ABSTRACT

WALSH, STEVEN JOSEPH. Engineering the Specificity of Riboswitches and Enzymes. (Under the direction of Gavin Williams).

Riboswitches are ligand dependent, cis-acting gene regulators located in the untranslated regions of mRNA which are found mostly in prokaryotes. Riboswitches have been shown to control gene expression through ligand binding to a genetically encoded RNA aptamer. This binding causes a conformational change in the expression platform region that can either trigger an ON or OFF switch for the transcription or translation of the gene. With the use of a photocaged theophylline ligand, spatiotemporal control of a riboswitch was shown for the first time using UV light. The work described here also displays that riboswitches are not always modular parts, as they are often described.

Additionally, high-throughput screening techniques that allowed for the detection of mutant theophylline and lysine riboswitches that recognize new ligands were developed. A mutant theophylline riboswitch that recognized allopurinol with an activation ratio of 15, compared to 1 for the wild-type, was discovered. The LysC lysine riboswitch was also engineered to recognize unnatural ligands and even convert the switching behavior from an OFF to an ON switch. These results show that new riboswitches can be created via the introduction of a relatively small number of mutations into existing riboswitches rather than requiring de novo construction. This will greatly increase the rate of new riboswitch discovery. Furthermore, an ultra-high-throughput, dual selection strategy was developed that could greatly increase the rate of screening for new riboswitch activity.
The malonyl-Coenzyme A (malonyl-CoA) synthetase MatB has been used to synthesize acyl-CoA extender units for probing the specificity of polyketide synthases (PKS). MatB has previously been engineered to recognize a wide array of different malonate substrates. Here, that library is expanded to include halogenated substrates. These new halogen functionalities could enable new chemistries to be applied to polyketide semi-synthesis. MatB was also engineered to recognize N-acetylcysteamine (SNAc) in place of Coenzyme A (CoA). The use of SNAc will lower future screening costs and open the door for development of new handles that could be recognized by PKS modules. Through a round of random mutagenesis followed by several rounds of saturation mutagenesis, a 5-fold increase in MatB activity with SNAc was achieved.

Finally, work developing screening techniques for detection of free oil from both condensed distillers syrup and milled corn germ will be discussed. These techniques provide fast, accurate determination of free oil concentrations after enzyme reactions. Additionally, the amount of hexane extractable oil from milled corn germ under differing conditions was found. These results show that enzyme treatments greatly affect the total amount of hexane extractable oil.
Engineering the Specificity of Riboswitches and Enzymes

by

Steven Joseph Walsh

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chemistry

Raleigh, North Carolina

2015

APPROVED BY:

_______________________________  ______________________________
Gavin Williams      Reza Ghiladi
Committee Chair

_______________________________  ________________________________
Christian Melander     Tatyana Smirnova
DEDICATION

I would like to dedicate this to my loving wife Angela and our two puppies Cooper and Nola.

It was their support that helped me achieve my goals.
BIOGRAPHY

Steve was born and raised in Massachusetts. Steve then earned his bachelor’s degree in chemical engineering from The Cooper Union in New York City. Then Steve followed his wife to North Carolina where he studied chemistry at NC State.
ACKNOWLEDGMENTS

I would first like to acknowledge my advisor Dr. Gavin Williams. His guidance and support were critical for my development into the scientist I am today. I would also like to thank the rest of my committee Dr. Reza Ghiladi, Dr. Christian Melander, Dr. Tatyana Smirnova, Dr. Greg Upchurch, and former committee member Dr. Alex Deiters.

I would not have had the same success without the help and support of the rest of the Williams Lab. Past and present labmates were always willing to lend a helping hand or provide advice when needed.

Finally, it would not have been possible without my family. Angela was always there for support and Cooper and Nola could always bring a smile to my face no matter what kind of day I had. My parents provided the base I needed to complete my thesis; instilling the values of hard work and dedication I needed, and also encouraging my pursuit of science. The puzzles my Father always had helped me develop the problem-solving skills I have today.
# TABLE OF CONTENTS

LIST OF TABLES...........................................................................................................................................v

LIST OF FIGURES...............................................................................................................................................vii

CHAPTER 1 Introduction to riboswitches.........................................................................................................1

1.1 Photocontrol of riboswitches.......................................................................................................................1

1.2 Riboswitch screening techniques................................................................................................................3

CHAPTER 2 Spatiotemporal control and modularity of riboswitches.................................................................8

2.1 Photo-control strategy ....................................................................................................................................8

2.1.1 Versatility of the photo-controlled riboswitch as a gene control element ........................................11

2.1.2 Spatial control of riboswitch activity....................................................................................................14

2.2 Conclusions and future work .....................................................................................................................15

2.3 Methods .....................................................................................................................................................16

CHAPTER 3 High-throughput screening techniques for riboswitch engineering............................................23

3.1 Modification of riboswitch ligand specificity................................................................................................23

3.1.1 High-throughput screening methods for mutation of the theophylline riboswitch ................................23

3.1.2 Ultrahigh-throughput screening of the theophylline riboswitch.......................................................32

3.1.3 Theophylline riboswitch in mammalian cells.......................................................................................37
3.1.4 Modification of ligand specificity in the lysine riboswitch.............................. 39

3.2 Development of an ultrahigh-throughput riboswitch screening method.............. 49

3.3 Conclusions and future work ............................................................................... 55

3.4 Methods.................................................................................................................. 57

CHAPTER 4 Introduction to polyketide synthases....................................................... 64

4.1 Introduction............................................................................................................... 64

CHAPTER 5 Engineering MatB to accept new substrates........................................... 72

5.1 MatB engineering for SNAc and pantetheine ......................................................... 72

5.2 MatB engineering with unnatural malonates....................................................... 86

5.3 Conclusions and future work ............................................................................... 96

5.4 Methods.................................................................................................................. 97

CHAPTER 6 Determination of free oil volumes from corn ........................................ 104

6.1 Introduction............................................................................................................. 104

6.2 Extractions of oil from corn ............................................................................... 107

6.3 Assay development for detection of free oil ....................................................... 116

6.3.1 OxiRed assay development ............................................................................ 116

6.3.2 Oil red assay development ............................................................................. 124

6.4 Conclusions and Future work ............................................................................ 130

CHAPTER 7 Future directions .................................................................................... 132
7.1 Riboswitch technology ............................................................................................... 132

7.2 Polyketide technology ............................................................................................... 133

REFERENCES .................................................................................................................... 135
LIST OF TABLES

**Table 2-1**: Activation ratios of the 12.1 riboswitch with different reporters and promoters. 

[a] Average activation ratio, ± standard deviation (n=3). [b] Includes N-terminal 14 amino acids of LacZ. [c] +/- n indicates insertion/deletion n nucleotides between promoter sequence and RBS of the riboswitch. .................................................................................................................................14

**Table 2-2**: Sequences of oligonucleotides. .................................................................................................................................21

**Table 2-2**: Sequences of riboswitches used in this study. Shown highlighted in each DNA sequence is the riboswitch portion (underlined), the ATG start codon (grey box), and promoter (bold). The reporter gene sequence immediately follows the ATG in each case. For the LacZ-gfp fusion, the LacZ leader sequence is shown in italics. .......................................................................................22

**Table 3-3**: Activation ratio of potential ligands with WT theophylline riboswitch. An activation ratio of one indicates no activation is observed. ............................................................................................................................25

**Table 3-4**: Activation ratio of the most active mutant with each new ligand. Experiments were conducted in triplicate and standard deviations were less than 10% of the activation ratio. .................................................................................................................................29

**Table 3-5**: Activation data for wild-type, 1st- and 2nd-generation mutant riboswitches with 1 and 3. The ON and OFF values represent the Miller units for each case and the activation ratio is the ratio of these two values. Experiments were conducted in triplicate and standard deviations were less than 10% of the activation ratio. .................................................................................................................................31
Table 3-6: Activation ratio data for 3rd generation mutant riboswitches screened against \textbf{1}. The ON and OFF values represent the Miller units for each case and the activation ratio is the ratio of these two values. Experiments were conducted in triplicate and standard deviations were less than 10% of the activation ratio..............................32

Table 3-7: Summary of results from initial dual selection vector screen assays..................52

Table 3-8: Primer sequences used for cloning of plasmid constructs.............................63

Table 5-9: Steady state kinetic parameters for WT and mutant MatB. ND signifies that the value could not be determined.................................................................91

Table 5-10: Mass spectrometry confirmation of the identity of MatB products...............91

Table 5-11: Mass spectrometry results for malonyl-CoAs used in single module polyketide reactions with diketide-SNAC. ND denotes that no product was detected.........................93

Table 5-12: Mass spectrometry results for Gcs reactions with different malonyl-CoA substrates. With \textbf{21} and \textbf{22} substrate reactions the hydrolysed product was observed at low levels. Reactions were run with enzymes in cell lysate.................................95

Table 5-5: Forward and reverse primer sequences for mutagenesis at each residue in MatB.........................................................................................................................99

Table 6-1: Reaction conditions for one-step, five enzyme OxiRed reaction. The fluorescence was normalized to the fluorescence from the four step reaction. The most active set of conditions is shown in red.................................................................120
LIST OF FIGURES

Figure 1-1: General strategy for riboswitch photo-control. A) Structures of theophylline (1) and the photocaged analogue 2. B) Schematic representation of the theophylline riboswitch 12.1. In the absence of UV irradiation, the caged ligand is unable to bind to the riboswitch aptamer domain, and gene expression is turned OFF. Upon UV irradiation, the ligand is now recognized by the riboswitch aptamer domain, and gene expression is turned ON. RBS, ribosome binding site. ........................................................................................................................................... 2

Figure 1-2: Overview of SELEX. Short pieces of RNA that tightly bind to ligand attached to a scaffold can be identified using this selection method. ................................................................................................................................. 5

Figure 1-3: Structures of adenine and azacytosine ............................................................................................................................................................... 7

Figure 2-1: In vivo light-activation of the theophylline riboswitch. A) Activity of the 12.1 riboswitch in a series of laboratory E. coli strains. Left axis (grey bars) represents the activation ratio of the riboswitch as the ratio of activities in the absence and presence of 1 (1 mM). Right axis represents the β-galactosidase activity (Miller units) of the riboswitch in the absence (closed circle) or presence (open circle) of 1 (1 mM). Dashed lines indicate standard deviation of activity (n=3). B) Dose dependent LacZ expression under control of the 12.1 riboswitch in the presence of 1. Dashed lines indicate standard deviation of activity (n=3). Axis as described for panel A. C) In vivo light-activation of 12.1-driven LacZ expression in liquid media. LacZ is only expressed in the presence of UV irradiation and 2 (1 mM in each case). Activation ratio is given above each grey bar. Error bars indicate standard deviation (n=3). ...............................................................11
**Figure 2-2:** In vivo spatial control of riboswitch activity. A mask was used to irradiate selected wells of a microplate with UV light. Each well contained an identical culture of E. coli TOP10 cells harboring the plasmid pSAL/12.1. The photocaged ligand 2 was included in all wells at 1 mM and a colorimetric LacZ assay carried out using X-gal. See Experimental Section for details.

**Figure 3-1:** Structure of the theophylline riboswitch. The left shows the structure in the no ligand OFF state and the right shows the riboswitch structure in the 1 bound ON state. The start codon is shown in red (58-60), Shine-Dalgarno sequence in green (43-47), and the aptamer in blue (25-29).

**Figure 3-2:** Panel of potential ligands chosen for the mutant theophylline riboswitch. 1 is shown for comparison.

**Figure 3-3:** Image of a plate used for blue/white screening. Colonies expressing LacZ appear blue while colonies not expressing the enzyme remain colorless.

**Figure 3-4:** Scheme of riboswitch screening using LacZ. The mutated riboswitches were spread on plates containing X-gal and no ligand. The colorless colonies show no LacZ expression so they were pooled for the next round of screening. Now the mutant riboswitches were spread on plates containing a non-native ligand and X-gal. This time the blue colonies were selected as they are expressing LacZ. These blue colonies were now either counterscreened on plates containing 1 and X-gal where the colorless colonies were selected, or they were secondary screened using the Miller assay.
Figure 3-5: FACS data for droplets containing the theophylline riboswitch. Droplets were first gated by GPF fluorescence (right) to select only those containing a cell. Cells with higher fluorescence (left) indicate higher levels of LacZ fluorescence. ......................................................... 34

Figure 3-6: FACS data for E. coli cells containing the theophylline riboswitch. Cells found within the M1 region show high concentrations of hydrolysed FUG, indicative of high LacZ production. ................................................................................................................................. 36

Figure 3-7: Microscope images of the mammalian 12.1DsRed construct co-transfected with GFP. The cells were incubated with varying concentrations of 1. ......................................................... 38

Figure 3-8: 2-D structure of the LysC riboswitch. The LysC riboswitch recognizes the amino acid lysine and functions as an ON switch. Highly conserved residues are shown boxed.... 40

Figure 3-9: Crystal structure of the LysC riboswitch. 57 ........................................................ 41

Figure 3-10: Binding interactions present in the LysC riboswitch. ............................................... 42

Figure 3-11: Plasmid map of the pLys vector. ....................................................................... 42

Figure 3-12: Scheme showing the initial screening strategy for LysC riboswitch mutants. First the counterscreen against the natural substrate, lysine is performed. Now that the mutants no longer recognize lysine, the next rounds of screening can be performed in LB media, significantly decreasing screening time. Then, repression of GFP expression screened for by selecting the non-fluorescent colonies when unnatural ligand is present. Finally, colonies expressing high GFP fluorescence in the absence on ligand are found. Additional secondary screening or mutagenesis is then conducted. ................................................................. 44

Figure 3-13: Example of an agar plate used for fluorescence screening .............................. 45
Figure 3-14: Lysine, 8, and analogues selected for initial screening of the mutant lysine riboswitch................................................................. 46

Figure 3-15: Activity with mutant lysine riboswitches with natural and unnatural ligand. Denotes that fluorescence with 8 was too low to display clearly. ................................. 48

Figure 3-16: Dual-selection strategy for mutant riboswitches. In the absence of ligand and IPTG, those riboswitches expressing the barnase-gfp fusion will die due to the barnase activity, resulting in a negative selection. Ligand and IPTG would then be added to the surviving cells. The IPTG would induce expression of barstar which would inhibit barnase, which was upregulated by the ligand. Levels of gene expression from the riboswitch could then be monitored by GFP fluorescence......................................................... 51

Figure 3-17: Modified dual selection platform using a two plasmid system. The negative selection remains the same but the positive selection now uses LacZ as the reporter......... 54

Figure 3-18: Droplet-based technique for selecting for cell-impermeable ligands. By subjecting cells enclosed in a droplet to a freeze-thaw cycle, lysing the cells, it may be possible to develop riboswitches for a wider array of ligands........................................ 56

Figure 4-1: Examples of clinically relevant polyketides......................................................... 65

Figure 4-2: 6-deoxyerthronolide B synthase (DEBS), produces the precursor to erythromycin A. This is an example of a type I PKSs and also illustrates the colinearity between a PKS’s protein sequence and the structure of the final product....................... 66

Figure 4-3: Scheme illustrating AT swapping using part of Mod2 as an example. By interchanging AT domains the substrate specificity could be altered. Here a malonyl-CoA is installed instead of the natural methylmalonyl-CoA. ................................................................. 67
**Figure 4-4:** Generic example of a type II PKS. .......................................................... 68

**Figure 4-5:** (Top) Native reaction of MatB. The ATP-dependent MatB catalyzed reaction of the malonate and thiol 13 forms a malonyl-CoA. (Bottom) Proposed reactions obtainable with mutant MatBs. Additional malonate analogues are accepted as well as incorporation of SNAc (12) and pantetheine (14) in place of CoA (13). .......................................................... 69

**Figure 4-6:** Malonate analogues accepted by mutant MatB enzymes. Additionally, MatB mutants that accept other hydrocarbon chains were found. .................................................. 70

**Figure 5-1:** MatB completely converts 13 to product within 5 minutes. Absorbance values were taken at 254 nm. ............................................................................................................. 73

**Figure 5-2:** MatB reactions with 12 fail to reach completion even after 3 hours. Absorbance data was collected at 235nm. .................................................................................................. 73

**Figure 5-3:** Scheme of the high-throughput screening method for detection of reaction of mutant MatB enzymes with 12 and 14. ................................................................. 75

**Figure 5-4:** Reaction of Ellman's reagent with a thiol. The amount of thiol can be monitored by reading absorbance at 412 nm. ............................................................................. 75

**Figure 5-5:** Optimized reaction conditions allow room for detection of MatB mutants with improved performance with 12. Colonies were ordered by descending order of absorbance and used to find the expected averages and standard deviations of WT MatB and negative control absorbance values from the screening method. The MatB reaction with CoA was used to indicate the minimal expected absorbance. ............................................................................. 76

**Figure 5-6:** MatB residues chosen for mutagenesis and screening with 12 and 14. The CoA arm is shown in grey. ............................................................................................... 77
**Figure 5-7:** Substrate binding pocket of MatB showing interactions with methylmalonyl-CoA.81 ...................................................................................................................................... 77

**Figure 5-8:** Flow chart illustrating secondary screening procedure to identify mutant MatBs with improved activity. ........................................................................................................................................................................ 80

**Figure 5-9:** Relative rate of reaction increase of mutant MatB protein with 12 ............... 82

**Figure 5-10:** Relative rate of reaction increase of mutant MatB protein with 12. .............. 83

**Figure 5-11:** Tree showing the order of mutations made to MatB that increased rate of reaction with 12 ........................................................................................................................................................................ 84

**Figure 5-12:** Locations of mutations yielding the highest increase in MatB activity with 12. Note that the locations are largely on the exterior of the enzyme and not in close proximity of the binding pocket (13 arm shown in blue). ........................................................................................................................................................................ 85

**Figure 5-13:** Activity of WT MatB with the halogenated malonates. Absorbance values show the results of an Ellman’s assay run either in the presence of or absence of WT MatB enzyme from lysed cells. Error bars show the standard deviation of eight experiments. ................. 87

**Figure 5-14:** Activity of T207X/M306X MatB mutants with halogenated malonates. Absorbance data shows the results of an Ellman's assay used to determine the concentration of 13 remaining after the reaction. Activity with WT MatB with each substrate is shown for comparison........................................................................................................................................................................ 88

**Figure 5-15:** HPLC analysis of the reaction of purified MatB enzyme with unnatural malonates. ........................................................................................................................................................................ 89

**Figure 5-16:** Scheme of the single module reaction of 19-22 with diketide-SNAc. ............ 92
Figure 5-17: Schematic of the Ery6 reaction with pentaketide and malonyl-CoA substrate to form the 10-dml product. ........................................................................................................ 93

Figure 5-18: Germicidine biosynthetic pathway................................................................. 95

Figure 6-1: Structure of the corn kernel. The germ, located at the center of the corn kernel, contains the majority of the oil enclosed within the kernel. The germ is 30-40% oil by weight. ............................................................................................................................................... 105

Figure 6-2: Mimic of the process of extracting oil from corn found in production plants. A washing step was added to the wet milling mimic as this was found to decrease background in further experiments. .......................................................................................................... 108

Figure 6-3: Hexane extractions of corn germ. Whole germ was milled through a 2 mm filter. Washing the germ with 15% ethanol mimics fermentation conditions.............................. 108

Figure 6-4: Amount of oil extracted from CDS by hexanes. Washing the germ with 15% ethanol mimics fermentation conditions................................................................. 109

Figure 6-5: Increase in extractable oil from milled, washed germ after enzyme treatment. Percentages are v/v............................................................... 110

Figure 6-6: Total extractable oil from whole, washed germ. Enzs 2 and 3 are components of enz 1. Percentages are v/v. ............................................................... 111

Figure 6-7: Picture of corn germ after enzyme incubation. The addition of enzs 1, 2, and 3 all break down germ structure and form a thicker emulsion layer......................... 112
Figure 6-8: Oil partitioning in corn germ after treatment with enzyme. The bound portion contains the solids and aqueous phase, the emulsion partition is comprised of the emulsion layer and the free oil partition is the oil that rises to the top when centrifuged. Percentages are v/v. ........................................................................................................................................................ 113

Figure 6-9: SEM images of corn germ surface after treatment with enzyme. .................. 115

Figure 6-10: Oil partitioning in CDS after treatment with enzyme. The bound portion contains the solids and aqueous phase, the emulsion partition is comprised of the emulsion layer and the free oil partition is the oil that rises to the top when centrifuged. Percentage is in v/v. ........................................................................................................................................ 116

Figure 6-11: Scheme of the OxiRed enzyme reaction pathway........................................ 118

Figure 6-12: The OxiRed assay shows a linear trend of fluorescence versus volume of corn oil. .......................................................................................................................................................... 119

Figure 6-13: Background fluorescence compared to experimental fluorescence for the OxiRed assay. Note that the germ treated with GK, GPO, and HRP would only account for the fluorescence due to the glycerol present. With no lipase, the triglycerides would not yield additional glycerol. .................................................................................................................................................. 121

Figure 6-14: Fluorescence from the emulsion and bound layers in the OxiRed assay. ...... 122

Figure 6-15: Validation of the OxiRed assay. Hexane extraction data was compared to OxiRed fluorescence for a 14 different enzymes. ........................................................................................................ 123

Figure 6-16: Structure of the oil red dye.............................................................................. 124
Figure 6-17: Scheme of the oil red assay. (Top) General protocol for the assay. (Bottom) expected results from the assay. A larger oil volume (B) will dilute the dye leading to a lower absorbance.

Figure 6-18: A linear trend is observed between absorbance and percentage oil red in a sample.

