/pET-17b fusion was expressed *in vitro* by using 3 different transcription and translation kits: PURExpress (New England Biolabs), EasyXpress (Qiagen), and the RTS 100 *E. coli* HY kit (5 Prime). All expression was done following the kits protocol. For those reactions that contained the addition of ligand, either 100 µM of SAH or 100 mM of lysine was added.

**General reporter protein production**

Production of proteins was done BL21 (DE3). In a 2.5 L flask, 1 L of Lysogeny Broth (LB) containing either ampicillin (100 µg/ml) or kanamycin (50 µg/ml) was inoculated (1:100) with an overnight culture made from a fresh transformation. The 1 L culture was grown until the OD_{600} = 0.6 and then induced with 1 mM IPTG. The culture was then grown overnight at 20 °C. The cells were then pelleted (30 minutes, 4000 rpm, 4 °C), resuspended in TNG (50 mM Tris pH 7.4, 0.1M NaCl, 10% glycerol) buffer,
and stored at -20 °C as necessary. Cells were then lysed by sonication and the soluble faction was obtained by centrifugation (1 hour, 4700 rpm, 4 °C).

**General His-tagged protein purification method**

Proteins were purified using a BioLogic DuoFlow FPLC machine (Bio-Rad) with a UV detection at 280nm and a HisTrap HP column (GE Healthcare). Holding the flow rate constant at 2 ml/min, the column was equilibrated with a low imidazole buffer (10 mM imidazole, 20 mM sodium phosphate, and 500 mM sodium chloride, pH 7.4). About 3.5-5 ml of soluble protein was then injected with the low imidazole buffer for a total volume of 15 ml followed by washing the column with an additional 10 ml of the low imidazole buffer. Elution was carried out by running a linear gradient of high imidazole buffer (200 mM imidazole, 20 mM sodium phosphate, and 500 mM sodium chloride, pH 7.4) from 0-50% for 20 ml, 50-100% for 10 ml, and holding the high imidazole buffer at 100% for an additional 10 ml, collecting 2 ml fractions. The column was subsequently washed with 10 ml of the low imidazole buffer. The purity of the fractions was analyzed by SDS-PAGE and Coomassie blue staining. Fractions that contained pure protein were pooled and buffer exchanged using an Amicon Ultra centrifugal filter (Millipore). Quantification of the protein was done by using a Bradford assay.

**Secondary emulsions**

The inner aqueous layer (50 µl) consisted of either a PURExpress *in vitro* transcription translation reaction (New England Biolabs) or 1µg of purified protein in PBS buffer. For all reactions that were triggered on DsRed, 1µg of DsRed was
added. The oil surfactant layer, which consisted of 1% (wt/vol) cholesterol and Span 60 in decane, was heated to 45°C until dissolved and then cooled to 30°C. A Mini-
Extruder (Avanti Polar Lipids Inc.) was fitted as pictured below with 12 µm
Nucleopore Track-Etch polycarbonate membrane (Whatman) with two Hamilton
syringes. The extruder was then washed three times by loading one of the Hamilton
syringes with 1 ml of decane and extruding back and forth each time. Once cooled,
200 µl of the oil layer and 50 µl the aqueous layer was loaded into 2 Hamilton
syringes, and the aqueous phase was then extruded though the membrane and into
the other syringe. Fourteen other extrusions were performed and the resulting
primary emulsion was injected into a clean sample vial and incubated for varying
times at 30°C. During the incubation step, the extruder was washed and fitted as
previously stated but with an 8.0 µm Nucleopore Track-Etch polycarbonate
membrane (Whatman). After the incubation, 250 µl of the primary emulsion and 750
µl of the secondary aqueous layer (PBS and 0.5% Tween 20) was loaded into each
syringe and the primary emulsion was then extruded though the membrane and into
the other syringe. Six additional extrusions were performed and then stored on ice
until FACS sorting or analysis. Prior to FACS, the emulsions were diluted 100-fold
with PBS.41
Analysis and sorting of secondary emulsions

Secondary emulsions were diluted 1000-fold in PBS buffer and run in a Dako Cytomation Molecular Flow Cytometer and run at about 1,000 events per second, with a 100 µm nozzle, exciting with a iCyte 200mw Blue Sapphire 488nm laser and measuring emissions passing a 530 nm bandpass filter. Single, non-aggregated droplets were triggered on DsRed and gated using forward and side scatter criteria. Results were analyzed using Summit software.

