Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this dissertation. I retain all ownership rights to the copyright of the dissertation. I also retain the right to use in future works (such as articles or books) all or part of this dissertation.

Signature:

_____ _____

Sean Lynch   Date

From Aptamer to Riboswitch: High-throughput Screens and Selections for the Identification and Creation of Synthetic Riboswitches

By

Sean Lynch

Doctor of Philosophy

Chemistry

_____

Justin P. Gallivan, Ph.D.

Advisor

_____

Stefan A. Lutz, Ph.D.

Committee Member

_____

David G. Lynn, Ph.D.

Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____

Date

From Aptamer to Riboswitch: High-throughput Screens and Selections for the Identification and Creation of Synthetic Riboswitches

By

Sean Lynch

B.A., State University of New York at Oswego, 2002

M.S., State University of New York at Oswego, 2004

Advisor: Justin P. Gallivan, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2009

Abstract

From Aptamer to Riboswitch: High-throughput Screens and Selections for the Identification and Creation of Synthetic Riboswitches

By Sean Lynch

The discovery of metabolite responsive riboswitches in bacteria supported the notion that ligand-binding RNA molecules could be used to control bacterial gene expression. Riboswitches are comprised of small-molecule binding sequences of RNA, known as an "aptamers", linked to "expression platforms" which translate the binding of a small molecule into a change in gene expression. While the number of engineered systems described in the literature utilizing RNA-small molecule interactions to control gene expression is limited, the principles driving their function are straightforward. Therefore, it is tempting to believe that in vitro selected aptamers could be readily used in the development of synthetic riboswitches that respond to new ligands.

In this thesis we describe a series of high-throughput screens and selections for the identification and creation of dynamic synthetic riboswitches in bacteria that respond to the small molecule, theophylline. We present a high-throughput, enzymatic assay that successfully identifies synthetic riboswitches with improved dynamic ranges. These new switches display essentially no background translation in the absence of their small-molecule effector, and large increases in its presence. Sequence data, coupled with in vitro and in vivo studies, enabled the development of a model describing their function where a transcript's secondary structure influences the translation of the downstream genes.

A second high-throughput screen capable of rapidly identifying synthetic riboswitches using fluorescent-activated cell sorting (FACS) is also described. The throughput of this screen approaches that of a genetic selection and was successfully used to identify riboswitches with capabilities that match or exceed those of most natural riboswitches. Characterization of these switches indicated the capacity of the ribosome binding site to dramatically alter the dynamic behavior of a synthetic riboswitch.

Using what we had learned about the function of theophylline riboswitches, we endeavored to select an RNA aptamers that binds the antibiotic erythromycin and subsequently attempted to integrate these aptamers into erythromycin-sensitive riboswitches using a genetic selection scheme. Despite limited success, the analysis of a single aptamer has revealed features that may guide the semi-rational design of a new erythromycin riboswitch and future efforts to select RNA aptamers for use in engineered riboswitches.

From Aptamer to Riboswitch: High-throughput Screens and Selections for the Identification and Creation of Synthetic Riboswitches

By

Sean Lynch

B.A., State University of New York at Oswego, 2002

M.S., State University of New York at Oswego, 2004

Advisor: Justin P. Gallivan, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2009

Acknowledgments

The long strange trip that has culminated with this dissertation would not have been possible without the kindness and encouragement of those who have kindly given their time and attention over the past five years. First and foremost, I would like to thank my advisor and mentor, Dr. Justin Gallivan. His seemingly infinite wisdom has helped me become a better scientist and his compassion and wit has seen me through some rather difficult times. To borrow a phrase from Homer Simpson, we've had some dizzying highs, some terrifying lows and some creamy middles both inside and away from the lab. Whenever I felt like my science or my life was in the tank, Justin's kindness and empathy would help put me at ease. For that, I send him my most sincere thanks.

I would like to also thank each of my committee members, Dr. David Lynn and Dr. Stefan Lutz for their encouragement and guidance. They are truly two of the wisest men I have ever had the pleasure to learn from. No matter what I have done over the past 5 years I could not help but feel they were always two steps ahead of me. I will consider it a great success if I can remain only two steps behind. Many thanks to you both.

As a Problems and Research to Integrate Science and Mathematics (PRISM) Fellow, I was given the incredible opportunity to work with Dr. Pat Marsteller. She has been a wonderful mentor and friend to me. She has shared in my accomplishments, empathized with my struggles, and through her words and deeds, challenged me to not be content with the status quo, rattle a few cages and carve my own niche. With the help of Jordan Rose and the many PRISM fellows, Pat has made me a better teacher, student and

citizen. Thank you, Pat and Jordan. I was also lucky enough to work with Dr. Kathryn Zuehlke, a science teacher at Chamblee High School. She, as much as anyone else, helped me to reach "escape velocity" with her reminder that no one was going to ask me to leave graduate school. Thanks, Dr. Z. Good luck!

As a Scholarly Inquiry and Research at Emory (SIRE) fellow, I had the pleasure to work with Dr. Leah Roesch whose enthusiasm for science and undergraduate research was infectious. She and the other graduate fellows, Alison Faupel, Aukje Kluge, Shana Kerr and Erin Sells helped reinforce my belief in the importance of intellectual curiosity. Many thanks and good luck to all of you.

I will always look fondly upon my time spent at SUNY Oswego. In an age when many chose to continue on into graduate school, not because of their academic experiences, but in spite of them, I have had the privilege to be mentored by the likes of Dr. Martha Bruch, Dr. Joseph LeFevre, Dr. Jeffery Schneider, Dr. James Seago and Dr. Norman Weiner. They, and many others at Oswego State, were the first to show me what a wonderful place academia could be. I would not have come this far without them. Thank you.

I would also like to thank my labmates, both past and present. While I have not kept a running tally, I think I can safely assume that I have spent more hours with my fellow lab members than I have with my wife. I joined the Gallivan Lab because of the people and they have never made me regret my decision. They have all made the long days and even longer nights in the lab incredibly enjoyable. I would especially like to thank Dr. Shawn Desai who was kind enough to teach me nearly everything I know about the fine art of Molecular Biology and progressive politics. I hope you enjoy Law school

as much as I have enjoyed your friendship. I would also like to thank Dr. Shana Topp.

Working on the same side of the lab bench for 5 years, Shana and I have shared many

long days and nights in the lab commiserating about our science, our boss and our

families. While we may never see eye to eye on issues like health care and gun control, I

will always consider her a dear friend and wish her nothing but the best of luck, though I

suspect she will never need it.

It has been far too long since my last encounter with many of my friends. For

that, I am truly sorry. Please do not let that diminish my appreciation for your friendship.

Special thanks to my good friend, Dave Schuster, who has been with me from our days in

Conklin to Oswego and all the way to Georgia. You helped get me started down this

Chemistry path by putting me on the ballot for Chemistry Club President at Oswego ….

without my consent. See what you did? I hope you're happy. Thanks, buddy.

I would be remiss if I did not also thank my partners in hilarity, Ian Goldlust and

Robin Higgins, for always reminding me to never take myself too seriously. I hope that

our paths will someday cross again.

I have needed every ounce of support provided by family and I hope I have

represented them well. There has been no shortage of hardships over the past few years

and my father, Andrew, has met them all with a quiet dignity and measured resolve that is

seldom seen. He has always reminded me of the value of a hard day's work while never

letting me forget to occasionally stop and enjoy myself. After all, we only live once. My

father is quick to tell anyone who will listen how proud he is of his sons. I very rarely get

the opportunity to tell him how proud I am of him. Dad, you are my hero. I love you.

Thank you for everything. Also, special thanks to my stepmother, Colleen, who has allowed me to sleep easy with the knowledge that my Dad is well taken care of. I am also especially grateful for the love and support of my Grandpa and Oma who taught me to reach for the stars and, most importantly, when I finally grab them to reach a little higher. Growing up, my mother, Ann, was my strongest supporter and I shall be eternally grateful for that. While she has been regrettably absent over the past few years, I still hold a special place in my heart for her and look forward to a time when she will read this.

Nothing motivates an older brother like his younger brother's accomplishments. I have, without a doubt, one of the best. Despite my two-year head start, Christopher has matched or surpassed nearly all of my successes. I have done all I can just to keep up with him. I couldn't be more proud. My brother was the best man at my wedding and despite our differences, he is, as the title suggests, still my *best* man. Through thick and thin he has and always will be there for me. Knowing how frail the bonds between siblings can sometimes be, this really means a lot. Thanks, kid.

My cousin, Collin, has been like a brother to me as well and his support has not gone unnoticed. I will continue to enjoy watching him and the rest of my cousins, Caitlin, Patrick, Kaleigh, Emily and Sarah grow into fine young men and women. Your futures are as bright as your minds. I am incredibly proud of each and every one of you. Thank you, all of you.

Finally, I would like to thank my wife to whom I dedicate this dissertation. You know as well as I that I could not have done this without you. For richer or poorer, in sickness and in health, I will spend the rest of my days trying to repay you for your

undying patience during this journey through graduate school. Without wavering, you have stood beside me through some of the toughest times in my life. Thank you. You will never have to question my love nor will you ever have to prove yours. You are my best friend and I love you. Yes I do. We made it!

Contents

List of Figures

Chapter 1:

Using Small Molecules to Control Gene Expression

1.1 Regulating Gene Expression through Small Molecule-Protein Interactions

The ability of a cell to regulate the expression of specific genes in response to chemical and physical changes in its environment is critical to the survival of the organism. The capability of organisms to differentially express genes in accordance with their need can be especially advantageous in conditions where metabolic resources are scarce. Given its importance, it should come as no surprise to find that evolution has provided myriad sophisticated systems for coordinating the expression of genes in response to small-molecule metabolites. Since the start of the 20th century, it has been well known that a variety of enzymes produced by microorganisms are produced in the presence of their small-molecule substrates that are typically intermittently found in the environment. The means by which this coordination of gene expression was controlled remained a mystery up until the 1960s.

Escherichia coli metabolize glucose to produce the ATP that drives most cellular processes. When necessary (i.e. when glucose is in short supply), *E. coli* possess the ability to transport and breakdown the disaccharide lactose into its monosaccharide components, glucose and galactose. It had been well known that these lactose metabolizing enzymes were only present in the bacteria when lactose was present in its environment; however, the mechanisms governing the synthesis and/or regulation of

these enzymes were unknown. In 1961, Jacob and Monod were the first to propose a model for the regulation of lactose metabolism[1]. Their model organized these genes into the lactose (lac) operon. Operons are composed of two distinct types of genes with distinct functions: (1)"structural" genes that give rise to non-regulatory proteins and (2) "operators", DNA sequences that dictate the expression of structural genes based on specific interactions with a repressor. In the case of the lac operon, the constitutively expressed lac repressor protein binds to the operator in the absence of lactose thus preventing the unneeded transcription of the downstream genes involved with lactose catabolism. In environments where glucose is scarce and lactose is abundant, the genes of the lac operon are up-regulated. This signaling mechanism begins with lactose entering the cell and subsequently being metabolized to the isomer, allolactose. Allolactose then binds to the lac repressor causing it to change conformation and dissociate from the lac operator. With the dissociation of the protein repressor, the operon's promoter becomes more accessible to RNA polymerase thus enabling the efficient transcription of the downstream lactose metabolizing genes. As lactose concentrations decrease, the lac repressor returns to its original conformation, down-regulating gene expression.

The isolation of the lac repressor in 1966 and its subsequent characterization provided the most compelling evidence for the small-molecule induction of lactose metabolizing genes through direct, small molecule-protein interactions[2]. Subsequent studies later revealed that even more complex repression mechanisms were present[3]. For example, in instances where excess glucose is present, operon expression is further repressed. These observations helped to explain the intricate mechanisms involved with the down-regulation of genes in response to small-molecules as is the case with

tryptophan biosysnthesis[4]. As tryptophan levels begin to rise in the cell, tryptophan binds

the trp repressor, which subsequently binds to the operator, obstructs RNA polymerase

and thus down-regulates the expression of genes responsible for tryptophan synthesis.

These protein-mediated models for the up- and down-regulation of gene expression in

response to small-molecules were quickly, and in some cases incorrectly, applied to the

regulation of nearly all bacterial genes. While these systems likely represent nature's

most ubiquitous method for controlling the transcription and translation of both

prokaryotic and eukaryotic genes in response to small molecules, they are certainly not

sole means of accomplishing this complex task.

1.2 Regulating Gene Expression with Small Molecule-RNA Interactions

As increasing amounts of evidence were collected in support of protein-mediated

models describing the regulation of gene expression in bacteria, Jacob and Monod's

speculation that these cellular processes may be under the direction of RNA or

"messenger" repression was largely forgotten. The roles played by RNA in controlling

cellular processes are becoming increasingly diverse. It is well known that the replication

of plasmids can be mediated by antisense RNA[5]. In addition to interacting with nucleic

acids, RNA can act in either a cis- or trans-fashion to bind proteins and influence the

actions of RNA polymerases[6]. RNA secondary structure is also sensitive to temperature

allowing it to act as a "thermosensor" of environmental conditions[7,8]. Given its diverse

capabilities, it is perhaps unsurprising that RNA can also function as a sensor of

metabolites and small molecules. The potential for RNA to adopt similarly complex

structural folds, bind a variety of complex small molecules and serve as allosteric

receptors for both natural and engineered RNA enzymes[9,10] (ribozymes) make it tempting to believe that nature could make use of RNA to accomplish the same regulatory functions as those once thought to be exclusively attributed to proteins.

Despite their prevalence, a number of protein cofactors responsible for the control of biosynthetic genes in bacteria remained elusive and suggested that the protein-mediated control of gene expression in bacteria may not paint as complete a picture as once thought. Prior to 2002, coenzyme B_{12} was known to be involved in the negative feedback inhibition of the *btuB* gene, which is responsible for the transport and biosynthesis of coenzyme B_{12} (AdoCbl) in *E.coli*. For years, the suspected protein factors mediating these processes could not be identified. The apparent absence of protein cofactors led Ron Breaker and coworkers to investigate the possible role played by 5' untranslated regions (5' UTRs) of bacterial mRNA in the regulation of gene expression. The highly conserved sequences present in the leader sequences mRNAs encoding for B_{12} biosynthetic genes, commonly known as the "B_{12}" box, were long known to be necessary for gene regulation but their function remained unclear. Breaker and coworkers speculated that the highly structured RNA of these conserved elements may directly interact with B_{12} metabolites and, in turn, influence gene expression without the requirement of a protein receptor.

Using equilibrium dialysis and structure probing assays that monitor the spontaneous self-cleavage of RNA, the leader sequence of the *btuB* gene was conclusively shown to directly bind AdoCbl[11]. Additionally, structure probing and in vivo expression of *btuB-lacZ* fusions revealed that binding of AdoCbl to this RNA could stabilize structural conformations, distinct from the unbound structure, capable of down-

regulating the translation of downstream genes through the obstruction of the ribosome binding site.

This discovery quickly led to the identification of several classes of similar elements, deemed "riboswitches", which coordinate the expression of multiple genes in response to metabolites that include flavin mononucleotide (FMN)[12], S-adenosylmethionine (SAM)[13], thiamine pyrophosphate (TPP)[14], glucosamine-6-phosphate (GlcN6P)[15], lysine[16], glycine[17] and purines[18,19]. With each new riboswitch reported, it becomes increasingly clear that riboswitches can serve many of the same functions once thought to be the responsibility of proteins. Through a variety of elegant mechanisms, riboswitches can control gene expression in response to signals that range from primary and toxic metabolites[20] to molecules involved with cell-to-cell communication[21].

Each riboswitch is composed of an RNA aptamer domain which acts as a selective receptor for the binding of a specific metabolite or small-molecule ligand; functioning similarly to its protein counterparts[22]. The binding of the ligand to aptamer induces a conformational change in an "expression platform" that is able to adopt one of two possible conformations in response to ligand binding leading to either an increase or decrease in the expression of the RNA transcript. The mechanisms by which these riboswitches function are nearly as diverse as the ligands to which they respond. The majority of riboswitches found thus far function by way of either transcriptional attenuation or by translational inhibition. However, ligand-induced changes in RNA stability[15] as well as intron splicing mechanisms in eukaryotes have also been observed [23-25].

Perhaps the best example of the diversity in mechanisms by which natural riboswitches may function is the thiamine pyrophosphate (TPP) sensing riboswitch. In Gram-negative *E. coli*, negative feedback control of TTP biosynthesis is accomplished through the inhibition of translation, similar to the AdoCbl riboswitch[14] (Figure 1.1A). TTP has been shown to directly bind to a conserved sequence of mRNAs responsible for its synthesis. Again, ligand binding induces a shift towards a conformation where translation is inhibited by the occlusion of the Shine-Dalgarno sequence of the RBS.
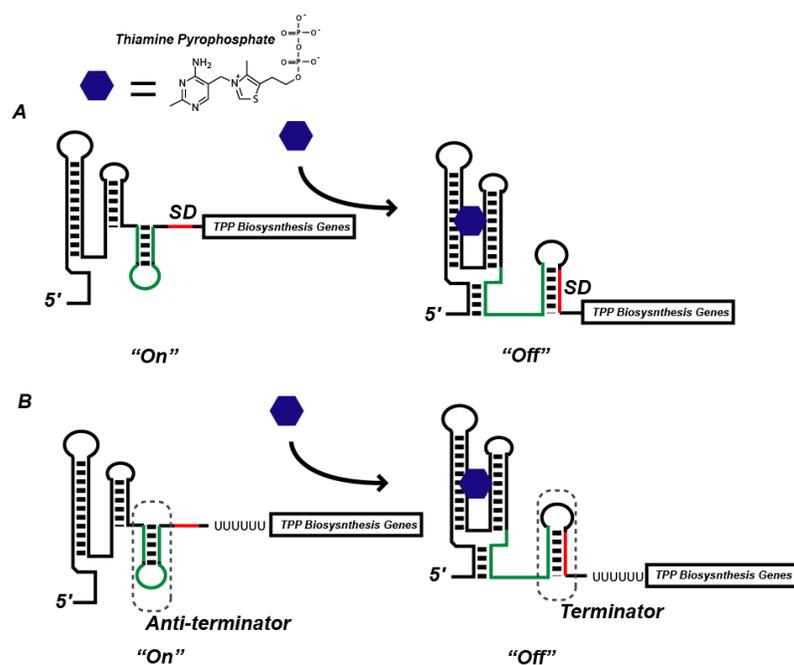
Figure 1.1 *The diverse mechanisms of the TPP riboswitch.* (A) TPP biosynthesis in Gram-negative *E. coli* is accomplished through the inhibition of translation. In the absence of TPP, genes that encode the enzymes for TPP biosynthesis are expressed. Upon ligand binding, the mRNA undergoes a conformational change occluding the Shine-Dalgarno (SD) sequence, preventing translation. (B) The TPP riboswitch of *B. subtilis* operates at a transcriptional level. In the absence of TPP, an anti-terminator stem loop structure is present allowing for the transcription of the downstream TPP biosynthesis genes. The binding of TPP to the 5' UTR stabilizes a conformation that favors the formation of a transcriptional terminator, causing premature termination of transcription.

The Gram-positive bacterium *Bacillus subtilis* also demonstrates the same TPP-dependent regulation of TTP biosynthesis as the Gram-negative *E. coli*; however, the mechanism underlying this control differs significantly (Figure 1.1B). Mironov et al. demonstrated through in vitro transcription studies of the "thi box" situated 5' to the genes responsible for thiamine biosynthesis in *B. subtilis*, that full-length transcripts were only observed in the presence of TPP[26]. These results indicated that this switch operates at a transcriptional level rather than at a translational level, where the binding of TPP to the 5' UTR stabilizes a conformation that favors the formation of a transcriptional attenuator, causing premature termination of transcription.

A number of variations on these mechanisms have also been identified. In some cases, the RBS can be buried within the RNA region tasked with ligand binding making conformational shifts in the mRNA unnecessary for changes in expression[27,28]. Alternatively, the RBS can be found sequestered within the secondary structure of the transcriptional terminator thus creating a system where even if transcription is not attenuated, the translation of downstream genes may still be inhibited.

The *glmS* gene encodes for an enzyme that produces glucosamine-6-phosphate (GlcN6P) expression in Gram-positive bacteria is also regulated by a riboswitch through direct, small molecule-RNA interactions[15]. Remarkably, the GlcN6P riboswitch responsible for the modulation of gene expression involves yet another unique component---the ribozyme (Figure 1.2). In this case, the ribozyme serves as an expression platform that translates GlcN6P binding to a change in gene expression. In the presence of GlcN6P, the ribozyme is activated, initiating the self-cleavage of the 5'

UTR of the glmS mRNA. This cleavage results in the generation of mRNA molecules harboring a free 5'-OH. This free hydroxyl serves as a recognition site for RNAse J1, a nuclease unique to Gram-positive bacteria, which destabilize the mRNA and shortening its lifetime in the cell.
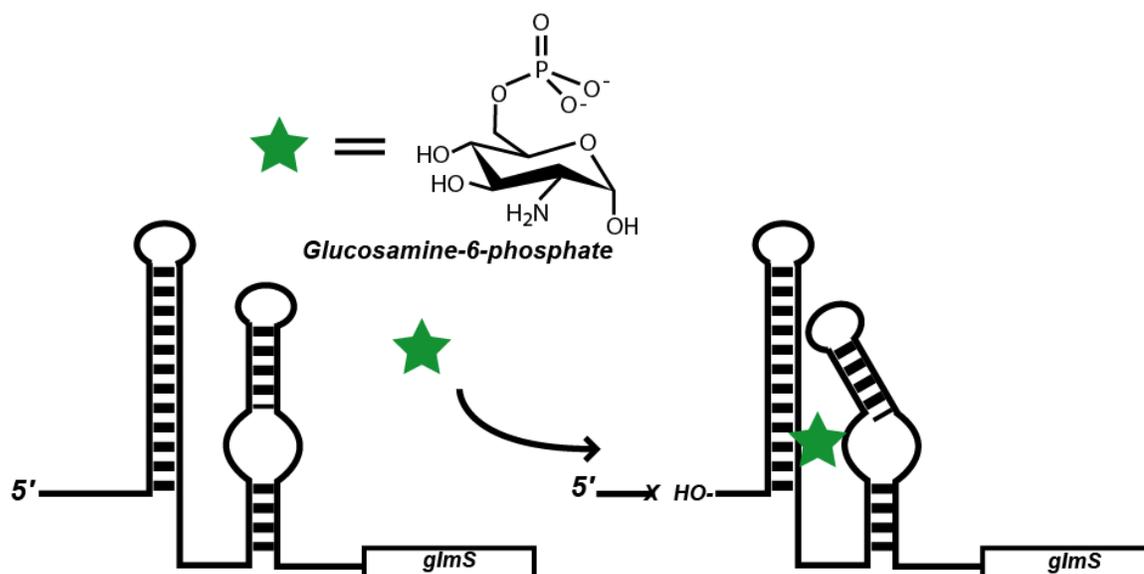
Figure 1.2 *The glmS riboswitch*. In the absence of GlcN6P, the downstream *glmS* gene is translated normally. Upon ligand binding, the ribozyme is activated, initiating the self-cleavage of the 5' UTR of the glmS mRNA. This cleavage results in the generation of mRNA molecules harboring a free 5'-OH. This free hydroxyl serves as a recognition site for RNAse J1, a nuclease unique to Gram-positive bacteria, ultimately destabilizing the mRNA and leading to a decrease in gene expression.

The glycine riboswitch exemplifies the power and complexity possessed by RNA regulators. In allosteric proteins, ligands often bind cooperatively. Ligand binding to an allosteric site of a protein can be translated through some structural rearrangement of the protein that gives rise to a remote binding site with increased affinity for a second ligand. The glycine riboswitch identified in *Bacillus subtilis* behaves in a similar fashion[17]. Aside from being the only known riboswitch to cooperatively bind its target ligand, the

glycine riboswitch is also unique in that it is one of the few riboswitches that activates gene expression in response to its effector. Perhaps the most striking feature of the glycine riboswitch is the simplicity of the glycine molecule itself. Compared to the ligands that interact with other riboswitches, glycine is incredibly small and lacks a great deal of functionality with which RNA can interact. In *Bacillus*, the glycine switch regulates the expression of a three-gene operon encoding for the production of enzymes responsible for the utilization of glycine as an energy source. Individually, these glycine aptamers require nearly a 100-fold increase in ligand concentration to reach complete saturation; however, in tandem, they only require a 10-fold increase in glycine levels. This cooperativity compares favorably with the cooperative binding of oxygen to hemoglobin and again illustrates the ability of RNA to perform functions once thought to only be possible with proteins[17].

The same glycine riboswitches are found upstream of genes encoding for glycine efflux pumps and used to activate gene expression when the cell is faced with toxic glycine levels. Still, these responses to chemical stress are not unique to glycine riboswitches. For example, in both Gram-negative and Gram-positive bacteria, the S-adenosylhomocysteine (SAH) riboswitch up-regulates genes responsible for the conversion of the toxic SAH molecule to methionine[20]. Amazingly, the SAH-binding aptamer of this riboswitch bears little resemblance to the closely related SAM riboswitch, yet can discriminate between SAH and SAM molecules that differ by only a single methyl group and a positive charge. Again, we see the abilities of RNA rival those of proteins.

While the bulk of riboswitches have been found in prokaryotes, riboswitches have also been discovered in eukaryotes. To find them, we again look to the regulation of TPP synthesis. In fungi, the TPP riboswitch is found in the intronic region of the 5'-UTR of genes involved in thiamine biosynthesis[23]. Here, ligand binding exerts its effects on gene expression by influencing the splicing of the transcript. In green algae, the TTP riboswitch similarly controls intron splicing; however, it surprisingly resides in a coding region[29]. Finally, an analogous TPP switch has been discovered in the 3' UTR of the thiamine biosynthesis genes in *Arabidopsis*, again controlling mRNA splicing[25].

As we begin to unveil new riboswitches that respond to new metabolites, we also begin to see the incredibly complex roles they play in cellular processes. For example, the recently discovered riboswitches of *Bacillus anthracis* and *Vibrio cholerae* respond to the secondary metabolite, cyclic diGMP, and indicate riboswitches function as much more than simple "molecular thermostats" that operate solely in response to primary metabolites[21]. Cyclic diGMP has been shown to be involved in cellular processes that range from cell motility and quorum sensing, as well as the expression of virulence factors indicating the role of riboswitches in regulating cellular processes may be far more expansive than we realize.

1.3 Engineering Small Molecule Control of Gene Expression

The extensive efforts made to understand the mechanisms by which cells coordinate the expression genes in response to small molecules is not simply an academic exercise. Advances in molecular biology and biochemistry have transformed the simple and elegant regulatory system of the *lac* operon into one of the most widely used tools to

science. Its power stems from allowing one to induce the expression of specific genes within the laboratory workhorse, *E. coli*, using a single, non-endogenous small molecule. Despite their many applications, the inducible expression tools used most often are limited by the small molecules they respond to. Conversely, the potential applications for systems that can be engineered to induce or repress gene expression in response to *any* desired small-molecule are limited only by one's creativity[30].

For example, the success of directed evolution experiments intended to evolve the substrate specificities of custom biocatalysts, hinge upon the availability of high throughput screens and selections proficient at recognizing the desired end-product and translating this chemical recognition into a measurable change in gene expression. The same chemical recognition events routinely take place inside the cell; however, the chemicals recognized are typically critical to the organism's survival. Using natural systems, such as the *lac* operon, as a model, one would need to engineer a system that involved a protein component capable of not only recognizing a small-molecule ligand, but also translating the recognition of said ligand to a change in gene expression.

Critical to these signaling systems are the direct interactions between the protein repressors and their orthogonal, small-molecule partners. Given the incredible amount of chemical diversity found in the cell, the utility of ligand-inducible or ligand-repressible expression systems employed in nature or in the laboratory is predicated upon their ability to discriminate between closely related molecules. Proteins, like the *lac* repressor, are well-suited for this task given their ability to adopt complex folds allowing for the tight, specific binding of complex small-molecules. The amazing specificity offered by

protein-mediated systems also makes any attempts to engineer new properties into these systems very challenging.

With this in mind, many have sought to engineer similar, small molecule-responsive systems that use the same ligand-binding properties displayed by natural riboswitches. The abilities to control gene expression through small molecule-RNA interactions were not entirely new concepts when the first riboswitches were discovered. Prior to the initial riboswitch discoveries, in vitro methods for selecting RNA molecules that bind small-molecule ligands with high affinity were well-established[31,32]. Commonly referred to as the Systematic Evolution of Ligands through EXponential enrichment (SELEX), this selection procedure involves passing a random pool of RNA over an affinity column with the small-molecule target appended to it. Unbound RNA molecules can then be washed away with the RNA molecules remaining bound to the column eluted off with a solution containing the free small-molecule target. Eluted RNA is then reverse transcribed into its cDNA which can then be PCR amplified. The iteration of these steps leads to the isolation of RNA molecules that can tightly and selectively bind the target ligand.

Since its initial report, SELEX has been successfully applied to isolate RNA aptamers to ligands such as biotin[33], cyanocobalamin[34], dopamine[35], tetramethylrosamine[36], tetracycline[37] and theophylline[38,39]. Analysis of these RNA aptamers quickly reveals RNA's abilities to adopt complex folds that allow for the tight, specific binding of complex small molecules much in the same manner as proteins. With this, it wasn't long before these selective binding abilities were exploited for the purposes

of controlling gene expression. Werstuck and Green demonstrated that the insertion of RNA aptamers selected to bind the Hoeschst dyes H33258 and 33342 into the 5' UTR of a *lacZ* reporter gene enabled the repressible control of translation in vitro and in Chinese hamster ovary cells[39]. Upon ligand binding, these highly organized aptamer structures presumably interfere with the ribosome scanning associated with eukaryotic translation initiation thereby inhibiting translation in the presence of the dyes.

In a similar approach, Grate and Wilson utilized aptamers with high affinity for malachite green to regulate the cell cycle of *Saccharomyces cerevisiae* in a ligand dependent manner[40]. By inserting these aptamers in the 5' UTR of a cyclin gene that directs the cell cycle in yeast, they were able to control the yeast's cell cycle based upon the presence or absence of the aptamer's ligand. RT-PCR of total yeast mRNA provided further evidence that changes in gene expression were due to transcript translatability rather than changes in mRNA concentration. Again, the presence of highly structured elements in the 5' UTR presumably interrupts the association of the ribosome with the transcript, therefore influencing the rate of translation. Also working in yeast, Suess and coworkers have used a tetracycline-binding aptamer to control gene expression by regulating translation efficiency[41] and pre-mRNA splicing[42]. By cloning the aptamer into an intron of eukaryotic mRNA, ligand binding interrupts mRNA splicing by disguising critical splice sites similarly to the TPP riboswitch of fungi.

Pelletier and coworkers took these ideas a step further and cloned theophylline- and biotin-binding aptamers into the 5' UTR, coding regions and 3' UTR of a variety of reporter genes including chloramphenicol acetyltransferase[43]. The most dramatic

inhibition of gene expression was observed when the ligand binding domains were engineered into the 5' UTR of mRNAs. When introduced into translation systems (wheat germ extracts, rabbit reticulocyte lysates, and *Xenopus* oocytes) these engineered constructs all exhibited decreased translation in the presence of the aptamer's ligand. The authors went on to confirm that ligand binding indeed inhibited the assembly of the 80S ribosome initiation complex using ribosome binding assays.