Figure 6-19: Relationship between oil red signal and total amount of oil added in CDS...

Figure 6-20: Oil red results after treatment of CDS with enzyme. The decrease in absorbance shows an increase in free oil. Error bars show the standard deviation of three experiments.

Figure 6-21: Validation of the oil red assay. Results from the assay were compared to results from hexane extractions.
CHAPTER 1

Introduction to riboswitches

1.1 Photocontrol of riboswitches

Riboswitches are a class of naturally occurring genetic regulatory devices, found in all three domains of life and widely distributed throughout bacteria. Riboswitches respond to a range of metabolites, including protein cofactors, amino acids, nucleobases, metal ions, and some natural products, often with exquisite specificity.\(^1,2\) Most riboswitches are located in the untranslated regions of metabolic genes and are usually involved in regulation of the associated pathway.\(^3\) In general, riboswitches consist of an aptamer domain, required for ligand recognition, and an expression platform. Ligand binding results in a conformational change in the expression platform that subsequently either switches gene expression from ON to OFF or from OFF to ON. Naturally occurring riboswitches control gene expression via transcriptional control, translational control, or by acting as cis-acting ribozymes.\(^4\) To date, only the \(glmS\) ribozyme is known to operate via the latter mechanism,\(^5\) although riboswitches that affect splicing in fungi, plants, and algae, are also known.\(^6,7,8,9\)
Given the remarkably small size of naturally occurring riboswitches, which do not require protein components for function, and the ability to easily insert them into untranslated regions of genes, there has been much interest in applying riboswitches to the conditional control of gene expression and to the sensing of small molecules. For example, riboswitches could be used in synthetic biology as parts of artificial genetic circuits for controlling cellular behaviour, and could be used to probe, interrogate, and manipulate biological processes in vivo for a variety of chemical biology applications.
In order to expand the applications of riboswitches, synthetic riboswitches with tailored ligand specificities and output functions have been engineered. However, the ability to control riboswitch activity inside cells in a spatial and temporal fashion remains severely limited. Light is an external input signal that can be used to control a broad array of biological processes with high spatio-temporal resolution, complete bioorthogonality, and simple equipment. Accordingly, light is a potentially powerful input that in combination with small molecule ligands could be used to control the activity of natural or engineered riboswitches in vivo with minimal invasion. Typically, photocaging groups are used to render small molecule ligands or biological macromolecules photosensitive. Upon irradiation with light, the caging group is removed, thus revealing the active small molecule or macromolecule and activating its function. Notably, while several ribozymes have been controlled using photocaging technologies, there are no reports of using light to control the activity of other types of riboswitches, such as those that operate at the transcriptional or translational level.

1.2 Riboswitch screening techniques

Riboswitches have the potential to act as antibiotics, chemosensors, regulators in metabolic pathways, controllers of cellular behavior, and logic gates. However, the discovery and development of new riboswitches remains slow, which currently limits the scope and utility of riboswitch-based technology.

When designing new riboswitches *de novo* they are often broken into their two major components, the aptamer region and the expression platform. The aptamer is responsible for the binding of small-molecule ligands while the expression platform, acting in response to
ligand binding, changes its conformation to regulate the gene expression. Separating the two components in early rounds of screening has made the design of each less cumbersome.\textsuperscript{33,34}

The systematic evolution of ligands by exponential enrichment (SELEX) is a process widely used for discovery of RNA aptamers (Figure 1-2).\textsuperscript{35,36,37,38} SELEX begins with a randomized pool of DNA that is transcribed into short pieces of RNA. The RNA is then incubated with ligand bound to a solid support for a predefined amount of time then washed until only the RNA bound to the column remains. The addition of free ligand elutes these RNAs which are then amplified and reverse transcribed back to DNA. The DNA can then be mutated again or transcribed and run through the column again with a higher selection pressure (usually shorter incubation time). Rounds of RNA binding that contain only the binding scaffold and no ligand are also run as a counter selection. In these rounds only the RNA that elutes in the washes is used for further rounds.\textsuperscript{39}
Figure 1-2: Overview of SELEX. Short pieces of RNA that tightly bind to ligand attached to a scaffold can be identified using this selection method.

SELEX can be used to select RNAs that bind the desired ligand out of very large pools ($10^{15}$). The throughput of SELEX allows for RNA sequences that tightly bind the aptamer, but further screening is still needed in order to form a functional riboswitch. Also, since SELEX is done in vitro, it fails to account for any off-target RNA binding of other molecules in the cell. While counter selections can be performed for some molecules, it would be nearly impossible to screen against all of the molecules found in intact cells. It has also been noted that riboswitches developed in vitro do not always function in vivo.

After finding an aptamer, either through SELEX or mutation of an existing riboswitch, a secondary screen is still required to test for riboswitch activity. An early dual-selection platform was described by the Yokobayashi group. In this platform the group took
advantage of the tetracycline resistance gene, which, when expressed allows cells to survive in the presence of tetracycline, but also makes them susceptible to Ni^{2+}. Therefore, selections for both the ON and OFF states are possible using this system. In this way, a riboswitch with an impressive activation ratio of 58 was found. The advantages of this method are that the dual selection is ultra-high-throughput and it is straightforward to modulate the selection pressure. A drawback of this method is that it only works with membrane permeable ligands, severely limiting the development of riboswitches for many metabolic pathways. This method is also limited to prokaryotic cells. Several other screens or selections have also been implemented for screening riboswitch activity.\textsuperscript{15,26,27} Many of these screens are performed \textit{in vivo} as they are often higher throughput and reduce challenges of transferring results from \textit{in vitro} to \textit{in vivo} later on.

The \textit{de novo} design of riboswitches can be labor intensive and slow. However, if riboswitches with specificity for a new ligand could be engineered from existing riboswitches, the need for SELEX and engineering of the expression platform would not be necessary, greatly reducing the time and labor required. As an example, by only saturating two nucleotide positions in the adenine responsive add A-riboswitch aptamer domain, the Micklefield lab successfully engineered add A to instead recognize azacytosine (Figure 1-3).\textsuperscript{45} Because the library consisted of only 15 mutants there was no need to develop a high-throughput screen. This study illustrates that mutation of existing riboswitches can be used to produce riboswitches with new specificities.
Current riboswitch screens and selections have at least one of three major downfalls: they are not generally applicable, they are low-throughput, or they require additional screening techniques and cloning after the initial screen. Designing a screening method for each individual riboswitch is slow and severely limits the potential for creating new riboswitches from existing ones. The same could be said for low throughput screening methods. While high-throughput methods like a tetracycline based screen would be useful for primary screening mutant riboswitches, it would still require modifications for additional screening and is limited to prokaryotes. By developing a generally applicable, high-throughput screening method riboswitches could be engineered in a faster, less labor intensive method. This could allow riboswitch incorporation into a wider array of biological applications. The goals of Chapter 3 are to show that mutation of existing riboswitches for altered ligand specificity is feasible with additional riboswitches and to develop a high-throughput screening method that could be applied to a wide range of riboswitches.
CHAPTER 2

Spatiotemporal control and modularity of riboswitches

This chapter is adapted from Walsh et al, ChemBioChem, 2014, 15, 1346-1351.

2.1 Photo-control strategy

Here, a photocaged analogue of a riboswitch ligand was used to afford spatial and temporal control of gene expression. Because the caged ligand (2) is cell permeable, non-toxic at active concentrations, and completely orthogonal to the host organism, this approach affords a convenient strategy to control gene expression in vivo. Furthermore, it was hypothesized that the simplicity and potential adaptability of this strategy might lead to the development of general tools for spatial and temporal control of gene expression in a wide variety of organisms.

An engineered riboswitch designed to respond to theophylline (1, Figure 1-1A), designated ’12.1’, was chosen as the prototype for this study because the switch is predicted to operate at the translational level via a simple RBS sequestration mechanism (Figure 1-1B). Accordingly, it was reasoned that the switch could potentially be utilized for the photo-control of biological processes in a wide variety of bacteria. The solution structure of the parent aptamer used to create the 12.1 riboswitch indicates that a uracil residue (U24) in the ligand-binding site of the aptamer is hydrogen bonded to N9 of 1. Disruption of this intermolecular bond likely destabilizes a set of stacking interactions that constitute the core of the aptamer structure, and could explain the remarkable discrimination that the aptamer displays between closely related small molecules. Accordingly, it was reasoned that a nitrobenzyl photocaging moiety located at N7 of 1 would provide an analogue (2, Figure 1-
1A)\textsuperscript{24} that would not be recognized by the aptamer portion of 12.1 and would therefore fail to
turn on gene expression. Conversely, irradiation should remove the caging moiety, revealing
the active ligand and switching on gene expression (Figure 1-1B). The synthetic riboswitch is
housed in the plasmid pSAL, which includes the IS10 promoter and terminator sequences, in
addition to a β-galactosidase reporter gene, lacZ.\textsuperscript{26} Determination of the activation ratio of
12.1 in response to 1 using a modified Miller assay in liquid cultures of \textit{E. coli} TOP10
verified the expected high activation response of this synthetic switch (Figure 2-1A). Indeed,
galactosidase activity of the theophylline-activated riboswitch was similar to that obtained by
constitutive LacZ expression from a control plasmid that lacked the riboswitch. Interestingly,
the activation ratio of 12.1 in response to 1 determined using other common laboratory
strains of \textit{E. coli} proved less impressive (Figure 2-1A), largely due to reductions in the ON
activity (presence of ligand), but nonetheless represented some of the most robust
riboswitches characterized to date. By varying the concentration of 1, the dose-dependency
of the 12.1 riboswitch was evaluated. As expected, the induction of LacZ expression
increases in a dose-dependent manner (Figure 2-1B) in response to increasing concentrations
of 1. Next, we investigated the ability of the caged theophylline 2 (Figure 1-1A) to induce
gene expression under the control of the 12.1 riboswitch. In the absence of 2, a very low level
of β-galactosidase expression is detected in the presence or absence of UV irradiation (Figure
2-1C), consistent with the low OFF activity of 12.1 – indicating that UV light alone does not
have an effect on reporter gene expression. Similarly, in the presence of 2 and the absence of
UV irradiation, essentially no β-galactosidase activity is detected, demonstrating that the
caged ligand is unable to induce a conformational change that exposes the 12.1 RBS to allow
gene expression. Notably, in the presence of 2 and UV irradiation, LacZ expression is activated to levels similar to that of natural theophylline (Figures 2-1A/C). Gratifyingly, 12.1-controlled LacZ expression in the presence of UV-irradiation and 2 was also linearly dependent on the concentrations of 2 (Figure 2-1D). Cumulatively, these data confirm the expected UV dependency of the 12.1 riboswitch in response to the photocaged ligand and demonstrates for the first time the ability to photo-regulate the activity of a simple riboswitch in vivo.
Figure 2-1: In vivo light-activation of the theophylline riboswitch. A) Activity of the 12.1 riboswitch in a series of laboratory E. coli strains. Left axis (grey bars) represents the activation ratio of the riboswitch as the ratio of activities in the absence and presence of 1 (1 mM). Right axis represents the β-galactosidase activity (Miller units) of the riboswitch in the absence (closed circle) or presence (open circle) of 1 (1 mM). Dashed lines indicate standard deviation of activity (n=3). B) Dose dependent LacZ expression under control of the 12.1 riboswitch in the presence of 1. Dashed lines indicate standard deviation of activity (n=3). Axis as described for panel A. C) In vivo light-activation of 12.1-driven LacZ expression in liquid media. LacZ is only expressed in the presence of UV irradiation and 2 (1 mM in each case). Activation ratio is given above each grey bar. Error bars indicate standard deviation (n=3). D) Dose dependent LacZ expression under control of the 12.1 riboswitch in the presence of 2 and UV irradiation. Activation ratio is given above each grey bar. Error bars indicate standard deviation (n=3).

2.1.1 Versatility of the photo-controlled riboswitch as a gene control element

Several riboswitches have been described as modular with regard to the reporter protein, and in at least some examples, with respect to the promoter system. To determine
whether 2 in combination with the 12.1 riboswitch could be used to photo-control the
expression of other reporter proteins, the lacZ gene was substituted with the gene encoding
GFP, yielding the plasmid pSAL/gfp. Subsequent quantification of GFP expression by
fluorescence measurements in the absence or presence of 1 revealed an activation ratio of
only four, an unexpected 20-fold lower than that with LacZ as the reporter (Table 2-1, entries
1 and 2). Similarly, when DsRed was used as the reporter gene (pSAL/DsRed), the activation
ratio was low (Table 2-1, entry 3). We hypothesized that the loss in activation efficiency
when lacZ is replaced with another reporter gene is a result of interactions between the lacZ
encoding sequence and the 12.1 riboswitch in the corresponding transcript. To test this
hypothesis, we fused the first 42 nucleotides of the lacZ gene sequence directly upstream of
the gfp gene in the plasmid pSAL/gfp, to afford the construct pSAL/lacZ-gfp. Subsequent
determination of the activity ratio of this lacZ-gfp fusion indicated it was unable to rescue
activation of the 12.1 riboswitch to efficiencies as high as that with full-length lacZ (Table 2-
1, compare entry 1 and 4), even though a purified LacZ-GFP fusion protein retained 95% the
fluorescent output of GFP alone (data not shown).

In a further effort to overcome the apparent dependency of gene reporter and
activation ratio, the relatively weak IS10 promoter of each of the pSAL series of vectors was
replaced with the stronger T7 promoter, while the IS10 terminator was substituted with the
T7 terminator sequence. The resulting lacZ, gfp, and DsRed constructs (pET17b/T7/lacZ,
pET17b/T7/gfp, and pET17b/T7/DsRed, respectively) were used to determine theophylline
activation ratios in E. coli BL21(DE3). However, none of these T7-based constructs
delivered activation ratios greater than five (Table 2-1, entries 5-7) and were 10- to 50-fold
lower than pSAL-based activation in the same strain (compare to Fig 2A). The lacZ-gfp
fusion also led to a poor activation ratio (Table 2-1, entry 8). Increasing or decreasing the
number of nucleotides between the promoter sequence and RBS failed to improve the
activation ratio when either GFP and IS10 were used as the reporter and promoter,
respectively (Table 2-1, entries 10-11). Similarly, sequence insertions or deletions failed to
improve activation ratios when combinations of lacZ/GFP and T7 were used as the reporter
and promoter, respectively (Table 2-1, entries 12-14).

In summary, the activation ratio of the 12.1 riboswitch was found to be dependent on
both the reporter gene and promoter sequence identity. Intriguingly, Mfold calculations of the
12.1 transcript do not support structural interaction between the promoter and 12.1 riboswitch
sequence (data not shown). This data suggest that a more complex mechanism might be
responsible for the poor activation ratio of the gfp- and/or T7-based constructs. Perhaps the in
vivo concentration of RNA transcript affects the riboswitch behaviour, a mechanism that
could be highly dependent on the vector copy number, promoter strength, and reporter
sequence identity. Some dependency between riboswitch activation ratio and promoter
identity has been reported previously for a theophylline riboswitch, although activation
ratios differed only ~10-fold with three different promoters in that study. In the context of
parts for synthetic biology applications, riboswitches possess a certain degree of portability
in terms of combining them with different promoters and gene reporters, however, they may
require at least some optimization for specific applications for which they were not originally
designed.
Table 2-13: Activation ratios of the 12.1 riboswitch with different reporters and promoters. [a] Average activation ratio, ± standard deviation (n=3). [b] Includes N-terminal 14 amino acids of LacZ. [c] +/- n indicates insertion/deletion n nucleotides between promoter sequence and RBS of the riboswitch.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Construct</th>
<th>Reporter/Promoter</th>
<th>E. coli host strain</th>
<th>Activation ratio [a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>pSAL/12.1</td>
<td>LacZ/IS10</td>
<td>TOP10</td>
<td>275.7 ± 26</td>
</tr>
<tr>
<td>2</td>
<td>pSAL/gfp</td>
<td>GFP/IS10</td>
<td>TOP10</td>
<td>3.6 ± 0.5</td>
</tr>
<tr>
<td>3</td>
<td>pSAL/DsRed</td>
<td>DsRed/IS10</td>
<td>TOP10</td>
<td>6.6 ± 0.8</td>
</tr>
<tr>
<td>4</td>
<td>pSAL/lacZ-gfp</td>
<td>GFP [b]IS10</td>
<td>TOP10</td>
<td>5.9 ± 0.6</td>
</tr>
<tr>
<td>5</td>
<td>pET17b/T7/lacZ</td>
<td>LacZ/T7</td>
<td>BL21 (DE3)</td>
<td>4.8 ± 0.5</td>
</tr>
<tr>
<td>6</td>
<td>pET17b/T7/gfp</td>
<td>GFP/T7</td>
<td>BL21 (DE3)</td>
<td>1.8 ± 0.3</td>
</tr>
<tr>
<td>7</td>
<td>pET17b/T7/DsRed</td>
<td>DsRed/T7</td>
<td>BL21 (DE3)</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td>8</td>
<td>pET17b/T7/lacZ-gfp</td>
<td>GFP [b]T7</td>
<td>BL21 (DE3)</td>
<td>1.8 ± 0.3</td>
</tr>
<tr>
<td>9</td>
<td>pSAL/tac/gfp</td>
<td>GFP/Tac</td>
<td>BL21 (DE3)</td>
<td>4.2 ± 0.6</td>
</tr>
<tr>
<td>10</td>
<td>pSAL/gfp+2[c]</td>
<td>GFP/IS10</td>
<td>TOP10</td>
<td>3.2 ± 0.7</td>
</tr>
<tr>
<td>11</td>
<td>pSAL/gfp-2[c]</td>
<td>GFP/IS10</td>
<td>TOP10</td>
<td>2.9 ± 0.6</td>
</tr>
<tr>
<td>12</td>
<td>pET17b/T7/lacZ+3[c]</td>
<td>LacZ/T7</td>
<td>BL21 (DE3)</td>
<td>4.2 ± 0.5</td>
</tr>
<tr>
<td>13</td>
<td>pET17b/T7/lacZ+1[c]</td>
<td>LacZ/T7</td>
<td>BL21 (DE3)</td>
<td>4.1 ± 0.7</td>
</tr>
<tr>
<td>14</td>
<td>pET17b/T7/gfp-3[c]</td>
<td>GFP/T7</td>
<td>BL21 (DE3)</td>
<td>1.8 ± 0.6</td>
</tr>
</tbody>
</table>

2.1.2 Spatial control of riboswitch activity

Based on the successful light-activation of riboswitch activity using bulk in vivo reactions, the effectiveness of using light to control riboswitch activity in a spatial fashion was investigated using liquid cultures. E. coli TOP10 cells harboring the plasmid pSAL were grown in each well of a microplate in the presence of 2. Selected wells of the 96-well microplate were irradiated with UV light using a mask. LacZ expression was only observed in the wells that were exposed to UV light, as judged by the distinct coloration of those wells.
as a result of the chromogenic galactosidase substrate X-gal (Figure 2-2). These data clearly demonstrates the ability to control riboswitch activity in a spatial and temporal fashion.

**Figure 2-2:** *In vivo spatial control of riboswitch activity. A mask was used to irradiate selected wells of a microplate with UV light. Each well contained an identical culture of E. coli TOP10 cells harboring the plasmid pSAL/12.1. The photocaged ligand 2 was included in all wells at 1 mM and a colorimetric LacZ assay carried out using X-gal. See Experimental Section for details.*

### 2.2 Conclusions and future work

The ability to dissect and manipulate biological processes in a spatial and temporal manner is greatly facilitated by the use of light. Here, we have demonstrated the ability to photo-control gene expression by combining a translational-controlled riboswitch with a light-activated version of its ligand. This strategy should prove a general approach for photo-controlling gene expression in vivo given (1) the breadth of ligands recognized by naturally occurring riboswitches,² (2) the emergence of efficient strategies for creating synthetic riboswitches,¹ and (3) the ease of installing photocaging moieties into diverse small
molecules. Moreover, riboswitches have been engineered to function in several Gram-negative and Gram-positive bacterial hosts, as well as plants, thus our photocaging strategy could potentially be applied in a broad range of organisms. Accordingly, we expect this photo-control strategy will be widely adopted for controlling gene expression, and will provide a valuable input for artificial genetic circuits.

2.3 Methods

**General:** All plasmids were verified by DNA sequencing. Purifications of all DNA were performed with kits from BioBasic. All reagents, including 1, o-nitrophenyl-β-D-galactopyranoside (ONPG), and ampicillin were purchased from Sigma. X-gal was purchased from MP Biomedicals. Synthetic oligonucleotides were purchased from IDT. The caged analog 2 was synthesized according to published protocols. All plate reader assays were performed on a BioTek Hybrid Synergy 4 plate reader (Winooski, VT, USA).

**Construction of pSAL plasmids:** The pSAL-based plasmids containing various reporter genes were constructed by PCR amplification of each *gfp*, *lacZ*, and *DsRed* reporter gene using forward primer oligonucleotides that included the 12.1 riboswitch sequence (aptamer domain, expression platform, and start codon) at the 5’-terminus of the oligonucleotide (see Table 2-2 for oligonucleotide sequences and template reporter plasmids). Each PCR product was digested with *KpnI* and *HindIII* and ligated into pSAL/12.1 that had been similarly digested, placing the riboswitch-reporter sequence immediately downstream to the IS10 promoter (see Table 2-3 for DNA sequences of each construct). The ligated DNA was then transformed into chemically competent *E. coli* Top10 cells.
**Construction of pET17b plasmids:** The pET17b series of riboswitch vectors containing various reporter genes were constructed by PCR amplification of each riboswitch-reporter fusion using forward primer oligonucleotides designed to replace the IS10 promoter sequence with the T7 promoter sequence (see Table 2-2 for oligonucleotide sequences). The template for each PCR amplification reaction was the corresponding pSAL vector, described above. Each PCR product was digested with *Bgl*II and *Hind*III and ligated into pET17b that had been similarly digested (see Table 2-3 for DNA sequences of each construct). The ligated DNA was then transformed into chemically competent *E. coli* TOP10 cells.