Breaking the secondary emulsions

The resulting double emulsion was broken by centrifuging at 13,000 rpm for 5 minutes at 22°C, and the top oil layer was removed. The remaining aqueous layer was then washed twice with 1 ml of water saturated diethyl ether, discarding the top organic layer each time. The resulting DNA was ethanol precipitated and PCR amplified.
Primary emulsions

For each primary emulsion, 1 ml of oil-surfactant mixture (4.5% Span-80, 0.5% Tween-80, 0.1% Triton X-100) and a flea micro 7X2 mm spinbar (VWR) was added to a CRYO.S tube (greiner bio-one) and kept on ice. A 50 µl aqueous layer consisting of either the Roche RTS 100 *E. coli* HY *in vitro* transcription/translation (IVTT) kit or purified protein in PBS buffer was prepared according to protocol and placed on ice. For the IVTT kit, each reaction was set up according to manufacturer’s instructions with between $10^8$ and $10^{10}$ molecules of DNA. The oil-surfactant mixture was then added to an ice bath and the stir bar was spun on high using a hot plate and allowed to cool for an additional minute. The aqueous layer was then added to the oil layer drop-by-drop over one minute and then allowing to stir for an additional 3 minutes.42

DNA capture assay

A 50 µl RTS 100 *E. coli* HY IVTT reaction was set up according to manufacturer protocols with 25 ng of either 5’ biotinylated or non-biotinylated DNA and was incubated for six hours at 30 °C. After the incubation, 20 µl of HIS-Select Nickel Magnetic Agarose Beads (Sigma) were washed three times in 500 µl TBST (tris-buffered saline tween-20 pH 7.6) and re-suspended in 150 µl of TBST. The reaction was then incubated at room temperature in an end-over-end rotator for 30 minutes after which the beads were pulled down using neodymium magnets. The beads were then washed three times with 400 µl of TBST and eluted with 400 µl of elution buffer (250 mM imidazole in TBST). The resulting DNA PCR amplified.42
Determining the activation ratio of the *E. coli* lysC riboswitch

A stock culture of BL21 pLysS cells containing the lysC/gfp/pUC19 fusion was grown overnight in LB media at 37 °C and 250 rpm. Each stock culture was then diluted 1000-fold in 1 ml of M9 media in the presence or absence of 100mM lysine. After growing the cells for an additional 4 hours at 37 °C and 250 rpm, the cells were centrifuged at 4000g at 4 °C for 7 minutes. Each culture was then resuspended in 200 µl of PBS and added to a 96-well plate. Fluorescence and optical density measurements were obtained from BioTek Hybrid Synergy 4 plate reader (Winooski, VT, USA).

**Sorting *E. coli* cells**

BL21 (DE3) cells containing either pET-17b or gfp/pET-28a vector were grown to an OD$_{600}$ = 1 in LB media. In order to better mimic the lower fluorescence that is expected with riboswitches, IPTG was not added. Therefore the fluorescence observed is due solely to the leaky expression of the T7 system. The positive control was *E. coli* containing the gfp/pET-28a plasmid, negative was pET-17b, and the samples were ratios of the positive to negative in the log scale up to 10. Right before sorting, the cells were then diluted to about $10^7$ cells/ml in PBS. Sorting was done using the same conditions as the emulsions.
2.5 References


Chapter 3: Altering the Substrate Specificity of Non-Ribosomal Peptide Synthetases by Directed Evolution

3.1 Introduction

Antibiotic resistance and the lack of new antibiotics to combat this is a known problem.1 Traditionally, most clinically used antibiotics are secondary metabolites that originated from various sources such as bacteria and fungi. In fact, of the new antibiotic classes that have recently come on the market, three are natural products or natural product derivatives: daptomycin (cyclic lipopeptide), retapamulin (semi-synthetic pleuromutilin), and fixaxomicin (polyketide macrolactone).1b Many of these metabolites are structurally complex, with multiple functionalities, chiral centers, and other decorations, which could be very difficult to synthesize by traditional synthetic approaches. Instead, current research is turning to synthetic biology to offer new approaches and methodology to solve the antibiotic problem.2