Each of these engineered systems influences gene expression by exploiting the ability of highly structured, ligand-bound aptamers to interrupt the complex eukaryotic translational machinery. It would stand to reason that these same principles could be applied to modulate gene expression prokaryotes. Surprisingly, only one example of this type of engineering has been reported. Topp and Gallivan revealed that the presence of a theophylline-binding aptamer in the coding region of a *lacZ* reporter gene could repress gene expression in the same ligand dependent fashion with the ligand-bound aptamer forming a complex that is stable enough to prevent the *E. coli* ribosome from translating through it[44].

Using the same theophylline aptamer, Wang et al. were able to influence viral replication in a ligand-dependent fashion by replacing a stem-loop structure needed for replication with the theophylline aptamer[45]. Ligand binding then stabilizes this structural element that is essential for the function of the viral replication machinery. As a result, the accumulation of the virus in the host cell becomes dependent on the presence of theophylline.

The inclusion of an expression platform in the design of these synthetic riboswitches has opened the door to even more complex methods for controlling gene expression with ligand-binding RNA aptamers. For example, Seuss and coworkers took advantage of a hairpin structure known for its "helix-slipping" mechanism[46]. By coupling a theophylline-binding aptamer to the hairpin structure in such a way that ligand binding would induce one base shift in the helix-slipping communication module. The authors carefully designed this structure in such a manner that the base freed upon theophylline binding was part of the RBS thus leading to an increase in gene expression.

Perhaps inspired by the glmS riboswitch, a number of groups have made use of the ribozyme to translate ligand binding to changes in gene expression. A number of ligand-dependent ribozymes have been created in vitro; however, their transition to in vivo systems has not been simple. Most frequently, the theophylline aptamer is integrated into the hammerhead ribozyme in such a manner that ligand binding initiates RNA self-cleavage. When integrated into a transcript in such a manner that the RBS of a reporter gene is near or buried within the modified ribozyme, ligand binding can initiate a self-cleavage reaction freeing up the RBS thus activating translation. Both Ogawa and Maeda[47] and Wieland and Hartig[45,48] utilized the functional component of the hammerhead ribozyme to transform a theophylline-binding aptamer into a function unit capable of controlling gene expression in *E. coli*.

With the incredible diversity among natural riboswitches and no shortage of methods for selecting RNA aptamers and allosteric ribozymes, why have there been so few examples of complex, engineered RNA regulatory elements? The preceding

examples of engineered RNA-based sensors are, in many ways, very rudimentary. With the exception of the integration of a "helix-slipping" motif and a fast-cleaving ribozyme, the engineered systems described rely only on ligand binding.   Furthermore, nearly all of the engineered systems involve the incorporation of the theophylline aptamer. Perhaps the best explanation for this is simply that the in vitro methods for selecting aptamers and allosteric ribozymes require conditions that are not amenable to in vivo applications. One thing that is certain is that, despite the simplicity of these systems "on paper", the best approach for taking an in vitro-selected, small molecule-binding aptamer and converting it into a ligand responsive riboswitch remains unclear.

Critical to the success of Weiland and Hartig[48] was the insertion of a randomized region into the structure connecting the aptamer to the ribozyme. By randomizing these nucleotides and screening for increased responses to theophylline, the authors were able create regulatory elements capable of generating 10-fold increases in gene expression based upon the presence of theophylline.   Underlying this approach is an acknowledgment of the inability to design or predict the means by which in vitro selected RNA-elements will exert their effects on gene expression. It is very possible that simple high specificity and affinity of RNA molecules selected to bind small-molecule targets in vitro may be insufficient for creating synthetic riboswitches. Therefore, approaches involving the creation of libraries consisting of candidate riboswitches, with a ligand binding aptamer and expression platform, followed by a screen for function hold the most promise for creating novel genetic control systems.

Compound	Structure	K_d (µM)	K_d (c)/ K_d (t)
Theophylline		0.32 ± 0.13	1
Caffeine		3500 ± 1,500	10,900
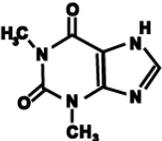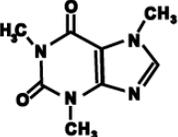
Figure 1.3 *Dissociation Constants for the mTCT4-8 Theophylline-binding Aptamer*. The aptamer binds theophylline tightly and selectively with nearly an 11,000-fold higher affinity for theophylline over caffeine.[38]

Previously, our lab has reported the creation of a synthetic, theophylline-sensitive riboswitch that functions in *E. coli* using the mTCT4-8 theophylline binding aptamer[49-52]. This aptamer, used in many of the above-mentioned works, binds theophylline with a high affinity (K_d = 0.32 µM). Furthermore, the aptamer is able to discriminate between theophylline and the similar caffeine molecule which differs by a single methyl group at the N7 position (Figure 1.3)[38]. By cloning this aptamer into the 5' UTR of an *IS10-lacZ* reporter gene, Desai and Gallivan were able to demonstrate a theophylline-dependent increase in β-galactosidase expression. The creation of this riboswitch was impressive; however, the riboswitch lacked the ability to completely suppress gene expression in the absence of theophylline and its mechanism of action was unknown. This synthetic riboswitch serves as an excellent representation of the potential power of synthetic, RNA-based genetic control elements and is the foundation upon which the subsequent work has been based. With the assumption that approaches involving the creation of candidate riboswitch libraries, followed by a screen for function

hold the most promise for creating novel genetic control systems, we present herein a series of genetic screens and selections that will be useful in the improvement of existing synthetic riboswitches and the identification of new ones.

1.4 References

(1)   Jacob, F.; Monod, J. *J. Mol. Biol.* **1961**, *3*, 318-56.

(2)   Gilbert, W.; Muller-Hill, B. *Proc. Nat. Acad. Sci. U.S.A.* **1966**, *56*, 1891-1898.

(3)   Makman, R. S.; Sutherland, E. W. *J. Biol. Chem.* **1965**, *240*, 1309-1314.

(4)   Bertrand, K.; Korn, L.; Lee, F.; Platt, T.; Squires, C. L.; Squires, C.; Yanofsky, C. *Science* **1975**, *189*, 22-6.

(5)   Wagner, E. G. H.; Simons, R. W. *Annu. Rev. of Microbiol.* **1994**, *48*, 713-742.

(6)   Klausner, R. D.; Harford, J. B. *Science* **1989**, *246*, 870-872.

(7)   Johansson, J.; Mandin, P.; Renzoni, A.; Chiaruttini, C.; Springer, M.; Cossart, P. *Cell* **2002**, *110*, 551-561.

(8)   Nocker, A.; Hausherr, T.; Balsiger, S.; Krstulovic, N. P.; Hennecke, H.; Narberhaus, F. *Nucl. Acids Res.* **2001**, *29*, 4800-7.

(9)   Wilson, D. S.; Szostak, J. W. *Annu. Rev. of Biochem.* **1999**, *68*, 611-647.

(10)   Famulok, M.; Jenne, A. In *Implementation and Redesign of Catalytic Function in Biopolymers* 1999, p 101-131.

(11)   Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R. *Chem. Biol.* **2002**, *9*, 1043.

(12)   Winkler, W. C.; Cohen-Chalamish, S.; Breaker, R. R. *Proc. Nat. Acad. Sci. U.S.A.* **2002**, *99*, 15908-15913.

(13)   Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R. *Nat. Struct. Biol.* **2003**, *10*, 701-7.

(14)   Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952-6.

(15)   Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* **2004**, *428*, 281-286.

(16)   Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R. *Genes Dev.* **2003**, *17*, 2688-97.

(17)   Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R. *Science* **2004**, *306*, 275-9.

(18)   Mandal, M.; Breaker, R. R. *Nat. Struct. Mol. Biol.* **2004**, *11*, 29-35.

(19)   Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chem. Biol.* **2004**, *11*, 1729-41.

(20)   Wang, J. X.; Lee, E. R.; Morales, D. R.; Lim, J.; Breaker, R. R. *Mol. Cell* **2008**, *29*, 691-702.

(21)   Sudarsan, N.; Lee, E. R.; Weinberg, Z.; Moy, R. H.; Kim, J. N.; Link, K. H.; Breaker, R. R. *Science* **2008**, *321*, 411-413.

(22)   Winkler, W. C.; Breaker, R. R. *Annu. Rev. Microbiol.* **2005**, *59*, 487-517.

(23) Cheah, M. T.; Wachter, A.; Sudarsan, N.; Breaker, R. R. *Nature* **2007**, *447*, 497-500.

(24) Sudarsan, N.; Barrick, J. E.; Breaker, R. R. *RNA* **2003**, *9*, 644-647.

(25) Wachter, A.; Tunc-Ozdemir, M.; Grove, B. C.; Green, P. J.; Shintani, D. K.; Breaker, R. R. *Plant Cell* **2007**, *19*, 3437-50.

(26) Mironov, A. S.; Gusarov, I.; Rafikov, R.; Lopez, L. E.; Shatalin, K.; Kreneva, R. A.; Perumov, D. A.; Nudler, E. *Cell* **2002**, *111*, 747-756.

(27) Rodionov, D. A.; Vitreschak, A. G.; Mironov, A. A.; Gelfand, M. S. *J. Biol. Chem.* **2002**, *277*, 48949-59.

(28) Vitreschak, A. G.; Rodionov, D. A.; Mironov, A. A.; Gelfand, M. S. *Trends Genet.* **2004**, *20*, 44-50.

(29) Croft, M. T.; Moulin, M.; Webb, M. E.; Smith, A. G. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 20770-5.

(30) Gallivan, J. P. *Curr. Opin. Chem. Biol.* **2007**, *11*, 612-9.

(31) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818-22.

(32) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-10.

(33) Wilson, C.; Nix, J.; Szostak, J. *Biochemistry* **1998**, *37*, 14410-9.

(34) Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33*, 973-82.

(35) Mannironi, C.; Di Nardo, A.; Fruscoloni, P.; Tocchini-Valentini, G. P. *Biochemistry* **1997**, *36*, 9726-34.

(36) Grate, D.; Wilson, C. *Proc. Nat. Acad. Sci. U.S.A.* **1999**, *96*, 6131-6136.

(37) Berens, C.; Thain, A.; Schroeder, R. *Bioorg. Med. Chem.* **2001**, *9*, 2549-56.

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

(39)   Werstuck, G.; Green, M. R. *Science* **1998**, *282*, 296-298.

(40)   Grate, D.; Wilson, C. *Bioorg. Med. Chem.* **2001**, *9*, 2565.

(41)   Suess, B.; Hanson, S.; Berens, C.; Fink, B.; Schroeder, R.; Hillen, W. *Nucl. Acids Res.* **2003**, *31*, 1853-8.

(42)   Weigand, J. E.; Suess, B. *Nucl. Acids Res.* **2007**, *35*, 4179-85.

(43)   Harvey, I.; Garneau, P.; Pelletier, J. *RNA* **2002**, *8*, 452-463.

(44)   Topp, S.; Gallivan, J. P. *RNA* **2008**, *14*, 2498-503.

(45)   Wang, S.; White, K. A. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 10406-11.

(46)   Suess, B.; Fink, B.; Berens, C.; Stentz, R.; Hillen, W. *Nucl. Acids Res.* **2004**, *32*, 1610-1614.

(47)   Ogawa, A.; Maeda, M. *Chembiochem* **2008**, *9*, 206-9.

(48)   Wieland, M.; Gfell, M.; Hartig, J. S. *RNA* **2009**, *15*, 968-76.

(49)   Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247-13254.

(50)   Zimmermann, G. R.; Wick, C. L.; Shields, T. P.; Jenison, R. D.; Pardi, A. *RNA* **2000**, *6*, 659-667.

(51)   Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J. P.; Pardi, A. *Nat. Struct. Biol.* **1997**, *4*, 644-9.

(52)   Zimmermann, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186-92.

Chapter 2:

A High-Throughput Screen for Synthetic Riboswitches[i]

2.1 Introduction

The theophylline riboswitches created by Desai and Gallivan provided yet another excellent example of the potential power of simple, engineered RNA-based methods for regulating gene expression[1]. While useful in many contexts, for the full utility of these switches to be realized, three issues needed to be addressed. First, the riboswitches lacked the ability to fully repress gene expression in the absence of theophylline. Second, only modest expression levels were observed in the presence of theophylline. For example, the most effective switch showed a signal to background ratio (activation ratio) of approximately 8 in the presence of 500 µM theophylline. While this 8-fold increase allowed us to perform both genetic screens and selections to detect the presence of theophylline, for more demanding screening applications, an increase in the ratio of signal to background is desirable. Finally, while it was known that these switches operated at a post-transcriptional level, a detailed mechanism explaining their function remained unclear.

Our approach to addressing these issues was guided by previous observations indicating that modifying either the length of the sequence separating the theophylline

---

[i] Figures and text used with permission from: S.A. Lynch, S.K. Desai, H.K. Sajja, J.P. Gallivan. *Chem. Biol.* **2007** *14*, 173−184. Copyright © 2007 Elsevier

aptamer and the ribosome binding site (RBS), critical components of the expression platform, or the sequence itself had dramatic effects on both the function and dynamic range of the synthetic riboswitches. Unfortunately, we could not determine precisely how these changes exerted their effects. We anticipated that by randomizing this region and screening the library for function, we could identify improved switches and possibly gain insight into their mechanisms of action.

The application of genetic screens and selections is often performed in the context of directed evolution experiments[2]. While the differences may be subtle, it is worth delineating the features that separate our combinatorial approach to improving our synthetic riboswitches from the directed evolution experiments described elsewhere. With each methodology, libraries of mutants are generated through a variety of methods, including PCR strategies using degenerate nucleotides, DNA shuffling[3], incremental truncation[4,5] and circular permutation[6]. Library creation is then followed by a screen or selection for desired phenotypes. The approaches diverge with directed evolution requiring the iteration of these steps while combinatorial methods do not. This slight difference has a major impact on the throughput required of the screening and selection strategies employed to identify ideal mutants.

Genetic selections offer elevated throughputs and enable the isolation of mutants from libraries consisting of as many as 10^{12} clones[13]. This throughput is achieved only in situations where the desired mutations confer a growth advantage over the other members of the library given a specific selection pressure (e.g. antibiotic resistance). Despite its advantages, genetic selections are often difficult to design, are inherently not quantitative,

and in situations demanding the counter-selection of one phenotype against another, as are the case with riboswitches, multiple cloning steps are typically required.

Genetic screens are distinct from selections in that all members of the library are typically viable, and the identification of desired mutants requires independent measures of specific phenotypes. These demands significantly lower the throughput of these methods. Even with the help of automation, typically no more than 10^6 clones can be screened in a given experiment. Despite these limitations, genetic screens can be extremely valuable methods for identifying and optimizing synthetic riboswitches. Screening experiments allows one to readily quantify expression levels in the presence and absence of a specific ligand thereby enabling the identification of functioning riboswitches. Furthermore, because the sequence space of RNA is small relative to proteins, the throughput offered by genetic screens is typically adequate as most, if not all, potential sequences can be sampled in the context of a simple genetic screening experiment. For example, experiments in which 8 RNA bases are randomized give rise to theoretical library sizes consisting of approximately 65,000 members. While libraries of 65,000 members are by no means "small", the screening of such libraries does not require the throughput offered by genetic selections in order to screen most, if not all, members of the library.

In this chapter, we present an automated high throughput screening method that identifies synthetic riboswitches that display extremely low background levels of gene expression in the absence of the desired ligand and robust increases in expression in its presence. We anticipate that the high throughput assay described herein will be generally

useful for discovering synthetic riboswitches with new ligand specificities and better performance characteristics.

2.2 Results and Discussion

2.2.1 Creation of a Library of Randomized Mutants

We previously created synthetic riboswitches by cloning a theophylline-binding aptamer at various locations upstream of the ribosome binding site of a β-galactosidase reporter gene (*IS10-lac*Z) that was controlled at the transcriptional level by a weak, constitutively active *IS10* promoter (Figure 2.1A). To increase the overall signal, we replaced the *IS10* promoter with the stronger P*tac*1 promoter. As expected, the stronger *tac* promoter enhances the level of β-galactosidase expression in the presence of theophylline, but also increases the background expression ~80-fold from ~10 Miller units to ~800 Miller units (Figure 2.1B). We used cassette-based PCR mutagenesis to create five different libraries in which the distance between the aptamer and the RBS was varied between 4 and 8 bases and the sequence was randomized fully (Figure 2.1C) because we previously observed that longer or shorter spacings resulted in poorly functioning switches[1].

The reporter gene used in these experiments contains a weak ribosome-binding site that limits translation, with the β-galactosidase gene fused to the C-terminus of an IS10 transposase sequence. Translation of this sequence is claimed to be limited further by the presence of infrequently used codons in its coding region thus allowing for the accurate measurement of small changes in β-galactosidase activity. It is important to note

that in the context of our switches, the IS10 sequence appears to do very little to limit

*lac*Z translation compared to the unmodified sequence.
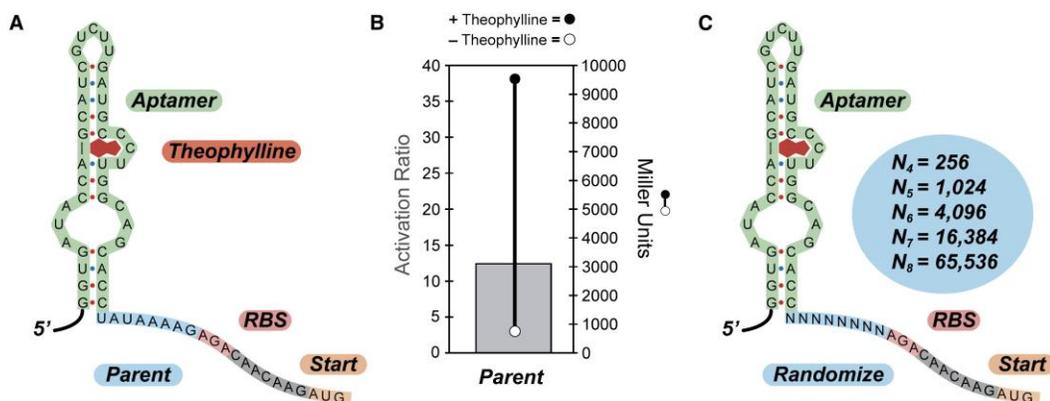
Figure 2.1 *Diagram of the 5' Region of a Synthetic Riboswitch, the Performance of the Synthetic Riboswitch, and the Randomization Strategy.* (A) Parent synthetic riboswitch created by Desai and Gallivan with 8 bases separating the mTCT4-8 theophylline aptamer (green) and the ribosome binding site (pink); the AUG start codon (peach) is also highlighted. The aptamer is shown in the secondary structure predicted for the theophylline aptamer by *mFold* and confirmed experimentally. (B) Measures of the activity of the synthetic riboswitch shown in (A) when cloned upstream of the *IS10-lac*Z gene fusion and expressed in *E. coli*. Right axis: β-galactosidase activity in the absence (open circle) or presence of theophylline (1 mM, closed circle). Activities are expressed in Miller units, and the standard errors of the mean are less than the diameters of the circles. Left axis: the activation ratio of the synthetic riboswitch (gray bar), which is determined by taking the ratio of the activities in the presence and absence of theophylline. (C) Sequence diagram of the randomized synthetic riboswitches. The aptamer is shown in green, the randomized regions in light blue, the ribosome binding site in pink, and the AUG start codon in peach. Theoretical library sizes different riboswitches are shown in the inset.

2.2.2 A High-throughput Screen for Optimally Functioning Riboswitches

To screen the libraries, *E. coli* were transformed with plasmids harboring randomized sequences of a given length and were grown on selective agar plates in the presence of X-gal (the substrate for β-galactosidase), but, more importantly, in the *absence* of theophylline. Most (~99%) clones appeared blue, indicating that they were expressing β-galactosidase in the absence of theophylline. However, a number of

colonies appeared white, suggesting little to no β-galactosidase expression. To identify whether these colonies harbored riboswitches that could activate protein translation in the presence of theophylline, we used a colony-picking robot to isolate the whitest colonies from each plate of approximately 4,000 colonies (Figure 2.2). These clones were inoculated into 96-well microtiter plates containing selective LB media. These cultures were grown overnight and were used to inoculate two new 96-well plates that either contained theophylline (0.5 mM) or did not. These plates were grown at 37 ℃ with shaking for 2–2.5 hrs, and β-galactosidase activity was assayed using an adaptation of Miller's method performed either by hand with a multi-channel pipettor, or using a
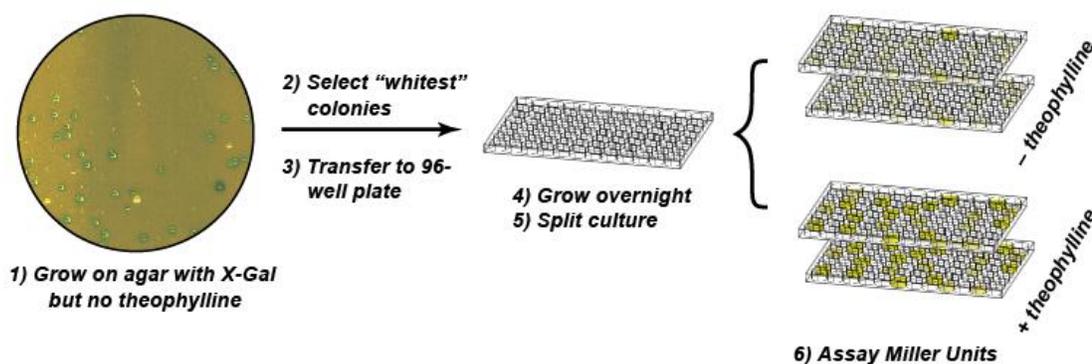
Figure 2.2 *Diagram of the High-Throughput β-Galactosidase Assay*. Potential riboswitches are identified by plating cells onto selective media containing X-gal but lacking theophylline. A robotic colony picker identifies the whitest colonies (lowest levels of β-galactosidase activity) and transfers the cells to a 96-well microtiterplate. The culture is grown overnight in selective media, split, and the clones are tested for β-galactosidase activity (Miller units) by using the automated assay described in the text.

robotic liquid-handling system[7-9]. Ratios of the Miller units for cultures grown in the presence of theophylline to those grown with or without theophylline were compared to identify functional switches.

To validate the screen, we chose clones that displayed an activation ratio of greater than 2 and assayed them individually in larger cultures, but we discovered that less than half of these clones functioned as switches. Further analysis indicated that these irregularities were due to a number of factors, including small but significant differences in the growth rates and lysis efficiencies of the cells in the microtiterplates. To improve the accuracy of identifying functional synthetic riboswitches, we performed each assay in duplicate, and analyzed the data using an empirically derived procedure through which we retained candidates that: A) showed an activation ratio of greater than 2.0 in two separate determinations; B) displayed a minimum level of β-galactosidase activity in the presence of theophylline (an $OD_{420} \geq 0.04$ in the Miller assay, regardless of cell-density); C) grew normally relative to others in the plate (as represented by OD_{600}); and D) showed consistent results between the two plates. This simple analysis significantly reduced the number of potential candidates, of which greater than 90% were confirmed as functional synthetic riboswitches when assayed individually in larger volumes of culture. For the small number of candidates that were not validated, sequencing often revealed mixed populations of plasmids that were likely introduced during colony picking and plating cells at lower densities minimized such events.

A.

	Theoretical Size	Colonies Plated	Clones Assayed	Confirmed Switches
N_4	256	~ 6,000	96	0
N_5	1,024	~ 4,000	192	4
N_6	4,096	~ 19,000	288	16
N_7	16,384	~ 28,000	480	>14
N_8	65,536	~ 59,000	864	>20
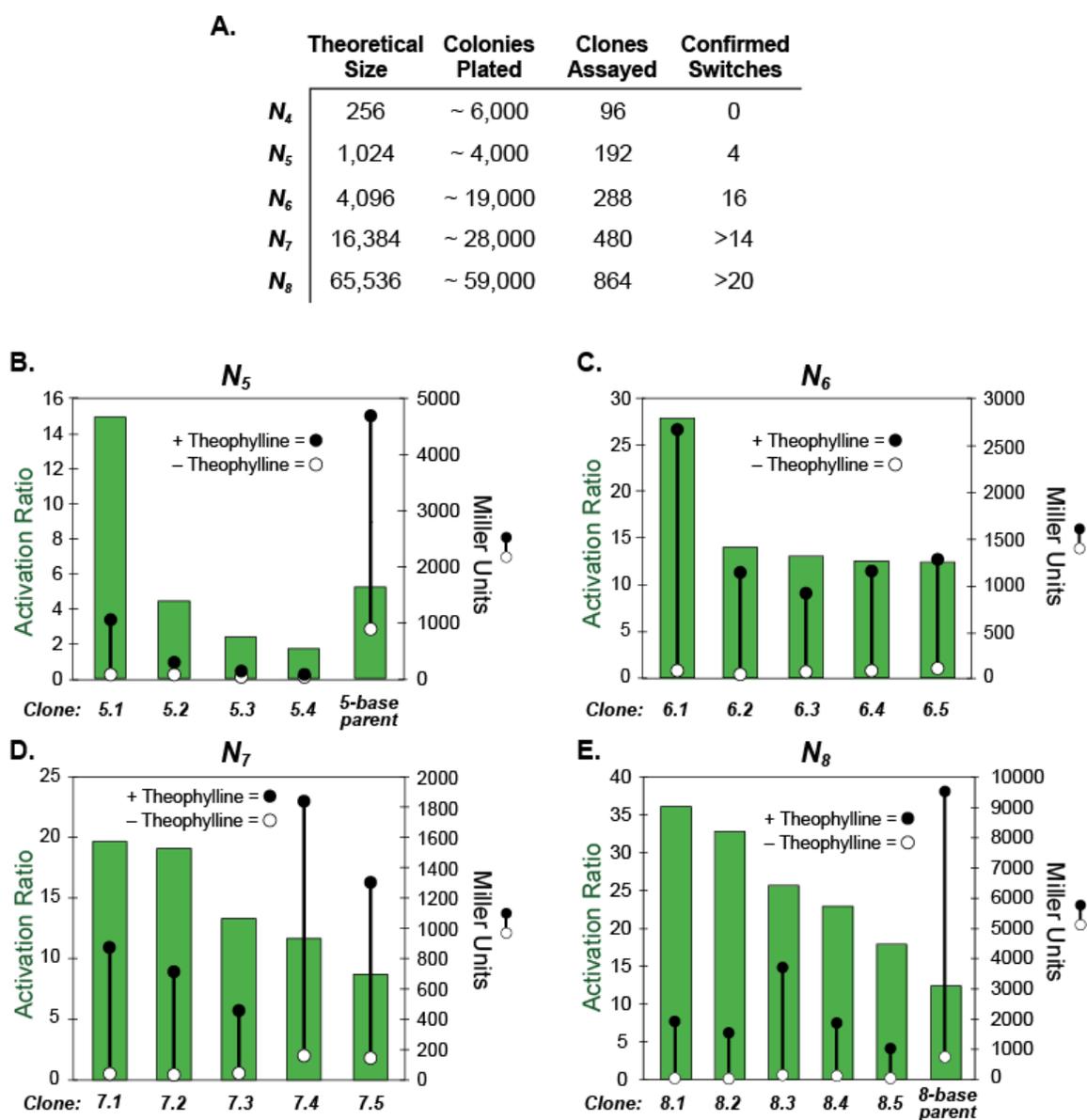
Figure 2.3. *Analysis of Libraries and Results from Screens.* (A) Statistics for libraries assayed. (B-E) Measures of the activities of the synthetic riboswitches identified in the high-throughput screens. Each measurement is done in triplicate. For all panels: right axis, β-galactosidase activity in the absence (open circle) or presence of theophylline (1 mM, closed circle). Activities are expressed in Miller units, and the standard errors of the mean are less than the diameters of the circles. Left axis, the activation ratio of the synthetic riboswitches (gray bar), which are determined by taking the ratio of the activities in the presence and absence of theophylline.

Figure 2.3 shows statistics for the libraries and the top validated hits identified in

each of these screens. For each switch, the absolute levels of β-galactosidase activity in

the presence and absence of theophylline (1 mM) are plotted, as well as the ratio of these numbers (activation ratio) for the switches that displayed the highest activation ratios. From these data, it is clear that the high throughput screening method is capable of identifying riboswitches that display both low background levels of protein expression in the absence of ligand, and strong increases in the presence of the ligand. Indeed, the best clone identified (8.1) displays a 36-fold increase in protein expression in the presence of theophylline, and a very low level of expression in its absence. To put this in perspective, natural riboswitches that regulate protein translation (though, most of these switches repress protein translation) show repression ratios of about 100 in the presence of the ligand[10]. Adenine-sensitive riboswitches from *Bacillus* show only 10 fold increases in gene expression[11]. The discovery of switches that function within a factor of three of natural genetic regulatory elements that have evolved over considerably longer time periods suggests that this screening method is quite effective, and that creation of new synthetic riboswitches based on different aptamers may be straightforward.

2.3 Conclusion

The screen described here is adept at identifying riboswitches with improved characteristics. Given the number of colonies screened relative to the theoretical library size, the switches we have identified likely represent some of the best riboswitches present in the library. Sequence results from our high performing riboswitches had revealed to us the sequence of expression platforms responsible for maintaining low levels of gene expression in the absence of theophylline and generating strong increases in its presence. With each riboswitch only differing by an expression platform, it is

reasonable to assume that switches with unified characteristics may have expression platforms with similar features. The following chapter describes how this wealth of sequence information led us to propose a mechanism of function and the subsequent series of experiments to probe this mechanism.

2.4 Experimental

General Considerations All plasmid manipulations utilized standard cloning techniques[12]. All constructs have been verified by DNA sequencing at the NSF-supported Center for Fundamental and Applied Molecular Evolution at Emory University. Purifications of plasmid DNA, PCR products, and enzyme digestions were performed using kits from Qiagen. Theophylline, *o*-nitrophenyl-β-D-galactopyranoside (ONPG) and ampicillin were purchased from Sigma. X-gal was purchased from US Biological. Synthetic oligonucleotides were purchased from IDT. All experiments were performed in *E. coli* TOP10 F' cells (Invitrogen) cultured in media obtained from EMD Bioscience.

Construction of Randomized Libraries. Libraries were constructed using oligonucleotide-based cassette mutagenesis. Mutagenic primers with degenerate regions were designed to create cassettes with randomized sequences of appropriate lengths between the mTCT8-4 theophylline aptamer and the RBS of the *IS10-LacZ* reporter gene.

Library screens. Library transformations were plated on large (241 mm x 241 mm) bioassay trays from Nalgene containing LB/agar (300 mL) supplemented with ampicillin (50 μg/mL) and X-Gal (25 mg dissolved in 4.0 mL of dimethyl formamide, final concentration 0.008%). Cells were plated to achieve a final density of

~4,000 colonies/plate. Cells were grown for 14 h at 37 °C, followed by incubation at 4 °C until blue color was readily visible.

The whitest colonies from each plate were picked using a Genetix QPix2 colony picking robot and were inoculated in a 96-well microtiter plate (Costar) which contained LB-media (200 µL/well) supplemented with ampicillin (50 µg/mL). The plate was incubated overnight at 37 ℃ with shaking (180 rpm). The following day, four 96-well plates (two sets of two) were inoculated using 2 µL of the overnight culture. The first set of plates contained 200 µL of LB supplemented with ampicillin (50 µg/mL). The second set of plates contained 200 µL of LB supplemented with both ampicillin (50 µg/mL) and theophylline (0.5 mM). Plates were incubated for approximately 2.5 hrs at 37 ℃ with shaking (210 rpm) to an OD_{600} of 0.085-0.14 as determined by a Biotek microplate reader. These values correspond to an OD_{600} of 0.3-0.5 with a 1 cm path length cuvette.