**Insertion/deletion nucleotides:** Insertion and deletion of nucleotides between the promoter and riboswitch sequence was accomplished using the QuickChange II mutagenesis kit (Stratagene) per the manufacturer’s instructions (see Table 2-2 for sequences of mutagenic oligonucleotides).

**Measurement of β-galactosidase activity and light-activation:** For LacZ reporter assays, each relevant pSAL- or pET17b-derived plasmid (see Table 2-1) was transformed into *E. coli* TOP10 or *E. coli* BL21(DE3), respectively. Single colonies were picked and grown overnight in LB media (3 mL) supplemented with ampicillin (100 µg/mL) at 37 °C, with shaking at 250 rpm. An aliquot (50 µL) of the overnight culture was used to inoculate fresh LB media (7 mL) supplemented with ampicillin (100 µg/mL) at 37 °C, with shaking at 250 rpm. The culture was grown until the OD$_{600}$ was 0.1, at which time the culture was divided into two portions (3 mL each). To one of these portions, 1 (10 mM stock in DMSO) or 2 (10 mM stock in DMSO) was added (to a maximum final concentration of 1 mM). IPTG was added to cells containing the pET17b constructs (to a final concentration of 1 mm). Samples
to be treated with light were irradiated for 3 min at 365 nm, using a hand-held 25 W UV lamp placed 10 cm from the samples. Cultures were then incubated in the dark for 5 h at 20 °C, with shaking at 250 rpm, and a sample (20 µL) was removed for determination of β-galactosidase activity using a protocol adapted from a previously described Miller assay. The culture aliquot was mixed with permeabilization solution (80 µL total volume, 100 mM sodium phosphate dibasic, 20 mM potassium chloride, 2 mM magnesium sulfate, 0.8 mg/mL hexadecyltrimethylammonium bromide, 0.4 mg/mL sodium deoxycholate, 5.4 µL/mL β-mercaptoethanol pH 7.0). The lysed culture sample was then mixed with the substrate solution (600 µL total volume, 60 mM sodium phosphate dibasic, 40 mM sodium phosphate monobasic, 1 mg/mL o-nitrophenyl-β-D-galactoside, 2.7 µL/mL β-mercaptoethanol, pH 7.0). The galactosidase was allowed sufficient reaction time to hydrolyze such that a slight yellow color is observed (typically between 2 and 60 min, at which point hydrolysis is linear over time) and quenched by the addition of 1 M sodium carbonate (700 µL) and centrifuged at 10,000 g. The absorbance was then read on a plate reader (Hybrid Synergy IV, BioTek) at 420 nm and another sample was used to read the cell densities at 600 nm. The Miller units for each sample were then found using the following equation:

Miller units = 1000 x Abs420 / (reaction time in min’s x mL of reaction volume x OD600)

Background LacZ activity was obtained by calculating the β-galactosidase activity of the host strain harboring empty vector. An average of the background was subtracted from the experimental data derived from the various riboswitch constructs. Activation ratios were calculated by dividing the Miller units in the presence of 1 or 2, by that in the absence of ligand. Each assay was performed with three independent cultures.
**Measurement of GFP expression and light-activation:** For GFP reporter assays, each relevant pSAL- or pET17b-derived plasmid (see Table 2-1) was transformed into *E. coli* TOP10 or *E. coli* BL21(DE3), respectively. Single colonies were picked and grown overnight in LB media (3 mL) supplemented with ampicillin (100 μg/mL) at 37 °C, with shaking at 250 rpm. An aliquot (50 µL) of the overnight culture was used to inoculate fresh LB media (7 mL) supplemented with ampicillin (100 μg/mL) at 37 °C, with shaking at 250 rpm. The culture was grown until the OD$_{600}$ was 0.1, at which time the culture was divided into two portions (3 mL each). To one of these portions, 1 (10 mM stock in DMSO) or 2 (10 mM stock in DMSO) was added (to a maximum final concentration of 1 mM). IPTG was added to cells containing the pET17b constructs (to a final concentration of 1 mM). Samples to be treated with light were irradiated for 3 min at 365 nm, using a hand-held 25 W UV lamp placed 10 cm from the samples. Cultures were then incubated in the dark for 2.5 h at 37 °C, with shaking at 250 rpm, and a sample (200 µL) was removed for determination of OD$_{600}$ and fluorescence intensity in a plate reader (Hybrid Synergy IV, BioTek). Fluorescence measurements were taken at an excitation and emission wavelength of 480 nm and 520 nm, respectively, and fluorescence intensities were normalized to the culture optical density. Background fluorescence was obtained by calculating the OD$_{500}$ normalized fluorescence of the host strain harboring empty vector. An average of the background was subtracted from the experimental data derived from the various riboswitch constructs. Activation ratios were calculated by dividing the OD$_{600}$-normalized and background-corrected fluorescence intensity in the presence of 1 or 2, by that in the absence of ligand. Each assay was performed with three independent cultures.
Spatial control of gene expression: An aliquot (10 μL) of *E. coli* TOP10 pSAL/12.1 starter culture was used to inoculate each well of a 96-well microplate that contained LB media (190 μL) supplemented with ampicillin (100 μg/mL). The microplate was incubated at 37 °C with shaking at 300 rpm for 1 h, at which point 2 was added (to a final concentration of 1 mM). The microplate was then covered with a mask and irradiated with UV light at 365 nm for 3 min, as described above. The cultures were incubated in the dark for 4 h at 20 °C, with shaking at 200 rpm. Then, the chromogenic reagent 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal, 2 μL of 20 mg/mL in DMSO) was added to each well. Following incubation for 10 min at room temperature in the dark, the microplate was photographed using a digital camera.
Table 2-14: Sequences of oligonucleotides.

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Oligonucleotides</th>
<th>Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSAL/gfp</td>
<td>FOR, SW03 (5'-GCATGGTACCGGTGATACCAGCATCGTCTTGATGCCCT-TGGCAGACCCCTGCTAAGGTAACAACAAAGATGAGCGAAAGGAGAAG-3') REV, SW04 (5'-GCATAAGCTTTTTGTAGAGCTCATCC-3')</td>
<td>pSAL</td>
</tr>
<tr>
<td>pSAL/DsRed</td>
<td>FOR, SW01 (5'-GCATGGTACCGGTGATACCAGCATCGTCTTGATGCCCT-TGGCAGACCCCTGCTAAGGTAACAACAAAGATGAGCGAAAGGAGAAG-3') REV, SW02 (5'-GAGTCAAGCTTTCTACTGGGAGCCGGTG-3')</td>
<td>pET17b/DsRed</td>
</tr>
<tr>
<td>pSAL/lacZ-gfp</td>
<td>FOR, SW08 (5'-GCATGGTACCGGTGATACCAGCATCGTCTTGATGCCCT-TGGCAGACCCCTGCTAAGGTAACAACAAAGATGAGCGAAAGGAGAAG-3') REV, SW04 (5'-GAGTCAAGCTTTCTACTGGGAGCCGGTG-3')</td>
<td>pET28a/gfp</td>
</tr>
<tr>
<td>pET17b/T7/lacZ</td>
<td>FOR 1, SW06 REV 07 (5'-GCTAAAGCTTTTATTTTTGACACCAG-3')</td>
<td>pSAL</td>
</tr>
<tr>
<td>pET17b/T7/gfp</td>
<td>FOR 1, SW03 REV 04 FOR 2, SW05</td>
<td>pET28a/gfp</td>
</tr>
<tr>
<td>pET17b/T7/DsRed</td>
<td>FOR, SW01 REV, SW02 FOR 2, SW05</td>
<td>pET17b/DsRed</td>
</tr>
<tr>
<td>pET17b/T7/lacZ-gfp</td>
<td>FOR 1, SW08 REV 04 FOR 2, SW06 FOR3, SW05</td>
<td>GFP in pET28a</td>
</tr>
<tr>
<td>pSAL/tac/gfp</td>
<td>FOR 1, SW03 REV, SW04 FOR 2, SW09 (5'-GCACAGATCTGTTGACAATTAATCATCGGCTCGTATAATGTTGTGGCCGGTGATACCAGCATCGTCTTGATGC-3')</td>
<td>GFP in pET28a</td>
</tr>
<tr>
<td>pSAL/gfp+2</td>
<td>FOR, SW16 (5'-TATGATCTACTAGGCGGTGATACCAG-3') REV, SW17 (5'-CTGGTATGATACCCCCATGTAACAACAAAGATGAGCGAAAGGAGAAG-3')</td>
<td>pET28a</td>
</tr>
<tr>
<td>pSAL/gfp-2</td>
<td>FOR, SW18 (5'-GATGATCTAACAACCAAGGTGATACCAG-3') REV, SW19 (5'-CTGGTATGATACCCCCATGTAACAACAAAGATGAGCGAAAGGAGAAG-3')</td>
<td>GFP in pET28a</td>
</tr>
<tr>
<td>pET17b/T7/lacZ+3</td>
<td>FOR, SW10 (5'-TCACTATGTTGAACCAGGTGATACCAG-3') REV, SW11 (5'-CTGTTGATCGGCGTACTAGTACATAAGGTAACAACAAAGATGAGCGAAAGGAGAAG-3')</td>
<td>pET17b/T7/lacZ</td>
</tr>
<tr>
<td>pET17b/T7/lacZ+1</td>
<td>FOR, SW12 (5'-TCACTATGTTGAACCAGGTGATACCAG-3') REV, SW13 (5'-CTGTTGATCGGCGTACTAGTACATAAGGTAACAACAAAGATGAGCGAAAGGAGAAG-3')</td>
<td>pET17b/T7/lacZ</td>
</tr>
<tr>
<td>pET17b/T7/gfp-3</td>
<td>FOR, SW14 (5'-TCACTATAGGCGGCGGTGATACCAG-3') REV, SW15 (5'-TCACTATAGGCGGCGGTGATACCAG-3')</td>
<td>pET17b/T7/gfp</td>
</tr>
</tbody>
</table>
**Table 2-15:** Sequences of riboswitches used in this study. Shown highlighted in each DNA sequence is the riboswitch portion (underlined), the ATG start codon (grey box), and promoter (bold). The reporter gene sequence immediately follows the ATG in each case. For the LacZ-gfp fusion, the LacZ leader sequence is shown in italics.

<table>
<thead>
<tr>
<th>Riboswitch Sequence</th>
<th>DNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pSAL/LacZ</strong></td>
<td>TGGTACGACTCTGTCTATAGTACATGACATCGTCTTGATGGTACAGCATCTGACGACTGTCAGCTTTGAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pSAL/GFP</strong></td>
<td>TGGTACGACTCTGTCTATAGTACATGACATCGTCTTGATGGTACAGCATCTGACGACTGTCAGCTTTGAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pSAL/DsRed</strong></td>
<td>TGGTACGACTCTGTCTATAGTACATGACATCGTCTTGATGGTACAGCATCTGACGACTGTCAGCTTTGAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pSAL/lacZ-gfp</strong></td>
<td>TGGTACGACTCTGTCTATAGTACATGACATCGTCTTGATGGTACAGCATCTGACGACTGTCAGCTTTGAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/lacZ</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/gfp</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/DsRed</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/lacZ-gfp</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pSAL/tac/gfp</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pSAL/GFP+2</strong></td>
<td>TGGTACGACTCTGTCTATAGTACATGACATCGTCTTGATGGTACAGCATCTGACGACTGTCAGCTTTGAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pSAL/GFP-2</strong></td>
<td>TGGTACGACTCTGTCTATAGTACATGACATCGTCTTGATGGTACAGCATCTGACGACTGTCAGCTTTGAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/lacZ+3</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/lacZ+1</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
<tr>
<td><strong>pET17b/T7/gfp-3</strong></td>
<td>TAAATACGACTCTACTATAGGCTGATAACGACATCGTCTTGATGCCCTTGGAACCCCTCGCAGGTAACAAAGATG</td>
</tr>
</tbody>
</table>
CHAPTER 3

High-throughput screening techniques for riboswitch engineering

3.1 Modification of riboswitch ligand specificity

3.1.1 High-throughput screening methods for mutation of the theophylline riboswitch

Having the ability to mutate existing riboswitches to recognize new ligands would allow for faster development of new riboswitches compared to *de novo* design. It was hypothesized that mutating the aptamer region of existing riboswitches could lead to the identification of mutant riboswitches that recognize small molecule ligands different from those recognized by the riboswitch *in vivo*. In order to find the most sensitive riboswitches a high-throughput, generally applicable screening method was desired. However, before delving too deep into the screen development, the hypothesis was tested utilizing previously developed screening methods. The lower throughput, non-generally applicable methods would provide insight into how probable it was to find mutant riboswitches that could recognize new ligands. To do this, a well-studied riboswitch was desired.

The theophylline riboswitch is one of the best understood riboswitches to date. The riboswitch, at 60 nucleotides in length, is short compared to most others and the exact location of the aptamer region has been well documented. Additionally, the expression platform is simpler than most other riboswitches, making it a good target for modification. The aptamer region comprises five sequential bases (shown in blue, Figure 3-1), allowing for a one step mutagenesis procedure that can cover the entire sequence space (1024 possible combinations). In addition, the original target ligand for the riboswitch (1, Figure 3-1) is cell
permeable and generally non-toxic to *E. coli*, although toxicity under some conditions is known.\textsuperscript{53}

**Figure 3-1:** Structure of the theophylline riboswitch. The left shows the structure in the no ligand OFF state and the right shows the riboswitch structure in the 1 bound ON state. The start codon is shown in red (58-60), Shine-Dalgarno sequence in green (43-47), and the aptamer in blue (25-29).

A panel of potential target ligands for testing the hypothesis stated above was selected on the basis of two criteria. First, the molecules had to be related in structure to 1 to increase the likelihood that the new ligand would be recognized by the mutant riboswitch. Planer, preferably aromatic compounds containing nitrogen and oxygen were desired as these compounds would be capable of ring stacking and hydrogen bonding, two common interactions in nucleotide binding. Second, the new ligands should not support activation of the WT theophylline riboswitch. To test the activation ratio with the 12.1 theophylline riboswitch a Miller assay was conducted with all of the non-native substrates (Table 3-1). Using these criteria an initial panel of potential ligands was chosen (Figure 3-2).
Figure 3-2: Panel of potential ligands chosen for the mutant theophylline riboswitch. 1 is shown for comparison.

Table 3-1: Activation ratio of potential ligands with WT theophylline riboswitch. An activation ratio of one indicates no activation is observed.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Activation Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0.9</td>
</tr>
<tr>
<td>7</td>
<td>1.1</td>
</tr>
</tbody>
</table>

In order to construct a library of RNA’s, the aptamer region (nucleotides 25-29) of the theophylline riboswitch gene sequence was mutated to NNNNN using site-directed mutagenesis, where N could be any of the four bases. Following transformation of the mutated library into *E. coli* and DNA sequencing of randomly selected clones, it was confirmed that the aptamer sequences were random (data not shown). In order to test all of
the possible 1024 mutants, a high throughput screen was required. Using LacZ as a reporter allowed for blue/white screening of the mutants. In blue/white screening 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) is spread across an agar plate prior to transforming colonies onto the plate. If LacZ is expressed, the colony will hydrolyze the sugar from X-gal allowing the indole aglycone to dimerize, resulting in a blue color. The colonies that do not express LacZ will remain colorless (Figure 3-3).

Figure 3-3: Image of a plate used for blue/white screening. Colonies expressing LacZ appear blue while colonies not expressing the enzyme remain colorless.

Screening the mutant riboswitch library required two steps (Figure 3-4). In the first step, *E. coli* cells harboring the mutant riboswitch libraries were spread onto plates in the absence of any added ligand. In this case the colorless colonies were picked, representing riboswitches in the OFF state when no ligand was present. The DNA from those colonies was
then extracted and used to transform *E. coli* cells that were then spread onto plates containing either 1 or a single non-native ligand, 3-7. Subsequently, the bluest colonies were selected for further analysis and represented ON riboswitches that express LacZ in the presence of ligand. Using this method, ~10,000 colonies were screened. This ten-fold over sampling increased the likelihood that every possible mutant would be screened at least once. To provide quantitative data and eliminate false positives, secondary screening was conducted using Miller assays in a 96-well plate format. After secondary screening a total of ~1,500 colonies, mutant riboswitches were identified that displayed robust activation ratios (>4-fold) towards 3 (Table 3-2). Interestingly, activation ratios of selected riboswitches towards 4-7 were not particularly high (<4-fold). The aptamer sequence of the top performing mutant with 3, A1-1, is GCCAU.
Create mutant library

Create mutant library

Pick colorless colonies

Pick colorless colonies

Pick blue colonies

Secondary screen with Miller assay

Figure 3-4: Scheme of riboswitch screening using LacZ. The mutated riboswitches were spread on plates containing X-gal and no ligand. The colorless colonies show no LacZ expression so they were pooled for the next round of screening. Now the mutant riboswitches were spread on plates containing a non-native ligand and X-gal. This time the blue colonies were selected as they are expressing LacZ. These blue colonies were now either counter-screened on plates containing 1 and X-gal where the colorless colonies were selected, or they were secondary screened using the Miller assay.
Table 3-2: Activation ratio of the most active mutant with each new ligand. Experiments were conducted in triplicate and standard deviations were less than 10% of the activation ratio.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Activation Ratio</th>
<th>Aptamer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analogue</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>2.7</td>
<td>129</td>
</tr>
<tr>
<td>5</td>
<td>3.4</td>
<td>112</td>
</tr>
<tr>
<td>6</td>
<td>1.8</td>
<td>140</td>
</tr>
<tr>
<td>7</td>
<td>2.1</td>
<td>153</td>
</tr>
</tbody>
</table>

The activation ratios of the best performing riboswitch towards 3-7 were small compared to that with 1, however, these ratios were on par with many other riboswitches. Ideally, a mutant riboswitch could be identified that displays modest activation toward a new ligand yet no longer recognizes 1. However, activation rations with 1 remained high. Since the entire theoretical sequence space was likely screened, an additional round of mutagenesis was not carried out.

It was observed that an unexpectedly large number of mutant colonies were blue during the negative screen (~90%), possibly indicating unstructured riboswitches that support access by the ribosomal protein translation machinery in the absence of ligand. A possible cause of this is illustrated in Figure 3-1. The aptamer region (blue, 25-29) binds to the SD sequence (green, 43-47) when the riboswitch is in the OFF state. Therefore, mutations to the aptamer may prevent base-pairing with four of the five bases of the SD sequence, thus exposing the SD and switching on expression of LacZ. In an attempt to minimize this possibility, the four bases of the WT SD sequence (nucleotides 44-47) that bound the aptamer were randomized (NNNN). Mutating the SD sequence and the aptamer region at the
same time would result in a library too large to exhaustively screen (262,144 possible mutants). Therefore, the SD sequence was mutated before the aptamer. Since the SD sequence is essential for ribosome recognition and only certain mutations would still allow for proper translation, the mutants were screened for gene expression before the aptamer region was also mutated.

To test for functioning mutant SD sequences blue-white screening was performed. *E. coli* harboring the mutant riboswitches were spread onto agar plates containing X-gal and 1. The resulting large number (~15%) of blue colonies showed that the mutant SD sequences were still functional as LacZ was being produced. Next, the plasmid DNA of the functional ON riboswitches was collected, and the aptamer region of the pooled SD sequence mutants was then also randomized. The goal of this two-step process is to arrive at new SD sequences that can better support sequestration by mutated aptamers in the absence of ligand. Thus, the SD/aptamer library was first screened in the absence of ligand and those OFF mutants were then screened in the presence of 3.