Of particular interest are non-ribosomal peptide synthetases (NRPSs). Not only do NRPSs produce clinically relevant antibiotics, but they also produce cytostatic agents, siderophores, immunosuppressants, and toxins3. NRPSs are large multidomain enzyme assemblies that produce secondary metabolite natural products called non-ribosomal peptides (NRPs). As the name states, the peptide products are produced in the absence of a ribosome and can often include canonical and non-canonical amino acids. Each non-ribosomal peptide synthetase contains many different protein domains, organized into modules of enzyme activities, in concert acting as an assembly line, whereby each module is responsible for the
incorporation of a single monomer (Figure 3.1). Each module has 3 core domains: the adenylating domain (A), condensation domain (C), and the thiolation domain (T). The A domain, known as the ‘gate keeper’\(^4\), is responsible for amino acid recognition and acyladenylate formation under ATP consumption. The T domain or peptidyl-carrier-protein (PCP) domain, which has been post-translationally modified with a 4'phosphopantetheine cofactor (Ppant) from Coenzyme A (CoA), attacks the adenylated product to form a covalent linkage. At this point, the substrate can undergo modifications such as epimerization (E domain) or N-terminal methylation (M domains). Peptide bond formation is then catalyzed between two adjacent T domains by the condensation domain (C domain), subsequently translocating the first peptide intermediate downstream. Termination of nonribosomal peptide biosynthesis is catalyzed by the thioesterase domain (TE) (Figure 3.1).
Given that each A domain of a given module selects the cognate amino acid added to be installed into the growing peptidyl product, it should be possible to create NRP analogs with novel biological activity by swapping A domains from one NRPS to another. Yet, these hybrids often have poor activity because of faulty intermodule communication, a problem that is hard to resolve because it is not that well understood. Still, in a combined research project between Christopher Walsh and David Liu’s lab they were successful at creating hybrid systems by coupling the
domain swapping with directed evolution. However, the screens used for the directed evolution experiments are specific to the individual NRPS and cannot be used universally.

Instead, researchers have turned to computational methods to alter the substrate specificity of endogenous A domains within a module. As it turns out, there are ten amino acids that line the binding pocket of the NRPS that are known to determine substrate specificity. These amino acids are known as the “specificity conferring code.” However, although this code gives some direction as to what amino acids should be mutated to shift specificity, this approach does not capture the intricacies underlying amino acid recognition, and usually only small changes in substrate specificities could be handled without a reduction in $k_{cat}/K_m$.

**Figure 3.2.** Altering the specificity of NRPSs. New non-ribosomal peptide analogs can be made by directed evolution, computational design, or domain splicing.
Because insufficient information is known about NRPSs to effectively execute module swapping or rational redesign, it has been proposed that a more successful strategy to currently obtain functional NRPS mutants that incorporate novel functionalities is through directed evolution. With directed evolution, proteins are mutated randomly, and those mutants with new or improved properties are identified by screening or selection. Hence little has to be known about the protein structure/function itself. However, in order to increase the chance of finding a functional mutant, searching large libraries is preferred. Yet, no general high-throughput screen has been reported to determine A domain activity. Instead, the screen typically used is a medium-throughput radioactive ATP/pyrophosphate (PPi)-exchange assay. This assay utilizes the reversibility of the amino acid adenylation step of the reaction (Figure 3.3). Here, if radioactive pyrophosphate is added, the reaction will proceed in the ‘reverse’ direction and radioactive ATP will be formed and captured. However, aside from the requirement to quantify radioactive samples in a medium-throughput manner, this assay only works with purified protein, which means that only 1000 mutants per week can be screened. Clearly, a higher throughput method needs to be developed, which is the focus of this work.
Previously, others have shown that the NRPS, Tyrocidine Synthetase A (TycA), has the ability to act as a coenzyme A (CoA) ligase. Here, instead of transferring the aminoacyl-adenosine monophosphate (aminoacyl-AMP) to the Ppant group of the T domain, the aminoacyl-AMP reacts with CoA to form the aminoacyl-CoA. Because the major specificity of amino acid incorporation in NRPSs is determined by the ability of the A domain to produce the aminoacyl-AMP, thiol consumption could be monitored as a way to determine A domain activity. One of the most common ways to quantify the number of thiol groups in a sample is to use Ellman’s reagent. Here, if CoA is consumed by the NRPS, Ellman’s reagent will have less available thiol to react with, which will result in a decrease in absorbance at 412 nm. Using this assay, no radiation will be used, and in earlier work, our lab has shown that protein purification is not required.
Figure 3.4. **TycA catalysis and quantification** (A) TycA-catalyzed nucleophilic thioesterification of an amino acid by the Ppant group bound to the T-domain (B) If the Ppant group is not bound, TycA acts as a CoA ligase where the activated aminoacyl-AMP (AA-AMP) will react with the available thiols, CoA or SNAC, to form the aminoacyl-CoA. (C) The rate of CoA ligation can be quantified by Ellman’s reagent which absorbs at 412 nm in the presence of thiol.