A high-throughput microtiterplate assay for β-galactosidase activity was adapted from previously described methods. Cultures were lysed by adding Pop Culture® solution (Novagen, 21 µL, 10:1, Pop Culture : lysozyme (4 U/mL)), mixed by pipetting up and down, and allowed to stand at room temperature for 5 min. In a fresh plate, 15 µL of lysed culture was combined with Z-buffer (132 µL, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH 7.0). This was followed by addition of ONPG (29 µL, 4 mg/mL in 100 mM NaH_2PO_4). ONPG was allowed to hydrolyze for approximately 20 min or until faint yellow color was observed. The reaction was quenched by the addition of Na_2CO_3 (75 µL of a 1 M solution). The length

of time between substrate addition and quenching was recorded and the OD_{420} for each well was determined. The Miller units were calculated using the following formula:

$$Miller\ units = OD_{420}/(OD_{600} \times hydrolysis\ time \times (volume\ of\ cell\ lysate/total\ volume)).$$

Ratios of the Miller units for cultures grown in the presence or absence of theophylline represent an "activation ratio". The initial pool of candidate switches comprised clones that showed an activation ratio of greater than 2.0 in two separate determinations. Candidates that did not display a minimum activity in the presence of theophylline (an $OD_{420} \geq 0.04$) in either determination were eliminated from consideration. As a final check, we visually inspected the data for aberrations, such as cultures that grew especially slowly or quickly (as represented by OD_{600}), or for cultures with dramatically different results between the two plates. Clones that were identified as potential switches were subcultured and assayed as previously described

2.5 References

(1)   Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247-13254.

(2)   Arnold, F. H.; Volkov, A. A. *Curr. Opin. in Chem. Biol.* **1999**, *3*, 54-59.

(3)   Stemmer, W. P. C. *Nature* **1994**, *370*, 389-391.

(4)   Lutz, S.; Ostermeier, M.; Moore, G. L.; Maranas, C. D.; Benkovic, S. J. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 11248-11253.

(5)   Lutz, S.; Ostermeier, M.; Benkovic, S. J. *Nucl. Acids Res.* **2001**, *29*, e16-20.

(6)   Qian, Z.; Fields, C. J.; Lutz, S. *ChemBioChem* **2007**, *8*, 1989-1996.

(7)   Bignon, C.; Daniel, N.; Djiane, J. *Biotechniques* **1993**, *15*, 243-244.

(8)   Griffith, K. L.; Wolf, J. R. E. *Biochem. Biophys. Res. Comm.* **2002**, *290*,
         397.

(9)   Jain, C.; Belasco, J. C. *Methods Enzymol.* **2000**, *318*, 309-332.

(10)   Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952-6.

(11)   Mandal, M.; Breaker, R. R. *Nat. Struct. Mol. Biol.* **2004**, *11*, 29-35.

(12)   Sambrook J, R. D. *Molecular cloning : A Laboratory Manual.*; Cold
         Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2001.

(13)   Link, A. J.; Jeong, K. J.; Georgiou, G. *Nat. Rev. Microbiol.* **2007**, *5*, 680-
         688

Chapter 3:

Investigations into the Mechanisms of Synthetic Riboswitches[i]

3.1 Introduction

Both natural and synthetic riboswitches can function by a variety of different mechanisms (see Chapter 1). The theophylline riboswitches that function in *E. coli*, first created by Desai and Gallivan[1], were initially designed to mimic the eukaryotic RNA-based sensors reported by the likes of Werstuck and Green[2], Grate and Wilson[3] and Pelletier and coworkers[4] where a ligand bound aptamer residing in the 5' UTR of reporter gene could hinder translation efficiency thus creating an "off" switch that represses gene expression in the presence of a ligand. While hypothesized to limit gene expression in the presence of its ligand, the insertion of a theophylline aptamer into the 5' UTR of a *lacZ* reporter gene in *E. coli* gave rise to a regulatory system that activated gene expression in the presence of theophylline rather than repressing it.

While these results were a bit of surprise, the potential utility of the theophylline riboswitch remained. By randomizing the sequences of expression platforms linking the

---

theophylline aptamer to the RBS of a *lacZ* reporter gene and employing the plate based assay for *lacZ* activity from Chapter 2, a number of riboswitches were isolated with improved characteristics (i.e. low expression levels in the absence of theophylline, and robust increases in its presence). Despite these successes, the mechanism of activation remained unclear.

Data collected from our lab's original theophylline riboswitches potentially fit models corresponding to those of natural riboswitches such as the glycine riboswitch and adenine riboswitch that activate gene expression at the level of translation and transcription, respectively. To determine whether the original switches were functioning to control transcription or translation, a polycistronic RNA was devised consisting of a theophylline riboswitch fused to the 61 amino acid N-terminal segment of the IS-10 transposase followed by three in-frame stop codons, a 28 base pair spacer connecting to a second gene comprised of a ribosome binding site followed by the wild-type *lacZ* sequence. If the riboswitch regulated expression at the level of transcription, the expression of the downstream *lacZ* gene would be dependent upon the presence or absence of theophylline. Alternatively, if the switch operated at a translational level, the expression of *lacZ* would be independent of theophylline concentration as the full length mRNA would be present at all times. When introduced into *E. coli*, *lacZ* expression proved to be independent of theophylline thus indicating the riboswitch operated at a post-transcriptional level.

Specific details of the mechanism of riboswitch function were only realized following the inspection of sequences from the expression platforms of improved switches identified from our high throughput screen; however, sequence data alone could

not explain many aspects of their function. For example, multiple natural riboswitches are known to be kinetically controlled[5,6] while others operate at a thermodynamic equilibrium[7]. The question remained as to whether our riboswitches are kinetically controlled, whereby each exists in two, mutually exclusive, non-interchangeable structures following transcription based upon the presence or absence of theophylline. Alternatively, our switches could operate at a thermodynamic equilibrium in which the ligand-free, or "off" conformation could be converted to a ligand-bound, or "on" conformation in the presence of theophylline. The following chapter describes both the in vivo and in vitro approaches to probing the mechanisms of our theophylline riboswitches. Additionally, a reexamination of our original riboswitch reveals the unique role played by the non-traditional *IS10-lacZ* reporter gene; specifically, the N-terminal IS-10 fusion and its implications for designing new synthetic riboswitches that respond to new small-molecule ligands.

3.2 Results and Discussion

3.2.1 Sequencing Suggests a Possible Mechanism of Action for Synthetic Riboswitch Function

In addition to identifying synthetic riboswitches that display excellent performance characteristics, the screen described in Chapter 2 provided a wealth of sequence information that allowed us to propose and test models for riboswitch function. Visual sequence analysis revealed several conserved or semi-conserved motifs that are complementary to different regions of the theophylline aptamer (Table 3.1).

Shown in Figure 3.1 are the *mFold*-predicted secondary structures[8,9] of two of the new synthetic riboswitches in the region extending from the 5'-end of the aptamer to the 3'-end of the AUG start codon of the *IS10-lacZ* gene. On the left-hand side are the two minimum energy structures—in both cases, the bases between the aptamer and the RBS are paired with the aptamer sequence, suggesting that translation might be inhibited in the
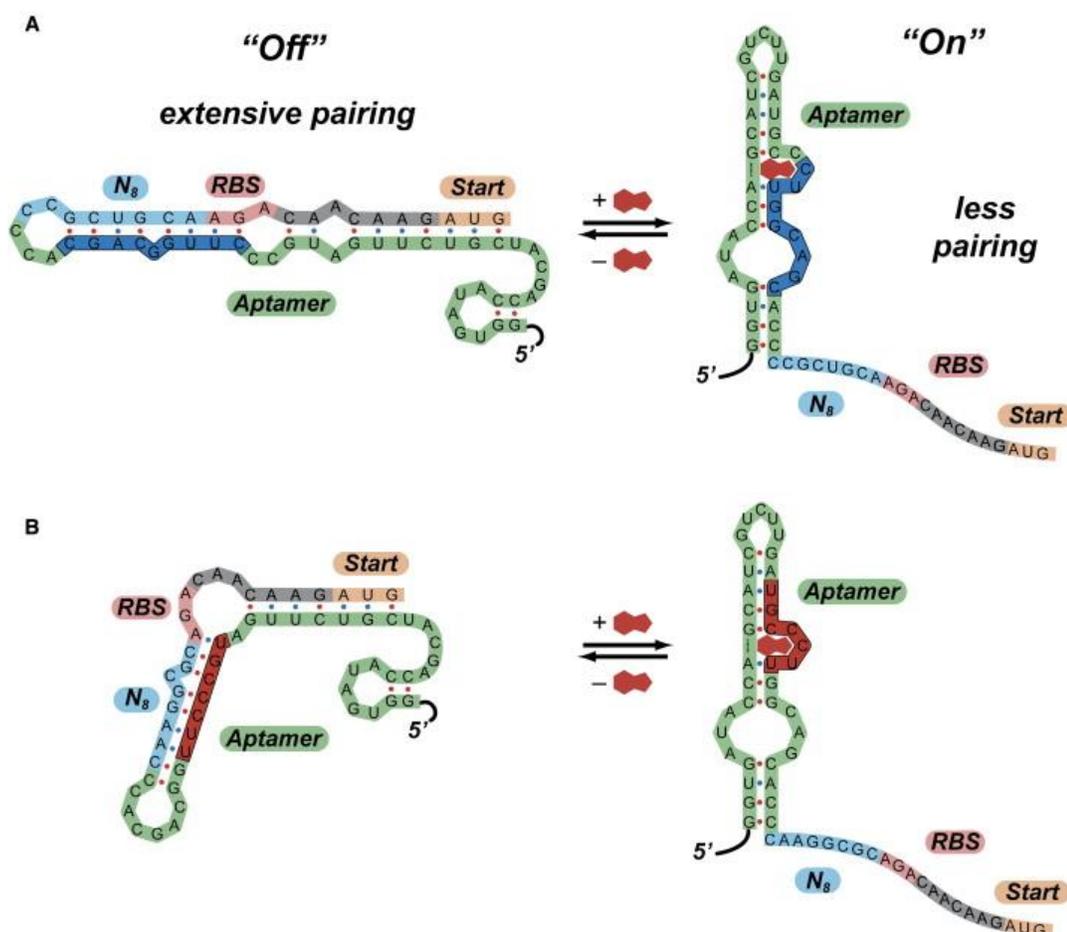
Figure 3.1 *Predicted Mechanisms of Action of Synthetic Riboswitches* (A) Predicted mechanism of action of clone 8.1. In the absence of theophylline (left), the 5' region can adopt a highly folded structure that extensively pairs the region that includes ribosome binding site to part of the aptamer sequence (shown in dark blue). In the presence of theophylline (right), the secondary structure shifts, such that the ribosome binding site is exposed. The secondary structure shown for the theophylline aptamer predicted by *mFold* and is identical to the structure determined by NMR. (B) Predicted mechanism of action of clone 8.2. In this case, the region containing the ribosome binding site pairs to a different region of the aptamer in the "off" state (shown in red). This demonstrates that several motifs may function as synthetic riboswitches.

absence of the ligand. On the right-hand side of Figure 3.1, the experimentally determined secondary structure for the theophylline-binding aptamer appears in the calculated structure and the RBS is not paired—in both cases, the calculated differences in free energy between the two structures are less than the free energy of theophylline binding (-9.2 kcal/mol), suggesting that conversion between the structures is thermodynamically favorable in the presence of theophylline. All of the synthetic riboswitches identified in our screens can be folded into one of the two folding motifs ("blue" or "red") shown in Figure 3.1.

The folds in Figure 3.1 suggest a possible switching mechanism in which extensive pairing in the region near the ribosome binding site prevents translation of β-galactosidase in the absence of theophylline. Such behavior is consistent with the studies of de Smit and van Duin who demonstrated that secondary structure near the ribosome binding site dramatically reduces the translation of the mRNA downstream[10-12]. Theophylline binding ($\Delta G_{bind} \sim -9.2$ kcal/mol)[13] could in principle drive the equilibrium toward the structures on the right, in which the known secondary structure of the aptamer is present and the ribosome binding sites are unpaired, which would increase the efficiency of translation.

To test whether base-pairing between the aptamer and a region near the ribosome binding site was important for function, we performed covariance experiments in which we mutated these bases to disrupt the putative base-pairing (Figure 3.2). Removal of several putative base pairs in clone 8.1 increased β-galactosidase expression 5-fold in the absence of the ligand, while restoring the base pairing by mutating the aptamer sequence

decreased the background level of β-galactosidase expression to its original level, consistent with the pairing hypothesis. Because these mutations also affect theophylline binding, these mutants were not active in the presence of theophylline.
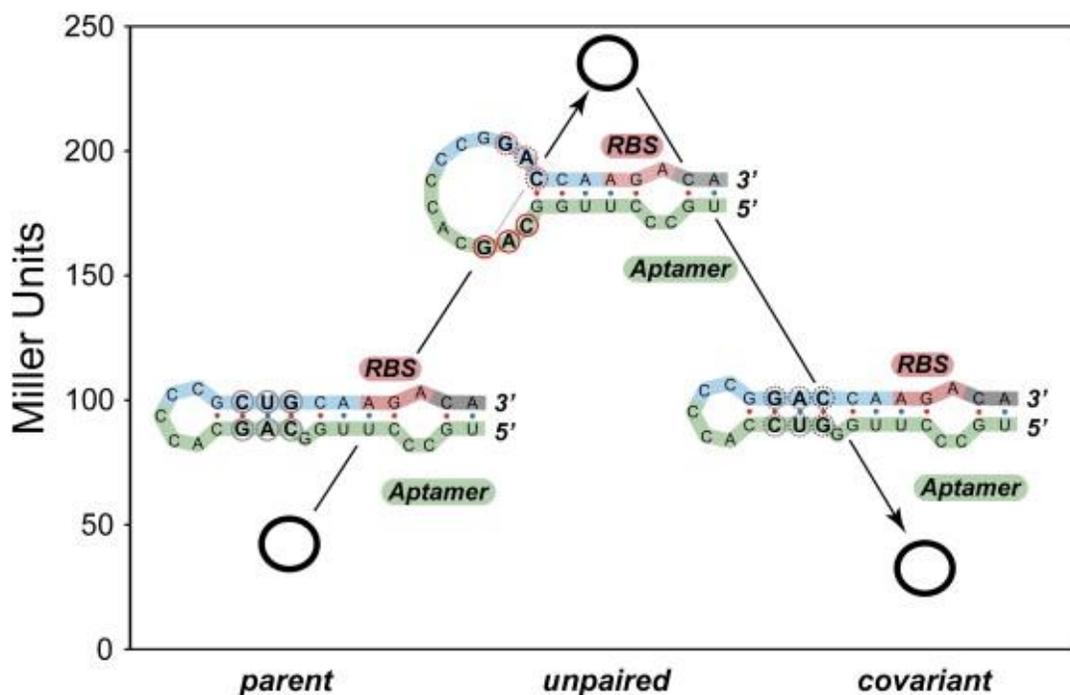
Figure 3.2 *Results from Covariance Experiments.* β-galactosidase activities in the absence of theophylline (open circles) for the parent synthetic riboswitch (clone 8.1) (left), a triple mutant that unpairs the region near the ribosome binding site (center, pSAL106), and a mutant that restores the pairing (right, pSAL104). Activities are expressed in Miller units (shown on the left axis); the standard errors of the mean are less than the diameters of the circles. For each construct, the predicted secondary structure is shown in the region near the ribosome binding site. The parent and covariant maintain the pairing and show low background levels of β-galactosidase activity. The mutant unpairs this region and leads to leaky expression in the absence of theophylline.

In light of these results that suggest that pairing in the region between the aptamer and the ribosome binding site is important for riboswitch function, we revisited the activation data from our previously reported synthetic riboswitches. In creating those switches, we did not explicitly engineer or screen for any particular sequence in this region. Reexamination of our previously reported riboswitch revealed that the expression

platform could conceivably pair with 2 separate regions of the mRNA transcript (Figures 3.3A and 3.4A). In the absence of theophylline, each of these pairings could potentially obstruct the ribosome binding site and limit translation. Because all of our newly identified switches appear to interact with the aptamer in repressing gene expression, it seemed reasonable to investigate the potential of the expression platform of our original switch to pair with the aptamer in a similar fashion.
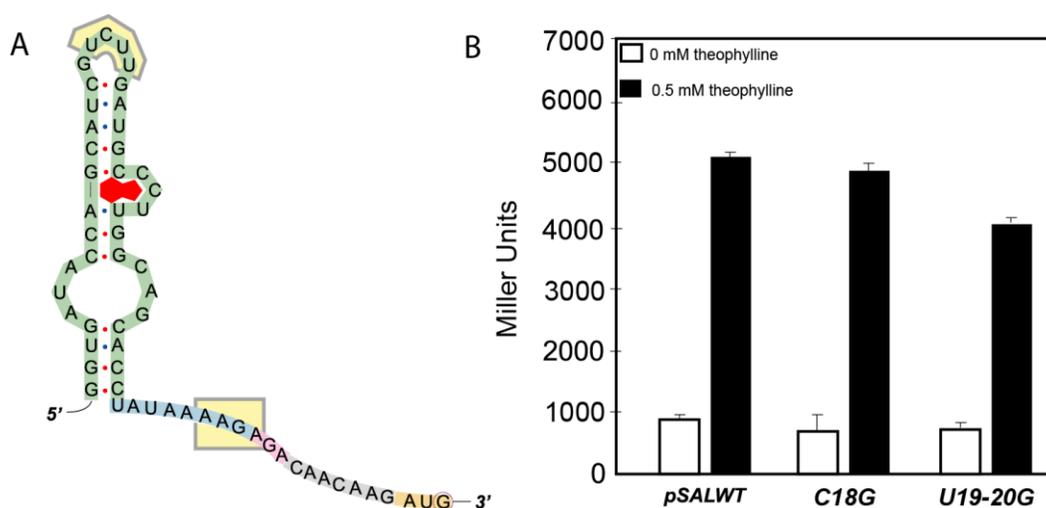
Figure 3.3 *Point mutations to probe the mechanism of our parent riboswitch.* (A) Parent riboswitch with potential base-paired regions of the aptamer and expression platform highlighted in yellow. Mutations were introduced into the aptamer (C and consecutive U residues), that would disrupt the possible base pairing in the absence of theophylline, increasing the level of gene expression in the absence of theophylline. (B) β-galactosidase activities in the absence (white) and presence (black) of theophylline for the parent synthetic riboswitch (pSALWT), a single (C18G) and double (U19-20G) mutant that potentially unpairs the region near the ribosome binding site. Activities are expressed in Miller units. For each construct, gene expression in the absence of theophylline remains consistent thus indicating the mutated residues do not participate in repressing gene expression.

Figure 3.4 *The Role of the IS10 Fusion in the Function of our Parent Riboswitch.* (A) Parent riboswitch with potential base-paired regions expression platform and the IS10 region of our reporter gene. (B) To investigate the need for the IS10 region, a construct was made containing the parent riboswitch but lacking the IS10 fusion (pSALWT-ΔIS10). β-galactosidase activities in the absence (white) and presence (black) of theophylline for the parent synthetic riboswitch (pSALWT), and the IS10 deficient mutant. Results indicate the IS10 region to be critical for the gene-dependent function of our parent switch. Activities are expressed in Miller units. (C) β-galactosidase activities in the absence (white) and presence (black) of theophylline for the best riboswitch found using our high-throughput screen (pSAL8.1), and the IS10 deficient mutant (pSAL8.1- ΔIS10). Activities are again expressed in Miller units. For each construct, gene expression in the absence and presence of theophylline remains consistent thus indicating the newly identified switch behaves independently of the reporter gene.

Point mutations introduced in the "tip" of the aptamer that would disrupt this potential

base pairing in the absence of theophylline, therefore disrupting the ability to represss

translation. These mutations did not appear to have any effect on the switches capabilities (Figure 3.3B) and could therefore be ruled out as bases participating in the switching mechanism. A close inspection of the *IS10lacZ* reporter revealed a 7 base sequence located 27 bases after the start codon (UUUCUCU) that is the precise reverse complement of the bases between the aptamer and the ribosome binding site. To test whether these regions pair to suppress gene expression in the absence of theophylline, we deleted the N-terminal IS10 fusion (pSALWT-ΔIS10). If bases within the IS10 sequence pair to suppress gene expression in the absence of theophylline, deleting this region should increase gene background levels of expression. Consistent with this model, the deletion of the IS10 sequence increased the β-galactosidase activity in the absence of theophylline approximately 15-fold (from ~800 to over 12,000 Miller units), confirming that the previously reported synthetic riboswitches depended on the presence of the IS10 sequence to function.

The models proposed in Figure 3.1 predict that all of the necessary elements to control gene expression are located in the 5'-UTR, suggesting that the sequence of the downstream gene should not matter. To test this, we replaced the entire 5'-UTR of the "leaky" pSALWT-ΔIS10 construct with the 5'-UTR from clone 8.1. Gratifyingly, this construct (pSKD8.1-ΔIS10) behaved nearly identically to clone 8.1, with low background levels of β-galactosidase expression in the absence of theophylline, and a strong increase in β-galactosidase expression in the presence of theophylline. Furthermore, we cloned the entire 5'UTR from clone 8.1 upstream of several reporter genes that lack the N-terminal

IS10 fusion, including, *dsRED*, and *cheZ*. These switches all function well and do not depend on the sequence of the gene downstream.

3.2.2 The Benefits of N-terminal Fusions in the Design of Synthetic Riboswitches

In revisiting the mechanism controlling our lab's original switches, we uncovered that the N-terminal IS10 protein fusion was required for their "gene-dependent" function. While we have since been able to identify switches that operate in a "gene-independent" fashion (i.e. the reporter gene sequence has no bearing on the switch's function) we believe the presence of N-terminal fusions can provide significant advantages in the engineering of new riboswitches. As our work suggests, gene dependent and gene independent mechanisms are equally plausible for the controlling gene expression. Using our screen, synthetic riboswitches isolated to eventually modulate metabolic pathways ultimately will not be used control the reporter gene (e.g. *lacZ*) that was used to identify it. If the switch identified operates in a gene dependent manner, it is highly unlikely that it will be transportable to another gene. This problem would be eliminated by screening for switches with a reporter gene containing a highly-tolerable N-terminal fusion that could be appended to any gene. Previous studies indicate that the N-terminal fragment of the IS10 transposase present in our reporter gene is post-translationally cleaved, making it an ideal fusion for these purposes. Additionally, cloning steps may be simplified by introducing a unique restriction site within the context of a these translational fusions. Figure 3.5 shows dose-response curves for our parent riboswitch controlling the expression of *IS10-lacZ, IS10-GFPuv* and *IS10-DsRed* reporter genes and demonstrates the utility of an N-terminal fusion.
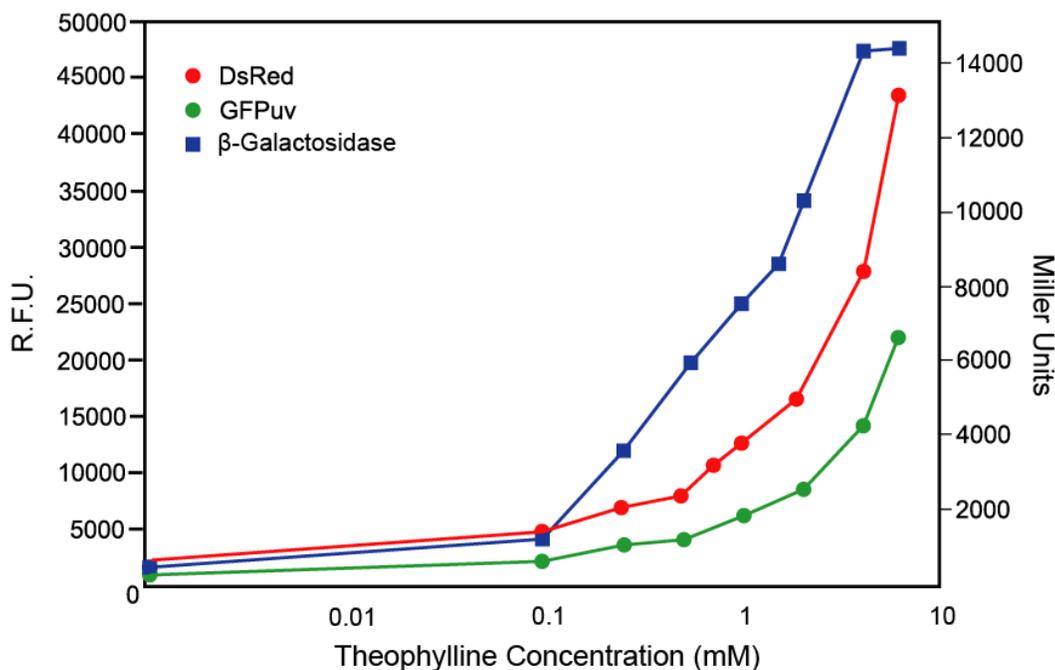
Figure 3.5. *Dose Response for Original Gene-Dependent Riboswitch with Multiple Reporter Genes.* The presence of the N-terminal IS10 fusion allows the gene-dependent parent riboswitch to modulate the expression of multiple reporters. Fluorescence is reported in relative fluorescence unit (R.F.U). Values were determined by subtracting out the background fluorescence of bacteria lacking any fluorescent protein. Measures of β-galactosidase activities are expressed in Miller units (right axis).

3.2.3 A Model for Synthetic Riboswitch Function

Our data suggest that these synthetic riboswitches display low background levels of translation in the absence of ligand and robust increases in the presence of the ligand for two reasons. The first is that in the absence of the ligand, the ribosome binding site is paired in such a way that translation is minimized. The second is that ligand binding drives the RNA to a conformation in which the RBS is unpaired. As described by de Smit and van Duin, sequences with pairing near the ribosome binding site typically display reduced translation rates because the 30S subunit of the ribosome binds most efficiently

to single-stranded regions of RNA[10-12]. Since translation is the slow step in protein production, the position of a pre-equilibrium between RNA structures that have the ribosome binding site paired or unpaired will factor directly into the rate expression for protein synthesis. Sequences that strongly favor pairing of the ribosome binding site are likely to show minimal levels of protein translation in the absence of ligand, since translation is very inefficient when the RBS is paired. Even if such sequences equilibrate with higher-energy structures where the RBS is unpaired (and translation is efficient), the relative population of these high-energy states, and thus the overall translation rate, will be low. The relative stabilities of these structures will dictate the background levels of translation in the absence of the ligand. However, if addition of the ligand shifts the population of mRNAs to a structure(s) where many of the ribosome binding sites are unpaired, this model predicts robust increases in protein translation from these efficiently translated mRNAs.

Figure 3.6A shows the predicted secondary structures and free energies of the region of the mRNA of clone 8.1 extending from the 5'-end of the aptamer to the 3'-end of the start codon. In the minimum energy structure ("Off", *left*), the RBS is paired. In the middle is the lowest energy structure ($\Delta\Delta G$=+5.5 kcal/mol) in which the ribosome binding site is predicted to be unpaired ("On"); this is also the lowest energy structure in which the secondary structure of the aptamer is present. On the right is the same calculated fold, but the free-energy has been lowered by 9.2 kcal/mol, the experimentally-determined free energy of theophylline binding ("Off•L").

An equilibrium model predicts that in the absence of the ligand, the majority of the RNA population adopts the "Off" structure in which the RBS is paired because the "On" structure is significantly uphill in free energy. Addition of theophylline provides the thermodynamic driving force to shift the equilibrium toward the "On•L" structure. These observations can be represented in the model shown in Figure 3.6B, where the pseudo-first order rate constants ($k_{obs\ 1}$, $k_{obs\ 2}$, and $k_{obs\ 3}$) include the rates of ribosome binding and the initiation of translation and the concentration of the 30S ribosomal subunit. The studies of de Smit and van Duin predict that the translation efficiency from the "Off" structure, in which the ribosome binding site is paired, would be significantly lower than the efficiency from the structures in which the RBS is unpaired ("On" and "On•L"), thus $k_{obs\ 1} <<< k_{obs\ 2} \approx k_{obs\ 3}$. Thus, in the absence of ligand, the overall translation rate would be low, and the rate would increase in the presence of the ligand.

While this model qualitatively fits our experimental observations, determining accurate rate constants and mRNA concentrations in vivo is very challenging. Nevertheless, we can test elements of the model, such as the ability of RNA structures to equilibrate in the presence of the ligand. Demonstrating that an mRNA transcript can undergo a ligand-inducible shift from an untranslated state into a translated state in vivo would provide evidence for the equilibrium model proposed above, and would also argueagainst a kinetic model, in which the committed step for translation occurs during transcription (i.e. a transcript is translated only if the appropriate ligand is present at the time of transcription; Figure 3.7A) [5,6].
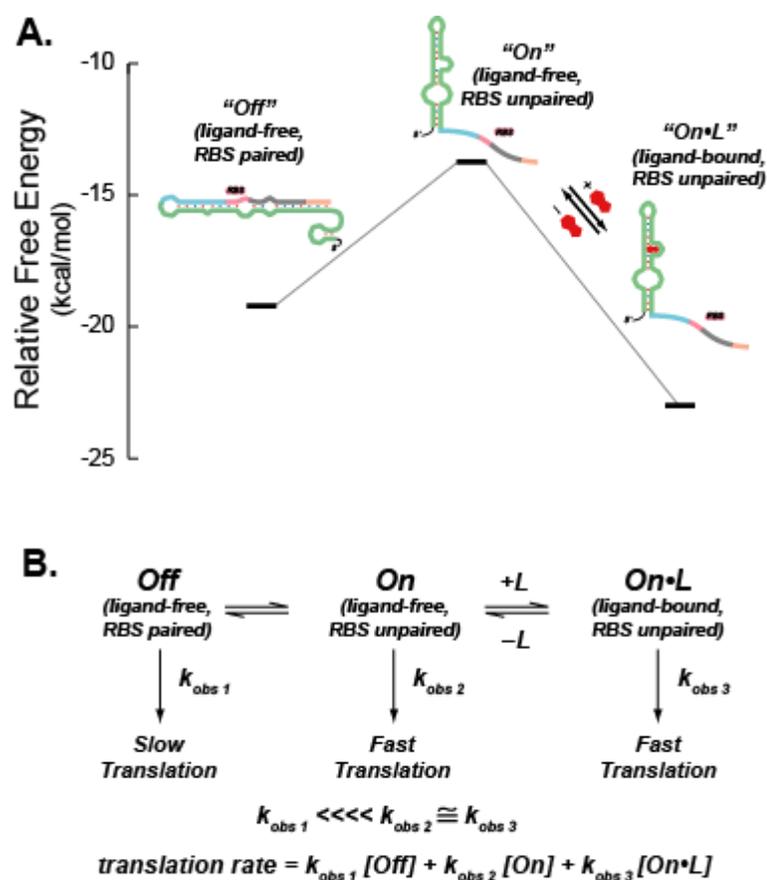
Figure 3.6 *Model for Synthetic Riboswitch Function* (A) Predicted structures and free energies of a portion of the 5′ region of clone 8.1 determined by mFold. To represent the free energies of the bound structures, we added the experimentally determined free energy of theophylline binding (−9.2 kcal/mol) to the free energy of the unliganded structure determined by mFold since the secondary structure of the mTCT-8-4 aptamer does not change upon ligand binding. (B) Kinetic model for synthetic riboswitch function. In the "Off" state, the ribosome binding site is paired and translation is slow (represented by the pseudo-first order rate constant $k_{1\ obs}$, which accounts for ribosome binding and translation). The "Off" state is proposed to be in equilibrium with the ligand-free "On" state, in which the ribosome binding site is unpaired and translation is relatively fast (represented by the pseudo-first order rate constant $k_{2\ obs}$, which is much greater than $k_{1\ obs}$). Since this state is not highly populated, background translation is minimal. Addition of ligand (L) shifts the equilibrium to the ligand-bound "On•L" state, in which the ribosome binding site is unpaired and translation is relatively fast (represented by the pseudo-first order rate constant $k_{3\ obs}$, which is also much greater than $k_{1\ obs}$).