Gratifyingly, after only primary screening ~50,000 colonies, activation ratios towards 3 were obtained that were several-fold higher than those with the previous riboswitch A1-1. As before, little or no increase in activation ratio was observed for any of the other ligands tested (Table 3-3). Sequencing revealed that mutants A2-1, A2-3, and A2-4 had the same SD sequence (UAUGG) while A2-2 had the SD sequence UUAGA. The aptamer for each mutant was different with the sequences for A2-1 to A2-4 being CCCAU, CUCAU, CCCAA, and GCCAU, respectively. Each aptamer sequence appears as though it would base pair with the corresponding SD sequence in at least three of the four possible positions.
Table 3-3: Activation data for wild-type, 1st- and 2nd-generation mutant riboswitches with 1 and 3. The ON and OFF values represent the Miller units for each case and the activation ratio is the ratio of these two values. Experiments were conducted in triplicate and standard deviations were less than 10% of the activation ratio.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>3 OFF</th>
<th>3 ON</th>
<th>Activation Ratio</th>
<th>1 OFF</th>
<th>1 ON</th>
<th>Activation Ratio</th>
<th>SD sequence</th>
<th>Aptamer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2-1</td>
<td>74</td>
<td>1296</td>
<td>17.5</td>
<td>74</td>
<td>9952</td>
<td>134.3</td>
<td>UAUGG</td>
<td>CCCAU</td>
</tr>
<tr>
<td>A2-2</td>
<td>102</td>
<td>1438</td>
<td>14.1</td>
<td>102</td>
<td>10148</td>
<td>99.2</td>
<td>UAAGA</td>
<td>CUCAU</td>
</tr>
<tr>
<td>A2-3</td>
<td>59</td>
<td>765</td>
<td>12.9</td>
<td>59</td>
<td>7637</td>
<td>129</td>
<td>UAUGG</td>
<td>CCCAA</td>
</tr>
<tr>
<td>A2-4</td>
<td>82</td>
<td>912</td>
<td>11.1</td>
<td>82</td>
<td>14285</td>
<td>174</td>
<td>UAUGG</td>
<td>GCCAU</td>
</tr>
<tr>
<td>A1-1</td>
<td>91</td>
<td>728</td>
<td>8</td>
<td>91</td>
<td>8736</td>
<td>96</td>
<td>UAAGG</td>
<td>GCCAU</td>
</tr>
<tr>
<td>WT</td>
<td>78</td>
<td>83</td>
<td>1.1</td>
<td>78</td>
<td>22284</td>
<td>285.7</td>
<td>UAAGG</td>
<td>CCCUU</td>
</tr>
</tbody>
</table>

While the activation ratio with 3 in A2-1 was increased an additional 2-fold compared to A1-1, the activation ratio with 1 was still higher than desired. In order to address this lack of specificity, a counter screen was implemented to eliminate those mutants that still recognized 1. Following the positive screen in the presence of 3-7, the resulting DNA extracted and pooled plasmid DNA was transformed into *E. coli* and spread onto plates containing X-gal and 1. The colorless colonies were selected as they showed the lowest activation with 1. A small panel of top-performing mutant riboswitches was chosen for analysis by Miller assay (Table 3-4). The SD sequences for A3-1 to A3-3 are UAUGG, UAAGU, and UAAGG (WT) and the aptamer sequences are GCCAU, CACUU, and CCCUA, respectively. Again, all of these SD sequences appear as though they would bind with their corresponding aptamer well in the absence of ligand.
The activation ratio of the best 3rd generation mutants A3-1, A3-2, and A3-3 with 1 was still higher than desirable, but were 5 to 14-fold lower than that of the WT riboswitch, and 3 to 9-fold lower than the previous set of riboswitches. Notably, the activation ratios of A3-1, A3-2, and A3-3 with 3 were also slightly (1.2-fold) lower than the previous set mutants, as expected. Because the most active mutants from previous screening also had a high activation ratio with theophylline, counter screening against 1 was expected to eliminate these mutants, leaving those with lower activation with 3. The goal of this screen was to create riboswitch with orthogonal specificity toward 3, not to maximize the activation ratio with 3. Thus, counter-screening riboswitch libraries against ligands closely related in structure to the desired target activator can improve orthogonality.

### 3.1.2 Ultrahigh-throughput screening of the theophylline riboswitch

In an attempt to increase the throughput of screening, and therefore the size of riboswitch libraries that could be searched, fluorescence activated cell sorting (FACS) was
also investigated. The use of FACS could be implemented in several different ways depending on the needs of the assay. The negative screen would remove the majority of the non-functioning riboswitches, eliminating any riboswitches that do not fold properly into the OFF state. This would result in a large reduction in library size that could then be further screened using plates or FACS. Positive sorts would then further decrease the number of mutants that would need to be secondary screened.

Two different reporter systems were pursued. The first used the existing LacZ reporter but with 3-carboxy-umbelliferyl β-D-galactopyranoside (CUG) as the substrate. CUG is hydrolyzed by LacZ forming a fluorescent product that is detected by FACS. However, CUG is not membrane permeable. Therefore, using CUG to determine the amount of LacZ expressed while maintaining the genotype-phenotype linkage is difficult in a high-throughput manner. At the time, Dr. John McArthur (Williams Lab) was working on in vitro compartmentalization techniques that rely on compartmentalizing single cells in water-in-oil-in-water (w/o/w) droplets. Dr. McArthur had shown that subjecting cells in water-in-oil droplets to a single freeze-thaw cycle would lyse the cells without destroying the droplets. In this way, it might be possible to maintain the genotype-phenotype linkage while allowing the CUG substrate to interact with the protein within a sortable droplet.
Figure 3-5: FACS data for droplets containing the theophylline riboswitch. Droplets were first gated by GPF fluorescence (right) to select only those containing a cell. Cells with higher fluorescence (left) indicate higher levels of LacZ fluorescence.

In order to test the ability of compartmentalized bacterial cells to be assayed for riboswitch activity by FACS, *E. coli* were prepared that harbored a plasmid that contained *lacZ* under the control of the WT theophylline riboswitch and *gfp* under the control of the T7 promoter. These cells were grown in the presence of IPTG to induce expression of GFP and either in the presence (1 mM) or absence of 1. The cells were then compartmentalized in a
water-in-oil emulsion, where the aqueous layer contained CUG. The droplets were then subjected to a single freeze-thaw cycle to lyse the cells. Following incubation at 37 °C, the droplets were further emulsified to form w/o/w droplets that could be sorted by FACS.

The droplets were first gated for GFP fluorescence, indicating that the droplet contained a cell (right, Figure 3-5). Those droplets were then scanned for coumarin fluorescence (left, Figure 3-5). The data show no significant difference between the ON and OFF states of the theophylline riboswitch. When a sample (100 μL) of the same cultures was analyzed by Miller assay, the activation ratio of the 12.1 riboswitch with CUG was found to be 178, in close agreement with previous data (Figure 1-2B). This showed that the CUG substrate and LacZ protein was a suitable reporter for riboswitch activity, and that the inability of the droplets to display theophylline activation was due to issues with the droplets and/or FACS. Repeated droplet experiments with differing conditions yielded no further improvement.

Because the droplets were not producing the expected activity, a new substrate, fluorescein di(β-D-galactopyranoside) (FUG), was tested. The fluorescence spectrum of FUG is similar to GFP, so gating using GFP fluorescence is not helpful. For this reason, and due to the problems using droplets in the experiments described above, the FUG assay was first attempted with live cells that were not compartmentalized. When the 12.1 theophylline riboswitch was analyzed for activity with FUG in the absence or presence of photocaged theophylline (2) and subjected to UV irradiation, activation was observed, as judged by the higher percentage of droplets in the M1 region (Figure 3-6).
Figure 3-6: FACS data for E. coli cells containing the theophylline riboswitch. Cells found within the M1 region show high concentrations of hydrolysed FUG, indicative of high LacZ production.

However, only ~36% of the cells show this increased fluorescence indicating that the majority of the cells show little or no change in the presence of 2. Similar results were obtained when 1 was used. Attempted optimization did not improve the fraction of cells in the M1 high GFP fluorescence region. The data show that approximately two-thirds of the mutants that expressed protein while in the OFF state would be sorted in with the genuine
ON riboswitches. Additionally, only about one-third of the cells expressing protein in the ON state would be collected. FACS was also tested with the theophylline riboswitch with GFP as the reporter. This was ineffective because the activation ratio with GFP was too small for the assay. For these cumulative reasons, FACS-based screening of riboswitch libraries was abandoned in favor of lower throughput screens.

### 3.1.3 Theophylline riboswitch in mammalian cells

There has been little success with non-ribozyme acting riboswitches in mammalian cells.\(^5\)\(^5\) One of the limitations of riboswitches is that they were mostly found in prokaryotes and do not easily transfer between species.\(^5\)\(^6\) Making the transition to mammalian cells would further increase the potential for riboswitches. With the aid of Dr. Jeane Govan from the Deiters’ lab, the function of the theophylline riboswitch was tested in HeLa cells. The cells were co-transfected with a vector containing a gene encoding for GFP and a second vector containing a vector harboring the theophylline riboswitch controlling the expression of DsRed. The activity of the riboswitch was monitored at different concentrations of 1 (Figure 3-7).
No riboswitch function was observed at any concentration. This result was expected because of how the riboswitch functions. The theophylline riboswitch functions by sequestering the ribosome binding site in the OFF state. However, this ribosome binding site (RBS) is the SD sequence used in many prokaryotes. Eukaryotes however do not have the
same ribosomes as prokaryotes, and these ribosomes recognize open reading frames in mRNA differently. Therefore, the RBS found in the theophylline riboswitch would not be recognized by the machinery in mammalian cells. Changing the RBS to match with the requirements of mammalian cells would require a complete reworking of the current riboswitch and was outside of the aim of this project. The results of this experiment further support the earlier findings that riboswitches are not easily interchangeable parts.

3.1.4 Modification of ligand specificity in the lysine riboswitch

With the success of mutating the theophylline riboswitch to recognize a new ligand, the aim was to show similar success could be obtained with additional riboswitches. A naturally-occurring, more complex riboswitch that worked at the transcriptional level was desired as this would show success through mutation was not limited to shorter, engineered riboswitches. The lysine riboswitch, LysC, is an OFF riboswitch (gene expression is down-regulated when the ligand is present). In nature the riboswitch is used to regulate genes responsible for the biosynthesis of lysine. Therefore, when high concentrations of lysine are present the expression of these genes is down-regulated to avoid wasting cellular resources. Previous work with the LysC riboswitch has shown its ability to recognize some molecules with structural similarity to lysine. Other work has shown that the riboswitch has antibiotic potential. When a lysine derivative was added into a culture it would lower expression of lysine producing proteins. The unnatural lysine derivatives could not be incorporated into proteins however, so the low lysine concentrations would slow growth and eventually kill the cells. Lysine riboswitches have been found in a variety of different organisms and display structural diversity (Figure 3-8).
Figure 3-8: 2-D structure of the LysC riboswitch. The LysC riboswitch recognizes the amino acid lysine and functions as an ON switch. Highly conserved residues are shown boxed.58

Figure 3-8 shows only a small number of nucleotides (boxed) are highly conserved and that much of the LysC riboswitch is poorly conserved, indicating that mutation at those poorly conserved locations would likely be ineffective at changing ligand specificity. Here, the goal was to create a mutant LysC riboswitch that displayed specificity for a new ligand. Mutations were therefore focused at the highly conserved residues because these residues were thought to be critical for ligand recognition and/or expression platform. Initially, eight residues were chosen for mutagenesis. The sites were chosen based on their proximity to the aptamer region of the riboswitch (Figure 3-9). The positioning of other conserved residues indicates that they are part of the expression platform, which is not the aim of mutagenesis as it may destroy the activity of the riboswitch with all ligands.
The eight mutation sites selected were G12, G14, C78, C79, G80, A81, G114, and G163 (Figure 3-8). These bases interact with or near the bound lysine, and are highly conserved in LysC riboswitches.

Previous studies had also investigated the binding interaction of the LysC riboswitch and made conclusions as to how the ligand was bound (Figure 3-10).
Figure 3-10: Binding interactions present in the LysC riboswitch.

The LysC riboswitch was first cloned upstream of the gfp gene, yielding the plasmid pLys (Figure 3-11). Altering the riboswitch too much at one time would likely cause it to lose all function. Thus, for the first round of mutagenesis, the mutations were broken into pairs, (G12, G14; C78, C79; G80, A81; G114, G163) creating four individual mutant LysC libraries. Initially, attempts were made to screen the activity of the mutant riboswitches using minimal media in order to reduce the endogenous concentration of lysine in favor of the target unnatural lysine analogues. However, this process proved too slow when multiple rounds of primary screening, followed by rounds of secondary screening were required.

Figure 3-11: Plasmid map of the pLys vector.
To increase the speed of screening a new strategy was devised (Figure 3-12). First the mutants were screened in the presence of lysine and those supporting GFP expression were collected. Given the LysC riboswitch is an OFF switch, such positive mutants should represent mutant riboswitches that no longer recognize lysine (or any other cellular component). The ultimate goal was to create a riboswitch that recognized a new ligand but not lysine. Thus, this screening step enabled future rounds of mutagenesis and screening to be conducted in LB media in the presence of lysine (which can now grow quickly), and improved the likelihood of discovering new riboswitch variants that respond to the target lysine analogue.
Figure 3-12: Scheme showing the initial screening strategy for LysC riboswitch mutants. First the counterscreen against the natural substrate, lysine is performed. Now that the mutants no longer recognize lysine, the next rounds of screening can be performed in LB media, significantly decreasing screening time. Then, repression of GFP expression screened for by selecting the non-fluorescent colonies when unnatural ligand is present. Finally, colonies expressing high GFP fluorescence in the absence on ligand are found. Additional secondary screening or mutagenesis is then conducted.

Accordingly, each of four mutant LysC riboswitch libraries were transformed into *E. coli* BL21(DE3) cells and spread onto LB agar plates supplemented with lysine, potassium, and IPTG. The brightest green colonies were pooled, cultured, and the plasmid DNA was extracted (Figure 3-13).
The next step of the assay was to screen for activity (repression) with unnatural ligands. Using the information in Figure 3-10, new ligands were selected for screening. It appeared that modifications to the carbon chain could not be too bulky and that a terminal amine was desired. For initial screening only small modifications to the structure were desired. Low cost ligands were also desired. A small panel of ligands was formed for screening (Figure 3-14).
However, after screening the pooled mutants that no longer respond to 8, clones that showed activity (repression) were not identified with the unnatural ligands, indicating that the recognition sequence in the aptamer was mutated too much, and likely no longer recognized the amine and carboxylic acid, or resulted in steric clashes.

A new screening strategy was required that would minimize the time required per round of screening while avoiding elimination of riboswitch activity towards lysine and the analogues. If the cell growth and GFP maturation rates could be improved on minimal media plates then the initial identification of riboswitches that no longer recognize 8 would not be required. Through a series of trials it was found that allowing the colonies to grow on minimal media containing no IPTG or ligand for ~20 hours, followed by addition of the
ligand and IPTG for 30 minutes, and incubation at room temperature was sufficient to enable
detection of GFP fluorescence in 2-3 days, in contrast to the 5+ days required previously.

Using this novel screening strategy, riboswitch functionalities were found from the
libraries of mutated LysC riboswitches. By first picking the fluorescent colonies from the LB
plates, where the cells were exposed to high 8 concentrations, then picking the non-
fluorescent colonies from the minimal media plates, where the 8 concentration is low, an ON
lysine riboswitch, designated L1.1, with an activation ratio of 2.3 with 8 was found. This
illustrates a shift in switching behavior compared to the natural riboswitch which is an OFF
switch with an activation ratio of 0.34 (note that an activation ratio less than one indicates
that the riboswitch acts in and OFF manner). This result is unexpected because the
expression platform is responsible for the change in activity, and only the aptamer was
mutated. Mutations to the aptamer region have not been shown to change an OFF riboswitch
to an ON riboswitch before. The nucleotide mutations to this riboswitch, G12A/G163C,
normally base pair to form a hairpin in the no ligand state. Therefore, this alteration may play
a role in not only the ligand binding, but also the function of the expression platform.

In addition, mutant OFF riboswitches were desired that turned off reporter expression
in response to the non-natural analogues 9-12. In principle, riboswitches identified by
screening for repression with the analogues could also be repressed by lysine 8 or lose
activity towards 8. To find the OFF riboswitches that recognize the new ligand regardless of
whether or not they respond to 8, colonies expressing the LysC mutant libraries were first
screened for GFP fluorescence in the absence of 8. The brightest green colonies (~20% of the
entire library) representing ON riboswitch mutants that do not respond to endogenous ligands
were then screened for GFP fluorescence in the presence of unnatural ligands 9-12 and the non-fluorescent colonies collected. Now, those colonies that were the least green were collected and represent mutant riboswitches that do not repress reporter expression in the presence of endogenous ligands but repress reporter expression in the presence of unnatural ligand. The response of these switches towards lysine has not been defined at this point. In contrast, to generate riboswitches that are genuinely orthogonal with respect to responding to the unnatural ligand but not 8, bright green colonies were collected after screening in the presence of 8. Next, those riboswitches unaffected by the presence of 8 were screened in the presence of low concentrations of 8 and the brightest colonies collected. Finally, the pooled mutants were screened in the presence of the unnatural ligand and the non-fluorescent colonies were collected. These final mutants should represent OFF riboswitches that repress reporter expression in response to the unnatural ligands but are unaffected by 8 (Figure 3-15).

*Denotes that fluorescence with 8 was too low to display clearly.

**Figure 3-15:** Activity with mutant lysine riboswitches with natural and unnatural ligand.
The early results from the lysine riboswitch mutants indicate that the riboswitch can be tailored to recognize new ligands. While the error bars were larger than desirable, riboswitches with different functionalities seem plausible. The results are still preliminary and will be further investigated. The activation ratios were interesting in that additional ON riboswitches were found. The mutant riboswitch G1.1 (G12A, G14C) showed no influence due to 8, while acting as an ON switch with 9. The mutation at G12 was also seen in L1.1, which also acts as an ON riboswitch. Riboswitch S1.1 (G14C, C78G) showed a much higher activation ratio with 8, than the WT riboswitch. As illustrated in Figure 3-9, G14 and C78 base pair in the wild-type LysC riboswitch. This G-C base pair was preserved in the S1.1 mutant, but interactions between these two nucleotides and A81 was altered. Further screening is needed to cover more sequence space to determine if additional interesting riboswitches will result from the pool.

3.2 Development of an ultrahigh-throughput riboswitch screening method

As evidenced above, one of the difficulties with designing assays for riboswitches is that they require both negative and positive screens. This can lower throughput and also add additional steps in the screening process. Blue/white and fluorescence screening are high-throughput methods for riboswitch screening, but can prove labor intensive and slow, as each step requires two days and multiple negative and positive screens are often required. A higher throughput, generally applicable screen could make mutating existing riboswitches more feasible than building them \textit{de novo}. In response to these criticisms, a generally applicable platform was developed in this section. Ideally, this platform could be used to
rapidly carry out negative and positive screens without requiring subcloning of candidate riboswitch mutants between each round.

The tight binding affinity between barnase and barstar proteins has often been leveraged for a variety of chemical biology applications. Barnase is known for its ability to degrade RNA and thus kill the host cell. It has been used for negative selections in several studies aimed at installing unnatural amino acids into proteins. However, in the presence of barstar, barnase is sequestered and loses its ability to degrade RNA. This protein system could form the basis of a powerful dual screen-selection platform if GFP could be fused to barnase (Figure 3-14). First, *E. coli* cells harboring a riboswitch library is grown in the absence of both ligand and IPTG. In this case, the inhibitor barstar, under control of the T7 promoter is not expressed. However, if the riboswitch is leaky in the absence of ligand, *barnase-gfp* gene will be expressed, killing the cell. In contrast, if the riboswitch is not leaky and is completely OFF then the *barnase-gfp* gene will not be expressed, and the host cell will survive. Therefore, after this simple selection in the absence of ligand, only the surviving colonies contain a riboswitch in the OFF state.

Next, the surviving cells are grown in the presence of both IPTG and ligand. Under these conditions, the inhibitor barstar is expressed and the ligand is available to potentially activate riboswitch dependent gene expression. Thus, colonies that contain riboswitches that activate gene expression in the presence of ligand produce both the barnase-GFP fusion and the inhibitor, barstar, and survive. Moreover, GFP fluorescence provides a quantitative read out of gene expression in the presence of ligand. These surviving cells can be analyzed by flow cytometry and sorted, leading to a dual selection platform where upwards of 10 million
mutants could be selected in a day. The process could also be carried out on a smaller scale using plate assays (Figure 3-16).

**Figure 3-16:** Dual-selection strategy for mutant riboswitches. In the absence of ligand and IPTG, those riboswitches expressing the barnase-gfp fusion will die due to the barnase activity, resulting in a negative selection. Ligand and IPTG would then be added to the surviving cells. The IPTG would induce expression of barstar which would inhibit barnase, which was upregulated by the ligand. Levels of gene expression from the riboswitch could then be monitored by GFP fluorescence.

A *barnase* gene containing two Amber stop codons was provided by the Deiters Lab. Site-directed mutagenesis was used to remove the two stop codons and then the 12.1
theophylline riboswitch was added to the 5’ terminus of the gene by PCR. This construct was then cloned into the commercial pT7-gem vector under the control of the constitutive SP6 promoter. Next, gfp was cloned into the 3’ end of barnase. Finally, barstar was synthesized and cloned into the vector under the T7 promoter. In order to test the ability of the dual-selection strategy to identify highly activate riboswitches, a number of control liquid cultures were prepared in the presence/absence of IPTG and with or without the theophylline 12.1 riboswitch. The screening platform appears to be working as expected in the absence of the riboswitch (-Riboswitch, Table 3-5). Gratifyingly, this illustrates that the barnase-GPF fusion maintains the toxicity of wild-type barnase and also the fluorescence of wild-type GFP.

Table 3-5: Summary of results from initial dual selection vector screen assays.