In order to demonstrate the feasibility of this study, the TycA NRPS module will be used. TycA is an NRPS module that activates the first amino acid (L-phenylalanine) of a cyclic, decapeptide, antibiotic, tyrocidine A (Figure 3.5). Kinetic data for phenylalanyl CoA formation has been previously determined and could be used as a proof of principle to determine the sensitivity of the Ellman’s assay.
3.2 Results and Discussion

The *tycA* gene was obtained from *Bacillus brevis* genome via PCR. *TycA* gene was then cloned into pET-28a using the restriction enzymes listed, affording the plasmid pET28a/TycA, and expressed in *E. coli* BL21 (DE3) pLysS. Figure 3.6 shows the Ni-NTA chromatography purified protein product with the expected mass of about 125 kDa and over 90% purity.
Figure 3.6. SDS-PAGE gel image of purified TycA. Based off of the ladder (left) it was determined that the purified protein was between 150 kDa (top most dark band) and 100 kDa (lower lighter band) which is in line with the expected mass at about 125 kDa.

HPLC and LC/MS analysis was used to determine if the enzyme was active with both the endogenous substrate (CoA) and N-acetylcysteamine (SNAC). In Figure 3.7, two new peaks appeared for both substrates. The peak at 10 minutes elutes at the same time for both substrates, and hence must be present in both reactions. The other new peak eluted at different times for each substrate (13 minutes for CoA and 14 minutes for SNAC) and therefore must be a different product. It was believed that the first peak was the AMP product and the latter peak aminoacyl-CoA or aminoacyl-SNAC product.
Figure 3.7. RP-HPLC analysis of TycA reactions. (A) TycA reaction with CoA (B) TycA reaction with SNAC: i) t= 0 ii) t= 5 hours without thiol source iii) t= 5 hours with thiol source

All product formation was verified by ESI LC/MS (table 3.1). Although a different column was used, it was determined that all the reactants and products eluted in the same order as listed for the HPLC analysis.

Table 3.1. Mass spectrometry analysis of TycA reaction.

<table>
<thead>
<tr>
<th></th>
<th>Elution time (min)</th>
<th>Calculated +H (m/z)</th>
<th>Experimental +H (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>2.2</td>
<td>508.00</td>
<td>508.0025</td>
</tr>
<tr>
<td>AMP</td>
<td>4.7</td>
<td>348.07</td>
<td>348.0692</td>
</tr>
<tr>
<td>CoA</td>
<td>6.3</td>
<td>768.12</td>
<td>768.1226</td>
</tr>
<tr>
<td>L-phenylalanine</td>
<td>6.9</td>
<td>166.08</td>
<td>166.0788</td>
</tr>
<tr>
<td>phenylalanine-CoA</td>
<td>8.3</td>
<td>915.18</td>
<td>915.1906</td>
</tr>
<tr>
<td>phenylalanine-SNAC</td>
<td>8.8</td>
<td>267.11</td>
<td>267.1150</td>
</tr>
</tbody>
</table>

A product standard for the aminoacyl-CoA/SNAC product is not available, subsequent interpretation of the HPLC product peak area is difficult. However, using
the peak area for phenylalanine or CoA, the overall conversion from substrate is 10%. The kinetics of the reaction catalyzed by TycA had been previously published. The $k_{cat}$ for aminoacyl-CoA formation was determined to be 0.25 min$^{-1}$ with a $K_M$ of 41.9 µm. Therefore, because the concentration of phenylalanine was 1 mM, the reaction should be almost complete within 8 hours. This would mean that for a 1mM reaction, the reaction should be almost at completion within 8 hours (the time frame of our HPLC assay). However, the aminoacyl-CoA half-life has been estimated to be only 2 hours under certain conditions.$^{11}$ Thus, product hydrolysis may contribute to the rather small HPLC product peak areas. At the same time, colorimetric assay using Ellman’s reagent failed to detect a significant change in CoA/SNAC concentration during these assays.