To test these models, *E. coli* that were constitutively transcribing a riboswitch-controlled *lacZ* gene were grown in the absence of theophylline. Theophylline was added and the β-galactosidase expression at various time points was assayed using Miller's method (Figure 3.7B). β-galactosidase activity (as measured by A_{420}) first appears approximately 200 seconds after the addition of theophylline and continues to increase at a steady rate. During the 200 seconds before β-galactosidase activity appears, several events must occur: 1) theophylline must enter the cell; 2) it must bind to the mRNA; 3) it must induce switching (through whatever mechanism); 4) the 'switched' mRNA must be translated; and 5) β-galactosidase must fold and become active. Liang et al. have shown β-galactosidase activity in *E. coli* appears within 60 seconds upon using IPTG to initiate *transcription* of *lacZ*. Their results show that the transcription rate of *lacZ* must be at least 57 nt/s (3420 nt/60 s)[14], and that the subsequent acts of translation, folding, and appearance of β-galactosidase activity must collectively take less than 60 seconds[14]. After accounting for the 60 s required for translation, folding, and the appearance of enzymatic activity, the fact that we do not observe β-galactosidase activity until 200 s after theophylline addition suggests that collectively theophylline entry, binding, and switching may take up to 140 s. Although this delay could be consistent with a kinetic barrier for interconversion between a non-translatable mRNA and a translatable mRNA, we cannot yet determine whether such conformational changes are responsible for the switching behavior.

An alternative to a theophylline-driven equilibrium model is a "kinetic" mechanism whereby the fate of an mRNA is determined at the time of transcription.

Much of the data thus far can also be explained in terms of a co-transcriptional kinetic model in which the majority of the mRNA pool folds into a non- or weakly-translatable conformation when theophylline is not present (e.g. the "Off" conformation shown in Figure 3.7A). If theophylline can induce some fraction of the mRNA to fold into the highly translatable "On-L" conformation, theophylline-dependent activation of translation could occur regardless of whether the translatable and non-translatable mRNA conformations can interconvert. Given that A) *lacZ* transcription is fast (~60 nt/s)[14]; B) both mRNA folding and translation occur co-transcriptionally in *E. coli*; C) the entire "functional unit" of the riboswitch is located within the first 100 transcribed bases; and D) the proposed "On" and "Off" conformations are both small hairpins, it is extremely likely that the fate (translatability) of an mRNA is determined within seconds of the initiation of transcription. As such, theophylline would have to be present during transcription for β-galactosidase activity to be observed.

To test whether β-galactosidase expression is dependent on the presence of theophylline while transcription is active; we used the antibiotic rifampicin to halt transcription and asked whether subsequent addition of theophylline could induce β-galactosidase expression. Rifampicin prevents RNA polymerase from entering the elongation phase following the initiation of transcription, but it has no effect on RNA polymerase once it has reached the elongation phase. The quick action of rifampicin (~5 s), coupled with the fast elongation rate of *E. coli* RNA polymerase (~60 nt/sec), suggests that within 10 s of rifampicin addition, no new transcripts are being produced and any elongating transcripts are ~100 nt beyond the 5'-UTR.
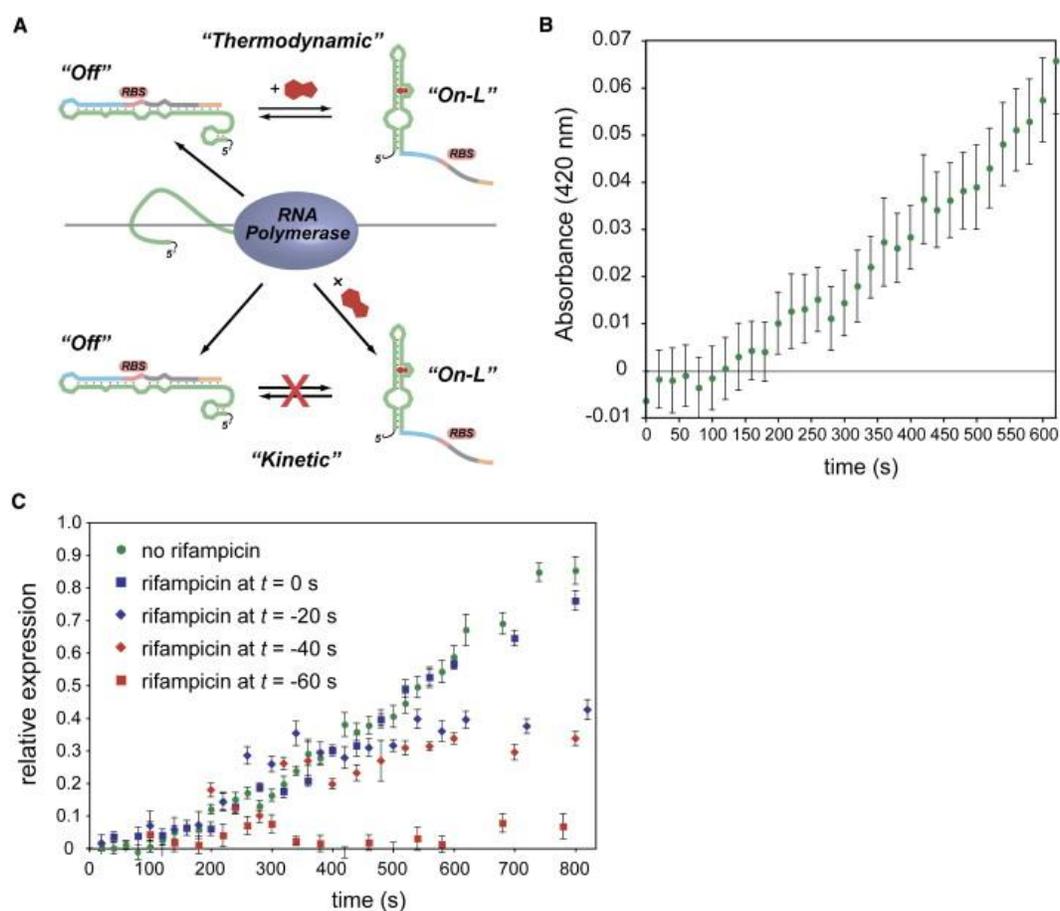
Figure 3.7 *Experiments to Determine Switching Mechanism* (A) Top, a thermodynamic model for riboswitch function. Newly synthesized RNA adopts a folded conformation that prevents translation ("Off"). Addition of theophylline can drive the conformation to the "On-L" conformation. Bottom, a kinetic model in which a newly synthesized RNA adopts either an "Off" or an "On-L" conformation depending on whether ligand is present during transcription; however, the "Off" and "On-L" conformations do not equilibrate. (B) Time course for appearance of β-galactosidase activity (measured by ONPG hydrolysis) for clone 8.1. Theophylline was added at t = 0 to a final concentration of 3 mM; significant activity appears at 200 s. (C) Time course for appearance of β-galactosidase activity (measured by ONPG hydrolysis) for clone 8.1. Theophylline was added at t = 0 to a final concentration of 3 mM; rifampicin was added at the times shown in the legend. Rifampicin does not eliminate β-galactosidase activity when added up to 40 s prior to the addition of theophylline, supporting the thermodynamic model shown in (A) in which theophylline can induce expression even in the absence of active transcription. Error bars are ± SD.

Figure 3.7C shows the theophylline-induced expression of β-galactosidase from cultures harboring switch 8.1 in which rifampicin was added 0, 20, 40, or 60 s prior to the addition of theophylline. While adding rifampicin 20 or 40 s prior to adding theophylline reduces the overall level of β-galactosidase activity, as would be expected from the

inhibition of mRNA synthesis, the onset of β-galactosidase activity occurs at the same time (~200 s) as it does in the absence of rifampicin. Assuming that rifampicin acts within 5 s and that *E. coli* RNA polymerase transcribes mRNA at 60 nt/s, by adding rifampicin 40 s prior to theophylline, no new transcripts should be synthesized, and any elongating transcripts should be >1800 nt long, far beyond the 100 nt 5'-UTR containing the riboswitch, at the time that theophylline is added. Though we cannot easily determine how quickly theophylline enters the cell and reaches an effective concentration, we can consider 2 limiting scenarios: theophylline enters essentially immediately, or theophylline entry is delayed by up to 200 s. If theophylline enters immediately, the delay before the onset of β-galactosidase activity is difficult to explain, as the co-transcriptional model predicts that the "decision" to translate occurs early in the process of transcription; since translation and folding occur within 60 s, leaving a 140 s gap. If theophylline entry is slow (up to 140 s), the co-transcriptional model becomes even less tenable because at +140 s, all transcripts should be completed, regardless of the time of addition of rifampicin. The fact that β-galactosidase activity continues to increase steadily long after the addition of theophylline, and > 300 seconds after the addition of rifampicin (Figure 3.7C), argues against a co-transcriptional model for the activation of expression and is more consistent with a model in which theophylline induces a conformational change in an existing mRNA (Figure 3.7A).

3.2.4 Possible Design Implications for Synthetic Riboswitches

While the generality of this equilibrium model will be investigated in future studies, in its current form, the model may provide guidelines for the design of synthetic

riboswitches that activate protein translation. In particular, the model suggests that it will be critical to balance the relative free energies of "off state", the ligand-free "on state", and the ligand-bound "on state". Since the energy differences of the ligand-free "on state" and the ligand-bound "on state" are dictated by the free energy of ligand binding to the aptamer, this puts thermodynamic constraints on potential structures for the "off state". If the free energy of the "off state" is close to that of the ligand-free "on state" (and there is interconversion), background translation in the absence of the ligand will be high. However, if the "off state" is very stable, binding of ligand may not result in a sufficient population of the ligand-bound "on state" to allow translation, even if this state is highly translatable. We anticipate that this model may help guide the design of future synthetic riboswitches.

3.2.5 Possible Evolutionary Implications for Natural Riboswitches

Our results show that beginning with an aptamer sequence that recognizes a small molecule, it is relatively straightforward to create synthetic riboswitches. Indeed, we have discovered a variety of high-performing synthetic riboswitches where both the length and composition of the sequence between the aptamer and the ribosome binding site vary considerably. Furthermore, we have shown two new functional motifs where the bases between the aptamer and the ribosome binding site pair to distinct regions of the aptamer, and retrospective analysis of a previously reported synthetic riboswitch reveals a third motif where these bases pair to part of the coding sequence to repress translation in the absence of the ligand. Taken together, our results show that there are many straightforward ways to generate synthetic riboswitches by inserting an aptamer sequence

into the 5'-UTR of a gene. With relatively few base changes, the dynamic ranges of these switches can be improved dramatically to give sensitive ligand-dependent genetic control elements. Additionally, these few base changes can take a switch from operating in a gene-dependent manner to one that operates independently of the gene it regulates. In turn, this allows the switch to potentially move about the genome while maintaining its ability to regulate expression.

It has been suggested that natural riboswitches may be molecular fossils from an RNA world, where RNA sequences that once served to bind to ligands (perhaps as cofactors for RNA catalysis), have since been co-opted for the purpose of gene regulation. Our results show that it is relatively straightforward on the laboratory timescale to convert an aptamer into a synthetic riboswitch. We suggest that using similar mechanisms, riboswitches may have evolved from pre-existing aptamers.

3.3 Experimental

General Considerations. All plasmid manipulations utilized standard cloning techniques[15]. All constructs have been verified by DNA sequencing at the NSF-supported Center for Fundamental and Applied Molecular Evolution at Emory University. Purifications of plasmid DNA, PCR products, and enzyme digestions were performed using kits from Qiagen. Theophylline, *o*-nitrophenyl-β-D-galactopyranoside (ONPG) and ampicillin were purchased from Sigma. X-gal was purchased from US Biological.

Synthetic oligonucleotides were purchased from IDT. All experiments were performed in *E. coli* TOP10 F' cells (Invitrogen) cultured in media obtained from EMD Bioscience.

Computational RNA Folding Protocol

Secondary structures of riboswitches were determined using the RNA mFold web server (http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi). Sequences stretching from 5'-end of the theophylline aptamer to 3'-end of the AUG start codon of riboswitches 8.1 and 8.2 were entered and secondary structures were calculated without constraints at 37 ºC with 50% suboptimality yielding 9 and 8 structures respectively. Extending the sequence to the transcription start site increased the number of suboptimal folds, but did not change the structure of the lowest-energy fold.

pSAL106

(Ptac1 - mTCT4-8 aptamer - CCGGACCA - *IS10LacZ*)

This plasmid is derived from pSAL8.1 and was made such that 3 mutations were introduced into the region between the mTCT4-8 aptamer and the RBS. The introduced mutations disrupt base pairing between the linker region and the aptamer in the absence of ligand. The plasmid was created using cassette mutagenesis. A PCR product was amplified by PCR from pSAL8.1 using forward primer SAL-39 which anneals to the mTCT4-8 aptamer and contains the desired mutations and reverse primer SKD 56. PCR product was gel purified digested with *Kpn*I and *HindIII*. The digested insert was again gel purified and ligated into vector pSKD445.1 which was previously digested with the same enzymes, dephosphorylated, and gel purified.

pSAL104

(Ptac1 – mutated mTCT4-8 aptamer - CCGGACCA - *IS10LacZ*)

This plasmid is derived from pSAL8.1 and was made such that 3 bases from the region between the mTCT4-8 and the RBS are exchanged with the complementary bases they are believed to be pairing with in the aptamer. The plasmid was created using cassette mutagenesis. A PCR product was amplified by PCR from pSAL8.1 using forward primer SAL-37 which anneals to the mTCT4-8 aptamer and contains the desired mutations and reverse primer SKD 56. PCR product was gel purified digested with *Kpn*I and *HindIII*. The digested insert was again gel purified and ligated into vector pSKD445.1 which was previously digested with the same enzymes, dephosphorylated, and gel purified.

pSALWT-ΔIS10

A cassette strategy was used to generate pSALWT-ΔIS10. A PCR product (A) was generated by using pSKD445.1 as a template with forward primer SKD-178 which anneals to pSKD445.1 upstream of the mTCT4-8 aptamer and the reverse primer SAL 024 which anneals to the expression platform of the parent riboswitch and contains an overlap with the full-length *lacZ* gene . A separate PCR product (B) was generated using forward primer SAL 028 and reverse primer ST008 using pRS415 as a template. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD 178 and reverse primer ST 088, to give PCR product C. PCR

product C was digested with KpnI and ClaI and cloned into those sites in pSKD445.1 to yield pSALWT-ΔIS10.

pSAL8.1-ΔIS10

The same strategy used to create pSALWT-ΔIS10 was used to generate pSAL8.1-ΔIS10 with the following exceptions. PCR product (A) was generated using forward primer SKD 178 and reverse primer SKD 328 using SKD445.1 as a template. PCR product (B) was generated using forward primer SKD 329 and reverse primer ST088 using pRS415 as a template. PCR products A and B mixed and amplified as before to generate PCR product (C) which was digested with KpnI and ClaI and cloned into those sites in pSKD445.1 to yield pSAL8.1-ΔIS10.

pSALWT-C18G

A cassette strategy was used to generate pSALWT-C18G. A PCR product (A) was generated by using pSKD445.1 as a template with forward primer SKD-178 which anneals to pSKD445.1 upstream of the mTCT4-8 aptamer and the reverse primer SAL 019 which anneals to mTCT4-8 aptamer and contains the C18G mutation . A separate PCR product (B) was generated using forward primer SAL 018 which also contains the appropriate mutations and reverse primer SKD 56 using pSKD445.1 as a template. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD 178 and reverse primer SKD 56, to give PCR product C. PCR

product C was digested with KpnI and HindIII and cloned into those sites in pSKD445.1 to yield pSALWT-C18G.

pSALWT-U19-20G

The same cassette strategy used to create pSALWT-C18G was used to generate pSALWT-U19-20G with the following exceptions. A PCR product (A) was generated by using the reverse primer SAL 023 which anneals to mTCT4-8 aptamer and contains the U19-20G mutations. PCR product (B) was generated using forward primer SAL 022 which also contains the appropriate mutations. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD 178 and reverse primer SKD 56, to give PCR product C. PCR product C was digested with KpnI and HindIII and cloned into those sites in pSKD445.1 to yield pSALWT-U19-20G.

Rifampicin Assay

An aliquot of a saturated overnight culture (500 μL) was used to inoculate 50 mL of LB supplemented with ampicillin (50 μg/mL). The fresh culture was grown with shaking (300 rpm, 37 ºC) to an OD_{600} of 0.4-0.5 (approx 2.75 h). The culture was split into two 16.6 mL fractions. Theophylline (50 mg/mL in DMSO, warmed to 37 ºC) and rifampicin (50 mg/mL in DMSO) were added as appropriate to one of the cultures such that the final concentrations were 3.0 M theophylline, 250 μg/mL rifampicin; the other culture was used as a control. Aliquots of the cultures (4 x 100 μL each) were collected at 20, 40, or 100 s intervals using a Freedom-EVO liquid handling system (TECAN) and were transferred to individual wells of a 96-well microtiter plate that were pre-filled with Pop

Culture® (10 µL). A 30 µL aliquot from each well of lysed cells was added to 119 µL of Z-buffer in a microtiter plate (see above). ONPG (29 µL, 4 mg/mL in 100 mM NaH$_2$PO$_4$) was added and was allowed to hydrolyze for 1 hr at 30 ºC, followed by quenching with Na$_2$CO$_3$ (75 µL of a 1 M solution). The OD$_{420}$ for each well was determined.

Plasmid SAL006.1

(Ptac1-mTCT4-8 aptamer-TATAAAAG–AGACAACAAG *-IS10-DsRedExpress*)

A cassette mutagenesis strategy was used to generate pSAL006. A PCR product (A) was generated by using pSKD445.1 as a template with forward primer SKD-178 which anneals to pSKD445.1 upstream of the mTCT4-8 aptamer and the reverse primer SKD-163 which anneals to the IS10 region of our reporter gene. A separate PCR product (B) was generated using forward primer SAL-002 and reverse primer SAL-003 using pDsRedExpress as a template. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD-178 and reverse primer SAL-003, to give PCR product C. PCR product C was digested with NdeI and AflIII and cloned into those sites in pSKD850.2 to yield plasmid pSAL006.1.

pSKD785.1

(Ptac1-mTCT4-8 aptamer- TATAAAAG–AGACAACAAG-*IS10-GFPuv)*

A cassette strategy was used to generate SKD 785.1. A PCR product (A) was generated by using pSKD 625mp3 as a template with forward primer SKD-160 which anneals to the transcriptional terminator of pSKD625mp3 and the reverse primer SKD-185 which

anneals to the pUC19 vector, 3' to the SapI site. A separate PCR product (B) was generated using forward primer SKD 186 and reverse primer SKD 187 using pGFPuv as a template. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD 187 and reverse primer SKD 185, to give PCR product C. PCR product C was digested with SapI and HindIII and cloned into those sites in pSKD445.1 to yield pSKD785.1.

3.4 References

(1)   Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247-13254.

(2)   Werstuck, G.; Green, M. R. *Science* **1998**, *282*, 296-298.

(3)   Grate, D.; Wilson, C. *Bioorg. Med. Chem.* **2001**, *9*, 2565.

(4)   Harvey, I.; Garneau, P.; Pelletier, J. *RNA* **2002**, *8*, 452-463.

(5)   Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M. *Biochemistry* **2005**, *44*, 13404-14.

(6)   Wickiser, J. K.; Winkler, W. C.; Breaker, R. R.; Crothers, D. M. *Mol. Cell* **2005**, *18*, 49-60.

(7)   Renate, R.; Kathrin, L.; Dagmar, G.; Ronald, M. *ChemBioChem* **2007**, *8*, 896-902.

(8)   Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911.

(9)   Zuker, M. *Nucl. Acids Res.* **2003**, *31*, 3406-3415.

(10)   de Smit, M. H.; van Duin, J. *Proc. Nat. Acad. Sci. U.S.A.* **1990**, *87*, 7668-7672.

(11)   de Smit, M. H.; van Duin, J. *J. Mol. Biol.* **1994**, *244*, 144-150.

(12)   de Smit, M. H.; van Duin, J. *J. Mol. Biol.* **1994**, *235*, 173-184.

(13)   Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(14)   Liang, S. T.; Ehrenberg, M.; Dennis, P.; Bremer, H. *J. Mol. Biol.* **1999**, *288*, 521-538.

(15)   Sambrook J, R. D. *Molecular Cloning : A Laboratory Manual.*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2001.

Chapter 4:

A Flow Cytometry-Based Screen for Synthetic Riboswitches[ii]

4.1 Introduction

While much is known a about the structure and function of natural riboswitches, the principles governing the creation of a synthetic riboswitch from existing RNA aptamers remain unclear. Although sequence data from our previously identified theophylline riboswitches have allowed us to propose and test a mechanism for how these new riboswitches functioned[1], and illuminated some principles that may ultimately enable the *de novo* design of new synthetic riboswitches from existing RNA aptamers, we believe that at present, genetic screens or selections remain the best strategy for discovering new synthetic riboswitches.

Current progress being made in the creation of new synthetic riboswitches has been reliant on innovative screening methods[2-4]. Screening for riboswitches presents a unique challenge in that one must seek out both high-levels of gene expression in the presence of ligand and low levels of expression in the absence. Ideally, a method for discovering riboswitches should be rapid, inexpensive, and very high-throughput. Additionally, the method should have the ability to quantitatively distinguish the switches with the best performance characteristics (i.e. for an 'on' switch, this would typically

mean low background expression in the absence of ligand and robust increases in gene expression in the presence of the ligand) from those switches that are leaky, or only weakly activate (or repress) gene expression in the presence of the ligand. Recently, our lab[1,4], and others[5,6] have reported a variety of methods to screen RNA libraries for robust-performing synthetic riboswitches. Using a yeast-based selection method, Suess and coworkers screened libraries of up to 50,000 members and isolated riboswitches that reduced gene expression 7.5-fold in the presence of the antibiotic neomycin[5]. Working in *E. coli*, Wieland and Hartig screened a 64-member library of theophylline-dependent aptazymes and discovered constructs that activated gene expression 10-fold[6]. Also working in *E. coli*, our lab has developed both a high-throughput screen[1] and a high-throughput selection[4] for theophylline-dependent riboswitches.

The enzymatic assays we have employed prove to be cumbersome, time consuming and labor intensive when libraries of sizes greater than 100,000 needs to be screened. These challenges are worsened if one wishes to exhaustively screen an entire library or isolate only rare members of the population such as those switches that are highly dynamic and display low levels of expression in the absence of ligand. Library sizes become even more unmanageable if one considers the lack of evidence for the strongest binding RNA aptamer as being the best candidate with which to create a riboswitch. If this is indeed the case, it may be necessary to screen libraries of aptamers in conjunction with libraries of randomized expression platforms. These issues present a major challenge in the creation of new synthetic riboswitches. Future successes in the design of new RNA-based genetic control elements will require new screening methods with larger throughputs than those currently available. Faced with these challenges, we

sought to develop a screen for synthetic riboswitches that would offer the throughput of a genetic selection and also provide the tunable nature of a genetic screen. To address these issues we have developed a method utilizing flow cytometry to enrich a library of mutant riboswitches facilitating the isolation of those switches with improved qualities.

To study the efficiency of gas mask filters used during World War II, Gucker and coworkers[7] developed a photoelectric apparatus capable of rapidly counting individual bacterial spores and smoke particles. Published in 1947, this work is often thought to be the first reported use of flow cytometry. In it, the authors speculated that this technology would have a wide variety of applications in the field of bacteriology. Despite these claims, flow cytometry has largely been confined to study of mammalian cells thanks in large part to the small size of bacteria and low concentrations of measurable components residing within cells. The incorporation of arc-lamps[8], as well as advances in optics technologies and improved fluorescent stains[9] helped open the door to useful analyses of bacteria using flow cytometry.

Flow cytometry allows one to sort a heterogeneous population of cells based on physical and chemical properties ranging from size and shape to intracellular pH levels[10]. Fluorescence-activated cell sorting (FACS) rapidly and quantitatively monitors the fluorescence characteristics of a population of bacteria while also enabling the user to separate individual cells with desired characteristic from the population. These capabilities are tremendously useful in genetic screening experiments requiring the isolation of interesting mutants from large libraries of clones.

In the past several years, advances in FACS and the development of new genetically encoded fluorophores[11] have led to remarkable increases in throughput and reductions in cost. Moreover, while FACS has most often been used to sort populations of eukaryotic cells, technological advances have opened the door to sorting large (~10^8 member) libraries of small cells, such as bacteria[12], and even cell-like compartments, such as oil-water emulsions[13,14]. Because FACS can readily distinguish small differences in fluorescence emission intensity, it has become a powerful screening technique in directed evolution experiments[15], where it is critical to quantitate small changes in gene expression or enzyme activity. We thus anticipated that FACS would be useful for discovering synthetic riboswitches that display low background levels of gene expression in the absence of a ligand and robust increases in gene expression in its presence.

Here we show that FACS can rapidly enrich a population of rare, dynamic riboswitches from a large library. The enriched population contained a high percentage of functioning riboswitches that displayed improved characteristics compared to those isolated using our previous screens and selections. Sequence data from these switches also sheds new light on the importance of the ribosome binding site, or Shine-Dalgarno sequence, on the function of synthetic riboswitches. Taken together, these results may provide new insights into the design of new synthetic riboswitches using *in vitro* selected aptamers.

4.2 Results

4.2.1 Screening of an N_8 Library

As a proof-of-principle experiment to determine whether FACS can identify synthetic riboswitches from large libraries, we reexamined a library we had screened using our previously reported *lacZ* assay[1]. This library consisted of the theophylline aptamer, followed by a sequence of 8 randomized bases, a ribosome binding site, and the red-fluorescent reporter gene *DsRed express*[16]. *E. coli* were transformed with the plasmid library (theoretically 65,536 members), plated onto selective LB-agar, and grown overnight. Approximately 65,000 transformants were scraped from the plate, grown in selective LB media in the absence of theophylline, and this culture was analyzed using flow cytometry (Figure 1 a). Analysis of the population indicated that a large percentage of clones displayed levels of fluorescence emission within the first decade of the log-scale. This was somewhat surprising as blue/white screening of the same library using the
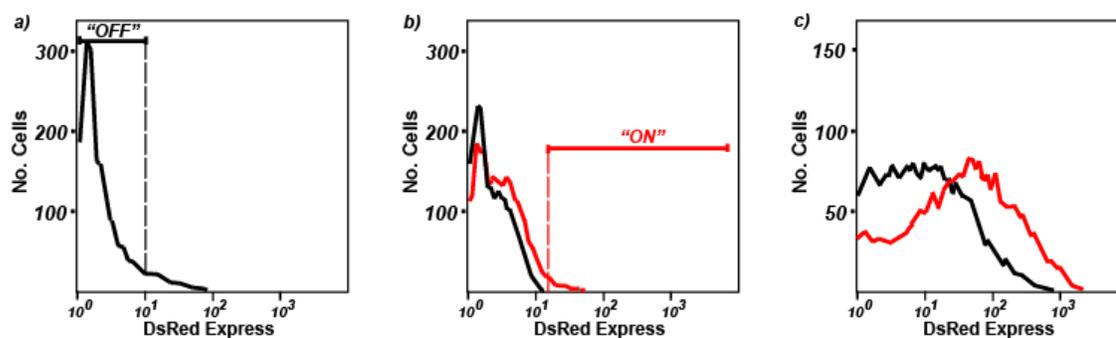
Figure 4.1 *FACS histograms representing a population of bacteria possessing a library of mutant theophylline riboswitches.* a) Profile of the N_8 library grown in the absence of theophylline. For the initial negative selection, a gate was set to collect bacteria falling in the "OFF" region. b) Profile of the N_8 bacteria collected in the initial negative selection grown in the absence and presence of theophylline (1 mM, red). Bacteria collected from the culture grown with theophylline were those that exhibited an increase in fluorescence when compared to the population of bacteria grown without. These are represented by the "ON" region and amount to approximately 2% of the culture. c) Profile of the N_8 bacteria following one round of negative selection and one round of positive selection. Riboswitches from this enriched population were subsequently cloned in front of a β-galactosidase reporter gene and screened enzymatically leading to the discovery of 8.1*. All histograms are on the same scale and were generated with identical instrument settings.

lacZ reporter gene revealed nearly all (>99%) of the population was blue in the absence of theophylline (i.e. most cells appeared to have high levels of gene expression, rather than the apparently low levels observed here). These results highlight a limitation of our plate-based *lacZ* assay, which only allowed qualitative ranking of gene expression and may have led us to overlook clones that appeared pale blue in the absence of theophylline, but nonetheless had acceptably low levels of background expression.

Encouraged by the initial flow cytometry results, we positioned a gate to collect bacteria displaying levels of fluorescence below the first decade of the log-scale (Figure 4.1 a). These bacteria were cultured overnight, and aliquots of the overnight culture were used to inoculate two separate cultures, one with theophylline (1 mM) and one without. Analysis of these cultures revealed that our initial attempt to isolate clones displaying low levels of fluorescence was successful, and that a small percentage of this sorted population exhibited a theophylline-dependent increase in fluorescence (Figure 4.1 b). These cells were collected, cultured overnight, and aliquots of the overnight culture were used to inoculate two separate cultures, one with theophylline (1 mM) and one without. These cultures were grown to an OD_{600} between 0.4-0.6 and were analyzed using flow cytometry, which revealed a dramatic theophylline-dependent shift in fluorescence (Figure 4.1c).

To quantify how individual riboswitches performed, the culture was grown to saturation, the plasmid DNA of the entire population was isolated, and the 5'-untranslated regions were subcloned upstream of a *lacZ* reporter gene. This library was used to transform *E. coli*, which were grown on selective LB-agar containing X-gal, but

no theophylline to give approximately 3,000 transformants. Using a robotic colony picker, we picked the 48 whitest colonies from this plate and performed our previously described enzymatic assay for riboswitch activity. To our satisfaction, greater than half of these clones were functioning riboswitches, the best of which, clone 8.1*, displayed a 59-fold activation ratio in the presence of 1 mM theophylline (Figure 4.3 a). By way of comparison, the best riboswitch isolated from our previous screen had an activation ratio of only 36, and its isolation required 846 separate β-galactosidase assays, compared to the 48 performed here.

Sequence analysis revealed that clone 8.1* harbored an expression platform identical to our previously reported switch (clone 8.1; 36-fold activation ratio), but had a point mutation that changed the Shine-Dalgarno sequence from AAG to the stronger AGGA[17]. It is interesting to note that this mutation occurred immediately 3' to the randomized 8-base region, suggesting that it was introduced by either an impurity in the oligonucleotides used to create the library, or by a replication error in *E. coli*. Sequencing of the PCR product used to create the library revealed that the vast majority of the population contained 8 nucleotides between the aptamer and the ribosome binding site (AGA). While this low resolution assay cannot detect a rare GGA sequence, it does rule out the possibility that the majority of the library was comprised of GGA ribosome binding site sequences. As such, we anticipate that this sequence would be underrepresented in the initial library.