<table>
<thead>
<tr>
<th></th>
<th>-Riboswitch</th>
<th>+Riboswitch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>No IPTG</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cell density</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>GFP fluorescence (absolute)</td>
<td>1823</td>
<td>1843</td>
</tr>
<tr>
<td>GFP fluorescence (normalized)</td>
<td>96150</td>
<td>61433</td>
</tr>
<tr>
<td>With IPTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell density</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td>GFP fluorescence (absolute)</td>
<td>40278</td>
<td>38539</td>
</tr>
<tr>
<td>GFP fluorescence (normalized)</td>
<td>46297</td>
<td>45340</td>
</tr>
</tbody>
</table>

However, when the riboswitch is present (+Riboswitch), the screen no longer works as expected, resulting in low barnase-GFP productions regardless of the presence of 1. This is consistent with previous data that showed that riboswitch activity can be dependent on the identity of the promoter and reporter gene.
To combat this problem a new vector screening platform was developed (Figure 3-17). In this second generation screening vector, the genes are distributed between two plasmids. The first vector, pStar, contains \textit{barstar}, under control of the T7 promoter. The other vector, pLB, was a modification of the pSAL vector, where \textit{barnase} is inserted at the 3’-terminus of the \textit{lacZ} gene. The pSAL vector includes a constitutive promoter and uses \textit{LacZ} as the reporter gene. The activation ratio observed from this vector is high at 275. Having \textit{LacZ} as the reporter would make the use of FACS more difficult compared to GFP, but using previously described agar plate screening techniques would still allow for screening of upwards of $10^5$ mutants per day.
Constructing pLB, in collaboration with Kyle Bingham, proved difficult. The barnase gene retained greater functionality when placed at the N-terminal end of the protein, as was seen in the GFP-barnase fusion. In this case the LacZ gene must be left at the N-terminal to achieve the proper riboswitch function as shown in Chapter 2. Cloning of different linker regions between the genes is still being performed so that the barnase functionality can be restored.
3.3 Conclusions and future work

Development of high-throughput screening techniques makes it possible to create new riboswitches from the mutation of existing ones. Agar plate-based screening methods were capable of altering ligand specificity of both the 12.1 theophylline riboswitch and the LysC lysine riboswitch. A mutation to the 12.1 theophylline riboswitch provided a riboswitch that could recognize 8 with an activation ratio of 17. In similar fashion, mutations to the LysC riboswitch provided a mutant that now acted as an ON riboswitch as opposed to the natural OFF functionality of LysC. This technology will make the development of new riboswitches faster than the current techniques of designing each from scratch. The slow rate of riboswitch development is one of the major hurdles that currently limits them from being used in a broader scale. This research takes a step towards making the evolution of riboswitches as feasible as that of proteins, which is widely done today.

If further advancements in the development of new riboswitch continue then they could be widely used in metabolic engineering. If riboswitches could be used to regulate the expression of individual genes in a pathway bottlenecks and overproduction of enzymes could be avoided. Riboswitches could be made to recognize the intermediates in a given pathway and used to regulate the next gene in the pathway. Since riboswitches work in a dose-dependent manner the genes would be up or down regulated depending on the amount of intermediate present.

One of the potential benefits of the barnase/barstar assay over previously reported assays is that it could be adapted to be used with cell-impermeable ligands. To do this the first part of the assay would remain unchanged, selecting against the riboswitches that
display leaky expression. Next, the cells would be encapsulated in water-in-oil emulsions containing the ligand and subjected to a freeze-thaw cycle. This step lyses the cell allowing the impermeable ligand to access the cellular machinery and activate the riboswitch, while still maintaining the genotype-phenotype linkage. The droplets can then be sorted by FACS, selecting the most active riboswitches by GFP or fluorescent LacZ products. Dr. John McArthur and Robert Kalkreuter in our group have worked to develop the droplet technology (Figure 3-18). While the use of FACS did not work well with the 12.1 theophylline riboswitch, a de novo riboswitch could yield better results. A common adage in directed evolution is that “you get what you screen for”63. Therefore, attempting to use FACS to screen a riboswitch that was originally found using a plate-based screen is likely to have complications. However, if initial screening is conducted using FACS some of the difficulties seen with the 12.1 theophylline riboswitch may not be present.

**Figure 3-18:** Droplet-based technique for selecting for cell-impermeable ligands. By subjecting cells enclosed in a droplet to a freeze-thaw cycle, lysing the cells, it may be possible to develop riboswitches for a wider array of ligands.
Utilizing this method, a major limitation of riboswitch development could be overcome. Many of the ligands that would be a target of riboswitches are produced in the cell and are membrane impermeable. These ligands could be intermediates in a metabolic pathway or targets of a riboswitch biosensor. Upon completion of the barnase/barstar dual selection platform, implementing this technology would increase the impact of the research and provide a wide array of targets for riboswitches.

3.4 Methods

General methods

All plate reader assays were performed on a BioTek Hybrid Synergy 4 plate reader (Winooski, VT, USA).

Riboswitch mutation

Mutation of the riboswitch aptamer and SD sequences was performed using Round the Horn PCR. Primers SW16 and SW17 were used for mutation of the aptamer while primers SW18 and SW19 were used to mutate the SD sequence. PCR was performed using Phire II Hotstart polymerase according to protocol. An annealing temperature of 63 °C was used.

Construction of GFP-based dual selection plasmid

All PCR reactions were performed with Phire II Hotstart polymerase with an annealing temperature of 63 °C and an extention time of 15 sec/kb PCR product unless otherwise noted. A PCR was run on the pET28a plasmid harboring sf-gfp. Primers SW20 and SW21 (Table 3-6) were used to produce a 12.1-gfp gene. This was cloned into pT7-gem via the HindIII and NotI restriction sites forming pT7-12.1. Site-directed mutagenesis was then performed using SW23 and SW24 on the double mutant barnase provided from the Deiters’
Lab to remove one of the mutations. The second mutation was removed with SW25 and SW26. This product was cloned into pT7-12.1 via the NsiI and NotI restriction sites forming p12.1B. Next, GFP was cloned from pET28a-gfp with SW21 and SW27. This PCR product was then used for the template in a PCR using SW21 and SW28, yielding gfp with the full-length linker region. This was cloned into p12.1B via the NotI and BamHI restriction sites yielding p12.1BG. Finally, barstar synthesized by Genewiz (South Plainfield, NJ) and cloned in via the SacI and XhoI restriction sites.

Construction of LacZ-based dual selection plasmids

Barstar was cloned into the pET28a vector via the SacI and XhoI restriction sites. Barnase was amplified from p12.1BG with SW29 and SW30 and cloned into p12.1BG via the NdeI and BamHI restriction enzymes. This provided the linker region from p12.1BG at the 5’-terminus of barstar. This construct was removed and subcloned into the pSAL plasmid at the 3’-terminus of LacZ via the NotI and BamHI restriction sites.

LysC riboswitch confirmation

The LysC riboswitch was previously cloned into pET17b by Nicole Coggins. The sequence was analyzed to confirm cloning had been completed successfully.

Blue/white plate screens

X-gal was dissolved in DMSO to 20 mg/mL and 20-50 μL (depending on selective pressure) was spread onto LB agar plates containing 100 μg/mL ampicillin. The X-gal was spread quickly with glass distillation beads to prevent uneven absorption of the substrate. For screens evaluating ligand response, 50 μL of 25 mM ligand (1, 3-7) was then spread on the plate and the plate was allowed to dry under flame. The mutant riboswitches were then
transformed into chemically competent Top10 cells and 50-100 μL of the transformation mixture was spread onto the plate which was then incubated overnight at 37 °C. The volume of transformation mixture plated was adjusted to obtain 1000-3000 colonies per plate. This maximized the number of colonies screened while still allowing them to be picked easily. After incubation either the white (for negative screen) or blue (for positive screen) were picked. If further screening was required all colonies were picked into a single 3 mL culture which was later minipreped and the DNA used for further transformations. If no further plate screens were needed the colonies were picked into wells of 96-deep well plates each containing 1 mL of LB supplemented with 100 μg/mL ampicillin.

**96-well plate screening**

Into deep-well plates containing 1 mL of LB supplemented with 100 μg/mL ampicillin per well mutant riboswitch colonies were picked. Each plate also contained two wells with an empty pUC vector and two well of the WT riboswitch. These plate were grown overnight at 37 °C, 350 rpm. These starter plates were used to inoculate new plates by adding 20 μL of the overnight culture to 980 μL of LB containing 100 μg/mL ampicillin. Each well of the starter plate was used to inoculate two side-by-side wells of the new plates. These plates were then grown at 37 °C, 350 rpm for ~1.5 hr, until the wells reached an OD₆₀₀ of 0.1. Then ligand (1, 3-7) was added to 1 mM to the even numbered wells and an equal amount of solvent was added to the odd numbered wells. The plates were returned to the shaker and incubated at 20 °C, 350 rpm overnight. A 10 μL aliquot of the overnight culture was then added to 40 μL of permeabilization solution (100 mM sodium phosphate dibasic, 20 mM potassium chloride, 2 mM magnesium sulfate, 0.8 mg/mL hexadecyltrimethylammonium
bromide, 0.4 mg/mL sodium deoxycholate, 5.4 µL/mL β-mercaptoethanol pH 7.0) in a clear 96-well plate. The mixture was allowed to sit for at least one hour then 80 µL of substrate solution (60 mM sodium phosphate dibasic, 40 mM sodium phosphate monobasic, 1 mg/mL o-nitrophenyl-β-D-galactoside, 2.7 µL/mL β-mercaptoethanol, pH 7.0) was added. The time of this addition was noted. Wells were quenched with 100 µL of 1 M sodium carbonate as soon as a faint yellow color was observed and the time of the quench was noted. When all cells were quenched the plates were read for absorbance at 420 and 600 nm. The Miller units of each well were then calculated as shown in Eqn 1 (Chapter 2). The activation ratio was determined by dividing the Miller units of the with ligand culture by that of the no ligand culture.

**In vitro FACS analysis**

BL21 (DE3) cells harboring GFP on a pET28 vector and the theophylline riboswitch were grown overnight in a 3 mL culture of LB media containing 50 µg/mL ampicillin and 50 µg/mL kanamycin. 50µL aliquots of this overnight culture was used to inoculate 2 – 3 mL cultures supplemented with 50 µg/mL ampicillin and 50 µg/mL kanamycin. The cultures were incubated at 37 °C, 250 rpm for 1 hour. IPTG was then added to both cultures at a final concentration of 1 mM. To one culture 1 was added to a final concentration of 1 mM and the two cultures were returned and incubated at 37 °C, 250 rpm for 3 hours. The cells were then centrifuged at 2500 rpm for 5 minutes and the pellet was twice washed with PBS buffer before being resuspended in 800 µL of PBS buffer. To 80 µL of the cells 3 µL of 25 mM CUG and 800 µL of oil phase (2.9% Abil EM90 in light mineral oil) were added. The sample was then homogenized on ice using a Fisher Scientific PowerGen 125 homogenizer with a
7x65 mm Fisher Scientific PowerGen disposable plastic generator at 9500 rpm for 5 minutes. Then the sample was submerged in ethanol at -78 °C for 1 minute, allowed to thaw and again placed in the -78 °C bath for 1 minute. The droplets were next incubated at 37 °C for 10 minutes. The droplets were cooled on ice and then 800 μL of aqueous layer (1.5% v/v medium viscosity carboxymethyl cellulose and 1% v/v Triton X102 in PBS buffer) was added. The mixture was homogenized at 8000 rpm for 3 minutes while on ice. The droplets were kept on ice until analysis. Just before FACS cell sorting, the droplets were diluted 10-fold in PBS buffer and passed through a 30 μm disposable Celtric filter. Droplets were triggered on the fluorescence of GFP (λ<sub>ex</sub> = 488 nm, λ<sub>em</sub> = 520 nm) to detect the presence of a cell within the droplet on a MoFlo XPD instrument. Those containing cells were analyzed for coumarin fluorescence at λ<sub>ex</sub> = 390 nm, λ<sub>em</sub> = 460 nm.

**In vivo FACS analysis**

Top10 cells harboring the theophylline riboswitch were grown overnight in a 3 mL LB media culture supplemented with 100 μg/mL ampicillin. This overnight culture was used to inoculate a new 7 mL culture supplemented with 100 μg/mL of ampicillin. After incubating at 37 °C, 250 rpm for 1 hour the culture was split into two 3 mL cultures and 1 was added to a final concentration of 1 mM to one of the cultures. Both were then returned to the shaker for incubation at 37 °C, 250 rpm for 3 hours. The cells were then twice washed with PBS buffer, spinning down at 2500 rpm for 5 minutes between washes. The cells were resuspended in 800 μL of PBS. To 100 μL aliquot 1 μL of 25 mM CUG of FUG was added and the mixture was incubated at 30 °C for 15 minutes. The cells were then spun down and twice washed with PBS buffer. The pellet was stored on ice until analysis at which time it
was resuspended in 1 mL of PBS buffer and analyzed ($\lambda_{\text{ex}} = 390$ nm, $\lambda_{\text{em}} = 460$ nm was used for CUG and FTIC was used for FUG).

**Barnase selections**

*E coli* BL21(DE3) cells harboring the pRS vector were grown in a 3 mL culture supplemented with 100 $\mu$g/mL ampicillin in the absence of ligand and IPTG overnight at 37 °C while shaking at 250 rpm. A 50 $\mu$L aliquot of the culture was used to inoculate a new 3 mL culture supplemented with 100 $\mu$g/mL ampicillin. The OD$_{600}$ of each of the overnight cultures was read. To this culture 1 mM IPTG and 1 mM theophylline was supplemented and the cells were allowed to incubate at 18 °C, 250 rpm overnight. 200 $\mu$L of the culture was then removed and the OD$_{600}$ and fluorescence ($\lambda_{\text{ex}} = 480$ nm, $\lambda_{\text{em}} = 520$ nm) were read on the plate reader.

**Mammalian cell work**

With the help of Jeane Govan from the Deiters’ Lab, DsRed under the control of the theophylline riboswitch and GFP were co-transfected into HeLa cells. The cells were incubated at 30 °C for two days. The wells were then imaged using filters for DsRed and GFP to look for fluorescence.
<table>
<thead>
<tr>
<th>SW</th>
<th>Primer sequences used for cloning of plasmid constructs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>5'-NNNNNTAAACAAACAGATG-3'</td>
</tr>
<tr>
<td>17</td>
<td>5'-GCAGGGTGTCGG-3'</td>
</tr>
<tr>
<td>18</td>
<td>5'-NNNNNGCAGCAGGGTGCTAAG-3'</td>
</tr>
<tr>
<td>19</td>
<td>5'-GCATCAAGACGATGC-3'</td>
</tr>
<tr>
<td>20</td>
<td>5'-GCATAAAGCTTGATACCAGCATCGTCTTGATGCCCTTGCGACACCCT</td>
</tr>
<tr>
<td></td>
<td>GCTAAGGGAACAAAAGAGTGCATAGCAAAGGAGAAAGAAGTTTTCAC-3'</td>
</tr>
<tr>
<td>21</td>
<td>5'-GCATGGATCCTTTATTTTGTAGAGCTCATCCATGCC-3'</td>
</tr>
<tr>
<td>22</td>
<td>5'-GCATAAAGCTTGATACCAGCATCGTCTTG-3'</td>
</tr>
<tr>
<td>23</td>
<td>5'-GGGTGGCTCAAAAAGGGAACCTTTCAGATCAGCTCGTCCGGGAAAA</td>
</tr>
<tr>
<td></td>
<td>GCATCGG-3'</td>
</tr>
<tr>
<td>24</td>
<td>5'-CCGATGCTTTTTCGGGAGCGACGTCTGCAAGGTTTCCCTTTTGATGC</td>
</tr>
<tr>
<td></td>
<td>CACCC-3'</td>
</tr>
<tr>
<td>25</td>
<td>5'-GATGTCGCTCCGGGGAAG-3'</td>
</tr>
<tr>
<td>26</td>
<td>5'-TGCAAGGTTCCCTTTTGTAGCCACC-3'</td>
</tr>
<tr>
<td>27</td>
<td>5'-GCAGAAGAAGCCGACGCAGAAGCCGACGAGTGAAGCAAGGAG</td>
</tr>
<tr>
<td></td>
<td>AAGGACTTTTTCAC-3'</td>
</tr>
<tr>
<td>28</td>
<td>5'-GCATGCGGCCGCCUGGCGGAAGCCGACGCAGCGAAGGACGCGCGAAGC</td>
</tr>
<tr>
<td></td>
<td>AAAGAAGCCGAG-3'</td>
</tr>
<tr>
<td>29</td>
<td>5'-GCATCATATGGATGTGCTCCGGGGAAAGC-3'</td>
</tr>
<tr>
<td>30</td>
<td>5'-GCATGGATCCTGGGTTCCCTTTTGATGCCACC-3'</td>
</tr>
</tbody>
</table>
CHAPTER 4

Introduction to polyketide synthases

4.1 Introduction

One of the overarching goals of synthetic biology is the manufacture of modular parts that can be interchanged to create new biological pathways for the synthesis of a wider array of final products. Evolution has provided enzymes capable of assembling and transforming common cellular substrates into a wide array of different final products. If enzymes from these pathways could be further tailored and interchanged, a vast library of natural products could be created.64

Modularizing the biosynthesis of polyketides is an attractive target for synthetic biology. Polyketides are a diverse group of natural products that include polyphenols, macrolides, polyenes, enediynes, and polyethers. Polyketides are important in pharmaceuticals as they serve as antibacterials, immunosuppressants, antiparasitics, cholesterol-lowering drugs, and antitumoral agents65 (Figure 4-1).
Polyketides are biosynthesized via condensation of acyl-CoA building blocks called extender units by polyketide synthases (PKSs). Type I PKSs are multifunctional enzymes organized into modules. Each module contains a set of non-iteratively acting enzymes that catalyze one cycle of polyketide chain elongation. An example of a type I PKS is 6-deoxyerythronolide B synthase (DEBS, Figure 4-2). PKSs such as DEBS use acyl carrier proteins (ACP) to tether the growing polyketide to the polypeptide during catalysis.\textsuperscript{66}
Figure 4-20: 6-deoxyerthronolide B synthase (DEBS), produces the precursor to erythromycin A. This is an example of a type I PKSs and also illustrates the colinearity between a PKS’s protein sequence and the structure of the final product.

Because all type I PKSs work in a similar manner, some groups hypothesized that whole modules or single enzymes could be swapped to produce new products.67,68,69 One of the areas often targeted for interchangeability is the AT domain, commonly referred to as the gatekeeper of the PKS as it selects which extender unit to install (Figure 4-3). Some limited success has been obtained with interchanging AT domains but the natural product was usually produced with higher yields.70,71
There are three classes of PKSs; types I, II, and III. Type I PKSs were discussed above and are illustrated by the DEBS pathway. DEBS consists of seven modules including the loading domain and is responsible for the synthesis of the erythromycin A scaffold. Each module in DEBS can be further broken down into domains including the acyl transferase (AT), keto-reductase (KR), enolreductase (ER), dehydrogenase (DH), ACP, β-ketoacyl-acyl carrier protein synthase (KS), and thioesterase (TE) domains. The size and complexity of the final polyketide are controlled by the number of acyl chain elongation steps, which is in turn controlled by the configuration of the modules making up PKSs. Thus, the structure of the
final polyketide can be predicted based on the identity and the location of the modules. The colinearity between PKS sequence and polyketide structure potentially allows for design of hybrid PKSs for the synthesis of novel polyketide analogues.

Type II PKSs are known for producing polyphenols with the use of discrete, monofunctional proteins (Figure 4-4). Unlike type I PKSs these type II PKS enzymes that act largely independent of one another. Type II PKSs are responsible for the production of anthracyclines, angucyclines, aureolic acids, tetracyclines, tetracenomycins, pradimicin-type polyphenols, and benzoisochromanequinones. Type II PKSs generally contain two KS domains, an ACP, and additional enzymes responsible for further tailoring of the product.

Type III PKSs produce smaller products than the type I and type II systems, creating simple aromatic compounds. Type III PKSs are defined by their lack of an ACP, instead interacting directly with the acyl-CoA. Type III PKSs are homodimers made up of a single ~40kDa domain. ACPs are not present in type III PKSs, instead one active site catalyzes the Claisen condensation, loads the starter unit and cyclizes the polyketide intermediate. Most importantly to this research, type III PKSs can display promiscuity towards acyl-CoA substrates.
The malonyl-Coenzyme A (malonyl-CoA) synthetase MatB has been used to synthesize acyl-CoA extender units for probing the specificity of PKSs (Figure 4-5) and ultimately providing a synthetic biology platform for diversifying polyketides. MatB from *Rhizobium trifolii* has been shown to accept a wide variety of malonic acid substrates\(^8^0\) while MatB from *Streptomyces coelicolor* was shown to be less promiscuous toward malonyl substrates.\(^8^1\) Interestingly, in addition to malonic acid promiscuity, previous work by Dr. Irina Koryakina (Williams Lab) and others showed that MatB is able to accept non-CoA thiol acceptors such as S-N-Acetylcysteamine (SNAC, 12) and pantetheine (14, Figure 4-5), but that the rates of SNAC and pantetheine incorporation were far slower than CoA.\(^8^0\)

**Figure 4-23:** (Top) Native reaction of MatB. The ATP-dependent MatB catalyzed reaction of the malonate and thiol 13 forms a malonyl-CoA. (Bottom) Proposed reactions obtainable with mutant MatBs. Additional malonate analogues are accepted as well as incorporation of SNAC (12) and pantetheine (14) in place of CoA (13).
The Williams group is interested in using MatB mutants to generate acyl-CoAs \textit{in vivo} by feeding non-natural malonate analogues to bacterial cultures that overexpress polyketide biosynthetic machinery. Previous work by Dr. Irina Koryakina illustrated that substitutions at positions Thr-207 and Met-306 lead to increased promiscuity towards various malonic acid analogues (Figure 4-6).\textsuperscript{82}

\begin{center}
\begin{tikzpicture}
\node at (0,0) {\text{\textbf{Figure 4-24: Malonate analogues accepted by mutant MatB enzymes. Additionally, MatB mutants that accept other hydrocarbon chains were found.}}}
\end{tikzpicture}
\end{center}

The use of SNAC-thioesters has several benefits over CoA-thioesters therefore, mutations that alter specificity of MatB towards the thiol acceptor substrate are also desired. For example, SNAC-thioesters are utilized by polyketide biosynthesis both \textit{in vivo} and \textit{in vitro} and are able to pass through bacterial membranes unlike CoA. Additionally, SNAC is significantly less expensive than CoA ($70/g compared to $8140/g, respectively, from
Sigma), and unlike CoA it is not provided by natural cell functions.\textsuperscript{83} Because SNAC is not native to the cell and thus has no native role, providing potentially high intracellular concentrations of 12 is less likely to disrupt metabolic activity within the cell. In contrast, overexpression of enzymes which disturb intracellular CoA concentrations often have deleterious effects on cell viability.\textsuperscript{84}
5.1 MatB engineering for SNAc and pantetheine

Evolving the substrate specificity of MatB to accept N-Acetylcysteamine (SNAc, 12) in place of Coenzyme A (CoA, 13) would allow for a dramatic cut in screening costs and would also prevent the reaction from using a substrate that is multifunctional in the cell, avoiding possible unwanted side-effects. The goal is to engineer MatB to provide mutants with improved specificity towards 12, and concomitant reduction in specificity towards CoA. This could provide MatB mutants with sufficient activity to efficiently generate acyl-SNAc extender units for in vitro PKS reactions. More significantly, the eventual goal is to use 12 as a more economical acceptor handle, and avoid problems associated with CoA toxicity in vivo.