3.3 Conclusions and Future Directions

Being able to alter the functional groups on NRPs can have many benefits which include engineering new pharmacologically relevant analogs or adding new synthetic handles. However, due to NRPs complex structure, this could be very difficult to accomplish using traditional organic synthesis. Because of this, research has turned to synthetic biology to alter the NRPS biosynthetic pathway, specifically the A-domain, to produce non-natural NRP analogs. Yet, no high throughput method exists to evolve NRPSs, hindering the development of novel NRPs. Here we reasoned that because the A-domain can also act as a CoA ligase, the CoA or thiol consumption could be quantitated using Ellman’s reagent. However, the ligation of
CoA to the amino acid is slow, and the quantification is further hindered by the short half-life of the product. Still, the reaction conditions of TycA and the Ellman’s assay could still be optimized. The maximum amount of product conversion taking into account the half-life is around 3 hours. The maximum amount of product conversion should be about 14%, which should be quantifiable. Also, by mutating the conserved serine on the T domain, which is normally post translationally modified with the ppant group, to an alanine, the catalytic turn over doubles. Increasing the turn over two-fold may be a large enough increase to help quantify thiol consumption.

3.4 Materials and Methods

Cloning:
TycA was cloned out of the *Bacillus brevis* genome using primers TycA-Nhe-1-Forward and TycA-EcoRI-Reverse. The primers incorporated the 5’ Nhe and the 3’ EcoRI restriction site. The linear TycA fragment along with the pET-28a vector were digested and ligated to form the TycA/pET-28a-His fusion.

Expression and Purification:
TycA/pET-28a-His was expressed in BL21 (DE3) pLysS cells. In a 2.5 liter flask, 1 L of 2 x YT medium containing kanamycin (50 µg/ml) was inoculated with in a 1:100 ratio from an overnight culture made from a fresh transformation. The cells were grown at 37 °C and 250rpm until the OD$_{600}$ was about 0.6. IPTG was added (0.5 mM final concentration) and the cells were allowed to express for 4 hours at 30 °C and 250 rpm. The cells were then pelleted (30 minutes, 4000 rpm, 4 °C) and
resuspended in binding buffer (0.5 M NaCl, 20 mM Tris-HCl, 5 mM imidazole, pH 7.8). Cells were then lysed by sonication and the soluble faction was obtained by centrifugation (1 hour, 18000 rpm, 4 °C). The protein was then purified by the General His-tagged protein purification method as previously described. The purified protein was concentrated and buffer exchanged with assay buffer using a 10000 MWCO centrifugal filter and stored -20 °C. Purity was verified by SDS-PAGE, and concentration was obtained using Bradford assay.

**In vitro TycA assay:**

In a 100 µl reaction, 100 µg of purified TycA was incubated in assay buffer (3 mM MgCl₂, 100 mM HEPES, pH 8.0) with 1 mM ATP, 1 mM thiol (CoA or SNAC), and 1 mM L-phenylalanine and incubated at 31 °C for varying time points. The reactions were then quenched with formic acid to a 1% final concentration. The reaction was centrifuged at 13,000 g for 10 minutes to remove all protein and the cleared supernatant was used for HPLC analysis.

**Ellman's assay:**

10 µl of the non-quenched TycA assay was added to 175 µl of assay buffer and 15 µl of 7.8 mM 5,5'-dithiobis(2-nitrobenzoic acid) in 0.1 M sodium phosphate pH 7. The reaction was incubated for 15 minutes at room temperature and then the absorbance was measured at 412 nm using BioTek Hybrid Synergy 4 plate reader (Winooski, VT, USA).
**HPLC analysis:**

A series of linear gradients was developed from 0.1% TFA in water (solution A) to 0.1% TFA in acetonitrile (HPLC grade, solution B) using the following method: 0-3 min isocratic flow 100% A, 3-25 min 80% B, 25-30 min isocratic flow 80% B, 30-33 min 100% A, 33-35 min isocratic flow 100% A. The flow rate was 1 ml/ min and the absorbance was measured at 210 nm.