4.2.2 Screening of an N$_{12}$ Library

This proof-of-principle experiment suggested three things: First, the FACS-based assay can discover new synthetic riboswitches; second, the assay has the ability to find rare sequences; and third, the sequence of the ribosome binding site is critical to the performance of this family of synthetic riboswitches. With these lessons in mind, we sought to optimize the entire expression platform, including the sequence of the ribosome binding site. To do this, we created a much larger "N_{12}" library, which theoretically contains 1.7 x 10^7 members, a number that would be virtually impossible to screen enzymatically. This N_{12} library was cloned into the 5'-untranslated region of *DsRed express* and the plasmid library was used to transform *E. coli*. This transformation yielded approximately 250,000 clones, which were subsequently sorted using the flow cytometer. We note, however, that the number of clones was limited by the transformation efficiency of this particular batch of electrocompetent cells. Although the cell sorting experiments were performed using only a fraction of the theoretical library size, this does not represent a limitation of the throughput of flow cytometry, which can approach 10^9 cells/day[12]. It is also important to note that our previously described *lacZ*-based assay proved only marginally successful in screening this library for functioning switches, as nearly all of the whitest clones selected displayed no enzymatic activity in the presence or absence of theophylline, presumably due to poor ribosome binding sites.

As with the N_8 library, we sorted our N_{12} library by positioning a gate to collect cells that exhibited levels of fluorescence emission below the first decade of the log-scale in the absence of theophylline (Figure 4.2 a). The collected cells were cultured overnight in the absence of theophylline and again analyzed with the cytometer. While clones with

low fluorescence dominated the population, a portion of the population remained highly fluorescent; this required an additional sort to collect only the weakly fluorescing clones. The clones collected following the second round of sorting were cultured overnight, and aliquots of the cultures were used to inoculate two new cultures: one with theophylline (1 mM) and one without. Analysis of these cultures revealed a small percentage of clones displaying high levels of theophylline-induced fluorescence (Figure 4.2 b). These highly fluorescent cells were collected and cultured overnight. Aliquots of this overnight culture were used to inoculate one culture containing theophylline (1 mM) and one without. Subsequent flow cytometry analysis of these cultures revealed the fluorescence of the population to be highly theophylline-dependent, with a significant portion of the population displaying levels of fluorescence in the presence of theophylline far greater
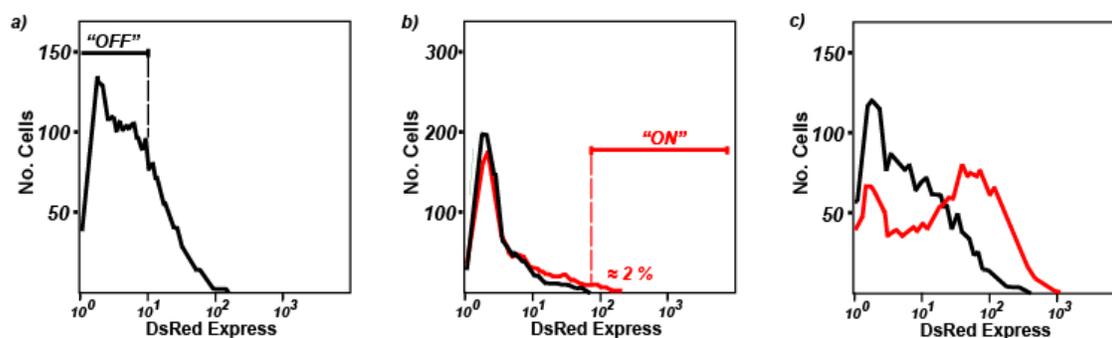
Figure 4.2 *FACS histograms representing a population of bacteria possessing a larger library of mutant theophylline riboswitches.* a.) Profile of the N_{12} library grown in the absence of theophylline. For the initial negative selection a gate was again set to collect bacteria falling in the "OFF" region. A two-step enrichment sort was required to collect only the weakly fluorescing clones. (b) Profile of the N_{12} bacteria collected in the initial negative selection grown following the two-step enrichment in the absence and presence of theophylline (1 mM). Clones displaying fluorescence levels falling within the indicated "ON" region were collected and again amounted to approximately 2% of the culture. (c) Profile of the N_{12} bacteria following one round of negative selection and one round of positive selection. Riboswitches from this enriched population were subsequently cloned in front of a β-galactosidase reporter gene and screened enzymatically.

than the same population grown in the absence (Figure 4.2 c). These highly fluorescent cells were collected and again cultured overnight.

The plasmid DNA from the saturated overnight culture was isolated, and the 5'-untranslated regions were again subcloned upstream of a *lacZ* reporter gene. This library was used to transform *E. coli*, which were grown on selective LB-agar containing X-gal, but no theophylline to give approximately 12,000 transformants. Using a robotic colony picker, we picked the 288 whitest colonies from this plate and performed our previously described enzymatic assay for riboswitch activity. Over 80 of these clones were verified as functioning theophylline-dependent synthetic riboswitches. The most impressive switch, clone 12.1, activated gene expression 96-fold in the presence of theophylline (1 mM), and produced greater than 10,000 Miller units of β-galactosidase activity, compared to 4800 Miller units for clone 8.1*, and 1820 Miller units for clone 8.1 (Figure 4. 3 a). To our knowledge, this represents the largest activation ratio of any synthetic riboswitch reported to date. Furthermore, the activation ratio of riboswitch 12.1 exceeds that of many natural riboswitches.

Figure 4.3 *Measures of reporter gene activity for switches identified using the high-throughput, FACS enrichment (12.1 and 8.1*) and best switch isolated using a previously reported β-galactosidase assay (8.1).* (a) *Right axis*: β-Galactosidase activities for switches in the absence (open circle) or presence of theophylline (1 mM, closed circle). *Left Axis*: Activation ratio (gray bar) is determined by the ratio of activities in the presence and absence of theophylline. Errors in the measurements (s.e.m.) are smaller than the diameter of the circles. (b-d) Histograms displaying the fluorescence profile for a population of *E. coli* harboring the indicated switches in the absence (black) and presence of theophylline (1 mM, red).

Histograms displaying the fluorescence profile of single *E. coli* clones using these riboswitches to control the expression of DsRed (Figures 4.3 b-d) shed additional light on the behavior of these switches and illustrate the need for caution when isolating single cells based on an observed phenotype. In the presence of 1 mM theophylline, nearly the entire populations of bacteria containing switches 8.1* and 12.1, display marked increases fluorescence when compared to those grown in the absence of theophylline. In somewhat of a surprise, a significant fraction of the population of bacteria containing switch 8.1 grown in the presence of theophylline appears to display the same fluoresce characteristics as those grown in the absence of theophylline. It is well-known that bacterial cultures comprised of a single genotype typically display a wide range of

phenotypes attributable to the stochastic nature of gene expression[18]. These results are noteworthy, as one considers that the selection criteria in flow cytometry experiments are determined by the phenotype of an individual cell, the inherent population heterogeneity of a bacterial culture, especially one containing a library of genotypes, makes the task of counter-selecting for one phenotype against another extremely challenging. For example, isolating a single bacterium displaying a low level of fluorescence does not guarantee that a culture grown from that single bacterium will possess, as a whole, those same fluorescence characteristics. With this in mind, the need for reinforcing a single-cell enrichment strategy with traditional assays, measuring average expression levels for an entire population, should be clear.

4.2.3 Exploring the Mechanisms of Improved Switches

Given the substantial improvements in how these switches functioned relative to our previously reported synthetic riboswitches, we were interested in exploring their mechanisms of action. Sequencing data suggested that these new switches likely regulated translation in the same manner as our previously reported synthetic riboswitches, and secondary-structure predictions provided by *mFold* suggested that these
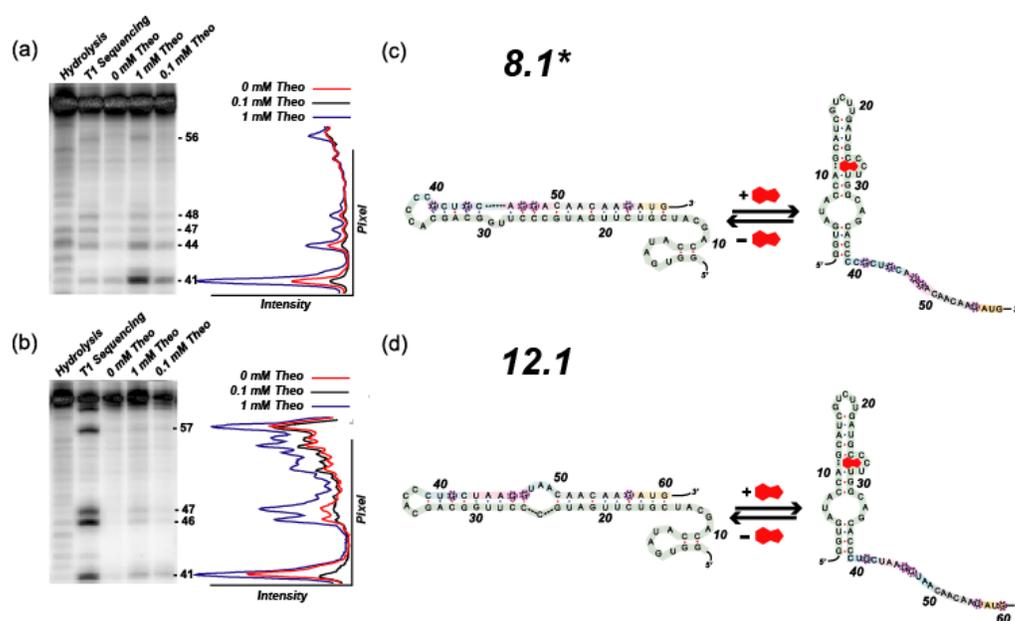
Figure 4.4 *RNaseT1 Digests and Predicted Mechanisms for Improved Switches.* Sequencing gels [7 M urea; 10% (29:1) acrylamide: bisacrylamide] indicating the cleavage pattern for 5' radiolabeled RNA from 5'-untranslated region of switches 8.1* and 12.1 digested by RNase T1 in the presence and absence of theophylline and under denaturing conditions (**a** and **b**). A graph of band intensity versus pixel location is shown to the right indicating a significant increase in cleavage at circled G residues. Cleavage patterns indicate G residues surrounding and included in the SD sequence are solvent accessible and single-stranded in the presence of theophylline and base-paired in its absence. Predicted mechanisms of riboswitch 8.1* and 12.1 function (**c** and **d**). The 5'-UTR of switch. 1 adopts a highly folded structure with extensive base-pairing in the region including the RBS. This base-pairing in the absence of theophylline prevents translation. In the presence of theophylline, the transcript adopts a secondary structure in which the RBS is exposed. The secondary structure shown for the theophylline aptamer predicted by *mFold* is identical to the secondary structure determined by NMR.

new switches adopted similar folds. We hypothesized that in the absence of theophylline, the switch adopts a conformation in which the ribosome binding site (Shine-Dalgarno sequence) is paired, and translation is inhibited. Upon binding of theophylline, the ribosome binding site becomes accessible, and translation occurs. Furthermore, because the RBS is stronger, translation is activated to a higher level than previously observed. To confirm that these newly identified switches behave as we had previously determined, we again performed *in vitro* T1 ribonuclease digestions of clones 8.1* and 12.1 both in

the absence and in the presence of theophylline (1 mM). Again, our *mFold* model[19,20] suggested that many of the G's near the RBS would become unpaired in the presence of theophylline thereby allowing for more efficient translation of the downstream gene. Results of the digestions are shown in Figures 4.4 a and b, where there are clear increases in the cleavage of four G residues near the RBS in the presence of theophylline, consistent with the models in Figure 4 c and d.

While the nuclease digestion studies suggest a switching mechanism, these experiments do not fully explain the differences in the activation ratios between the switches. Inspection of the expression platforms revealed that the switches with the highest levels of expression in the presence of theophylline all have at least 4 consecutive bases that are complementary to the anti-Shine-Dalgarno sequence of *E. coli* 16S RNA. Furthermore, our best switch, clone 12.1, not only possessed a longer Shine-Dalgarno sequence (UAAGG) than the other switches, but it also appeared to be spaced near optimally, 4 nucleotides upstream from the AUG initiation codon[21] (Figure 4.5).

To test whether the Shine-Dalgarno sequences were responsible for the different behaviors of the switches, we removed the aptamer sequences from clones 8.1, 8.1*, and 12.1 and determined the overall levels of gene expression. Because removing the aptamer drastically increases gene expression and leads to cell death due to the toxicity of β-galactosidase, we also replaced the Ptac promoter in all three constructs with a weaker *IS10* promoter to reduce the overall level of gene expression. Figure 4.6 shows that the Shine-Dalgarno sequences from clones 8.1* and 12.1 both lead to higher expression levels than the sequence from clone 8.1.

```
anti-SD   3'-------AUUCCUCCA|----------5'
   12.1   5'---CUGCUAAGGUAAC|AACAAGAUG-3'
   8.1*   5'--CCGCUGCAGGACAA|CAAGAUG---3'
   8.1    5'-CCGCUGCAAGACAAC|AAGAUG----3'
```
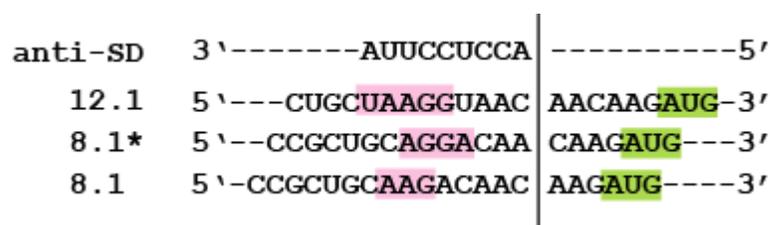
Figure 4.5 *Nucleotide sequence of putative anti-Shine-Dalgarno sequence from the 3'-end of the 16s rRNA aligned with expression platforms from high-performing theophylline riboswitches.* Expression platforms include the Shine-Dalgarno sequence (pink) and start codon (green). The putative Shine-Dalgarno sequence (UAAGG) or switch 12.1 (95-fold activation ratio) is longer than that of the other switches and is nearly optimally spaced 3'-to the AUG start codon.

These results are consistent with the predicted strengths of the Shine-Dalgarno sequences and their positions relative to the initiation codon. In this context, we define the aligned spacing as the number of bases separating the AUG start codon and the 5′-A of the anti-Shine–Dalgarno sequence[21].

When placed in the context of the thermodynamic equilibrium model for we have explaining the function of our riboswitches, the importance of the RBS in the function of a riboswitch becomes even more evident.   Our equilibrium model predicts that in the absence of the ligand, the majority of the RNA population adopts the "Off" structure in which the RBS is paired with the addition of theophylline providing the thermodynamic driving force to shift the equilibrium toward the "On" structure in which the RBS is unpaired. Because the free energies for the most stable predicted secondary structures of the 5' UTR of all of riboswitches are similar (~-20 kcal/mol) and the

experimentally-determined free energy of theophylline binding should remain constant in all switches (~9 kcal/mol)[22], it is reasonable to assume that the populations of RNA
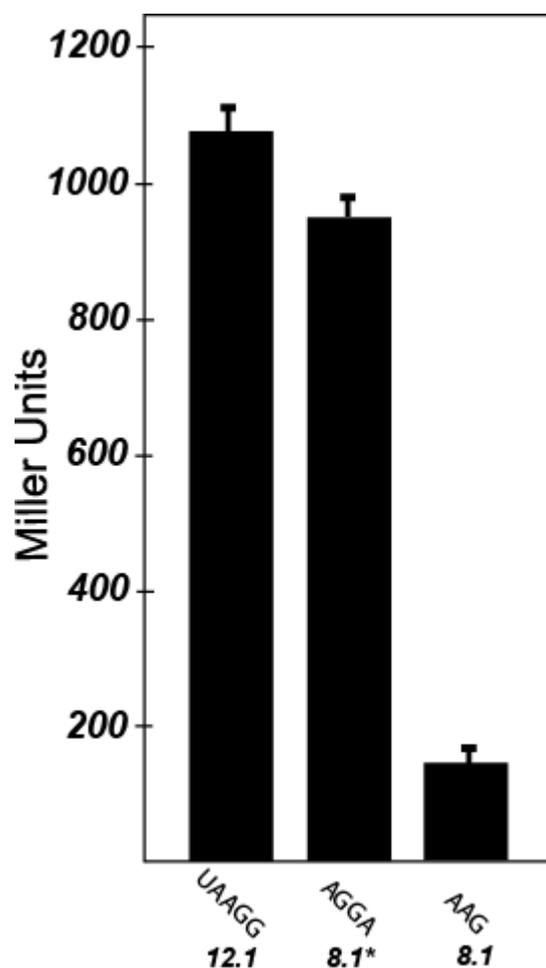
Figure 4.6 Measures of β-galactosidase activity (in Miller units) for constructs with 5'-untranslated regions lacking aptamers, containing only the expression platforms of switches 12.1, 8.1* and 8.1. Results indicate expression platforms with longer Shine-Dalgarno sequences give rise to higher expression.

adopting the "Off" structure in the absence of theophylline and the "On" structure in its presence should also be similar. With the RBS being a critical determinant of the "translatability" of transcript and population of RNA adopting the "On" structure likely remaining constant upon the addition of identical concentrations of theophylline, the most dynamic response should be seen from those switches with the strongest RBS.
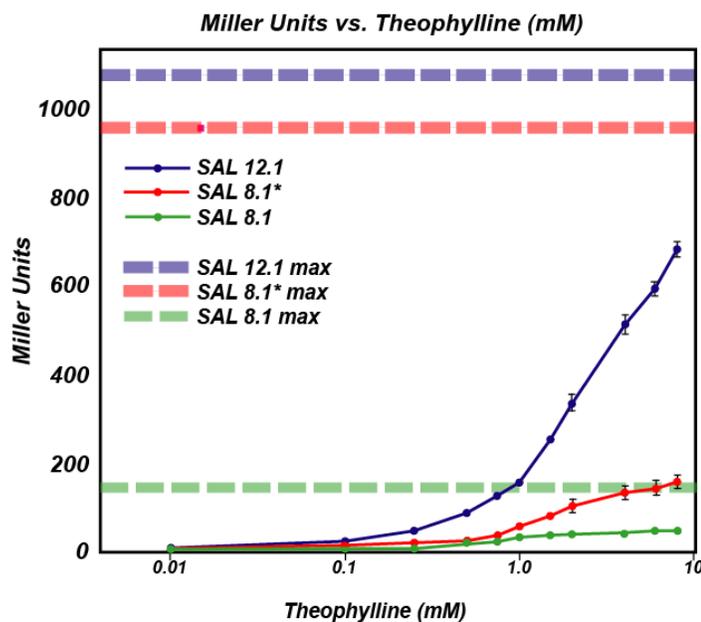
Figure 4.7 *Dose response data reported for switches 8.1, 8.1* and 12.1.* Each switch begins to respond at a theophylline concentration of ~0.01 mM, switch 12.1, with its strong, optimally-spaced RBS shows a rapid increase in expression that approaches the maximum expression level (dashed line).

The dose response data reported in Figure 4.7 demonstrates that while each switch begins to respond at a theophylline concentration of ~0.01 mM, switch 12.1, with its strong, optimally-spaced RBS shows an increase in expression that rapidly approaches the maximum expression level that can be expected from such a construct.

4.3 Discussion

Our results demonstrate that FACS is capable of isolating excellent synthetic riboswitches from large libraries. The screen is very high throughput, and can, in principle, examine libraries of greater than 10^8 members. Using the screen, we have discovered a new theophylline-dependent synthetic riboswitch, clone 12.1, that displays

low levels of background gene expression in the absence of theophylline, and a 95-fold increase in gene expression in its presence. To our knowledge, this is the largest activation ratio of any synthetic riboswitch reported to date. The key to these improvements involved optimizing the expression platform, including the ribosome binding site. In a previous study, the randomized expression platforms were limited to a sequence length of 8 bases while sequence of the putative, non-specific RBS (AGA) was held constant, thus restricting the number of accessible Shine-Dalgarno sequences. This was necessary because randomizing the RBS would lead to libraries that would be too large to screen using an enzymatic assay. While our recently reported motility-based selection method was capable of screening such large libraries ($>10^6$ members)[4], it remained challenging to discover the switches that most strongly activated gene expression. The FACS-based assay described here offers the best of both worlds, in that its throughput is competitive with a genetic selection, but it also has the ability to semi-quantitatively determine gene expression to reveal the best synthetic riboswitches.

To perform well as an "on" switch, a synthetic riboswitch must carefully balance the ability to repress gene expression in the absence of the ligand, while strongly activating gene expression in its presence. For synthetic riboswitches that regulate gene expression post-transcriptionally, this typically involves base-pairing between the expression platform (which often includes the Shine-Dalgarno sequence) and the aptamer. If these interactions are too strong (or if ligand binding is weak), the switch will not turn on; if they are too weak, the switch will be leaky in the absence of the ligand. Furthermore, the activation ratio will depend on other factors, including, but not limited to the strength of the Shine-Dalgarno sequence, and the relative stabilities of the mRNA

in the presence and absence of the ligand. Because so many of these parameters are variable, some in not entirely intuitive ways, it should come as no surprise that there are several different ways to create a synthetic riboswitch from a known aptamer. Furthermore, aptamer selection experiments typically isolate multiple aptamers that bind the target ligand. The creation of new synthetic riboswitches will likely require the screening of pools of aptamers enriched for ligand binding, along with screening randomized expression platforms, which makes the need for reliable high-throughput screening methods even greater. While it is possible to design switches *de novo*, the new riboswitches reported here perform better than any rationally designed synthetic riboswitch.

4.4 Conclusions

We have reported a screen to discover synthetic riboswitches using fluorescence activated cell sorting. The screen is very high throughput, and can, in principle, examine libraries of $>10^8$ members. Using the screen, we have discovered a new theophylline-dependent synthetic riboswitch, clone 12.1, that displays low levels of background gene expression in the absence of theophylline, and a 95-fold increase in gene expression in its presence. To our knowledge, this is the largest activation ratio of any synthetic riboswitch reported to date. While the results achieved using this flow cytometry based screen are impressive, we must emphasize that flow cytometry is unlikely to serve as a stand alone method for discovering synthetic riboswitches. The ability of a FACS-based screen employed in tandem with a β-galactosidase reporter gene to assay large libraries

($>10^8$ members theoretically) allows one to optimize the expression platform to discover outstanding synthetic riboswitches.

Using nuclease mapping, we have confirmed the likely mechanism of action of the switch, which involves a ligand-dependent rearrangement of the RNA secondary structure to reveal the ribosome binding site. We also confirmed that the strength of the ribosome binding site, which is dictated by its length and distance from the initiation codon, plays a critical role in how the best switches function. Because FACS can sort through large libraries and reveal outstanding synthetic riboswitches, we anticipate that it will be a useful screening tool in the field.

4.5 Experimental

General Considerations

All plasmid manipulations utilized standard cloning techniques and all constructs were verified by DNA sequencing. Purifications of plasmid DNA, PCR products, and enzyme digestions were performed using kits from Qiagen. Theophylline, o-nitrophenyl-β-D-galactopyranoside (ONPG) and ampicillin were purchased from Sigma. X-gal was purchased from US Biological. Synthetic oligonucleotides were purchased from IDT. All PAGE supplies were purchased from Bio-Rad Laboratories. All experiments were performed in *E. coli* TOP10 F' cells (Invitrogen) cultured in media obtained from EMD Bioscience.

Plasmid SAL006.1

(Ptac1 - mTCT4-8 aptamer - TATAAAAG – AGACAACAAG *-IS10-DsRedExpress*)

A cassette mutagenesis strategy was used to generate pSAL006. A PCR product (A) was generated by using pSKD445.1 as a template with forward primer SKD-178 which anneals to pSKD445.1 upstream of the mTCT4-8 aptamer and the reverse primer SKD-163 which anneals to the IS10 region of our reporter gene. A separate PCR product (B) was generated using forward primer SAL-002 and reverse primer SAL-003 using pDsRedExpress as a template. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD-178 and reverse primer SAL-003, to give PCR product C. PCR product C was digested with NdeI and AflIII and cloned into those sites in pSKD850.2 to yield plasmid pSAL006.1.

N_8 Library Construction – Expression Platforms with 8-base Randomized Region

(Ptac1 - mTCT4-8 aptamer - NNNNNNNN – AGACAACAAG *-IS10-DsRedExpress*)

Cassette A for the 8 base library (N=8) was created using forward primer SKD-178, which anneals to pSAL006.1 upstream of the mTCT4-8 aptamer, and reverse primer SKD-148. SKD-148 contains a randomized region of 8 bases that is flanked by constant sequences that anneal to the region including the ribosome binding site and the ATG start codon (5') and the mTCT4-8 aptamer (3'). Cassette B was created using forward primer SKD-147 and reverse primer SAL61 which anneals 3' to the NcoI site of our *IS10-DsRed-Express* reporter gene. SKD-147 also contains a randomized region of 8 bases flanked by constant sequences that anneal to the mTCT4-8 aptamer (5') and the region from the ribosome binding site and the ATG start codon (3') PCR fragments were amplified and gel purified before assembly. Constant sequences from SKD-147 and SKD-148 overlap and allowed for cassette pieces A and B to be assembled and amplified

using primers SKD-178 and SAL-061. The assembled PCR product was gel purified, digested with NdeI and NcoI and again gel purified. Digested insert was then ligated into vector pSAL006.1 digested with the same restriction enzymes, dephosphorylated and gel purified.

N_{12} Library Construction – Expression Platforms with 12-base Randomized Region

(Ptac1 - mTCT4-8 aptamer – NNNNNNNNNNNN – CAACAAG -*IS10-DsRedExpress*)

Library was created as described above with the following exceptions. For cassette 1, reverse primer SAL-045 was used and for cassette 2, forward primer SAL-044 was used. Primers SAL-044 and SAL-045 contain regions of 12 randomized bases with overlapping constant sequences that anneal to the mTCT4-8 aptamer and the region between and including ribosome binding site and start codon of the *IS10-DsRedExpress* reporter gene.

Flow Cytometry

To determine the number of transformed bacteria, library transformations were first plated on large (241 mm x 241 mm) bioassay trays from Nalgene containing 300 mL of LB/agar, supplemented with ampicillin (50 μg/mL). Plates were scraped using 2 mL of liquid media with 500 μL of the cultivated bacteria being used to inoculate a 50 mL culture of LB supplemented with ampicillin. This culture was incubated for 14 h at 37 °C with shaking (250 rpm). The following day, 50 μL of the overnight culture was used to inoculate 5 mL of LB supplemented with ampicillin and incubated at 37 °C with shaking for approximately 3 h to an OD_{600} between 0.3 and 0.5. At this time, 750 μL of the culture was centrifuged at 5,000 rpm for 10 min, the media was removed, and the cell

pellet was resuspended in 1.5 mL of PBS buffer (177 mM NaCl, 2.7 mM KCl, 5.3 mM $Na_2HPO_4 \cdot 7H_2O$, 1.8 mM KH_2PO_4, pH = 7.4). Cultures were immediately analyzed on a Becton Dickinson FACSVantage SE flow cytometer using an Innova70 spectrum laser tuned to 568 nm for excitation. Fluorescence was detected through a 630/22 bandpass filter.

For all negative selections, at least 2×10^5 clones were collected. For all positive selections, at least 2×10^4 clones were collected at greater than 90% purity. Sorted bacteria were cultured overnight at 37 °C in LB supplemented with ampicillin. Plasmids were isolated from the overnight culture of sorted bacteria and used to clone the enriched pool of riboswitches in front of an *IS10-LacZ* reporter gene. PCR reactions were set up using the isolated plasmids as a template with forward primer SKD-178 and reverse primer SKD-56. The resulting PCR product was gel purified, digested with KpnI and HindIII and again gel purified. The digested insert was then ligated into vector pSKD445.1 digested with the same restriction enzymes, dephosphorylated and gel purified. Bacteria were transformed with the resulting ligation and screened enzymatically with Miller units determined as previously described [1,23].

Structure Probing

Synthetic oligonucleotides SAL-073 and its reverse complement, SAL-074 were mixed, heated to 95 °C for 2 h and cooled 1 °C per minute to 4 °C to yield a double-stranded DNA pool with a T7 RNA polymerase promoter fused directly 5' to the mTCT4-8 aptamer and expression platform of switch SAL-12.1. Using 1 ng of the double-stranded DNA pool, a 20 µL *in vitro* transcription reaction was prepared using the AmpliScribe™

T7-Flash™ Transcription Kit from Epicentre Biotechnologies and incubated at 37 °C for 1 h. Following transcription, 1 μL of DNaseI was added to the reaction mixture to remove the DNA template. Transcribed RNA was diluted to a volume of 100 μL, phenol:chloroform extracted and ethanol precipitated. The RNA was resuspended in 44 μL of nuclease-free water and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs). Again, the reaction was diluted to 100 μL, phenol:chloroform extracted, ethanol precipitated and the RNA was resuspended in 43 μL of nuclease-free water. Resuspended RNA was 5' end-labeled using T4 polynucleotide kinase (New England Biolabs) and 0.5 μL [γ-^{32}P]ATP (7000 Ci/mmol, 150 mCi/ml, MP Biomedicals). Radiolabeled RNA was purified using denaturing gel electrophoresis. The RNA band was then excised from gel and eluted in 5 mL of nuclease-free water overnight at 37 °C with shaking. The eluted RNA was ethanol precipitated and resuspended in 50 μL of nuclease-free water.

The secondary structure of the end-labeled RNA was probed using T1 nuclease (Ambion). Structure analysis reactions were performed in the absence of theophylline, presence of 1 mM theophylline, and under denaturing conditions. To each structure analysis reaction mixture (9 μL: ~1 pmol RNA, 1x structure buffer, 6 mM MgCl$_2$, 0 or 1 mM thoephylline) and sequencing reaction (9 μL : 1 pmol RNA, 1x sequencing buffer) 1 μL of T1 RNase (1U) was added and incubated at room temperature for 15 minutes. Reactions were stopped by adding 10 μL stop buffer (8 M Urea:50 mM EDTA) and separated using denaturing gel electrophoresis (7 M urea; 10% (29:1) acrylamide: bisacrylamide), imaged using a phosphorimager and the data analyzed using ImageQuant software.

Analyses of Expression Platforms

To investigate the effects of the Shine-Dalgarno sequence on the expression levels of riboswitches, the theophylline aptamer was deleted from each construct. To accomplish this PCR products were made using forward primers SAL-085, 086, and 087 which contain KpnI restriction sites and anneal to the expression platforms of switches 8.1, 8.1* and 12.1 respectively, and reverse primer SKD 056. PCR products were digested with KpnI and HindIII, gel purified and cloned into vector pSKD1248.2 digested with the same restriction enzymes, dephospholyated and gel purified. pSKD1248.2 is nearly identical to pSKD445.1 with the following exceptions. Oligonucleotide mutatgenesis was used to insert a multiple cloning site 5' to the theophylline aptamer and replace the Ptac promoter with the weaker *IS10* promoter.

pSAL 8.1-DsRed, pSAL8.1-DsRed and pSAL 12.1-DsRed*

Riboswitches from pSAL8.1, pSAL8.1* and pSAL12.1 were PCR amplified using forward primer SKD 178 and reverse primer SKD 56. and gel purified This PCR product was the digested using KpnI and HindII and cloned into vector pSAL205mp3 cut with those same restriction enzymes. pSAL205mp3 is identical to pSAL006.1 with the following exceptions. Oligonucleotide mutatgenesis was used to remove the HindII cut site from DsRed.

4.7 References

(1)   Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. *Chem. Biol.* **2007**,

14, 173-184.

(2)   Muranaka, N.; Sharma, V.; Nomura, Y.; Yokobayashi, Y. *Nucl. Acids*

         Res. **2009**, *37*, e39.

(3)   Nomura, Y.; Yokobayashi, Y. *Biosystems* **2007**, *90*, 115-20.

(4)   Topp, S. G., J.P. *ChemBioChem* **2008**, *9*, 210-213.