Before initiating a program of protein engineering, activity of wild-type MatB with both 12 and 13 was determined by high-performance liquid chromatography (HPLC, Figures 5-1 and 5-2).
Figure 5-25: MatB completely converts 13 to product within 5 minutes. Absorbance values were taken at 254 nm.

Figure 5-26: MatB reactions with 12 fail to reach completion even after 3 hours. Absorbance data was collected at 235nm.
HPLC analysis confirms that the truncated acceptor 12 is a poor substrate for MatB, compared to 13. Only 39% conversion of 12 to malonyl-SNAC is observed within three hours while complete conversion is achieved with 13 in less than five minutes. The poor inefficiency of 12 as an acceptor substrate needs to be overcome in order to make MatB-catalyzed synthesis of acyl-SNACs a viable option in vitro or in vivo. In addition to screening for activity with 12, mutant MatB activity with 14 was also examined. The longer chain length of 14 more closely mimics 13, which may aid in detecting a more active MatB mutant. Like 12, 14 is less expensive than 13, is not native to the cell, and exhibits low activity with WT MatB.

Critical to the success of enzyme engineering is the availability of a suitable high-throughput screen or selection to identify improved enzymes from libraries of at least several hundred variants. Dr. Irina Koryakina had previously designed and executed a method for detecting activity of MatB mutants with unnatural malonates (Figure 5-3). A mutant library was prepared and individual colonies were picked into 96-well plates. The cells were grown to an optical density at 600 nm (OD$_{600}$) of 0.6 and then protein expression was induced. After expressing the mutant MatB protein the cells were lysed through a single freeze-thaw cycle in the presence of lysozyme. The cleared cell extracts were then added to a reaction mixture containing the necessary substrates and cofactors and incubated for a set time before being quenched with Ellman’s reagent and read on the plate reader (Figure 5-4). Here, a similar screening strategy was employed in a high-throughput microtiter plate screen to report the activity of MatB mutants towards the thiol acceptors 12, 13, and 14.
Much of the process of preparing the lysed cells for the reaction had been optimized previously, so the focus of this optimization was instead on the incubation time and adjusting the amount of Ellman’s reagent used to quench the reaction. After several attempts an optimized method was obtained. This optimized method allowed enough time for the reaction with \( \textit{12} \) to occur while still being short enough to observe which mutants had the fastest rate of reaction. Control plates run with WT MatB and no enzyme controls (pET28a) were used to test the method (Figure 5-5). The conditions used provide a clear difference between the positive and negative controls while still leaving a wide range for detection of improvements as illustrated by the \( \textit{13} \) controls, showing the results expected from complete conversion of thiol substrate.
Figure 5-29: Optimized reaction conditions allow room for detection of MatB mutants with improved performance with 12. Colonies were ordered by descending order of absorbance and used to find the expected averages and standard deviations of WT MatB and negative control absorbance values from the screening method. The MatB reaction with CoA was used to indicate the minimal expected absorbance.

With an optimized screening method in hand, MatB residues were chosen for mutagenesis. Upon analysis of the MatB crystal structure, Leu-183, Pro-184, His-189, Val-230, Met-233, Arg-236, Thr-243, Ser-261, Arg-283, Gly-285, Gly-392, Gly-393, Tyr-394, Gly-404, Asp-423, Leu-424, His-457, and Arg-461 were selected for saturation mutagenesis (Figure 5-6). The entire binding pocket was targeted for mutagenesis to increase the odds of finding an effective MatB mutant. Mutagenesis was focused on those residues known to bind the thiol (Figure 5-7). In another attempt to increase the probability of finding a hit mutant, screening for increased activity with 14 was also tested.
Figure 5-30: MatB residues chosen for mutagenesis and screening with 12 and 14. The CoA arm is shown in grey.

Figure 5-31: Substrate binding pocket of MatB showing interactions with methylmalonyl-CoA.81
The wild-type codon of each residue selected for mutagenesis was substituted with NNK (N=A,T,G,C; and K=G or T) by saturation PCR mutagenesis. The NNK libraries limit the number of possible codons to 32. The libraries also eliminate two of the three possible stop codons. Some amino acids are still oversampled with this library as they are in the full NNN library. However, an NNK library requires only 172 and 224 library members to have 95 and 99% certainty of screening the entire library, respectively. The NNN library would require screening 240 and 338 colonies to achieve the same certainties. Therefore, screening 200 mutants from each library provided >97% certainty of screening all of the possible mutants in each library. For each residue, forward and reverse oligonucleotides were designed to include NNK at the desired site and were used to generate each saturation library using the T207S/M306I MatB mutant (previously shown by Dr. Irina Koryakina to be promiscuous with malonate derivatives) as template DNA according to the Invitrogen Site Directed Mutagenesis II Kit. Following transformation of each PCR product mixture into E. coli BL21(DE3), DNA sequencing of randomly chosen transformants confirmed that the site-directed saturation mutagenesis reaction successfully mutated the target codon in a random manner (data not shown).

A total of 300 colonies from each saturation library were primary screened. Then the two to three most active mutants from each 96-well plate were picked for secondary screening (Figure 5-8). First the colonies were again tested using the same 96-well plate format as before, but this time each mutant was tested in triplicate. This removed false positives before additional screening. Next, the hits that remained were grown on a 100 mL scale, and the mutant MatB was expressed and purified using either nickel beads or fast
protein liquid chromatography (FPLC). These purified enzymes were then used to catalyze *in vitro* reactions with 12 and analyzed by HPLC.

Problems arose while screening MatB activity with 12 and 14. Low resolution mass spec showed that the 12 and the 14 were forming dimers, leaving them unable to complete the MatB reaction. New stocks were ordered and the problem persisted. The reactions could not be run with a reducing agent present as that led to high background in the Ellman’s assay. When the reagents were reduced and purified before screening high background in the Ellman’s assay was still present, though in reactions with 12 it was within the acceptable limit. Another problem with the substrates was in the secondary screen. Both the 12 and 14 had much lower extinction coefficients than 13, making detection of changes in activity by HPLC more difficult to quantify. Increasing the concentration of 12 or 14 was not an option as this would alter the reaction conditions, thus making the screening methods inconstant.
Select top performing mutants from primary screen

Perform second round of plate based Ellman's assay to remove false positives

Pick top performing mutants from secondary assay

Culture cells and express protein at 100 mL volume

Purify protein with Ni beads or FPLC

Run in vitro reaction and analyze by HPLC

**Figure 5-32:** Flow chart illustrating secondary screening procedure to identify mutant MatBs with improved activity.

Unfortunately, MatB activity with 12 and 14 failed to improve significantly with mutations at the sites shown above. Given this panel of saturation mutagenesis libraries failed
to give the desired results, whole gene random mutagenesis was utilized. This was first avoided due to the moderate throughput of the screening method, which allows only a small percentage of the total number of mutants to be screened. Random mutagenesis was obtained by PCR amplification of the T207S/M306I MatB gene using Mutazyme DNA polymerase, which is used due to its high error rate. The PCR reaction conditions were adjusted according to the manufacturer’s instructions to obtain an ideal mutation rate of 1-2 amino acid substitutions per gene product,\textsuperscript{86} which was confirmed by DNA sequencing randomly selected transformants.

However, screening 30,000 members of the error prone PCR library failed to yield mutations that showed improved activity with 12. However, useful information was extracted from this library by DNA sequencing the top 50 performing library members. This analysis revealed that mutations at residues A15, V180, A365, and L388 were more frequently observed than would be expected by random chance. This suggests that substitution at residues A15, V180, A365, and L388 are the most important for at least retaining activity with 12. Thus, given it was unlikely that every other substitution at these positions had previously been screened, each residue (A15, V180, A365, and L388) was subjected to saturation mutagenesis and the subsequent three libraries each screened for activity with 12 as the thiol acceptor. Gratifyingly, several library members with improved activity compared to the wild-type MatB were found, the most active single mutant displayed \(\sim3\)-fold increased activity with 12 (Figure 5-9) compared to the WT enzyme.
Combinations of the most active single mutations were created to test for further improvement with 12. Each single mutant from Figure 5-9 was used as the template for saturation mutagenesis at each of the other three locations, creating 12 new libraries. For each library, 200 mutants were primary screened for improved activity with 12. The top 7-10 mutants from each library were then selected for secondary screening and sequencing. Screening of these libraries yielded several mutants with improved activity towards 12, with the most active mutant, A15L/L388M, showing ~5-fold improvement over WT MatB. A similar strategy was used for additional mutation sites that were uncovered while screening the randomly mutated library (Figure 5-10). A mutation tree shows how each of the mutations in Figure 5-10 was found (Figure 5-11).
In addition to the saturation libraries, additional randomly mutated MatBs were created using the mutants shown in Figure 5-10 as the template for further mutation. A library of 5000 mutants was primary screened, 1000 mutants from each of the top 5 mutants found previously. The most active 3 mutants from each plate were selected for secondary screening. However, no additional increase in activity was observed from these mutations.
As illustrated above, the mutant found from the original randomly mutated library was often not the most active mutant. Figure 5-11 shows the relative increase in reaction rate of each set of mutations found. Surprisingly, most of the mutation sites found were not located near the binding site (Figure 5-12).
Figure 5-36: Locations of mutations yielding the highest increase in MatB activity with 12. Note that the locations are largely on the exterior of the enzyme and not in close proximity of the binding pocket (13 arm shown in blue).

Though the 5-fold improvement in activity with 12 are significant, compared to the rate MatB has with 13, 12 is still slow. Previous MatB engineering performed by Dr. Irina Koryakina showed a increases in activity with unnatural malonate substrates ranging from <2 all the way to a >460-fold increase. Because this is such a broad range and different substrates are being examined it is difficult to make comparisons between her work and the work presented here.

It is important to note that most mutants did not show a decrease in activity with 13, and competition assays showed that it was still the preferred substrate. Additionally, the T207S/M306I MatB mutant that was used as the initial template was more stable than the top performing mutants. These new mutant MatBs did not survive freeze-thawing as well and also precipitated out of solutions more easily than the template MatB enzyme.
5.2 MatB engineering with unnatural malonates

In addition to improving MatB activity with unnatural thiol acceptors, improving activity towards new malonate analogues was an important target. Unlike the thiol acceptors, the malonates are incorporated into the final polyketide product. Therefore, having the ability to incorporate new functionalities in these side chains will have a direct effect on the performance of the polyketide product. Previously, Dr. Irina Koryakina had found that mutations at T207 and M306 had dramatic effects on the promiscuity of the enzyme for non-natural malonates. However, these studies were limited to analogues with alkyl, azido, alkynyl, and allyl side chains, with little exploration of hetero-atom containing functional groups. Using this information, libraries of T207X, M306X mutants were prepared and screened for activity with hydroxy- (15), fluoro- (16), chloro- (17), and bromomalonate (18). Each malonate was prepared through saponification of the corresponding diester, which are each commercially available. The overall goal is to further broaden the specificity of MatB to accept these halogenated malonic acids. There are approximately 4000 known halogenated compounds produced by living organisms. Both the structure and the function of these compounds are diverse with functions ranging from antibiotics, to antitumor compounds, to proteasome inhibitors. Additionally, 20 to 30% of drugs contain at least one fluorine atom. This makes creation of halogenated products a promising approach for the discovery of biologically active polyketides.
Previous work showed that MatB displays poor activity with 15 and 16. Therefore initial testing was conducted to establish whether such low activity could be detected using Ellman’s assay (Figure 5-13). Gratifyingly, the colorimetric assay was able to report low level of activity towards 15 and 16 using the wild-type MatB, while little or no activity was detected with 17. Unexpectedly, 18 was a detectable substrate for the WT MatB, as judged by the decrease in $A_{420}$ compared to that in the absence of enzyme. After a 4 hour incubation time, 18 showed approximately 15 percent yield, while the natural substrate, malonic acid, is completely converted to product within 5 minutes.

![Figure 5-37: Activity of WT MatB with the halogenated malonates. Absorbance values show the results of an Ellman’s assay run either in the presence of or absence of WT MatB enzyme from lysed cells. Error bars show the standard deviation of eight experiments.](image)

Using two mutant libraries, T207X and M306X, previously prepared by Dr. Irina Koryakina, microplates of crude extracts containing the mutants were screened via the Ellman’s assay, in the same manner the thiol acceptors were screened (Figure 5-14).
Remarkably, library members with improved activity compared to wild-type MatB were identified with all three substrates, 16-18. After secondary screening to eliminate false positives, as described previously, the resulting top mutant with each substrate was assayed by HPLC analysis of reaction mixtures using purified enzyme (Figure 5-15). MatB T207A was the most active mutant with both 17 and 18, while 16 showed highest activity with the double mutant T207L/M306A.

**Figure 5-38:** Activity of T207X/M306X MatB mutants with halogenated malonates. Absorbance data shows the results of an Ellman’s assay used to determine the concentration of 13 remaining after the reaction. Activity with WT MatB with each substrate is shown for comparison.
The activity of the MatB T207A mutant, showed an increase in activity with all three substrates. However the MatB T207L/M306A mutant showed the highest activity with 16. To more accurately describe potential changes to substrate specificity, a HPLC-based end-
point assay was used to collect steady state kinetic parameters for WT MatB, T207A, and T207L/M306A with 16-18 (Table 5-1). The WT enzyme shows a 41 and 410-fold lower catalytic efficiency for 16 and 18, respectively, when compared to methylmalonic acid, as judged by $k_{\text{cat}}/K_m$. Kinetic parameters of the WT enzyme with methylmalonic acid and 16 were in close agreement with literature values \textsuperscript{81,88} Notably, introduction of the single amino acid mutation T207A led to a 1.8- and 11-fold improvement in catalytic efficiency towards 16 and 18, respectively. These improvements were largely due to an increase in the $k_{\text{cat}}$ for each substrate. In agreement with Ellman’s assay of crude extracts containing WT MatB, activity of WT MatB with 17 was not detectable by HPLC, Remarkably however, activity of the MatB mutant T207A could be detected, although the catalytic efficiency of this mutant towards 17 remains 70-fold lower than the WT enzyme with methylmalonate. Nevertheless, this represents at least a 78-fold improvement in catalytic activity towards this substrate based on the estimated detection limit of $<0.004 \text{mM}^{-1} \text{s}^{-1}$. The T207L/M306A double mutant showed little activity with 17 or 18. In every case, product identity was confirmed by LC-MS analysis of the product mixtures (Table 5-2).
Table 5-7: Steady state kinetic parameters for WT and mutant MatB. ND signifies that the value could not be determined.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Parameter</th>
<th>Methylmalonate</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>2.29 ±0.3</td>
<td>0.27 ±0.02</td>
<td>ND</td>
<td>0.17 ±0.05</td>
</tr>
<tr>
<td></td>
<td>$K_M$ (mM)</td>
<td>0.11 ±0.02</td>
<td>0.54 ±0.05</td>
<td>ND</td>
<td>3.25 ±0.34</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_M$ (mM$^{-1}$ s$^{-1}$)</td>
<td>20.82</td>
<td>0.5</td>
<td>ND</td>
<td>0.05</td>
</tr>
<tr>
<td>T207A</td>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>2.12 ±0.23</td>
<td>0.41 ±0.04</td>
<td>0.71 ±0.08</td>
<td>0.94 ±0.04</td>
</tr>
<tr>
<td></td>
<td>$K_M$ (mM)</td>
<td>0.14 ±0.01</td>
<td>0.46 ±0.03</td>
<td>2.31 ±0.35</td>
<td>1.72 ±0.21</td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_M$ (mM$^{-1}$ s$^{-1}$)</td>
<td>15.14</td>
<td>0.89</td>
<td>0.31</td>
<td>0.55</td>
</tr>
<tr>
<td>T207L/M306A</td>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>2.19 ±0.27</td>
<td>0.48 ±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_M$ (mM)</td>
<td>0.12 ±0.02</td>
<td>0.41 ±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_{\text{cat}}/K_M$ (mM$^{-1}$ s$^{-1}$)</td>
<td>18.25</td>
<td>1.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5-8: Mass spectrometry confirmation of the identity of MatB products.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Calculated mass (Da)</th>
<th>Experimental mass (Da)</th>
<th>ΔDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>870.1183</td>
<td>870.1174</td>
<td>0.0009</td>
</tr>
<tr>
<td>20</td>
<td>872.114</td>
<td>872.1137</td>
<td>0.0003</td>
</tr>
<tr>
<td>21</td>
<td>888.0844</td>
<td>888.0811</td>
<td>0.0033</td>
</tr>
<tr>
<td>22</td>
<td>932.0339</td>
<td>932.0371</td>
<td>-0.0032</td>
</tr>
</tbody>
</table>

In addition, mass spectrometry revealed that 17 and 18 were hydrolyzing over time. After a few weeks the amount of hydroxymalonate in the stocks increased from not detectable to more than half the total malonate concentration. The acyl-CoA stocks were kept
frozen except for when they were in use and this problem persisted. For this reason it is recommended that small scale saponifications be run as needed instead of doing larger scale reactions and saving the product. Otherwise product may be stable if stored as a dry solid. Hydrolysis of 16 was not observed.

The malonyl-CoAs (19-22) were then tested for acceptance by polyketide synthases. Sarah Schultheis provided purified modules from the erythromycin (Ery) and pikromycin (Pik) PKS pathways. Activity with modules Ery3, Ery5, Ery6, PikIII, and PikIV from these PKS pathways was tested with 19-22 (Figure 5-16). Previous work showed that 20 was accepted by Ery3 and Ery6 when in the presence of the trans-AT from the disorazole PKS.\(^8\) Reactions were analyzed by mass spectrometry (Table 5-3). It was found that 19 is accepted by PikIII and Ery3, and that no other unnatural substrate was accepted. These trials are being repeated in the presence of trans-AT. The other three substrates were not shown to be accepted by any of the modules.

\[ \text{Figure 5-40: Scheme of the single module reaction of 19-22 with diketide-SN Ac.} \]
Table 5-9: Mass spectrometry results for malonyl-CoAs used in single module polyketide reactions with diketide-SNAC. ND denotes that no product was detected.

<table>
<thead>
<tr>
<th></th>
<th>Methylmalonyl-CoA</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>PikII</td>
<td>1.03E+05</td>
<td>1.00E+05</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>PikIV</td>
<td>1.24E+05</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Ery3</td>
<td>7.53E+04</td>
<td>4.78E+03</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Ery5</td>
<td>1.46E+05</td>
<td>N.D.</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Ery6</td>
<td>1.38E+05</td>
<td>N.D.</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

Activity of 19-22 was also tested with a panel of Ery6 mutants that had previously been shown to utilize unnatural substrates by Dr. Irina Korykina and Christian Kasey. The reactions yield a 10-deomythylethynolide (10-dml) product (Figure 5-17).91 Again, reactions were analyzed by mass spectrometry.

Unfortunately, none of the mutants were capable of accepting the halogenated substrates. The reactions were repeated and the initial results were confirmed. These preliminary results were not surprising as previous work required a trans-AT. Additionally, the mutations on Ery6 were not selected to perform with these substrates. Further
experiments are being conducted to increase the likelihood of a PKS module being able to accept 20-22.

Phung Nguyen had previously worked with the enzymes responsible for making germicidin, which is produced by a type III PKS. Three enzymes are used to synthesize the germicidine derivative product: FabD loads a malonyl-CoA onto a holo-ACP, FabH then catalyzes the condensation of malonyl-ACP with isobutyryl-CoA, and finally, germicidine synthase (Gcs) catalyzes the condensation and cyclization of this product with a substituted malonyl-CoA to form the germicidine product (Figure 5-18). FabD is not required for the reaction, but is believed to increase the rate of reaction.92 These enzymes were found to be promiscuous with malonyl-CoAs, so 19-22 were tested for incorporation. Cell lysates containing each enzyme were tested for activity with 19-22 as well as with the natural methyl- and ethylmalonyl-CoA substrates (Table 5-4).
Figure 5-42: Germicidine biosynthetic pathway.

Table 5-10: Mass spectrometry results for Gcs reactions with different malonyl-CoA substrates. With 21 and 22 substrate reactions the hydrolysed product was observed at low levels. Reactions were run with enzymes in cell lysate.

<table>
<thead>
<tr>
<th>Methylmalonyl-CoA</th>
<th>Ethylmalonyl-CoA</th>
<th>19</th>
<th>20</th>
<th>21*</th>
<th>22*</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.41E+06</td>
<td>1.80E+06</td>
<td>3.34E+05</td>
<td>3.80E+04</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>
Remarkably, mass spectrometry results show that substrates 19 and 20 are both utilized by Gcs. However, 21 and 22 failed to incorporate. Low concentrations of the hydrolyzed form of these products were observed, likely resulting from Gcs accepting the low amount of 19 present due to the hydrolysis of the 21 and 22.