**Mass spectrometry analysis:**

Samples were subjected to positive ESI LC/MS using the same gradients as listed using 0.1% formic acid in water (solution A) and 0.1% formic acid in acetonitrile (LC/MS grade, solution B).
3.5 References:


Appendix
## Appendix A. Primers

### Table A.1. Primer names and associated sequence

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ahcY assembly PCR</strong></td>
<td></td>
</tr>
<tr>
<td>ahcY-1</td>
<td>CGATTCTCTAGACCCTCCC CGCA CACC</td>
</tr>
<tr>
<td>ahcY-2</td>
<td>GAGCCCTGACAGGGTTAACCTCTGTC</td>
</tr>
<tr>
<td>ahcY-3</td>
<td>GGTTCACCTGTCAGGCTCAGGATGG</td>
</tr>
<tr>
<td>ahcY-4</td>
<td>CGTTGTCGGTTAAGGCTTGCCGGCT</td>
</tr>
<tr>
<td>ahcY-5</td>
<td>AAGCCTTAAACGCAACGGCGCCCA</td>
</tr>
<tr>
<td>ahcY-6</td>
<td>TACTACGAGATCCATGGAGAGTCTCC</td>
</tr>
<tr>
<td>ahcY-F</td>
<td>CGATTCTCTAGACCCTCCC</td>
</tr>
<tr>
<td>ahcY-R</td>
<td>TACTACGAGATCCATGGAGAGTCTCC</td>
</tr>
<tr>
<td><strong>pET-17b/ahcY/gfp</strong></td>
<td></td>
</tr>
<tr>
<td>ahcYhindIII-r</td>
<td>ATTATAAGCTTTGAGAGTCTCCAT</td>
</tr>
<tr>
<td>GFP HindIII</td>
<td>GCATAAGCATTGCAAAGGAAGAAGAC</td>
</tr>
<tr>
<td>GFP REV</td>
<td>GGCCTGATCTGTAGAGCTCTAC</td>
</tr>
<tr>
<td><strong>lysC genome cloning</strong></td>
<td></td>
</tr>
<tr>
<td>lysine switch-F</td>
<td>NNNNTCTAGAATGTACTACCTGC</td>
</tr>
<tr>
<td>lysine switch-R</td>
<td>NNNNAAAGCTTTGACTACTCCGTG</td>
</tr>
<tr>
<td>lysCdownstream</td>
<td>CATCAAAGCTTTACGCGGTACCGCC</td>
</tr>
<tr>
<td><strong>lysC translational construct</strong></td>
<td></td>
</tr>
<tr>
<td>lysC reg trans Insert-F</td>
<td>CATCAGCGGCAGCAGCATCATCATCAT</td>
</tr>
<tr>
<td>lysC reg trans Insert-R</td>
<td>CATCACCATGTTTGTAGAGCTCATC</td>
</tr>
<tr>
<td>lysC reg trans Vector-F</td>
<td>ATGACATGGTGAGATCCACTAGTA</td>
</tr>
<tr>
<td>lysC reg trans Vector-R</td>
<td>CATCAGCGGCAGCAGCTCATCATCAT</td>
</tr>
<tr>
<td><strong>thiM genome cloning</strong></td>
<td></td>
</tr>
<tr>
<td>ThiM Forward</td>
<td>TCATTCTAGAATGTGCCGTTTTCTCTCG</td>
</tr>
<tr>
<td>ThiM Reverse</td>
<td>TCATAAGCTTTGAGGCGAGGTCCA</td>
</tr>
<tr>
<td><strong>Linear IVTT PCR primers</strong></td>
<td></td>
</tr>
<tr>
<td>Forward biotin</td>
<td>TCAACGCAAGGACGACGATC</td>
</tr>
<tr>
<td>Reverse biotin</td>
<td>GGCCTTTCGCTCTCCAGAA</td>
</tr>
<tr>
<td>Forward biotin</td>
<td>5’/-5’BioTEG/ TCAACGCAAGGAGCAGCATC</td>
</tr>
<tr>
<td><strong>TycA cloning</strong></td>
<td></td>
</tr>
<tr>
<td>TycA-Nhe-1-Forward</td>
<td>ATTGCGCTAGCAGTGTAAATCGGCAATCTCA TCGA</td>
</tr>
<tr>
<td>TycA-EcoRI-REVERSE</td>
<td>ATTGCGAATTCTTACGCGCAGTG TATTTGCAAGCAAT TCGA</td>
</tr>
</tbody>
</table>
Appendix B. Constructs

Figure A.1. DNA constructs used to obtain a functional riboswitch