(5)   Weigand, J. E.; Sanchez, M.; Gunnesch, E. B.; Zeiher, S.; Schroeder, R.;

         Suess, B. *RNA* **2008**, *14*, 89-97.

(6)   Wieland, M., Jörg S Hartig *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 2604-

         2607.

(7)   Gucker, F. T.; O'Konski, C. T.; Pickard, H. B.; Pitts, J. N. *J. Am. Chem.*

         Soc. **2002**, *69*, 2422-2431.

(8)   Harald, B. S.; Erik, B. *Cytometry* **1980**, *1*, 32-36.

(9)   Diaper J. P., T., K. and Edwards C. *Appl. Microbiol. and Biotechnol.*

         1992, *38*, 268-272

(10)   Elizabeth, M.; Catherine, R.; David, H. *Cytometry* **1986**, *7*, 347-355.

(11)   Shaner N.C, Steinbach P.A., Tsien R.Y. *Nat. Methods* **2005** *2*, 905–909

(12)   Link, A. J.; Jeong, K. J.; Georgiou, G. *Nat. Rev. Microbiol.* **2007**, *5*, 680-
         688.

(13)   Aharoni, A.; Amitai, G.; Bernath, K.; Magdassi, S.; Tawfik, D. S. *Chem.*

         Biol. **2005**, *12*, 1281-1289.

(14)   Hai, M.; Bernath, K.; Tawfik, D.; Magdassi, S. *Langmuir* **2004**, *20*, 2081-

2085.

(15) Aharoni, A.; Thieme, K.; Chiu, C. P. C.; Buchini, S.; Lairson, L. L.; Chen, H.; Strynadka, N. C. J.; Wakarchuk, W. W.; Withers, S. G. *Nat. Methods* **2006**, *3*, 609-614.

(16) Bevis, B. J.; Glick, B. S. *Nat. Biotechnol.* **2002**, *20*, 83-87.

(17) Barrick, D.; Villanueba, K.; Childs, J.; Kalil, R.; Schneider, T. D.; Lawrence, C. E.; Gold, L.; Stormo, G. D. *Nucl. Acids Res.* **1994**, *22*, 1287-1295.

(18) Elowitz, M. B.; Levine, A. J.; Siggia, E. D.; Swain, P. S. *Science* **2002**, *297*, 1183-1186.

(19) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911-940.

(20) Zuker, M. *Nucl. Acids Res* **2003**, *31*, 3406-3415.

(21) Chen, H.; Bjerknes, M.; Kumar, R.; Jay, E. *Nucl. Acids Res.* **1994**, *22*, 4953-4957.

(22) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(23) Jain, C.; Belasco, J. C. *Meth. Enzymol.* **2000**, *318*, 309-332.

Chapter 5:

Selection of RNA Aptamers that Bind Erythromycin

5.1 Introduction

Polyketides are a diverse class of natural products that exhibit a number of medicinal properties including anti-tumor[1], anti-infective and immunosuppressive[2] traits. Beginning with its discovery in 1952[3], the broad spectrum antibiotic and polyketide natural product, erythromycin A (Figure 5.1) along with its synthetic derivatives[4], have become some the most widely administered drugs in production. Originally isolated from the Gram-positive *Saccharopolyspora erythraea* (previously known as *Streptomyces erythraeu*), the production of erythromycin remains dependent on the fermentation of large volumes of *S. erythraeu*.
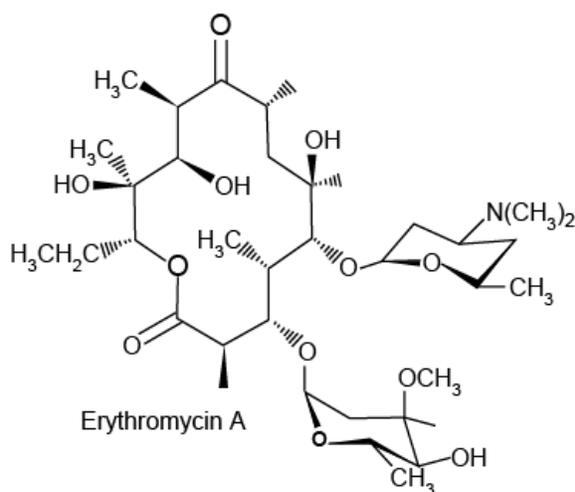
Figure 5.1 *Erythromycin A*

Chemically, erythromycin is comprised of a highly-functionalized 14 member macrolide ring that is decorated with the unusual sugars, desosamine and cladinose. The

total synthesis of erythromycin was posthumously reported in 1981 by R. B. Woodward his team of researchers[5]. With 10 of the 13 carbons of the macrolide ring being stereocenters, its synthesis from simple starting materials proved to be non-trivial. Like most antibiotics, bacteria can readily acquire resistance to erythromycin making the search for bioactive, erythromycin derivatives increasingly important.

The semisynthesis of novel erythromycin derivatives has, in the past, been the most common approach to producing novel therapeutics that outperform the original biomolecule itself with respect to potency and stability. For example, numerous methods for the regioselective, 6-OH alkylation of erythromycin A have been described[4] that give rise to a number of new antibiotics including, clarithromycin[6]. Recently, the engineering of biosynthetic pathways in charge of producing polyketide natural products has greatly simplified the process of generating novel derivatives with potential medicinal value[7-9]. Modular polyketide synthases (PKS) are multienzyme structures responsible for the microbial biosynthesis of precursors to such molecules as erythromycin A and have been the focus of a great deal of research over the past 20 years[10]. The most widely studied PKS is the 6-deoxyerythronolide B synthase (DEBS) from *S. erythraea[11]* and is responsible for the production of the 14-membered macrolide precursor to erythromycin, 6-deoxyerythronolide B (6-dEB) (Figure 5.2). Molecular biology and a combination of genetic and chemical tools have indicated DEBS to be comprised of 3 large proteins, each consisting of 2 modules with a series of catalytic domains present in each module. At the very minimum, each module contains a β-keto acylthioester synthase, acyl transferase and acyl carrier domain along with some combination of a keto reductase, dehydratase, enoyl reductase or thioesterase. The entire synthase is organized in such a

manner that it functions as a "molecular assembly line" responsible for the stepwise synthesis of the final macrolide product from 2-, 3- and 4-carbon precursors such as acetyl-CoA and propiolyl-CoA and methyl malonyl-CoA.
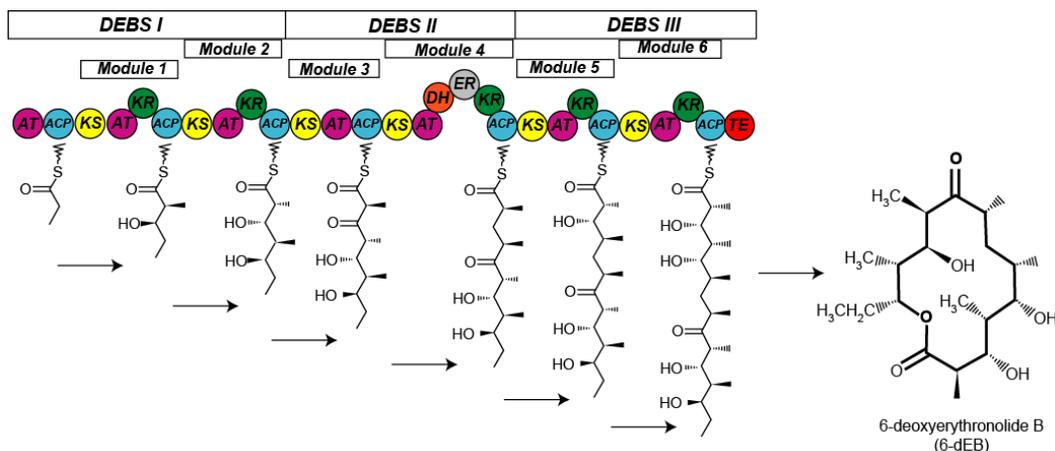
Figure 5.2 *Biosynthesis of 6-dEB from polyketide synthase, DEBS*. Domain organisation of the erythromycin polyketide synthase. Starting with malonyl-CoA, the polyketide chain is extended in a step-wise fashion using 2-carbon units. Each catalytic domain is represented by a circle. Each module includes the essential β-keto acylthioester synthase (KS), acyl transferase (AT) and acyl carrier domain (ACP) along with some combination of a keto reductase (KR), dehydratase (DH), enoyl reductase (ER). Cyclization of the final product is carried out by the thioesterase (TE). Adapted from Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380-416.

The identification of the DNA sequences encoding the various components of these synthases[11] along with the modular nature with which they are organized has allowed many to successfully manipulate and heterologously express these multienzyme structures in hosts, such as *E. coli*, producing an abundance of polyketide structures[12]. For example, the relocation of the thioesterase domain responsible for the final cyclization of the 14-membered ring to various positions along the synthase leads to the production of lactone rings of various sizes[13,14] (Figure 5.3). Also, the addition or removal of modules responsible for the reduction of the polyketide backbone, as well as

the stereochemistry at centers carrying alkyl and hydroxyl substituents can also be manipulated[15].
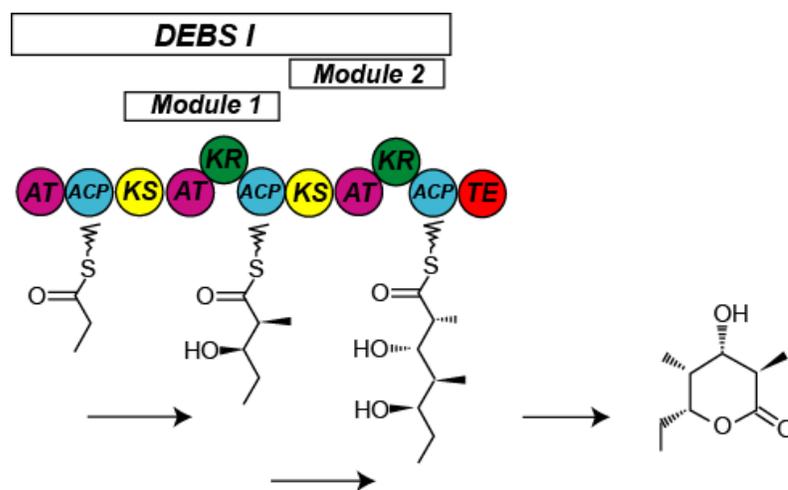
Figure 5.3 *Engineered PKS to produce novel polyketides*. The insertion of a thioesterase domain into module 2 yields 6-membered lactone ring. Adapted from Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380-416.

As with most therapeutic drugs derived from secondary metabolites, their bioactivity is contingent on varying degrees of glycosylation[16]. Compared to PKS, attempts at the reconstitution of biosynthetic pathways responsible for the glycosylation of molecules, such as erythronolide B, have been met with lesser degrees of success, with significantly fewer examples described in the literature[17]. To accomplish this, one must not only express the enzymes responsible for the biosynthesis of amino- and deoxysugars, such as desosamine and cladinose respectively, but also express the appropriate glycotransferases tasked with appending these sugars to the 14-membered ring along with the appropriate cytochrome P450 oxidative enzymes[18] (Figure 5.4). These challenges are exacerbated by the glycosyl donor and acceptor specificities of many glycotransferases making the enzymatic glycosylation of polyketide derivatives difficult.
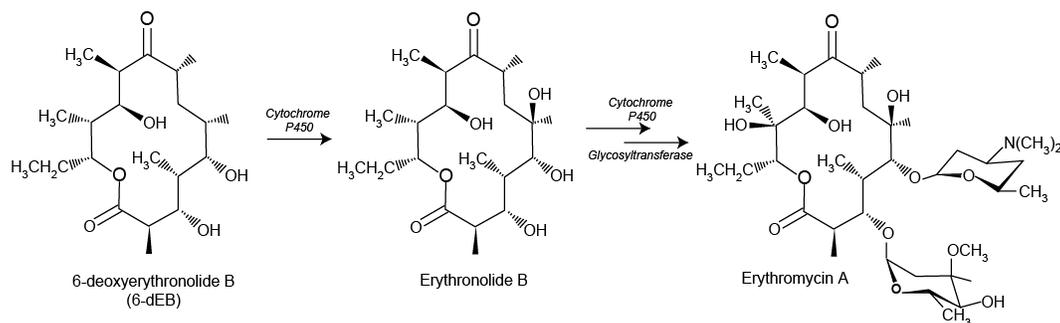
Figure 5.4 *Biosynthesis of Erythromycin A from 6dEBS.* Tailoring of the macrolide ring is carried out by cytochome P450 with glycosyltransferases append the deoxysugar cladinose and aminosugar desosamine.

Custom catalysts that can glycosylate novel erythromycin derivatives will likely require a significant amount of protein engineering. Efforts to expand the substrate specificity of biocatalysts like glycosyltransferases often employ directed evolution techniques[19]. In a typical directed evolution experiment, a library of mutant enzymes is created and the preferred enzymatic activity is identified using some kind of high throughput screen or selection. While methods for detecting enzymatic glycoslyation, such as HPLC, MS and NMR are useful, they are often very expensive and do not offer the throughput necessary for most directed evolution approaches. The throughput of these experiment can be dramatically increased by employing a genetic selection where the desired end product of an enzyme-catalyzed reaction can be linked to the survival or growth of an organism[20]. These selections are straightforward when the product is an essential molecule. Unfortunately, essential metabolites are rarely the product of engineered enzymatic reactions.

At the heart of in vivo screens and selections for evolved biocatalysts is a chemical recognition event that must eventually be translated into a change in phenotype. Given their abilities to adopt complex folds allowing for the tight, specific binding of complex small molecules, proteins are well-suited for this task of chemical recognition. Working in *E. coli*, Witholt and coworkers used a mutant transcription activator, *Nah*R, which responds to benzoate and 2-hydroxybenzoate to screen and select for benzaldehyde dehydrogenases from *Pseudomonas putida*[21]. Along these same lines, working in yeast, Schwimmer *et al.* have utilized structure-based approaches to create functional variants of the human retinoid X receptor (RXR), a ligand-activated transcription factor. While wild-type RXR receptors respond to 9-cis retinoic acid, the engineered receptors described respond to synthetic, retinoid-like compounds and can be used to screen for such compounds in vivo[22]. Unfortunately, these efforts require the additional engineering of existing protein receptors to not only bind a closely related molecule but translate that binding into a change in gene expression. To our knowledge, no such candidate protein receptors exist for erythromycin.

New methods to screen for the glycosylation of natural products, like macrolide antibiotics, would have a number of implications for the creation of new antibiotics with novel pharmacological properties. The recent reports of a high-throughput screen for detecting the glycosylation of fluorescent substrates[19] has facilitated a series of directed evolution experiments intended to evolve the substrate specificities of glycosyltransferases, which have yielded promising new enzymes. Currently, the best method for the detecting the production of fully glycosylated macrolide antibiotics in *E. coli* involves a colony-based bio-assay in which a thin layer of solid media inoculated

with *B. subtilis* is overlaid upon a recombinant strain of *E. coli* expressing the biosynthetic pathways that give rise to the bioactive macrolide 6-deoxyerythromycin D[23]. Glycosylation activity is realized by the inhibition of *B. subtilis* growth. The ability to identify mutants with enhanced abilities to produce the fully glycosylated macrolide relies on the function of macrolide export pumps proficient at exporting macrolide antibiotics bearing cladinose sugar moieties. Since it is uncertain if novel erythromycin derivatives will contain the same sugar moiety, the utility of such a screen may be limited.

Our lab has previously demonstrated that using a theophylline-binding RNA aptamer, synthetic riboswitches can be used in genetic screens and selections to detect theophylline[24-26]. The screens we have developed to identify improved synthetic riboswitches have also revealed a number of insights into their function and provide strong evidence of the ability to create new switches from small molecule-binding aptamers. With this in mind, we would like to engineer a synthetic riboswitch capable of recognizing and responding to the fully glycosylated erythromycin A molecule that could be employed to screen for enzymatic glycosylation in vivo.

To accomplish this, one must first select an RNA aptamer with a high affinity for erythromycin. Multiple methods for selecting RNA molecules which bind to target ligand have been reported, with the most common method being the Systematic Evolution of Ligands by EXponential enrichment (SELEX) (Figure 5.5)[27,28]. SELEX begins with the creation of random libraries of RNA molecules that are passed over a matrix containing a small-molecule bound to a solid support. Bound RNA is eluted with free small molecule

following a wash to remove non-binding or weakly associated RNA. The eluted RNA is then reverse transcribed, PCR amplified and then used to transcribe the RNA for the following round of selection. Using this method, numerous RNA molecules have been selected to bind a variety of small molecules with high affinity and selectivity. The well-characterized mTCT4-8 theophylline-binding aptamer[29-32] used to create our previously described riboswitches was selected using this approach and has been shown to bind theophylline tightly (K_d = 0.32 μM) and selectively with a 10,000-fold greater affinity for theophylline over the closely related caffeine molecule.

Herein we describe our efforts to select a tight-binding, erythromycin aptamer using SELEX for use in the creation of a synthetic riboswitch. The subsequent characterization of a single aptamer isolated from our in vitro selections has revealed some potentially interesting variations between the unbound and ligand-bound structure that could possibly be exploited to create a new erythromycin sensitive riboswitch.
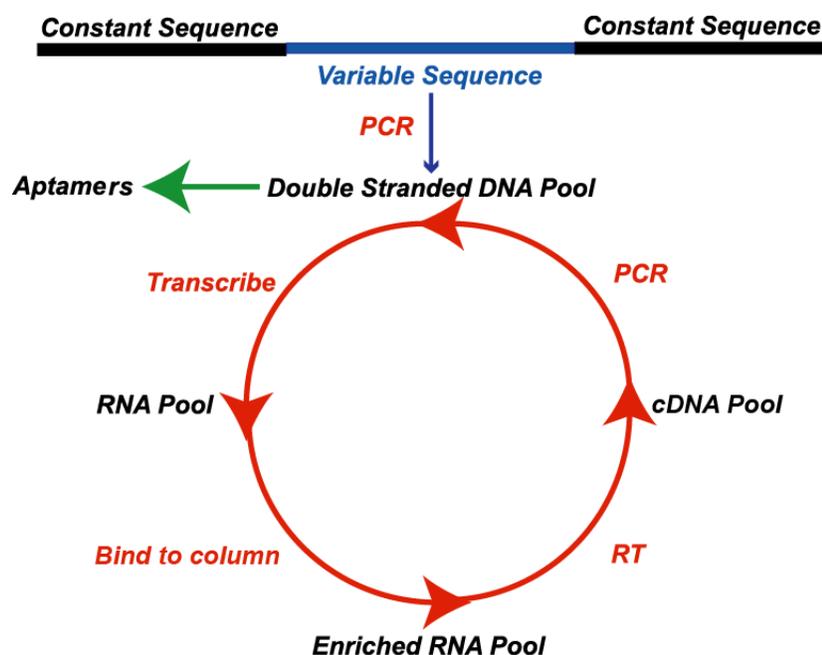
Figure 5.5 *Systematic Evolution of Ligands by EXponential enrichment (SELEX).*

5.2 Results and Discussion

5.2.1 SELEX

RNA aptamers were selected to bind the polyketide antibiotic, erythromycin, using previously established procedures. Briefly, aptamers were selected from a pool of RNA molecules harboring a randomized region of 40 nucleotides flanked by 2 constant sequences that were used for reverse transcription and PCR amplification. The DNA template used to generate the RNA pool was designed such that the 5' -forward priming region was identical to the same priming sequence from the template previously used in the selection of a codeine aptamer[33]. The priming sequence located 3' to the N40 region was identical to same region of the template reported in the selection thyroxine

aptamers[34]. This modification was necessary to eliminate the unmanageable amount of primer dimer amplification in the absence of a DNA template. The PCR amplification of the DNA pool was accomplished using a forward primer containing a T7 promoter and a reverse primer annealing to the 3' constant region of our template DNA. The resulting PCR product was purified, concentrated and used to transcribe the RNA pool for our aptamer selections.

The total complexity of this RNA pool was not determined; however, similar experiments estimate the diversity as being approximately 10^{15} molecules[35]. Aptamers were selected by passing the ^{32}P-radiolabled pool of RNA molecules over a solid support consisting of agarose beads derivitized with erythromycin. Prior to beginning the SELEX process, a negative selection step was taken to remove non-specific binding molecules. Initially, the pool of RNA was passed over an unmodified agarose column with the retained RNA being discarded and the unbound RNA being applied directly to the erythromycin-agarose affinity column prepared by Dr. Sam Reyes. Following 10 minute incubation at room temperature, the RNA remaining unbound to the erythromycin-affinity column was removed by washing with binding buffer. RNA molecules remaining on the affinity column were eluted with erythromycin (5 mM in binding buffer). This eluted RNA represented less than 0.1% of the pool following the negative selection step.

Eluted RNA was reverse transcribed to complete the first round of SELEX with the cDNA being PCR-amplified. The RNA pool for the second round was transcribed from this double-stranded DNA pool and subsequently loaded onto an erythromycin-

affinity column, washed with buffer and again, RNA bound to the column was eluted with free erythromycin and reverse-transcribed. After 3 rounds of selection and amplification, a significant amount (~40%) of the RNA was selectively eluted from the column therefore; for the 4th round, we increased the wash volume to 50 column volumes in hopes of removing weakly binding RNA molecules. A steady increase in the amount of RNA selectively eluted was observed until round 8 where nearly half of the RNA applied to the column could not be eluted with free erythromycin. At this point the pool was cloned and 10 colonies from both the 7th and 8th rounds were sequenced. Figure 5.6 displays a round-by-round account of the total RNA washed and eluted from the column as well as the RNA remaining bound to the column.
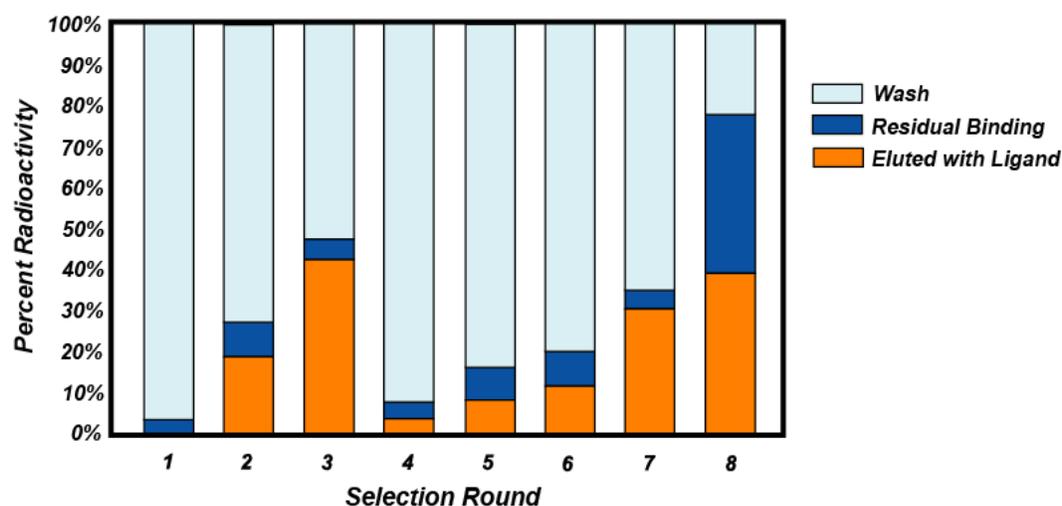
Figure 5.6 *Round-by-round summary of the selection of erythromycin-binding aptamers.* For each round, the percentage of RNA present in the wash (light blue), elution (orange) and remaining bound to the column (dark blue) was determined by scintillation counting. The decrease in eluted RNA for round 4 was due to the increased volume of wash buffer applied prior to eluting RNA.

Representative sequences of aptamers selected from rounds 7 and 8 are displayed in Figure 5.7. Visual inspection of these sequences revealed that the RNA pool appeared to have converged upon 2 separate sequences. Given the diversity of the starting pool of

RNA, it was deemed unlikely that sequences, such as 4617 and 4624 that differ by only a single base could be isolated randomly. Representative clones 4617 and 4625 were selected for further analysis, including isocratic elution and structure probing, to determine dissociation constants and the nucleotides involved with ligand binding.

```
4617   --TGACGAGCAGGCGAA-CCCGGTGTTTCTTGGACGCGTGCGA--
4624   --TGACGAGCAGGCGAA-CCCGGTGTTTCTTGGACGCGTGCGC--
4626   --TGACGAGCAAGTGAAAACCCG---TTCCTGGAGCGCAATGTGT

4625   CTTGACACAGTGTCGATGCG-CTTATCACGGGTGCCGTCCA---
4627   CTTGACACAGTGTCGATGCG-CTTATCACGGGTGCCGTCTT---
```

Figure 5.7 *Representative sequences isolated from rounds 7 (4617) and 8 (4624-4627) of SELEX.* Sequence converges on 2 separate motifs. 4617 and 4625 were chosen for further structural analyses.

5.2.2 Analysis of Putative Aptamers

Isocratic elution[35,36] was used to determine a K_d value for aptamer 4617 (Figure 5.8). Isocratic elution, sometimes referred to as competitive affinity chromatography, determines a dissociation constant by comparing the elution volume required to wash a fixed amount of RNA off an affinity column using buffer that contains no ligand to the elution volume required using a buffer with a known ligand concentration ([L]). To begin, 300 pmol of radiolabled RNA (50 µL; V_n) transcribed from 4617 was loaded onto an erythromycin-affinity column and allowed to equilibrate for 30 min. The RNA was then washed off the column using binding buffer containing erythromycin (5 mM). Fractions (500 µL) were collected until no radiation remained on the column and the total elution volume determined (~6.5 mL; V_{el}). A second 300 pmol aliquot of RNA was

loaded onto the same column and, again, allowed to equilibrate for 30 min. RNA was then washed off the column in 500 μL fractions using only binding
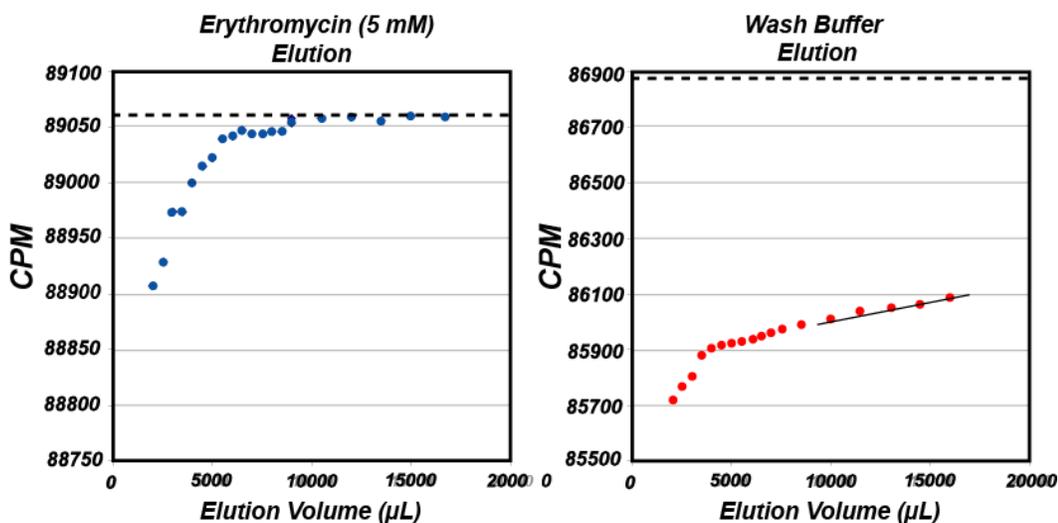
Figure 5.8 *K_d Determination for aptamer 4617 using isocratic elution.* Data points are measures of radiation contained in each fraction collected. Dashed line represents the total amount of RNA loaded onto the column determined by scintillation counting. For wash buffer elution, the total elution volume was determined by a linear extrapolation of the final 5 data points.

buffer. After 18 mL, a small but detectable amount of RNA remained on the column. At this point, the elution was stopped and the total elution volume (~75.4 mM; V_e) was determined from a linear extrapolation of the data points. This is, admittedly, an imperfect experiment; however, the value obtained should be a conservative estimate of the elution volume. Using the volumes obtained and the following equation (eq.1), a K_d of ~430 μM was calculated.

$$\text{eq. 1} \quad K_d = [L]\,\frac{(V_{el}-V_n)}{(V_e-V_{el})}$$

This calculated dissociation constant is slightly outside the range of expected K_d values for aptamers obtained from SELEX experiments[29-32] (low μM to nM) indicating

very weak or non-specific binding. Furthermore, we remain suspicious as to the validity of this experiment as greater than 99% of the RNA was eluted from the column with 5 mL of buffer regardless of the erythromycin concentration. These results are atypical of isocratic elution experiments and are cause for concern. Additionally, in-line probing and nuclease structure probing studies of the same aptamer did not reveal any structural modulation upon the addition of erythromycin. Taken together, we chose to set this putative aptamer aside and shift our focus to aptamer 4625.

With only a limited quantity of our erythromycin-affinity column remaining, we were unable to repeat any isocratic elution experiments using aptamer 4625. In-line probing[37] experiments have previously been used to determine dissociation constants and offer the added benefit of illuminating some of the residues that may be involved with ligand binding. In-line probing is a technique that exploits the natural instability of RNA. In mildly alkaline conditions, the phosphorus center of RNA's phosphodiester backbone is prone to nucleophilic attack by the 2' OH of the ribose sugar. Regions of an RNA molecule that are less structured have a greater likelihood of adopting the in-line configuration necessary for this cleavage reaction to take place. Changes in these cleavage patterns between RNA aptamers incubated in the absence of a ligand and those incubated in its presence can be used to determine ligand binding information.

RNA radiolabled at the 5' end (>10,000 cpm) was incubated in an in-line probing buffer (final concentrations: 50 mM Tris-HCl (pH 8.3 @ 25 °C), 20 mM $MgCl_2$, 100 mM KCl) with increasing concentrations of erythromycin for 40 h at room temperature and the cleavage products separated by gel electrophoresis (Figure 5.9). Small, but detectable
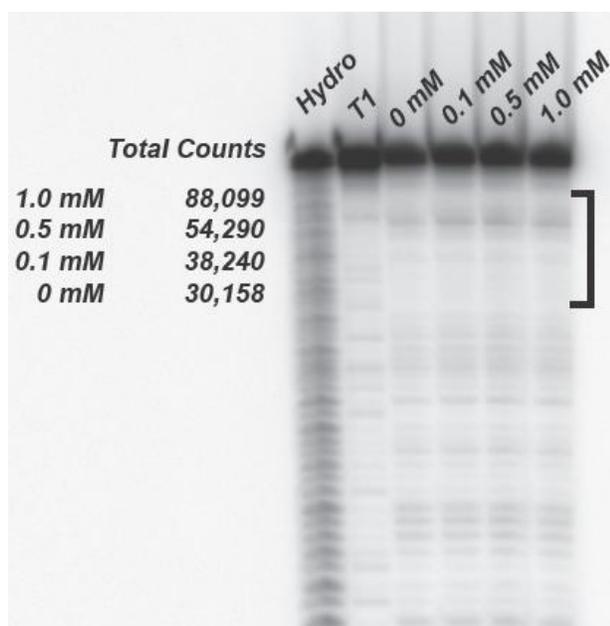
Figure 5.9 *In-line probing of an erythromycin aptamer.* Total counts of radiation within the bracketed region corresponding to the 3' region of our aptamer for each lane were determined by ImageQuant software and are shown on the left. Small, but detectable increases in cleavage were measured indicating this region becomes less structured upon ligand binding.

differences in the cleavage patterns could be seen in the 3' region of our transcript corresponding to the constant region of our RNA pool. In increasing concentrations of erythromycin, an increase of the amount of cleavage was observed indicating this area of the RNA molecule becomes less structured upon ligand binding. Unfortunately, TLC experiments revealed that even in mildly alkaline solution, erythromycin breaks down within a few hours, presumably through the hydrolysis of the lactone present in the macrolide ring, making any calculations of a K_d from this experiment very difficult.