5.3 Conclusions and future work

The results indicate that the benefits of developing a MatB mutant that can accept 12 or 14 over 13 are outweighed by the difficulty of finding such a mutant. Screening in excess of 30,000 mutants could not yield a mutant that showed preference of 12 or 14 over 13 and showed only small increases in 12 utilization in comparison to 13. Screening MatB with additional 12 and 14 mutants may still prove fruitful. Being able to create mutant MatBs that recognize a variety of thiol handles could be useful after additional AT domain engineering. If the AT domains could be engineered to recognize these different handles it may be possible to create full length polyketides with modifications at specific locations.

Mutating MatB to accept different malonates continues to prove fruitful with the ability to make malonyl-CoAs with halogens. These malonyl-CoAs have the potential to add chemical diversity to polyketides. Rachael Hall will continue work on incorporating these unnatural substrates into germicididine analogues with mutant Gcs.
5.4 Methods

General methods and materials

The MatB gene from *R. trifolli* was previously synthesized by GeneScript and cloned into pET28a with *NdeI* and *HindIII* restriction sites. All plate reader assays were performed on a BioTek Hybrid Synergy 4 plate reader (Winooski, VT, USA).

Purification of MatB protein

The MatB in pET28a was transformed into *E. coli* BL21 (DE3) cells and plated on LB agar plates containing 50 µg/mL kanamycin. A single colony was then selected and grown overnight at 37 °C, 250 rpm in 3mL LB media supplemented with 50 µg/mL kanamycin. This culture was used to inoculate a 500 mL culture of LB media containing 50 µg/mL kanamycin. The culture was allowed to grow at 37 °C, 250 rpm to an optical density of 0.6 at 600 nm at which time it was induced with IPTG (final concentration of 1mM) and incubated for approximately 18 hours at 18 °C, 250 rpm. The cells were centrifuged at 10,000 g and resuspended in lysis buffer (50mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication. The debris removed by twice centrifuging at 10000 g. The soluble extract was then run on a fast protein liquid chromatograph (FPLC) with a nickel column to bind the His-tagged protein using a gradient of wash buffer (20 mM sodium phosphate, pH 7.4 containing 0.5 M NaCl and 20 mM imidazole) and elusion buffer (20 mM sodium phosphate, pH 7.4 containing 0.5 M NaCl and 200 mM imidazole). A 30000 MWCO centrifugal filter was then used to concentrate the purified protein and buffer exchange it to a 100 mM Tris-HCl containing 10% glycerol. An SDS-PAGE gel was used to
confirm the protein purity and the protein concentration was found via a Bradford Assay using BSA as a protein standard.

**HPLC analysis of MatB activity**

MatB activity was tested via HPLC analysis. Reaction mixtures containing sodium phosphate (100 mM, pH 7), MgCl₂ (2 mM), ATP (0.4 mM), coenzyme A (0.2 mM), malonate (0.8 mM), and MatB (2 µg per 200 µL) were allowed to react at room temperature and then quenched with ice-cold methanol. Reactions were centrifuged at 10000 g to remove the protein and then run on an HPLC using gradients of 0.1% TFA in water (solution A) and methanol (solution B) by using the following gradient: 0-32 min, 80% B; 32-35 min, 100% A. The flow-rate was 1 mL/min. Absorbances were read at 254 nm as sample eluted from a Pursuit XRs C18 column. Similar reactions were conducted with an equal concentration of 12 in the place of 13 which were read at 200 and 235 nm.

**Plate reader assays of MatB activity**

Ellman’s reagent was prepared by dissolving 77.1mg of 5,5’-dithiobis(2-nitrobenzoic acid) in 25 mL of 0.1 M sodium phosphate buffer and adjusting the pH to 7. A calibration curve for the Ellman’s reagent was made with dilutions of both 13 and 12. 15 µL of Ellman’s reagent was added to 200 µL reaction mixtures in sodium phosphate buffer containing MgCl₂ (2 mM), ATP (0.4 mM), malonate (0.8 mM), either 12 (0.2 mM) or 13 (0.2 mM) and either purified MatB protein or soluble extract from MatB producing cells. Absorbance at 412 nm was then read on the plate reader.
Kinetics determination for MatB

MatB kinetics for \textbf{13} and \textbf{12} tested via HPLC analysis. Reaction mixtures containing sodium phosphate (100 mM, pH 7), MgCl\textsubscript{2} (2 mM), ATP (0.4 mM), malonate (0.8 mM), MatB (0.5 \mu g per 200 \mu L), and varying amounts of either \textbf{13} were allowed to react at room temperature for 10 minutes and then quenched with ice-cold methanol. Reactions were centrifuged at 10000 g to remove the protein and then run on an HPLC using gradients of 0.1% TFA in water and methanol. Absorbances were read at 254 nm as sample eluted from a C18 column. Similar reactions were conducted with an equal concentrations of \textbf{12} in the place of \textbf{13}. These were allowed to react for 30 minutes and were read at 200 and 235 nm by HPLC analysis using the gradient above.

Expression in 96-well plates

For expression in 96-well plates an Eppendorf epMotion liquid handling machine was utilized for liquid transfer steps. To each well 1 mL of LB media supplemented with 10 \mu g/mL of ampicillin was added. Colonies were then picked from LB agar plates with \textit{E. coli} BL21 cells containing pET22B, pET22B-CloL, or pET22B-mutant cloL and used to inoculate the wells. These plates were then incubated at 37 °C, 350 rpm overnight. New plates were prepared with wells containing 1 mL LB media supplemented with 10 \mu g/mL ampicillin and these were inoculated with 100 \mu L of the overnight plate. This was incubated at 37 °C, 350 rpm for approximately 3.5 hours at which time IPTG was added to a final concentration of 1mM and incubation was continued at 20 °C, 350 rpm for approximately 18 hours. Plates were centrifuged at 5000 g and the supernatant poured off. The cells were then resuspended in 100 \mu L of 50 mM Tris-HCl containing 10% glycerol and 2 mg/mL of
lysozyme and subjected to one freeze-thaw cycle. The cellular debris was then removed by centrifugation at 5000 g. 25 µl of the supernatant was then transferred to a microtiter plate containing 100 µL of reaction mixture containing ADHM, (0.25 mM), p-anisic acid, (1 mM), ATP (0.5 mM), MnCl₂ (5 mM), and Tris-HCl (50 mM). The reaction was incubated at 30 °C for 30 minutes then quenched with an equal volume of ice-cold methanol. The fluorescence was then read on a plate reader with λ_{ex}= 355 nm and λ_{em}=460 nm.

**Saponification of malonate diesters**

Into 10 mL of water NaOH (2.4 g, 60 mmol) and diester malonate (6 mmol) was stirred overnight at 65 °C. The solution was then cooled on ice before cold 12 M HCl was slowly added until the pH dropped to 1.5-2. The solution was then extracted 4 times with diethyl ether. The combined organic extracts were dried over MgSO₄ then filtered. The ether was removed under vacuum to yield white solid.

**Synthesis of diketide SNAc**

Meldrum’s acid (2.88 g) was dissolved in 40 mL of distilled DCM and stirred in an ice bath. Pyridine (3.25 mL) was then added and the solution stirred for 20 min. A solution of propionyl chloride (1.75 mL) in 10 mL distilled DCM was slowly added. The reaction was stirred on ice which warmed to room temperature overnight. The reaction was washed with 0.1 M HCl 3 times, and the aqueous fractions were extracted with DCM then dried over MgSO₄ and filtered. The solvent was then removed under vacuum and recrystallized in petroleum ether to yield yellow solid, acylated meldrum’s acid.
The acylated meldrum’s acid (1 g) was dissolved in 5 mL toluene and SNAc (0.6 g) was added. The reaction was stirred at 80 °C under N₂ overnight. Solvent was removed under vacuum. Solid was washed with hexanes and then the solvent was removed.

This product was dissolved in 10 mL of distilled THF with potassium tert-butoxide (0.3 g) and stirred at 0 °C for 10 minutes. Iodomethane (0.85 mL) was slowly added and the reaction was stirred at 0 °C warming to room temperature overnight. The reaction was quenched with 50 mL of 0.1 M HCl and extracted with 50 mL ethyl acetate 3 times. The combined organic extracts were washed with brine, dried over MgSO₄ then filtered. The solvent was removed under vacuum and the diketide SNAc was column purified.

**DEBS holo-Mod6TE reactions**

DEBS holo-Mod6TE reactions were prepared with by adding 90 µg DEBS holo-Mod6TE to 35 µL of 50 mM Tris-HCl (pH 7) containing 2 mM MgCl₂, 5 mM diketide SNAc, and 4 mM acyl-CoA (from MatB reactions). Reactions were incubated at room temperature overnight and then analyzed by LC-MS.

**Pentaketide reactions**

Reactions were run with a total volume of 80 µL in 100 mM KH₂PO₄ (pH 7) containing 5 mM MgCl₂, 1 mM pentaketide, 3 mM unnatural malonyl-CoA, 5 mM glucose-6-phosphate, 0.5 mM NADP⁺, 0.3 U glucose-6-phosphate dehydrogenase, and 25 µg mutant Ery6 enzyme. Reactions were run overnight at room temperature then quenched with 80 µL ice-cold methanol and then analyzed by LC-MS.
**Germicidine reactions**

*In vitro* Gcs reactions were run in 100 µL containing 1 mM isobutyryl-CoA, 2 mM malonyl-CoA, 2 mM malonyl-CoA analogue, 2 mM TCEP, 10 mM magnesium sulfate, 50 mM HEPES (pH 7), 150 mM NaCl, 10 µg FabH, and 10 µg Gcs. Reactions were run overnight at room temperature then analyzed by HPLC at 290 nm and LC-MS.

**Mutagenesis of MatB**

Random mutagenesis was performed using the Genemorph Mutazyme II Kit according to the protocol. T7 forward and T7 reverse primers were used for the PCR reaction. Site-directed mutagenesis was performed using the Invitrogen Site-Directed Mutagenesis Kit according to the protocol. Primers used for mutations at each location are shown in Table 5-5.
**Table 5-11:** Forward and reverse primer sequences for mutagenesis at each residue in MatB.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>L183X</td>
<td>5'-GCTGTCCATGCCNNKACGCGTCTCG-3'</td>
<td>5'-CACGCAGCGTMNNNGCATTGGACAGC-3'</td>
</tr>
<tr>
<td>P184X</td>
<td>5'-CACGCCCTGNNKATCTTCTCATACC-3'</td>
<td>5'-GGTATGAAAGATMNNCAGGCCGTG-3'</td>
</tr>
<tr>
<td>H189X</td>
<td>5'-CGGCCCTGATTNNKCGCGCCCTG-3'</td>
<td>5'-CGGCACGGGCMNNAAACAGACCGG-3'</td>
</tr>
<tr>
<td>V230X</td>
<td>5'-GCCGACGAANNKGTAGCTGTATG-3'</td>
<td>5'-CATCAGACTAACMNNTCTGTCGG-3'</td>
</tr>
<tr>
<td>M233X</td>
<td>5'-GTAGTCTGTNNKCCCGCAGCAAC-3'</td>
<td>5'-GTTGCCTGCAGGGMNACAGACTAAC-3'</td>
</tr>
<tr>
<td>T243X</td>
<td>5'-CGCAGGCAANNKATGCTGTATG-3'</td>
<td>5'-CCATCAGCATMNNTGCTCGG-3'</td>
</tr>
<tr>
<td>S261X</td>
<td>5'-GTCTGCTGCAANNKCCCGCCTG-3'</td>
<td>5'-CAGGCCGGGMNNTTGCACGAC-3'</td>
</tr>
<tr>
<td>R283X</td>
<td>5'-GAATTCACGCTTNNKACCATTGCAG-3'</td>
<td>5'-CGTACCGGTMNNAGGTGAATTTC-3'</td>
</tr>
<tr>
<td>G285X</td>
<td>5'-CTCGTACCNNKACGCGGATTC-3'</td>
<td>5'-GAATCGGCTGMNNNGTACGAG-3'</td>
</tr>
<tr>
<td>G392X</td>
<td>5'-CATATTGTCCNNKCGTGTAAAG-3'</td>
<td>5'-CTTTACCACGGMNNACAGAC-3'</td>
</tr>
<tr>
<td>G393X</td>
<td>5'-GTGGCCCCTTNKOAAAGATCTGGG-3'</td>
<td>5'-CCAGATCTTTMNNACCGGACG-3'</td>
</tr>
<tr>
<td>Y394X</td>
<td>5'-CCGCCGAGGCGNNKGTTCATATGG-3'</td>
<td>5'-CAATATGAACMNNGCCTTCGGG-3'</td>
</tr>
<tr>
<td>G404X</td>
<td>5'-GGTATTTTCNNKKGTTAACCAC-3'</td>
<td>5'-GAATGTTTAAACCMMNTGAAATAACC-3'</td>
</tr>
<tr>
<td>D423X</td>
<td>5'-GGTGAATTTNKNKAATCGAAGGG-3'</td>
<td>5'-CGCTGATTGTTGMNNAATTTCAACC-3'</td>
</tr>
<tr>
<td>L424X</td>
<td>5'-GAAATCTGACAAKNKGAAGCGCTCG-3'</td>
<td>5'-CGACGCGTTCMTMNNTGATCAAATTC-3'</td>
</tr>
<tr>
<td>H457X</td>
<td>5'-GGTGTCCGGNNKCCGGGATTTTGG-3'</td>
<td>5'-CCAAAATCCGGGMNNCGGAACC-3'</td>
</tr>
<tr>
<td>R461X</td>
<td>5'-CTGCAGGATNNKCTGGCTGTAT-3'</td>
<td>5'-ATAGCGGAGCCGMNNATCCTGAG-3'</td>
</tr>
</tbody>
</table>
CHAPTER 6

Determination of free oil volumes from corn

This research was conducted at Novozymes in Franklinton, NC. Due to agreements with the company, enzyme names will not be used, methods will not be described in detail and future directions will be kept brief.

6.1 Introduction

In attempts to increase profits, production facilities that generate ethanol from corn have invested in building corn oil extraction facilities. Currently, ethanol producing facilities are producing approximately 3 billion pounds of corn oil annually, which translates to about $1 billion dollars per year. However, current extraction techniques only remove around one-third of the oil contained within the kernel. Current techniques used for the removal of higher percentages of oil are not cost effective as they require large initial investments and high energy costs. However, for every one-percent increase in extractable oil, profits increase by ~$30 million. For this reason several companies are competing to develop technology to increase oil yields.
Figure 6-43: Structure of the corn kernel. The germ, located at the center of the corn kernel, contains the majority of the oil enclosed within the kernel. The germ is 30-40% oil by weight.

The majority of oil found in the corn kernel is found in the germ, located in the center of the kernel (Figure 6-1). The germ is 30-40% oil by weight depending on the breed and growing conditions of the corn. There are two common techniques for producing ethanol from corn; one uses dry grinding and the other wet milling. In dry grinding the kernel is ground in the absence of water and the components of the kernel are processed together. With this method oil is extracted after fermentation. Oil is removed from the resulting thin stillage after it has been concentrated into condensed distiller’s syrup (CDS). Wet milling steeps the corn kernels before grinding, allowing the kernel to be separated into its components. Here the germ is separated before fermentation and the oil is removed, most commonly, by hexane extraction.

Hexane extractions fail to remove the majority of the corn oil because the oil remains bound in some form. The exact nature of this trapping of the oil is unknown, but theories point to one of four possibilities: 1) oil is bound in the form of oil-in-water emulsions which
are stabilized by protein, free fatty acids, mono and di-glycerides, and/or phospholipids; 2) as oil attached to hydrophobic material; 3) as oil bodies within intact cell structures or within larger particles; 4) and as intact oil bodies within broken cellular structures. The trapped oil is likely bound in some combination of the four, but very little literature exists on this subject. This lack of knowledge has made it difficult to design a strategy to extract the oil.

Previous published results are limited and often contradict one another. The literature has shown that the addition of proteases increase oil yield. Other research groups have shown similar results from other crops with high-oil percentages as well. These results indicate that some of the oil that remains in the kernel after extraction is stabilized by proteins, and likely zein. Therefore, disruption of protein bound oils may further increase the amount of oil that can be extracted by hexanes. The use of cellulases has shown differing effects. In one case the use of cellulases has shown to increase the amount of extractable oil, while another group found that there was no change. Additionally, finer milling of corn germ was found to increase oil yield.

The first goal of the research presented in this chapter was to determine how the unextractable oil from corn germ and CDS was bound. This knowledge was critical in order to determine how oil yields could be increased. The second goal was to develop a high-throughput screening method for determination of the amount of free oil released from corn germ and CDS after treatment with enzymes. For this analysis free oil was defined as the unbound corn oil that rested on the top of the mixture after centrifugation for 10 minutes at 3000 g.
6.2 Extractions of oil from corn

In order to determine how oil remained in corn after hexane extraction it was first important to develop a mimic of this process in the lab (Figure 6-2). To mimic conditions in a manufacturing plant, the dried corn germ was milled to 2 mm using a knife mill. The 2 mm milling was less than the 3 mm found at most production facilities, but was the largest diameter obtainable with the equipment available. Following the milling of the corn germ, both the germ and CDS could be analyzed using the same protocol. Next, to simulate the fermentation process and wash away background free oil, the milled germ or CDS was washed with 15% ethanol three times. While the milled germ would normally not be used in the fermentation process, adding this step to the process was later found to help eliminate background while screening. Therefore, this step was used for all experiments to maintain consistency. Now that the processing of the manufacturing plant was roughly mimicked, the total amount of hexane-extractable oil from germ and CDS was determined (Figures 6-3 and 6-4, respectively). Note that a higher percentage of oil was extracted using this method than would be expected from large scale extractions due to greater agitation, smaller particle size and the use of three rounds of extraction. The amount of oil extracted represents the highest volume of extractable oil expected without enzyme treatment.
Dry grind plant

CDS is removed after fermentation

Oil is removed from CDS by hexane extractions

Dry grind mimic

CDS is washed with 15% EtOH (3x)

Oil is removed from CDS by hexane extractions (3x)

Wet milling plant

Germ is separated from kernal and milled to 3 mm

Dried corn germ is milled to 2 mm

Oil is removed from CDS by hexane extractions (3x)

Wet milling mimic

Corn oil is extracted from germ using hexane extractions

Corn oil is extracted from germ using hexane extractions (3x)

**Figure 6-44:** Mimic of the process of extracting oil from corn found in production plants. A washing step was added to the wet milling mimic as this was found to decrease background in further experiments.

**Figure 6-45:** Hexane extractions of corn germ. Whole germ was milled through a 2 mm filter. Washing the germ with 15% ethanol mimics fermentation conditions.
Figure 6-46: Amount of oil extracted from CDS by hexanes. Washing the germ with 15% ethanol mimics fermentation conditions.

In both systems approximately half of the oil is lost in the wash, agreeing with information received from industrial plants, which estimate oil loss during fermentation to be 50%. These data illustrated that washing the starting material would also help eliminate high background signal in the free oil screen.

To determine how the remainder of the oil was bound the washed corn germ was incubated with various enzymes, each of which was hypothesized to break down one of the four theoretical ways oil is found bound in corn. Also, because previous studies had shown that grinding the germ finer allowed for higher yields of oil from hexane extractions, an enzyme known to breakdown the kernel was used (enz 1). The hypothesis was that enz 1 would transfer oil from the bound state to the free oil state, and in the process slightly increase the total amount of extractable oil. Surprisingly, results showed that enz 1 increased the amount of extractable oil significantly, even when only present at 0.01% v/v (Figure 6-5).
At 1% v/v enz 1 the amount of extractable oil doubled. Enz 4 is known to degrade emulsions, but no significant increase in extractable oil was observed over samples run with only enz 1.

**Figure 6-47:** Increase in extractable oil from milled, washed germ after enzyme treatment. Percentages are v/v.

However, using 1% v/v enz 1 is not feasible on a large scale. Therefore, because the enzymes are known to be stable and active for long periods of time, longer incubation times were also examined. Longer incubation times should allow for further degradation of the corn germ and reduce the amount of enzyme required. Longer times yielded similar results to the higher concentration of enzyme (Figure 6-6). Long incubation times with further reduction in enzyme concentration showed that high oil yields persisted even with 100-fold reduction in concentration (data not shown).
The overnight incubation with enzyme helps shed light on how the enzymes were able to allow for greater hexane extraction. Visualizing the corn showed that the samples incubated with enzyme resulted in finer germ particles and a thicker emulsion layer (Figure 6-7). Finer milling of germ was already known to increase oil yields so it appears as if the enzymes are performing a similar function. Because enz 1 is a mixture of two related but different classes of enzymes, it was broken into its components. It appears enz 3 is the more important of the two, but that both are needed to reach the full potential of the enzymes. These results indicate that a large percentage of the trapped oil is likely enclosed within intact cells and germ particles.

**Figure 6-48:** Total extractable oil from whole, washed germ. Enzs 2 and 3 are components of enz 1. Percentages are v/v.
Figure 6-49: Picture of corn germ after enzyme incubation. The addition of enzs 1, 2, and 3 all break down germ structure and form a thicker emulsion layer.

Again, since enz 4 was known to breakdown emulsions it was then added with the other enzymes to test for synergistic properties. The observed emulsion layer in this case was thinner which prompted the question: If enz 4 was breaking down the emulsion layer, where was the oil from this layer going if it wasn’t being extracted? (Figure 6-4). The results indicated that little oil was trapped in the emulsion layer, which contradicted previous results. To investigate this, the samples were divided into three layers. The first layer consisted of the free oil, the second layer contained the emulsions/fat-pack, and the final layer contained the remainder, namely the solids and aqueous components. After carefully separating the layers each was hexane extracted and the total oil contained within was extracted (Figure 6-8).
Figure 6-50: Oil partitioning in corn germ after treatment with enzyme. The bound portion contains the solids and aqueous phase, the emulsion partition is comprised of the emulsion layer and the free oil partition is the oil that rises to the top when centrifuged. Percentages are v/v.