With the in-line probing results, we speculated that the 3' region of our ligand-bound aptamer is likely single stranded and could therefore be investigated using V1

nuclease. V1 nuclease cleaves only double stranded RNA molecules and only involves a 5 min incubation period compared to the 40 h incubation required for in-line probing making the erythromycin degradation problems associated with in-line probing irrelevant. 5' end-labeled RNA (>10,000 cpm) was incubated in a manufacturer-provided buffer with increasing concentrations of erythromycin, digested with V1 nuclease and the digest products separated with gel electrophoresis. As expected, the cleavage of the 3' region of our aptamer (residues 54-76) decreased in the presence of erythromycin indicating a double-stranded to single-stranded transition upon ligand-binding. The decrease in band intensity seen in the 5' region (residues 13-29) of the transcript may be attributable to ligand binding. The total counts measured for each of the 0, 0.1 and 0.5 mM lanes are very similar indicating the differences seen are due to changes in RNA structure rather than experimental error. The low intensity seen in the 1 mM lane was likely caused by RNA loading error or by an insufficient amount of RNA being used in the structure probing reaction as the total counts determined for this lane was far less that the 0, 0.1 and 0.5 mM lanes.
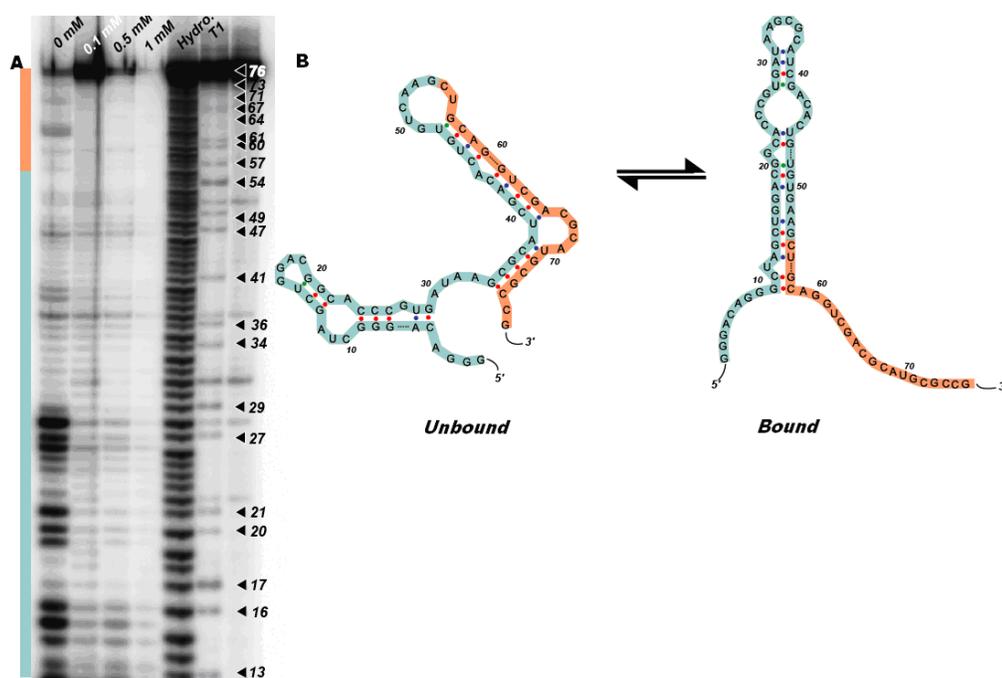
Figure 5.10 *Structure probing analysis of putative erythromycin aptamer.* Cleavage of the 3' region of the transcript (residues 54-76) decreased in the presence of erythromycin indicating a double-stranded to single-stranded transition upon ligand-binding. The decrease in band intensity seen in residues 13-29 of the transcript may be attributable to ligand binding. The low intensity seen in the 1 mM lane was likely caused by RNA loading error. The bound structure is the minimal energy structure (25.4 kcal/mol) determined by mFold. The unbound structure was determined by mFold also; however, the folding protocol used prohibited the 3' residues from pairing (13.9 kcal/mol).

These results are potentially very interesting when viewed in the context of how our theophylline riboswitches function[25]. RNA transcripts containing our theophylline riboswitch in the absence of theophylline show significant base-pairing in the region immediately 3' to the theophylline aptamer that contains the ribosome binding site. Theophylline binding then drives the base-paired structure surrounding the RBS to become unpaired which ultimately leads to translation of the downstream gene. If a similar structural rearrangement is seen with our aptamer, our approach to turning it into an erythromycin-sensitive riboswitch should be relatively straightforward.

5.3 Conclusion

We have successfully performed 8 rounds of SELEX to isolate an RNA aptamer with high affinity for erythromycin. Sequence analysis of aptamers isolated from rounds 7 and 8 revealed the pool was converging on 2 separate sequences, 4617 and 4625. Isocratic elution experiments were conducted on 4617 and the results indicated the aptamer likely has a very weak affinity for erythromycin. In-line probing and V1 nuclease mapping of 4625 revealed an apparent structural change in the 3' end of the aptamer upon ligand binding at ligand concentrations greater that 0.1 mM. In-line probing results showed small, but detectable structural changes in the same region of the aptamer in the presence of the ligand; however, the alkaline reaction conditions cause the breakdown of erythromycin making the results unusable for K_d determination. It is entirely possible that the breakdown of erythromycin on the affinity column used for our SELEX experiments was the cause of high amounts of RNA remaining bound to the column in round 8 of our SELEX experiments. With the column stored in buffer with a pH of 7.4, over time, the lactone of the macrolide rig could have hydrolyzed causing us to inadvertently select for aptamers that bind to the hydrolyzed product. Elution steps using intact erythromycin may have eluted only those RNA molecules binding to the cyclized target leaving behind those bound to the linear polyketide.

5.4 Experimental

Initial DNA Pool Construction. Template DNA

(5'-GGGACAGGGCTAGC-N40-CTGCAGGTCGACGCATGCGCCG-3')

contained a 40 nt randomized region. This template was used to generate initial DNA

pool by PCR amplification using forward primer SKD 493 and reverse primer SKD 535.

The PCR amplification was conducted for 30 cycles using 100 pmol of template, 100

pmol each of forward and reverse primer. The PCR product was used for the transcription

reaction after column purification using a MinElute PCR purification kit (Qiagen) and

concentrated by eluting with 10 μL of water.

In Vitro Transcription (internal radiolabeling of RNA). The DNA pool generated from

PCR amplification was transcribed into an initial RNA using the AmpliScribe™ T7-

Flash™ Transcription Kit (Epicentre) as per the manufacturer's instructions.

Radiolabeling was accomplished through the incorporation of 70 μCi [α-^{32}P]-UTP (3000

Ci/mmol, MP Biosciences). After 1 h incubation at 37 °C, the template DNA was

degraded by adding 20 U of DNase I (Epicentre) and incubating at 37 °C for 30 min. The

transcription reaction was mixed with an equal volume of gel loading buffer (95%

formamide, 20 mM EDTA, bromophenol blue). The RNA was purified on a 10%

denaturing polyacrylamide gel, products were visualized by UV-shadowing and the

bands corresponding to the full length RNAs were excised. Gel slices were crushed and

soaked in water and RNAs were passively eluted from the gel slices overnight at 4 °C.

The resulting solution was passed through a 0.2 μm filter (VWR) to remove polyacrylamide gel particles. The RNA was further purified by ethanol precipitation by adding 3 volumes of 95% ethanol, 0.1 volumes of 3 M Sodium Acetate (pH 5.2) and 0.5 μL of glycogen (20 mg/mL). RNA was precipitated by incubating at -80 ℃ for 30 min and then by centrifuging at 18,000 rcf for 30 min. Precipitated RNA was dissolved in binding buffer (20 mM Tris-HCl, 5 mM $MgCl_2$, 250 mM NaCl; pH 7.4) and the quantity of RNA was determined by measuring the absorbance at 260 nm.

Column-based Selection. Affinity column was prepared by Dr. Sam Reyes. An erythromycin A-(*E*)-9-oxime derivative[38-40] bearing a short, inert alkyne handle and "clickable" agarose beads[41] bearing a terminal azide were coupled through Cu(I)-catalyzed cycloaddition to give rise to an erythromycin-affinity column. The radiolabeled RNA pool (300 pmol) in binding buffer was denatured by incubating at 70 ℃ for 2 min and then allowed to refold by slow cooling to room temperature. To remove RNA transcripts that bind agarose, the pool was then loaded on a column containing 250 μL of CarboxyLink coupling resin without erythromycin and incubated for 30 min at room temperature. The column was rinsed with 2 mL of binding buffer, and the unbound RNAs were then loaded on a second column containing a 250 μL of erythromycin-agarose. The column was incubated at room temperature for 30 min and then washed with 2.0 mL of binding buffer. Bound transcripts were then eluted with 1 mL of 5 mM of erythromycin solution.

At all steps, the total amount of RNA was accounted for. All aliquots of the solutions from the wash steps were recovered after loading onto the erythromycin-affinity column along with the erythromycin-affinity column itself and the radioactivity measured using a scintillation counter. A small aliquot of the eluted RNA was used to measure radioactivity and used to account for the final portion of RNA.

Reverse Transcription. The RNA eluted from the column was recovered by ethanol precipitation and dissolved in 20 μL of water. An aliquot (16 μL) was reverse transcribed using Superscript III reverse transcriptase (Invitrogen) as per the manufacturer's instructions. A 10 μL aliquot of the reverse transcription reaction was used as a template in a PCR reaction to generate the DNA pool for the subsequent selection cycles. The PCR amplification was carried out in a 100 μL volume containing 50 pmol each of forward and reverse primer (PCR conditions: initial denaturation of 94 °C for 2 min; 94 °C for 30 sec, 55 °C for 30 sec, extension of 72 °C for 2 min; 20 cycles). Transcription and purification procedures for subsequent cycles were as described for the generation of initial RNA pool. For subsequent selection cycles, RNA pool was refolded by heating at 70 °C for 2 min, cooled down to room temperature. RNA pool was then directly loaded on an erythromycin-affinity column containing 250 μL of agarose-erythromycin. The selection conditions were as described above for rounds 1-3. For round 4-8 the wash volume was increased to 50 column volumes of binding buffer.

In Vitro Transcription (5' end-labeling of RNA) for Structure Probing. DNA templates for aptamers 4617 and 4625 were generated by PCR amplification using forward primer SKD 493 and reverse primer SKD 535 with templates SAL 468 mp10 and 469 mp8 respectively. The PCR product was used for the transcription reaction after column purification using a MinElute PCR purification kit (Qiagen) and concentrated by eluting with 10 μL of water. Using 1 μg of the double-stranded DNA pool, a 60 μL in vitro transcription reaction was prepared using the AmpliScribe™ T7-Flash™ Transcription Kit from Epicentre Biotechnologies and incubated at 37 °C for 1 h. Following transcription, 1 μL of DNaseI was added to the reaction mixture to remove the DNA template. Transcribed RNA was purified using denaturing gel electrophoresis with RNA being visualized by UV shadowing. The RNA was excised from the gel and eluted overnight in 400 μL of elution buffer at 4ºC. Eluted RNA was precipitated with ethanol as before, resuspended in 50 μL of nuclease-free water and its concentration determined using absorbance at 260 nm. Approximately 1 μg of RNA was dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) for 30 min at 37 ºC. The reaction was diluted to 100 μL, phenol:chloroform extracted, ethanol precipitated and the RNA was resuspended in 13 μL of nuclease-free water. Resuspended RNA was 5' end-labeled using T4 polynucleotide kinase (New England Biolabs) and 4 μL [γ-^{32}P]ATP (7000 Ci/mmol, 150 mCi/ml, MP Biomedicals). Radiolabeled RNA was again purified using denaturing gel electrophoresis. The RNA band was then excised from gel and eluted in 400 μL of nuclease-free water overnight at 4 °C. The eluted RNA was then ethanol precipitated and resuspended in 50 μL of nuclease-free water.

The secondary structure of the end-labeled RNA was probed using V1 nuclease (Ambion). Structure analysis reactions were performed in the absence of erythromycin and in the presence of increasing concentrations of erythromycin. To each structure analysis reaction mixture (9 μL: ~1 pmol RNA (>10,000 cpm), 1x structure buffer, 5 mM $MgCl_2$ and 1 μL of the appropriate concentration of erythromycin) 1 μL of V1 RNase (0.1U) was added and incubated at room temperature for 5 minutes. For T1 sequencing reactions, the same amount of RNA was added to a 10 μL reaction mixture containing final concentrations of 50 mM sodium citrate (pH 5), 10 M urea and 1.5 mM EDTA to which 1 μL of T1 nuclease (1U/μL) was added and incubated at 55 °C for 5 min. For hydrolysis sequencing reactions, ~1 pmol RNA (>10,000 cpm) was incubated at 90 °C for 5 min in 1x hydrolysis buffer (50 mM Sodium Carbonate; pH 9.0). All reactions were stopped by adding 10 μL stop buffer (8 M Urea:50 mM EDTA) and separated using denaturing gel electrophoresis (7 M urea; 12% (29:1) acrylamide: bisacrylamide), imaged using a phosphorimager and the data analyzed using ImageQuant software.

5.5 References

(1)   Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B.
       Science **2000**, *287*, 640-642.

(2)   Schwecke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.;
       Haydock, S. F.; Oliynyk, M.; Caffrey, P.; CortÃ©s, J.; Lester, J. B. *Proc.
       Nat. Acad. Sci. U. S. A.***1995**, *92*, 7839-7843.

(3)   McGuire, J. M.; Bunch, R. L.; Aanderson, R. C.; Boaz, H. E.; Flynn, E.
       H.; E. H. Powell; Smith, J. W. *Antibiot. Chemother.* **1952**, *2*, 281-283.

(4)   Ma, Z.; Clark, R. F.; Brazzale, A.; Wang, S.; Rupp, M. J.; Li, L.;
       Griesgraber, G.; Zhang, S.; Yong, H.; Phan, L. T.; Nemoto, P. A.; Chu, D.
       T. W.; Plattner, J. J.; Zhang, X.; Zhong, P.; Cao, Z.; Nilius, A. M.;
       Shortridge, V. D.; Flamm, R.; Mitten, M.; Meulbroek, J.; Ewing, P.;
       Alder, J.; Or, Y. S. *J. Med. Chem.* **2001**, *44*, 4137-4156.

(5)   Woodward, R. B.; Logusch, E.; Nambiar, K. P.; Sakan, K.; Ward, D. E.;
       Au-Yeung, B. W.; Balaram, P.; Browne, L. J.; Card, P. J.; Chen, C. H. *J.
       Am. Chem. Soc.* **2002**, *103*, 3215-3217.

(6)   Brunet, E.; Muñoz, D. M.; Parra, F.; Mantecón, S.; Juanes, O.; Rodríguez-
       Ubis, J. C.; Carmen Cruzado, M.; Asensio, R. *Tet. Lett.* **2007**, *48*, 1321-
       1324.

(7)   Khosla, C. *Chem. Rev.* **1997**, *97*, 2577-2590.

(8)   McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. *Proc. Nat. Acad. Sci. U. S. A.* **1999**, *96*, 1846-1851.

(9)   Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380-416.

(10)   Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468-3496.

(11)   Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *252*, 675-679.

(12)   Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. *Science* **2001**, *291*, 1790-1792.

(13)   Gokhale, R. S.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Science* **1999**, *284*, 482-485.

(14)   Kao, C. M.; Luo, G.; Katz, L.; Cane, D. E.; Khosla, C. *J. Amer. Chem. Soc.* **2002**, *117*, 9105-9106.

(15)   Böhm, I.; Holzbaur, I. E.; Hanefeld, U.; Cortési, J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **1998**, *5*, 407-412.

(16)   Weymouth-Wilson, A. C. *Nat. Prod. Rep* **1997**, *14*, 99 - 110.

(17)   Salvador, P.; Eduardo, R.; Hugo, G. M.; John, R. C.; Hugo, G. *Microb. Biotechnol.* **2008**, *1*, 476-486.

(18)   Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. *Nat. Prod. Rep.* **2002**, *19*, 542-580.

(19)   Williams, G. J.; Zhang, C.; Thorson, J. S. *Nat. Chem. Biol.* **2007**, *3*, 657-662.

(20)   Parikh, M. R.; Greene, D. N.; Woods, K. K.; Matsumura, I. *Protein Eng. Des. Sel.* **2006**, *19*, 113-119.

(21)   van Sint Fiet, S.; van Beilen, J. B.; Witholt, B. *Proc. Nat. Acad. Sci. U.S.A.* **2006**, *103*, 1693-1698.

(22)   Schwimmer, L. J.; Rohatgi, P.; Azizi, B.; Seley, K. L.; Doyle, D. F. *Proc. Nat. Acad. Sci. U.S.A.* **2004**, *101*, 14707-14712.

(23)   Lee, H. Y.; Khosla, C. *PLoS Biol.* **2007**, *5*, e45.

(24)   Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247-13254.

(25)   Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. *Chem. Biol.* **2007**, *14*, 173-184.

(26)   Topp, S. G., J.P. *ChemBioChem* **2008**, *9*, 210-213.

(27)   Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818-822.

(28)   Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.

(29)   Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(30)   Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J. P.; Pardi, A. *Nat. Struct. Biol.* **1997**, *4*, 644-9.

(31)   Zimmermann, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186-92.

(32)   Zimmermann, G. R.; Wick, C. L.; Shields, T. P.; Jenison, R. D.; Pardi, A.
         RNA **2000**, *6*, 659-667.

(33)   Win, M. N.; Klein, J. S.; Smolke, C. D. *Nucl. Acids Res.* **2006**, gkl718.

(34)   Levesque, D.; Beaudoin, J.-D.; Roy, S.; Perreault, J.-P. *Biochem. J.* **2007**,
         403, 129-138.

(35)   Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33*, 973-82.

(36)   Arnold, F. H.; Blanch, H. W. *J. Chrom. A* **1986**, *355*, 13-27.

(37)   Regulski, E. E.; Breaker, R. R. In *Post-Transcriptional Gene Regulation*
         2008, p 53-67.

(38)   Proctor, A. J.; Beautement, K.; Clough, J. M.; Knight, D. W.; Li, Y. *Tet.
         Lett.* **2006**, *47*, 5151-5154.

(39)   Xu, R.; Sim, M.-K.; Go, M.-L. *J. Med. Chem.* **1998**, *41*, 3220-3231.

(40)   Pandey, D.; Katti, S. B.; Haq, W.; Tripathi, C. K. M. *Bioorg. Med. Chem.*
         2004, *12*, 3807-3813.

(41)   Punna, S.; Kaltgrad, E.; Finn, M. G. *Bioconjugate Chem.* **2005**, *16*, 1536-
         1541.

Chapter 6:

Using a Dual-Selection Strategy to Identify

Erythromycin Riboswitches

6.1 Introduction

In chapters 2 and 4, we have outlined 2 powerful screening methods that can be used to identify synthetic riboswitches with improved characteristics. Furthermore, the theophylline sensitive riboswitches we have isolated using these screens has elucidated many of the principles underlying their mechanism of function[1]. Despite these results, we remain unable to rationally design a riboswitch from an existing RNA aptamer. These efforts are further complicated when one considers the lack of evidence suggesting that the tightest binding, in vitro-selected aptamers with high specificity for a target ligand are suitable for use in creating a synthetic riboswitch.

In the previous chapter we described a series of SELEX experiments that have resulted in the selection of a pool of aptamers enriched for binding erythromycin. With the features of in vitro selected aptamers necessary for creating synthetic riboswitches yet to be discerned, it appears likely that the best approach to creating new synthetic riboswitches will likely require the combination of in vitro selections for candidate aptamers with in vivo screens or selections for function. This approach was best demonstrated by Wiegand et al[2]. Working in yeast, the authors cloned a pool of neomycin-binding aptamers into the 5' UTR of a GFP reporter gene. Clones exhibiting

fluorescence when grown on minimal media were isolated and subsequently grown in the absence and presence of neomycin identifying functioning riboswitches being identified by a comparative loss of fluorescence in the presence of neomycin. While this approach was effective, our previous efforts to design theophylline riboswitches lead us to believe that, working in *E. coli*, a library of expression platforms would need to be screened in concert with an enriched pool of aptamers. To identify riboswitches from the significantly larger library generated in this type of experiment, it will be beneficial to use a genetic selection that provides a larger throughput.

We have previously described a FACS-based strategy for screening a library of riboswitches[3] that has a throughput which approaches that of a genetic selection; however, the high cost of such experiments and limited access to a flow cytometer has limited their usefulness. Our lab has also developed a motility-based selection using *cheZ* as a reporter gene[4]. The CheZ protein is one of many proteins in *E. coli* that is required for cell chemotaxis and motility. Bacteria not expressing CheZ "tumble" in place and therefore do not move. When expressed at an appropriate level, cells begin to travel on semisolid media. By modulating the expression of *cheZ* with a library of riboswitches, one can readily identify functioning switches by selecting for mutants that do not move in the absence of the ligand, and subsequently selecting for mutants that move in the presence of the ligand. While very powerful given the extremely high throughput it offers, this motility-based selection can only be used to identify switches responding to ligands that do not interfere with natural bacterial chemotaxis as is the case with erythromycin (unpublished observations).

Selections for riboswitches are complicated in that they require the counter-selection of an "on" phenotype against an "off" phenotype. While the selection of "on" phenotypes can easily be accomplished using an antibiotic resistance gene, the detection of "off" phenotypes is far more challenging. For example, the lethal barnase gene can be used to identify an "off" phenotype as the expression of it will kill the cell; however, in the context of selecting for riboswitches, this would require multiple cloning steps. Furthermore, a single point mutation in this gene can render it harmless, circumventing the entire negative selection step.

A potent counter-selection scheme for the identification of engineered riboswitches was recently reported utilizing tetA[5,6], a tetracycline resistance gene that encodes for a tetracycline export pump, which exports the antibiotic at the expense of proton uptake. TetA is a unique reporter gene in that it offers two distinct phenotypes and can act as both a positive ("on" phenotype) and negative ("off phenotype") selectable marker. When expressed, TetA confers resistance to the antibiotic tetracycline. Additionally, cells expressing TetA are also sensitive to a variety of compounds including toxic metal salts, such as $NiCl_2$, and fusaric acid, a lipophilic chelating agent[7-9]. Using this single selectable marker with multiple distinct phenotypes that are inextricably linked not only simplifies the selection procedure, but also decreases the likelihood of false positives being selected.

Using this dual selection method, Yokobayashi and coworkers have transformed a natural thiamine pyrophosphate (TPP)-responsive riboswitch that represses gene expression in the presence of TPP to one that activates gene expression in the presence of

the same ligand[10]. Additionally, this method has been used to create a dual-input riboswitch that only activates gene expression in the presence of *both* theophylline and TPP but not either ligand individually[11].

In this chapter, we explore the possibility of using this *tetA* counter-selection scheme to identify erythromycin-sensitive riboswitches constructed from a pool of in vitro selected aptamers with an affinity for erythromycin.

6.2 Results and Discussion

6.2.1 Selection to Identify Erythromycin Riboswitches

The dual selection strategy is outlined in Figure 1. To begin, a pool of candidate aptamers from round 7 of our previous SELEX experiments (see Chapter 5) was cloned into the 5' UTR of our *tetA* reporter. Separating the aptamers and the start codon of our *tetA* reporter was a sequence of 12 randomized bases followed by the CAACAAG sequence from our previously reported theophylline riboswitches. As seen in Chapter 4, the RBS can be an important determinant in riboswitch function. The increased throughput offered by this selection strategy affords us the luxury of not specifying a RBS and, instead, allowing the selection to identify RBS sequences which offer the optimal levels of gene expression required for survival.

This library of riboswitches was used to transform the electrocompetent, minimal-genome strain of *E. coli*, MDS42. A library of approximately 100,000 clones was harvested and the plasmids were isolated. The library of plasmids was then used to transform an erythromycin resistant strain of *E. coli*, N281[12]. In wild-type *E. coli*,

erythromycin binds to the 50S subunit of the ribosome and prevents ribosomal assembly. Strain N281 contains a mutated ribosomal protein, L22, which prohibits binding of erythromycin and therefore enables ribosome assembly[13]. Regrettably, N281 does not transform well, and a simple ligation reaction cannot be used to efficiently transform this strain.
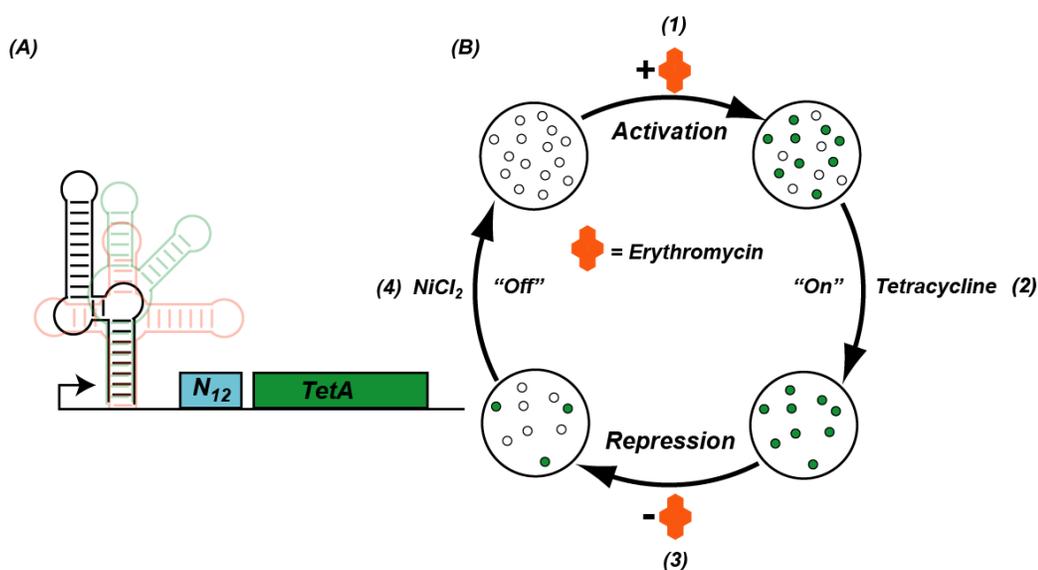
Figure 6.1 *Dual genetic selection scheme to identify erythromycin riboswitches*. (A) A library of candidate riboswitches is constructed by cloning a library of erythromycin aptamers into the 5' UTR of a *tetA* reporter gene. Separating the aptamers and reporter gene for in the library are randomized 12 base expression platforms. (B) Dual genetic selection. 1) A library of candidate riboswitches controlling *tetA* gene expression is constructed in *E. coli* under in nonselective conditions in the presence of erythromycin to activate tetA expression. 2) Positive selection: Following *tetA* activation the library of clones is then grown in the presence of erythromycin and tetracycline to select only switches that have successfully activated *tetA* expression. 3) Surviving clones are harvested and grown under non selective conditions in the absence of erythromycin to allow tetA expression to readjust. (4) Negative selection: Clones are then grown on media containing erythromycin and in the absence of erythromycin. Only clones displaying low levels of *tetA* expression can survive the negative selection step.

The positive selection for "on" switches began with cells transformed with our library of candidate riboswitches and randomized expression platforms being plated on solid LB media and grown overnight. The media was supplemented with ampicillin to maintain the plasmid, and erythromycin (200 μg/mL) to activate the expression of the TetA resistance protein. The following day, cells were harvested by overlaying the agar plate with 2 mL of LB liquid media. Harvested cells (100 μL) were then used to inoculate a 50 mL culture of LB supplemented with ampicillin and erythromycin to maintain tetA expression levels and grown to mid-log phase (OD 600 ~ 0.4-0.6). This culture was then diluted to a concentration of ~100,000 cells/mL with 1 mL of the dilution being plated on solid media supplemented with ampicillin, erythromycin and tetracycline (30 μg/mL) and grown overnight. For this selection step, rich media was necessary as minimal media containing 3 antibiotics (tetracycline, ampicillin and erythromycin) inhibits the growth of *E. coli*.

To counterselect for "off" switches, the clones surviving the positive selection were harvested as before and used to inoculate a culture containing only ampicillin that was again grown to mid-log phase. This culture was then diluted (~100,000 cells/mL) and plated on solid LB media containing ampicillin and $NiCl_2$ (3 mM) and grown overnight. Surviving clones were again harvested and the tetracycline and $NiCl_2$ selections were repeated for one additional round as described above. Previously, $NiCl_2$ has been used to identify tetracycline sensitive strains of *E. coli* grown on rich media containing 3 mM $NiCl_2$ and should, in principle, be able to identify members of our library now exhibiting low expression of the TetA resistance protein.

Following two rounds of positive and negative selection, 96 surviving clones were picked using a colony-picking robot and transferred to a 96-well microtiter plate and grown overnight. Each well of the microtiter plate contained 200 μL of LB media supplemented with ampicillin and erythromycin. To identify clones that displayed erythromycin-dependent growth in the presence of tetracycline, the overnight cultures were used to inoculate 2 new microtiter plates, one with LB supplemented with ampicillin and tetracycline and one with LB supplemented with ampicillin tetracycline *and* erythromycin. Cell growth was monitored by OD_{600} for 8 hours with no differences in growth observed. Faced with these negative results we chose to investigate the parameters of the selection scheme to determine the riboswitch qualities (i.e. expression levels) selected for using this scheme.

6.2.2 Determining the Parameters of the Dual Selection System

To determine the parameters for the dual selection scheme described above, we opted to use a previously identified theophylline riboswitch as a positive control for the purposes of 1) Confirming the ability of TetA to confer tetracycline resistance and 2) Verifying the ability of TetA expression to confer $NiCl_2$ sensitivity. By controlling TetA expression with a well-characterized switch, we reasoned that we would be able to relate the levels of tetA expression required to survive each of the selection steps to a previously determined measure of gene expression previously determined using a *lacZ* reporter (Miller units). We chose the riboswitch from pSAL8.1* to investigate the tetA selection system given the large dynamic range observed with a lacZ reporter (from ~70 Miller units in the absence of theophylline to ~5000 in its presence) and because a

theophylline dose response curve describing its expression levels had already been obtained.

A plasmid with *tetA* under the control of switch 8.1* was used to transform *E. coli* strain N281. After plating on solid selective media (LB-amp), a single clone was picked and used to inoculate 5 mL of LB supplemented with ampicillin.   To allow TetA expression to stabilize, 4 separate cultures were inoculated the following day using the overnight culture, with each containing ampicillin, erythromycin and a unique concentration of theophylline (0, 0.1, 0.25 and 0.5 mM). It is important to clarify that the erythromycin present in the media should have no effect on the theophyline riboswitch and is instead present to mimic the conditions of our previous selections. Cultures were grown to mid-log phase, diluted as before, plated on solid media supplemented with ampicillin, erythromycin, the appropriate concentration of theophylline and tetracycline and grown overnight.

The results shown in Figure 6.2 indicate a theophylline concentration of 250 μM induces a sufficient amount of TetA expression to allow *E. coli* to survive on the tetracycline concentration (30 μg/mL) used in the positive selection steps described above.   Previous dose response experiments indicated that this switch produced expression levels of ~1000 Miller units at these concentrations of theophylline (Figure 6.4 A). These results agreed nicely with previous TPP switches identified using this method as nearly all identified switches produced expression levels of at least 1000 Miller units in the presence of TPP at the concentrations used in the selection.