Results show that the amount of extractable oil increases in every partition when the germ is treated with enz 1. Again, this was unexpected as the increase in extractable oil was originally thought to come from increases in free oil and oil within emulsions at the expense of the oil in the bound state. The emulsion layer was visibly thicker after the germ was incubated with enz 1, so this increase in extractable oil was expected. Now it was hypothesized that enz 1 was breaking down the germ structure, increasing the surface area of the germ and allowing for hexane to contact more of the germ and penetrate deeper than it would have with intact germ particles.

The results obtained from enz 4 were as expected. The oil in the bound state was unchanged and only a small increase in total extractable oil was observed. The amount of
free oil increased, and is thought to be from the emulsion layer. In separate experiments only
the emulsion and free oil layers were incubated with enz 4 and the amount extractable oil was
increased from both layers. It is believed that enz 4 frees oil from the emulsion layer while
also destabilizing it. This makes it more energetically favorable to extract oil from the
emulsions. When both enz 1 and 4 are used, a combination of the effects is observed. Also,
the total increase in extractable oil in the two enzyme system is higher than the sum of the
increase from the two single enzyme reactions, showing synergistic effects.

To test the hypothesis that enz 1 was breaking down germ and increasing the
surface area, scanning electron microscopy (SEM) was used to visualize the surface of the
germs (Figure 6-9).
The SEM images show that the starch encased particles on the surface of the germ have been removed and an oil layer is visible on the surface after treatment with enz 1. Enz 4 has a smaller effect but shows fewer enzyme encapsulated modules. The combination of enzs 1 and 4 appears similar to enz 1 alone. SEM also showed that the average particle size was smaller after treatment with enz 1 compared to the no enzyme control.

A similar set of hexane extraction experiments was conducted using the CDS and the results were more similar to what was expected for the germ (Figure 6-10). The total amount of extractable oil showed no significant change, but the partitioning of the oil was affected. The CDS contains little germ, as the germ mostly settles into the wet cake layer prior to
separation. Therefore, the solids present are mostly starches and protein compared to the high lipid content of the germ.

**Figure 6-52:** Oil partitioning in CDS after treatment with enzyme. The bound portion contains the solids and aqueous phase, the emulsion partition is comprised of the emulsion layer and the free oil partition is the oil that rises to the top when centrifuged. Percentage is in v/v.

Novozymes is currently pursuing a patent on this data.

6.3 Assay development for detection of free oil

6.3.1 OxiRed assay development

The other goal of the summer was to formulate a high-throughput method of determining the amount of free oil in corn germ and CDS. Free oil is the most desirable form of oil as it can be easily removed using existing equipment at most production plants, thus eliminating the need for costly new equipment. A high-throughput assay was required to test
the effects of different enzyme treatments on free oil release. Ideally, one assay would work for both dried germ and CDS, two common, but different systems used in the industry. It was quickly determined that since corn germ and CDS are such different systems it would be difficult to create one assay that worked well with both systems. Rather than finding one assay that was a compromise between the two systems, two different assays were developed.

Previous work showed that the free oil from corn was almost entirely triglycerides (Novozymes, unpublished). Since the medical field has been developing methods to detect the levels of triglycerides in blood for decades, this was the first area to be researched. Many of the techniques are specific to blood serum and would not apply to needs of this assay. One older method of detection of triglycerides levels showed promise as the company already had access to all of the necessary enzymes (Figure 6-11).\textsuperscript{101}
A screen utilizing OxiRed was modified to fit the needs of this project. Through a series of enzymatic reactions, the triglycerides were used to form peroxides which were then reacted with OxiRed to produce the fluorescent product resorufin. There were doubts about how effective this assay could be. Each enzyme in the process required different conditions with temperatures ranging from 25 °C to 55 °C and pHs ranging from 6 to 9.5. This would require a four step reaction procedure which would lower throughput and increase error of the assay. Also, the assay uses glycerol, a common biological molecule, as an intermediate which could lead to high background signal. Another obstacle was that since the oil is not...
soluble in the aqueous phase, where the enzymes are present, the effectiveness of the assay as a whole was in doubt. The series of reactions had also not been used on a system as complex as milled corn germ so there was concern that the reaction may either not work, or work too well and react with the oil contained in the emulsion or bound states.

To test whether vigorous shaking would be sufficient to allow the enzymes to react with the corn oil in an aqueous system, a mixture of buffered aqueous phase and store-bought corn oil was tested for activity. The procedure was carried out in four steps with each enzyme treatment conducted under its ideal conditions in terms of pH and temperature. Results showed a linear trend of fluorescence versus volume of corn oil (Figure 6-12). These preliminary results illustrated that the assay should be investigated further.

![Graph showing linear trend of fluorescence versus volume of corn oil.

Figure 6-54: The OxiRed assay shows a linear trend of fluorescence versus volume of corn oil.}
To make the assay feasible, the enzyme reaction procedure still needed to be simplified. A one-step procedure was desired, so all of the enzymes were run in single reactions under varying conditions. Some of the enzymes, such as horseradish peroxidase, were known to be more active than others, so the pHs and temperatures selected were chosen to be closer to the ideal conditions of the less active enzymes. An optimized temperature and pH was found for the one-step incubation (Table 6-1). The ideal conditions were found to be pH = 8 at 35 °C where 97% activity was shown when compared to the four step procedure.

Table 6-1: Reaction conditions for one-step, five enzyme OxiRed reaction. The fluorescence was normalized to the fluorescence from the four step reaction. The most active set of conditions is shown in red.

<table>
<thead>
<tr>
<th>pH</th>
<th>Temp (C)</th>
<th>Normalized Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>25</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.81</td>
</tr>
<tr>
<td>7.5</td>
<td>25</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.97 (red)</td>
</tr>
<tr>
<td>8.5</td>
<td>25</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.94</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Now that the assay showed a linear trend and could be done in a one-step reaction, it was important to investigate the background fluorescence that resulted from the natural levels of glycerol in corn as well as the glycerol contained within some of the relevant enzymes the assay would be tested against. It was found that the background glycerol from the enzymes being assayed could be nearly eliminated if lyophilized protein was used (data not shown). However background from the corn germ was present (Figure 6-13).

*Figure 6-55: Background fluorescence compared to experimental fluorescence for the OxiRed assay. Note that the germ treated with GK, GPO, and HRP would only account for the fluorescence due to the glycerol present. With no lipase, the triglycerides would not yield additional glycerol.*

When the washed germ was incubated in the presence of the final three enzymes in the OxiRed screen, which will go from glycerol to product, but will not hydrolyze the triglycerides, there was a modest signal. The signal approximately doubles when lipase is
also added. This shows that the signal from the residual free oil is approximately equal to the signal from the background glycerol. Adding enzymes that had previously been shown to increase free oil from corn germ showed a large increase in fluorescence. Upon further optimization the signal to noise ratio was decided to be acceptable.

The final test of the assay before validation was to check if the assay was reacting solely with the free oil, or if the assay was also detecting oil in the emulsion and bound states. To test this, reactions with enz 1 and/or enz 4 were run and then all of the free oil was removed. The remaining emulsions, solids, and aqueous components were then run in the OxiRed reaction and the signal was shown to be small compared to the signal from the samples containing the free oil (Figure 6-14).

**Figure 6-56:** Fluorescence from the emulsion and bound layers in the OxiRed assay.
Now that the initial concerns surrounding this assay were shown to be manageable, a validation experiment could be run for the assay. For this experiment an extraction method that removed only the free oil from corn germ was developed and utilized. This allowed for comparison of the amount of free oil extracted and the OxiRed signal (Figure 6-15).

![Figure 6-57: Validation of the OxiRed assay. Hexane extraction data was compared to OxiRed fluorescence for a 14 different enzymes.](image)

The results of the experiments compared well. While there was not perfect alignment between the assay and the extractions, the results are consistent with what was expected from a primary screen. The assay was able to correctly predict the amount of oil extracted for one enzyme where the initial hexane extraction not performed properly. A previously developed
method using a densiometer and HPLC measurement could then be used to secondary screen the most active enzymes. The assay was shown to work in 96-well plate format with a pipetting robot allowing for up to 5000 different enzyme mixtures to be tested each day.

6.3.2 Oil red assay development

The OxiRed assay would not however work with CDS because of the high glycerol levels in the substrate. As expected, a second assay needed to be developed for this system. Early in the project it was observed that other scientists used a staining dye, oil red (Figure 6-16), to stain lipids in cells. The dye bound to uncharged lipids, did not stain phospholipids, and was water insoluble. The dye could also be detected by absorbance and was inexpensive.

![Figure 6-58: Structure of the oil red dye.](image)

Based on the properties of this dye it was hypothesized that it could be used to detect the volume of free oil using an absorbance assay (Figure 6-17).
Add a set volume of oil red

Mix with gentle shaking

Centrifuge

Read absorbance of a set volume of top oil layer

**Figure 6-59:** Scheme of the oil red assay. *(Top) General protocol for the assay. *(Bottom) expected results from the assay. A larger oil volume *(B) will dilute the dye leading to a lower absorbance.*

Using this assay, the CDS would be incubated with the enzyme of interest under the ideal conditions for the enzyme. After this incubation a set amount of oil red, dissolved in corn oil, would be added to each well and allowed to mix for ten minutes with moderate
shaking. The mixture would then be centrifuged, allowing the now stained oil to rise to the top. Now a set volume of oil could be removed from each well and read for absorbance. A lower absorbance indicates a higher volume of oil as the dye has been diluted into a larger volume.

To test the validity of this simple assay, first, it had to be shown that the dye would stain the oil and that the absorbance trends would behave as expected. Store-bought corn oil was added to water and mixed. A set amount of oil red, dissolved in corn oil, was then added and it was again mixed. After centrifugation a set volume of the top oil layer was removed and read for absorbance. The expected absorbance trend was observed (Figure 6-18) and it was also observed that the dye was insoluble in water.

Figure 6-60: A linear trend is observed between absorbance and percentage oil red in a sample.
Because the CDS is a more complicated system than the model system used above, it still needed to be seen that the same trend would exist in the CDS system. CDS was added to water and mixed prior to addition of oil red. After mixing in the dye, the mixture was centrifuged and the absorbance of a set volume of the oil layer was detected (Figure 6-19).

![Absorbance vs Weight of CDS](image)

**Figure 6-61:** Relationship between oil red signal and total amount of oil added in CDS

The results appear as expected with a larger volume of the CDS, and thus a larger volume of oil, yielding a lower absorbance. A similar set of experiments was conducted on CDS after reaction with enzymes (Figure 6-20).
Figure 6-62: Oil red results after treatment of CDS with enzyme. The decrease in absorbance shows an increase in free oil. Error bars show the standard deviation of three experiments.

These results agree with previous data obtained from the amount of free oil present after incubation with enzyme.

Again the assay was validated by comparing the results of the assay to hexane extractions of the free oil layer (Figure 6-21).
Figure 6-63: Validation of the oil red assay. Results from the assay were compared to results from hexane extractions.

The assay performed well, with the results of the oil red assay aligning with the results of the hexane extractions. The low absorbance values from the oil red assay show a high level of oil, so the highest values in the hexane extractions should correspond to the lowest values in the oil red assay. Again the results of the oil red assay correctly predicted the amount of oil extracted from a number of enzymes in addition to those shown here.

This assay is not as high throughput as the OxiRed assay mainly because the CDS contains less oil, so a larger volume is required. This limits the assay to 24 well plates, but even at this size up to 1000 enzyme mixtures could be analyzed each day. This assay is less expensive and more versatile than the OxiRed assay. It was shown that this assay could also
work on germ samples, but the OxiRed assay proved to be slightly more accurate on that system. The oil red assay was also used for detection of free oil obtained in fermentations, replacing an older, slower, less accurate method that had previously been used.

Using both of these assays a number of interesting results were obtained. A class of enzymes that was not expected to have an effect on freeing oil from germ or CDS was shown to be the most active class. When these enzymes were then combined with enz 1 and enz 4 a large increase in free oil was observed leading Novozymes to pursue a patent on the results. HPLC analysis was routinely used to confirm the composition of the free oil, with little change being seen for the different enzyme reactions.

**6.4 Conclusions and Future work**

Enz 1 showed the ability to nearly double the amount of hexane extractable oil from corn germ. This large increase in oil would result in nearly $1 billion per year increase in corn oil produced while requiring only a relatively small cost of enzymes. Based on the increase in extractable oil after incubation with different enzyme classes, it was determined that the trapped oil is mostly in bound within cells and intact germ. A smaller percentage of oil is found in the emulsion layer.

Both the oil red and OxiRed assays were validated and used for high-throughput screening. Both assays can be run in a single day and can be automated. The use of these assays has already proven fruitful, as a class of enzymes that was not predicted to have an effect on increasing free oil was found to be among the most active. Increasing free oil would result in cost effective oil removal from industrial plants that do not contain the equipment
necessary for hexane extraction. It has also been hypothesized by many in the field that removal of more oil would increase ethanol yields in fermentations.

With these assay developments it will be possible to test enzymes mixtures to look for synergistic effects. Prior to these assays it was too labor intensive to look for these improvements and the field was largely ignored. Results showed that different classes of enzymes need to be used in conjunction in order to show increased free oil yields. Therefore early experiments on the subject using only single enzymes lead some to believe little could be done to increase free oil from corn.
7.1 Riboswitch technology

Data described in Chapter 2 illustrates that riboswitches are not yet the modular parts that many in the field wish them to be. In my opinion, this is still the largest drawback of riboswitch technology. Because the activation ratio is affected greatly by the promoter and reporter gene, screening of riboswitches is also limited. During screening, a reporter gene with a detectable output is required. If that riboswitch is designed with the aspiration of eventually being used with a different gene, the screening may prove ineffective. Riboswitches are often treated as though they are unaffected by anything in their surroundings, with the exception of its ligand. In reality they are very sensitive to their environment, with activation ratios varying greatly with even minute changes to conditions. I believe this issue is often swept under the rug rather than being researched. Until these issues can be resolved, riboswitches will not become widely used, and their great potential will not develop into reality.

To eliminate this problem, high-throughput screening methods are critical. Part of the problem is that, currently, the sample size of known riboswitches is too low, and much of the information we do know about them is not compiled into a single source. Through the development of high-throughput screening methods it would not only select for the riboswitch with the highest activation ratio for a particular ligand, but it would also likely provide several mutants with improved activity for that ligand. Now each of these mutant riboswitches could provide new information in regard to how the riboswitch functions.
Mutations to the expression platform may provide insight into not only how the riboswitch functions, but also how it interacts with the promoter region and reporter gene. While the Micklefield lab did well to demonstrate that natural riboswitches can be mutated to recognize new ligands, the small library size and lack of a high-throughput screening technique limit the utility of the research. Investing additional mutants may have provided more insight into the riboswitch’s mode of action and the role the aptamer plays in this.

If the ultimate goal of this research, an ultra-high-throughput dual selection strategy capable of selecting for both membrane permeable and impermeable ligands, is achieved, mutation of existing or creation of new riboswitches could be accomplished quickly. This technology, along with the increased knowledge obtained about riboswitches, could bring the idea of using riboswitches for metabolic engineering closer to a reality. Riboswitches could provide exact control over each enzyme in a pathway based on the concentration of its substrate. This could serve to increase yields by avoiding improper expression of enzyme, which results in either bottlenecks or poor utilization of cellular resources.

7.2 Polyketide technology

Mutation of MatB for a 5-fold improved performance with 12 served multiple purposes. First, as stated above, it served to provide an inexpensive alternative to 13 that would prevent negative side effects when used in in vivo assays. This reduction in cost could one day help make scale up of these reactions more feasible. In addition, the mutations found from this assay could be useful if the William’s Lab wanted to one day engineer specificity into the AT domains of PKSs. If the specificity was built around recognition of the thiol acceptor, less mutagenesis of the AT domain would be required. Instead of needing to mutate
an AT for each malonate derivative, it would instead be mutated be promiscuous with the malonate and to be specific to the thiol acceptor, if possible. This would require MatB to be promiscuous with additional thiol acceptors. The mutation sites found through this research may serve to expedite future research.

Having the ability to add halogens into polyketides could improve or alter their function. Halogens are shown to be present in natural products and prevalent in drugs. These halogen containing drugs are often more lipophilic and less water soluble than the non-halogenated analogues. This allows these halogenated drugs to better penetrate the membrane. Additionally, the added bulk of the larger halogens are hypothesized to occupy more volume in the binding pocket of molecular targets.\textsuperscript{102} Incorporation of halogens into germicidine derivatives could provide the product with new functionalities. Additionally, continued work with modules from Type I PKSs may allow for the incorporation of 19-22.
REFERENCES


5Winkler, W; Nahvi, A; Roth, A; Collins, J; Breaker, R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 2004, 428, 281-286.

Wachter, A; Tunc-Ozdemir, M; Grove, B; Green, P; Shintani, D; Breaker, R. Riboswitch control of gene expression in plants by splicing and alternative 3’ end processing og mRNAs. *Plant Cell* **2007**, *19*, 3437-3450.


Dixon, N; Duncan, J; Geerlings, T; Dunstan, M; McCarthy, J; Leys, D; Micklefield, J. Reengineering orthogonally selective riboswitches. *Proc Natl Acad Sci. USA* **2010**, *107*, 2830-2835.

13Fowler, C; Brown, E; Li, Y. Using a riboswitch sensor to examine coenzyme B(12) metabolism and transport in E. coli. ACS Chem Biol. 2010, 17, 756-765.


Lee, H; Larson, D; Lawrence, D. Illuminating the chemistry of life: design, synthesis, and applications of “caged” and related photoresponsive compounds. *ACS Chem Biol.* **2009**, *4*, 409-427.


28 Sparkman-Yager, D; Correa-Rojas, R; Carothers, J; Chapter Sixteen - Kinetic Folding Design of Aptazyme-Regulated Expression Devices as Riboswitches for Metabolic Engineering, Burke-Aguero, D Editor(s), Meth Enzy, 2015, 550, 321-340.


30 Sudarsan, N; Hammond, M; Block, K; Welz, R; Barrick, J; Roth, A; Breaker, R. Tandem riboswitch architectures exhibit complex gene control functions. Science, 2006, 314, 300-304.

31 Berens, C; Suess, B. Riboswitch engineering – making the all-important second and third steps. Curr Opin Biotech. 2015, 31, 10-15.


34 Beisel, C; Smolke, C. Design principles for riboswitch function. *PLOS Comp Biol.* **2009**, *5*, e1000363.

35 Stoltenburg, R; Reinemann, C; Strehlitz, B. SELEX – A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolec Eng*, **2007**, *24*, 381-403.

36 Ceres, P; Trausch, J; Bately, R. Engineering modular ON RNA switches using biological components. *Nuc Acid Res*, **2013**, *41*, 10449-10461.


42 Carothers, J; Oestreich, S; Szostak, J. Aptamers selected for higher-affinity binding are not more specific for the target ligand. J Am Chem Soc. 2006, 128, 7929-7937.

43 Weigand, J; Schmidtke, S; Will, T; Duchardt-Ferner, E; Hammann, C; Wohnert, J; Suess, B. Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity. Nuc Acid Res. 2011, 39, 3363-3372.


45 Dixon, N; Duncan, J; Geerlings, T; Dunstan, M; McCarthy, J; Leys, D; Micklefield, J. Reengineering orthogonally selective riboswitches. Proc Natl Acad Sci, 2010, 107, 2830-2835.


Topp, S; Reynoso, C; Seeliger, J; Goldlust, I; Desai, S; Murat, D; Shen, A; Puri, A; Komeili, A; Bertozzi, C; Scott, J; Gallivan, J. Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl. Environ. Microbiol.* **2010**, *76*, 7881-7884.


Rudolph, M; Vockenhuber, M; Suess, B. Synthetic riboswitches for the conditional control of gene expression in Streptomyces coelicolor. *Microbiology* **2013**, *159*, 1416-1422.


60 Lee, L; Tidor, B. Barstar is electrostatically optimized for tight binding to barnase. *Nat Struct Biol*. **2001**, *8*, 73-76.


Jiang, M; Pfeifer, B; Metabolic and pathway engineering to influence native and altered erythromycin production through *E. coli*. *Metab Eng*. 2013, 19, 42-49.

Reeves, C; Murli, S; Ashley, G; Piagentini, M; Hitchinson, C; McDaniel, R. Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochem*, 2001, 40, 15464-15470.

Cane, D; Walsh, C; Khosla, C. Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* 1998, 282, 63-68.


75Hertweck, C; Luzhetshyy, A; Rebets, Y; Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep*, 2007, 24, 162-190.


77Austin, M; Saito, T; Bowman, M; Haydock, S; Kato, A; Moore, B; Kay, R; Noel, J. Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat Chem Biol.* 2006, 2, 494-502.


79Chemler, J; Buchholtz, T; Geders, T; Akey, D; Rath, C; Chlipala, G; Smith, J; Sherman, D. Biochemical and structural characterization of germicidin synthase: analysis of a type III polyketide synthase that employs acyl-ACP as a starter unit donor. *J Amer Chem Soc.* 2012, 134, 7359-7366.


89 Walker, M; Thuronyi, B; Charkoudian, L; Lowry, B; Khosla, C; Chang, M. Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science, 2013*, 6, 1089-1094.