With the expression levels required to survive the tetracycline step established, we attempted to determine the expression levels of TetA that would prove toxic when grown in the presence of $NiCl_2$. A single clone harboring the same construct comprised of *tetA* under the control of switch 8.1* was again used to inoculate selective media and the culture grown overnight. The overnight culture was used to inoculate 4 cultures containing increasing concentrations of theophylline (0, 0.1, 0.25 and 0.5 mM). Once grown to mid-log phase, cultures were diluted and plated on selective media containing 3 mM $NiCl_2$ and the appropriate concentration of theophylline.
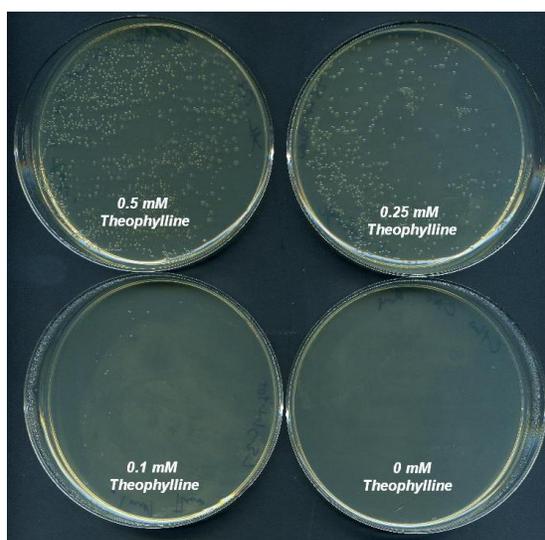
Figure 6.2 *Growth assay to determine parameters of positive selection step from tetA selection scheme on rich media.* Bacteria harboring the *tetA* gene under the control of our previously described theophylline riboswitch 8.1* were grown in the presence of increasing concentrations of theophylline and tetracycline (30 μg/mL). Growth was observed between the theophylline concentrations of 0.1 and 0.25 mM theophylline. With switch 8.1*, these theophylline concentrations induce β-galactosidase activities of approximately 1000 Miller units determined with a *lacZ* reporter.

Unfortunately, bacteria expressing tetA were not sensitive to $NiCl_2$ at any theophylline concentration. Even in the presence of 1 mM theophylline, *tetA* expression did not prove to be lethal to bacteria grown on 3 mM $NiCl_2$. Using a β-galactosidase reporter in the presence of 1 mM theophylline, switch 8.1* activated gene expression to a level that corresponds to ~6000 Miller units (Figure 6.4 A). Any additional increases in $NiCl_2$ concentration proved to be lethal to all bacteria regardless of the level of *tetA* expression. Similar results were obtained from experiments carried out using fusaric acid, a compound that has previously been used to identify tetracycline-sensitive bacteria, in place of $NiCl_2$. Also, clones surviving this concentration of $NiCl_2$ were tested to confirm the expression of an active TetA protein. Surviving clones were again grown in the presence of theophylline and subsequently shown to maintain the tetracycline resistant phenotype in the same theophylline-dependent manner

Thus far, all of the experiments described were conducted using rich, LB media. Previous efforts employing this selection scheme required the use of minimal media, and we reasoned that the rich media used thus far may enable the bacteria to undermine either the positive or negative selection steps. To investigate the influence of the rich media on this selection scheme, we carried out the same set of experiments using minimal media. Due to the toxicity of media containing multiple antibiotics, erythromycin was removed from the media as it is not required for the activity of a theophylline riboswitch.

Again, 4 separate cultures were inoculated using an overnight culture of *E. coli* with containing *tetA* under the control of switch 8.1* with each having a unique concentration of theophylline (0, 100, 250 and 500 μM). Cultures were grown to mid-

log phase, diluted as before, plated on solid M9 minimal media supplemented with
ampicillin, the appropriate concentration of theophylline and either tetracycline
(30 μg/mL) or NiCl$_2$ (0.3 mM).   The nickel chloride concentration was lowered to
0.3 mM, which was the concentration previously used on minimal media to select for
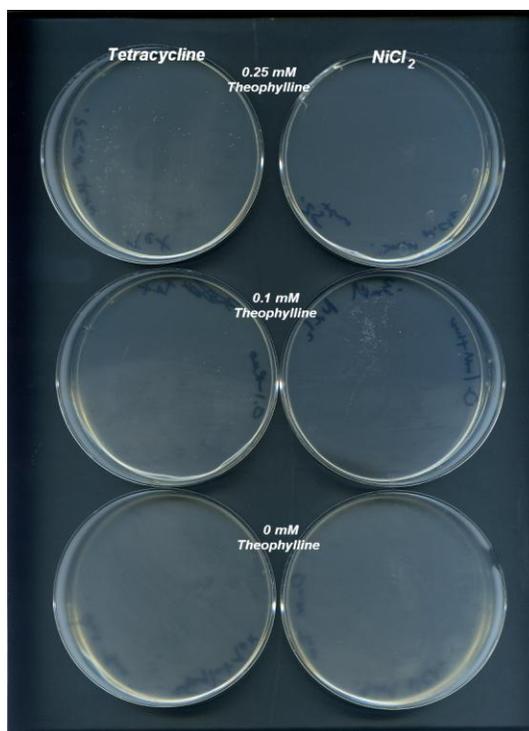
Figure 6.3 *Growth assay to determine parameters of positive selection step from tetA selection
scheme on minimal media.*   Bacteria harboring the *tetA* gene under the control of our
previously described theophylline riboswitch 8.1* were grown on mimimal media in the
presence of increasing concentrations of theophylline, tetracycline (left; 30 μg/mL) and NiCl$_2$
(right; 0.3 mM).   Growth was observed on tetracycline plates between the theophylline
concentrations of 0.1 and 0.25 mM theophylline.   The same concentrations of theophylline
were also observed to confer NiCl$_2$ sensitivity.   In this image, the plate containing NiCl$_2$
appears to be bare; however, this is a result of poor imaging.   Small, visible colonies were
picked from this plate, cultured and the theophylline-induced tetracycline resistant phenotype
verified.

tetracycline sensitive bacteria[5,6,10]. Figure 6.3 shows that, after a 48 h incubation period, theophylline concentrations of 250 μM were sufficient to induce adequate TetA expression to confer tetracycline resistance. NiCl$_2$ sensitivity was observed at approximately the same theophylline concentration that conferred tetracycline resistance.

Again, surviving clones were grown on minimal media in the presence of theophylline and subsequently demonstrated that they had maintained the tetracycline resistant phenotype. Taken together, these results indicate the rich media from our previous experiments had compromised the negative selection steps.

6.3 Conclusion

Using a pool of putative erythromycin aptamers from our in vitro SELEX experiments, we created a library of possible riboswitches by cloning the pool of aptamers into the 5' UTR of a *tetA* reporter gene. A selection scheme exploiting the 2 distinct phenotypes (tetracycline resistance and NiCl$_2$ sensitivity) produced by TetA expression, was employed to identify functioning switches that activated gene expression in the presence of erythromycin. Following 2 rounds of selection, we discovered no erythromycin-sensitive switches and chose to examine the parameters of our selection scheme using a previously discovered, well-characterized theophylline riboswitch[3].

Miller Units vs. [theo]
Switch 8.1*
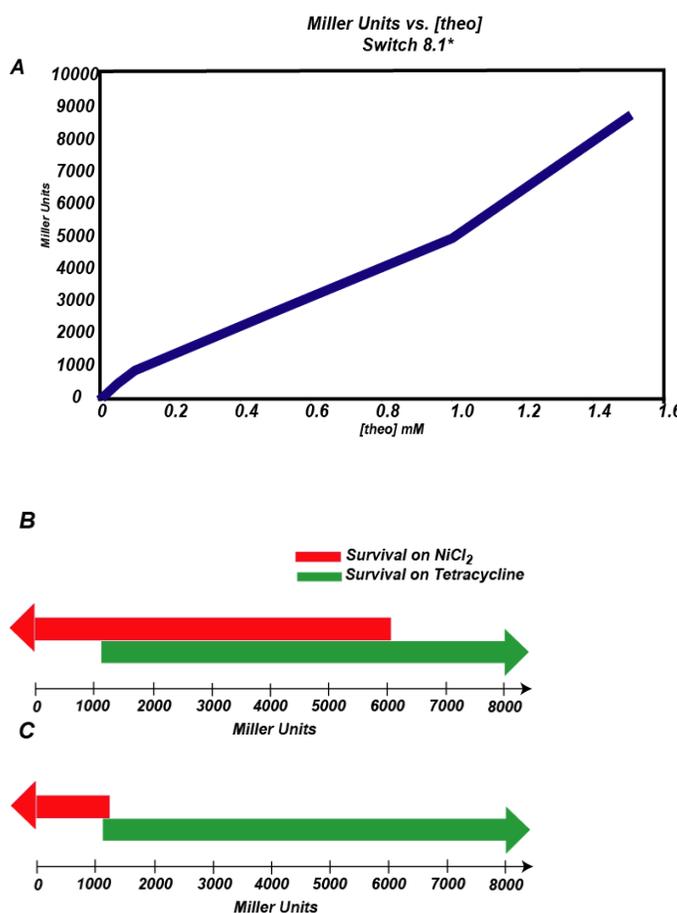
Figure 6.4 *Correlation of expression levels required for survival on NiCl₂ and tetracycline.* (A) Dose response of β-galactosidase activity (Miller units) for switch 8.1* as function of theophylline concentration. Theophylline concentrations of 0.1-0.25 mM give rise to expression levels corresponding to ~1000-1500 Miller units. (B) On rich media, theophylline concentrations giving rise to expression levels greater than ~1000-1500 Miller units were required to survive on tetracycline (30 μg/mL). Survival on NiCl₂ (3 mM) was observed at theophylline concentrations giving rise to expression levels of at least 6000 Miller units. (C) On minimal media, theophylline concentrations giving rise to expression levels greater than ~1000-1500 Miller units were required to survive on tetracycline (30 μg/mL). Survival on NiCl₂ (0.3 mM) was observed at theophylline concentrations giving rise to expression levels of less than 1000-1500 Miller units. Taken together these results indicate the need for minimal media in the negative selection step.

Specifically, we chose to examine the selection parameters when bacteria were grown on rich media versus minimal media. These investigations revealed the requirement of using minimal media in the negative selection steps and confirmed the selection scheme's utility for identifying riboswitches by counterselecting for "on" phenotypes against "off" phenotypes.

6.4 Future Directions

Given what we now know about the media requirements of the dual selection strategy, it would be prudent to repeat the selection for erythromycin-sensitive riboswitches. Due to the toxicity of multiple antibiotics in the media, rich media must be used for the positive selection steps. This should not present a problem given the similar results observed in control experiments using rich and minimal media. Negative selection steps should be performed on minimal media as it provides a condition in which high levels of *tetA* expression give rise to $NiCl_2$ sensitivity. Also, combining this selection scheme with a high throughput screening method, such as FACS, may also be worthwhile.

By tagging the TetA protein with a GFPuv translational fusion, one can potentially use a single reporter to select for high levels of expression in the presence of a ligand using tetracycline and subsequently screen for low levels of gene expression in its absence using FACS, or vice versa. A similar approach was recently reported by Yokobayashi and coworkers[10, 11]. A vector has been created for these purposes (SAL 773.1) and contains the *tetA* gene fused to GFPuv through a long, flexible linker. The expression of this *tetA-GFPuv* gene is under the control of switch 8.1*.

6.5 Experimental

A vector containing the tetA gene (pLac+thiM#2tetA) was kindly provided by the Yokobayashi group at the University of California, Davis. A library of candidate riboswitches was created by cloning a library of aptamers from round 7 of our SELEX

experiments into the 5' UTR of the *tetA* resistance gene. Separating the aptamers and the start codon of the *tetA* gene is a 12 base randomized sequence followed by the same CAACAAG sequence from our theophylline switches. To begin, a PCR product (A) was generated using forward primer SAL 95, which anneals to the 5' priming region of our DNA pool of aptamers and contains a KpnI cut site, and reverse primer SAL 157 which anneals to the 3' constant region of our DNA pool and contains a 12 base randomized region as well and an overlap with the *tetA* gene. PCR product (B) was generated using forward primer SAL 156, which anneals to the *tetA* gene and overlaps with the aptamer pool, and reverse primer SAL 159 which anneals to the 3' end of *tetA* and contains a SacI restriction site. PCR products A and B were mixed and assembled in a PCR reaction containing outer primers SAL 95 and SAL 159 to generate PCR product (C). This PCR product was then digested with KpnI and SacI and cloned into the same sites of vector pSAL8.1*. The ligation to generate this plasmid library was butanol precipitated and used to transform an electrocompetent strain of *E. coli*, MDS42. Top10F' *E. coli* should be avoided in these experiments as they are already tetracycline resistant. Approximately 100,000 clones were grown and the plasmid library harvested. These plasmids were then used to transform the erythromycin-resistant strain of *E. coli*, N281.

To begin, a PCR product (A) was generated using forward primer SKD 178, which anneals to pSAL8.1* vector 5' to the mTCT4-8 aptamer, and SAL 155 which anneals to the expression platform of switch 8.1* and contains an overlap with the tetA gene. PCR product (B) was generated using forward primer SAL 154, which anneals to the *tetA* gene and overlaps with the expression platform of switch 8.1*, and reverse primer SAL 159 which anneals to the 3' end of *tetA* and contains a SacI restriction site. PCR products A

and B were mixed and assembled in a PCR reaction containing outer primers SKD 178 and SAL 159 to generate PCR product (C). This PCR product was then digested with KpnI and SacI and cloned into the same sites of vector pSAL8.1* to yield SAL 765.2.

A second plasmid with GFPuv translationally fused to TetA was also created using the following strategy. A PCR product (A) was generated with forward primer SAL 174, which anneals to the *tetA* gene 5' to the EagI restriction site, and SAL 175 which anneals to the 3' end of *tetA* and contains a (Gly-Ser-Ser-Ser)$_4$ overhang. A second PCR product (B) was generated using forward primer SAL 176, which anneals to the 5' end of the GFPuv gene immediately following the start codon and contains the (Gly-Ser-Ser-Ser)$_4$ overhang, and reverse primer SAL 177 which anneals to the 3- end of GFPuv and contains a SacI restriction site. PCR products A and B were combined and assembled in a PCR reaction using forward primer SAL 174 and reverse primer SAL 177 to yield PCR product (C). This PCR product was the digested with EagI and SacI and cloned into those same sites in vector SAL 765.2 yielding pSAL773.1.

Solid M9 minimal medium containing 0.8% glycerol and 0.1% casamino acids supplemented with ampicillin (50 μg/mL) was used to grow N281 *E. coli* cells transformed with plasmid pSAL773. When appropriate, tetracycline (30 μg/mL), erythromycin (200 μg/mL, Sigma Aldrich) and NiCl$_2$ (0.3 mM) were added to media. Cells were grown at 37 °C.

6.6 *References*

(1)   Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. *Chem. Biol.* **2007**, *14*, 173-184.

(2)   Weigand, J. E.; Sanchez, M.; Gunnesch, E.-B.; Zeiher, S.; Schroeder, R.; Suess, B. *RNA* **2008**, *14*, 89-97.

(3)   Lynch, S. A.; Gallivan, J. P. *Nucl. Acids Res.* **2009**, *37*, 184-192.

(4)   Topp, S. G., J.P. *ChemBioChem* **2008**, *9*, 210-213.

(5)   Muranaka, N.; Sharma, V.; Nomura, Y.; Yokobayashi, Y. *Nucl. Acids Res* **2009**, *37*, e39.

(6)   Nomura, Y.; Yokobayashi, Y. *Biosystems* **2007**, *90*, 115-20.

(7)   Bochner, B. R.; Huang, H. C.; Schieven, G. L.; Ames, B. N. *J. Bacteriol.* **1980**, *143*, 926-933.

(8)   Griffith, J. K.; Buckingham, J. M.; Hanners, J. L.; Hildebrand, C. E.; Walters, R. A. *Plasmid* **1982**, *8*, 86-88.

(9)   Maloy, S. R.; Nunn, W. D. *J. Bacteriol.* **1981**, *145*, 1110-1111.

(10)   Nomura, Y.; Yokobayashi, Y. *J. Am. Chem. Soc.* **2007**, *129*, 13814-13815.

(11)   Sharma, V.; Nomura, Y.; Yokobayashi, Y. *J. Am. Chem. Soc.* **2008**, *130*, 16310-16315.

(12)   Wittmann, H. G.; Stöffler, G.; Apirion, D.; Rosen, L.; Tanaka, K.; Tamaki, M.; Takata, R.; Dekio, S.; Otaka, E.; Osawa, S. *Mol. Gen. Genetics* **1973**, *127*, 175-189.

(13)   Chittum, H. S.; Champney, W. S. *J. Bacteriol.* **1994**, *176*, 6192-6198.

Chapter 7:

Controlling Gene Expression with Visible Light

7.1 Introduction

In addition to controlling bacterial gene expression with small molecules, we have also explored the possibility of using visible light to accomplish this same task. Using light to control gene expression in bacteria has a variety of applications in the investigation of biological processes as it allows for the spatiotemporal control of bioactivity. Light also offers the advantage of being relatively harmless and likely limits the many unintended effects on gene expression caused by chemical inducers[1].

The most common approach to the photoinducible control of gene expression in bacteria involves a practice known as "caging" where small molecules needed for the induction of gene expression are inactivated using a photoprotecting group. Upon the absorption of light, the photoprotecting group is removed thus enabling the small molecule to function normally and activate gene expression. One of the most exciting applications of this approach is seen in the work of Lawrence *et al*[2]. By caging agonists of the insect ecdysone receptor[3], the authors were able to achieve photoinducible control of gene expression in mammalian cells. Yeast constitutively expressing a protein consisting of an ecdysone receptor fused to the DNA binding domain of a transcriptional activator was grown in the presence of cell-permeable, photocaged ecdysone molecules. After irradiation with light, the photocaged agonists were released, restoring their bioactivity. Binding of the free ecdysone to its orthogonal receptor triggers the transcriptional activator thus turning on gene expression. The variety of exogenous, non-

toxic ecdysone agonists available[4], along with the transportability of the ecdysone receptor into many eukaryotes[5] makes this approach very powerful.

Photosynthetic bacteria and algae must coordinate a number of cellular processes in response to light including cell motility and photosynthesis. The sensing of both the quality and quantity of light is accomplished using light-sensitive, membrane-bound photoreceptor proteins[6], known as phytochromes, that have evolved to respond to specific wavelengths of light. Using nature as an inspiration, Levskaya *et al.* engineered a bacterial photoreceptor to induce gene expression in *E. coli* using red light[7]. Summarized in Figure 7.1, a chimeric protein was created by coupling the photoreceptor of the bacterial phytochrome, Cph1 from the cyanobacterium *Synechocystis*[8], which absorbs red and far-red light, to the histidine kinase domain of the *E. coli* EnvZ protein via an N-terminal translational fusion. The two-component EnvZ-OmpR regulatory system of *E. coli* directs porin gene expression in response to changes in environmental osmolarity and has been used for the construction of functional chimaeras[9,10]. In this system, changes in environmental solute concentration influence the phosphorylation of the EnvZ histidine kinase. When phosphorylated, EnvZ can transfer the phosphoryl group to the OmpR protein.   Upon phosphorylation, OmpR binds to the *ompC* promoter, thus activating the transcription of downstream genes. When fused to the Cph1 photoreceptor, the histidine kinase of EnvZ functions normally in the absence of red light. When exposed to red light, the photoreceptor changes conformation and disrupts the function the histidine kinase thus prohibiting the phosphorylation of OmpR, "turning off" gene expression.
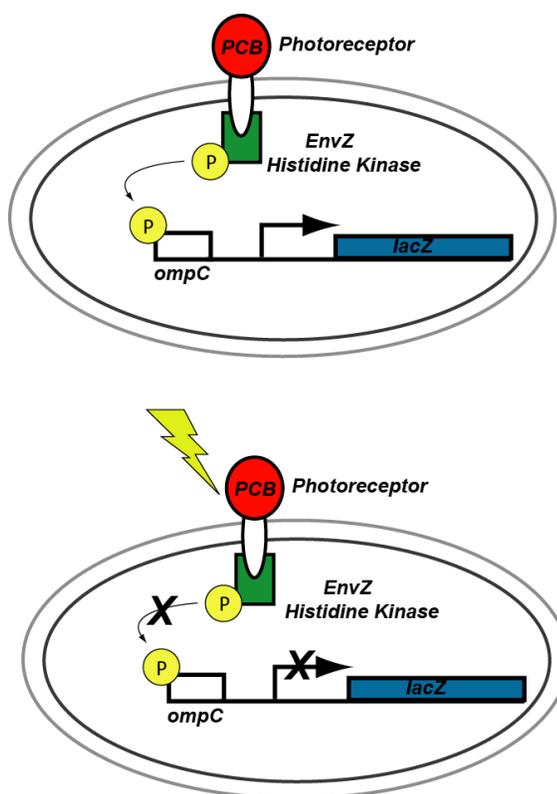
Figure 7.1 *Engineered photoreceptor to control the expression of specific genes with red light.* The Cph1 from the cyanobacterium *Synechocystis,* was attached to the histidine kinase domain of the *E. coli* EnvZ protein via an N-terminal translational fusion. When phosphorylated, EnvZ can transfer the phosphoryl group to the OmpR protein which activates transcription of genes downstream from the ompC promoter. When fused to the CphI photoreceptor, the histidine kinase of EnvZ functions normally in the absence of red light. When exposed to red light, histidine kinase function is interrupted thus prohibiting the phosphorylation of OmpR, "turning off" gene expression.

In a similar approach, Quail *et al.* fused a different red-light photoreceptor to the DNA binding domain of the GAL4 transcriptional activator[11]. The absorption of red light is translated to a change in gene expression through communication with PIF, a basic helix—loop-helix protein that selectively interacts with the light induced conformation of the red-light photoreceptor. Working in yeast, the authors were able to achieve 1000-fold increases in gene expression upon irradiation with red light.

With the variety of known natural bacterial proteins that absorb light of differing wavelengths[12,13], it may be possible to take a similar photoreceptor-based approach to engineering photoresponsive systems that will respond to other wavelengths of light. For example, the well-studied PYP-phytochrome-related (Ppr) photoreceptor from *Rhodospirillum centenum* absorbs blue light (434nm) through a *p*-hydroxycinnamic acid chromophore[14-16] (Figure 7.2). The distinctive photoactive yellow protein (PYP) of the Ppr from *R. centenum* separates this phytochrome from the closely related elements of plants and cyanobacteria. Phytochomes are typically composed a chromophore-binding domain linked to a histidine kinase domain whose function and signaling partner are often unknown. In plants and cyanobacteria, the phytochrome chromophore (most often a bilin) is covalently attached to the chromphore-binding domain through a conserved cystiene residue that is not present in the Ppr photoreceptor. Instead, the *p*-hydroxycinnamic acid chromophore is attached to the PYP domain through a thioester linkage where, upon the absorption of 434 nm light, the *p*-hydroxycinnamic acid chromphore undergoes a trans-cis isomerization, leading to a conformational change in the protein that ultimately disrupts the histidine kinase activity of the Ppr.

Using the system created by Levskaya *et al.* as a model, it is reasonable to suggest that by replacing the Cph1 photoreceptor with the Ppr from *R. centenum,* one may be able to create a new photoresponsive chimera. Much like riboswitches and their small-molecule activators, the mechanism by which light induces a change in gene expression in engineered systems is not always intuitive. Therefore, we believe a genetic screen may be useful in creating photoreceptor chimeras proficient at modulating gene expression in
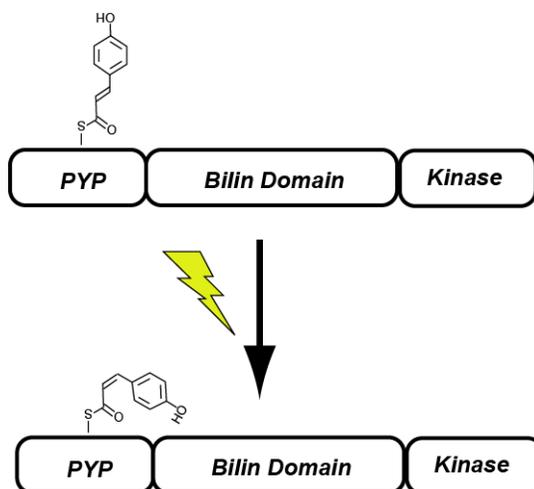
Figure 7.2 *PYP-phytochrome-related (Ppr) photoreceptor from* Rhodospirillum centenum. Upon absorption of blue light, the *p*-hydroxycinnamic acid chromophore undergoes a trans-cis isomerization that is translated to a change in gene expression through a kinase. In *R. centenum*, absorption of light down regulates chalcone synthase gene expression.

bacteria. Levskaya *et al.*[7] created a small library of chimeras consisting of the EnvZ histidine kinase fused to a series of photoreceptors of varying lengths with the lengths of the photoreceptor chosen based on ClustalW2 sequence alignments of the photoreceptor and EnvZ histidine kinase. Similar attempts to align Ppr with the EnvZ histidine kinase failed to indicate any areas of overlap likely to produce functioning chimeras. With this, we chose to create a library of chimeras consisting of a full-length photoreceptor linked to the EnvZ histidine kinase through a randomized, 3 amino acid linker. Herein we describe a method using the *envZ* knockout strain of *E. coli*, CP919[10], which harbors a *lacZ* reporter gene on its chromosome under the control of the *ompC* promoter, which

enables one to screen for functioning chimeras by monitoring the activation of β-galactosidase expression in either the presence or absence of light.

7.2 Results and Discussion

To begin, we cloned the histidine kinase portion of *envZ* into a plasmid using a strategy that will allow us to easily create a library of photoreceptors of varying lengths. Previously it has been shown that the number of amino acids from the photoreceptor that are included in a photoreceptor chimera can significantly influence its function[7]. We obtained a plasmid containing the PYP and bilin domain of the blue-light Ppr photoreceptor and created a library of chimeric photoreceptors by varying the composition of amino acids joining the photoreceptor to the envZ histidine kinase. Once our library of protein chimeras had been constructed, they were introduced into the *E. coli* strain CP919 harboring a second "helper" plasmid, pACYC(TAL,pCL)[15], which contains the tyrosine ammonia lyase and *p*-hydroxycinnamic acid ligase genes necessary for the in vivo production and attachment of the *p*-hydroxycinnamic acid chromophore, and screened for the disruption of histidine kinase function upon the absorption of light using a *lacZ* reporter gene. CP919 is an *E. coli* strain that is deficient in *envZ* and contains a *lacZ* reporter gene on its chromosome downstream of the *ompC* promoter. The bacteria were grown in the absence of light on selective media containing S-gal with functioning chimeras promoting the expression of *lacZ* appearing black. S-gal is a light-insensitive β-galactosidase substrate similar to the more common light-sensitive X-gal; however, it appears black in the presence of β-galactosidase rather than blue. Using a colony picking robot, 96 black clones were isolated from a library of ~5,000 clones,

transferred to a 96-well microtiter plate and subsequently grown to saturation. The following day, the 96 cultures were "stamped" in duplicate on solid media, again, containing S-gal with one set of clones grown under direct light and other grown in the dark. Photoresponsive bacteria could then be identified by qualitative differences in color between those grown in the light and those grown under direct light. Multiple clones displaying repressed β-galactosidase activity in the presence of light were identified, cultured in the presence and absence of light and assayed using Miller's method to quantitatively establish the extent of gene expression. Unfortunately, quantitative analysis of β-galactosidase activity did not reveal any modulation of gene expression based upon exposure to light.
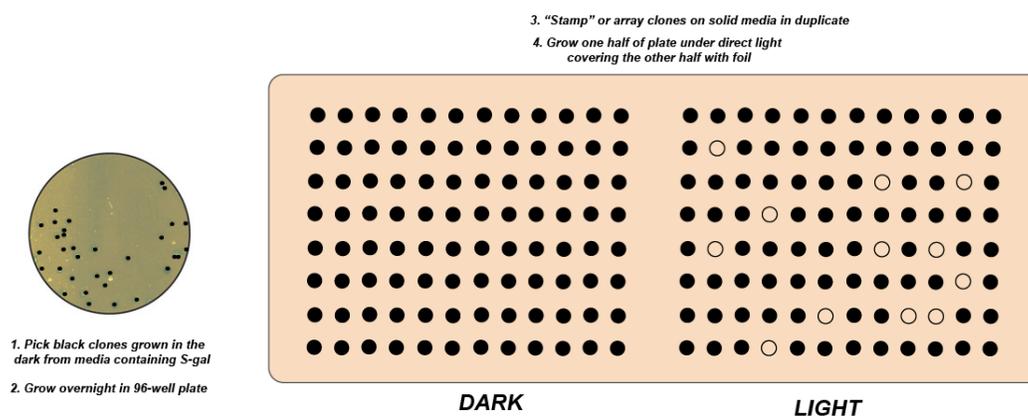
Figure 7.3 *Summary of β-galactosidase screen to detect functioning photoreceptor chimeras.*

We were suspicious of the ease with which we were able to identify constructs that appeared to function using our screen. At least 10 of the 96 clones appeared to show a photoinduced change in gene expression. To determine if the S-gal media or nonspecific effects of light on *E. coli* gene expression were playing a role in the

qualitative differences of gene expression we had observed, we chose to inoculate selective LB-agar containing S-gal with a culture of bacteria prior to solidifying. The bacteria used to inoculate the media contained a "leaky" riboswitch controlling *lacZ* which in the absence of theophylline generated β-galactosidase activities of approximately 800 Miller units[17]. By covering one half of the plate with foil and exposing the other half to direct light, we would be able to visualize the change in gene expression similar to the "bacterial photography" experiments of Levskaya *et al.*[7] Unfortunately, this control plate revealed that the half of the plate grown in the light displayed markedly less color that the half grown in the dark. These results called into question all of the qualitative differences we observed in our initial screen.

One anomaly (RH008-A12) was identified from our screen that appeared to activate gene expression in the presence of light. This was a potentially exciting discovery as many natural phytochromes appear to deactivate their target genes upon exposure to light. In an attempt to verify the phenotype, again we chose to inoculate selective media containing S-gal with a saturated culture of RH008-A12 by inoculating the media prior to solidifying. Indeed, it appeared that the phenotype was genuine; however, quantitative determinations of Miller units determinations were again inconculsive and indicated gene expression may be repressed by the light rather than activated (Figure 7.4).

Sequencing revealed the anomalous clone to contain a full length histidine kinase fused to only the first 81 amino acids of our photoreceptor. While we could only qualitatively verify a phenotype, we decided to redesign our chimera with a library of

linker sequences such that it contained only the PYP domain rather than the full length photoreceptor. Also, since RH008-12 appeared to behave opposite to our control experiments (i.e. the anomalous clone appeared to turn on gene expression whereas the control plate showed a decrease in gene expression in the presence of light) further screening for this phenotype was warranted. The PYP domain is the only portion of the photoreceptor of which a crystal structure has been solved and indicates that the protein, specifically an N-terminal π helix, undergoes a significant structural rearrangement upon exposure to light. By shortening the photoreceptor to only this domain, we reasoned that we may be able to exploit this structural change to produce a light-dependent increases in gene expression.

A plasmid library of a short photoreceptor fused to the histidine kinase through a randomized linker was used to transform strain CP919. The library of clones was again plated on selective media containing S-Gal and grown 37 ºC in the dark overnight. The following day, the whitest clones were picked and the stamping assay repeated. Two clones that showed possible increases in gene expression were selected and the Miller units assayed as before. Again, the quantitative assay showed no differences in β-galactosidase activity.
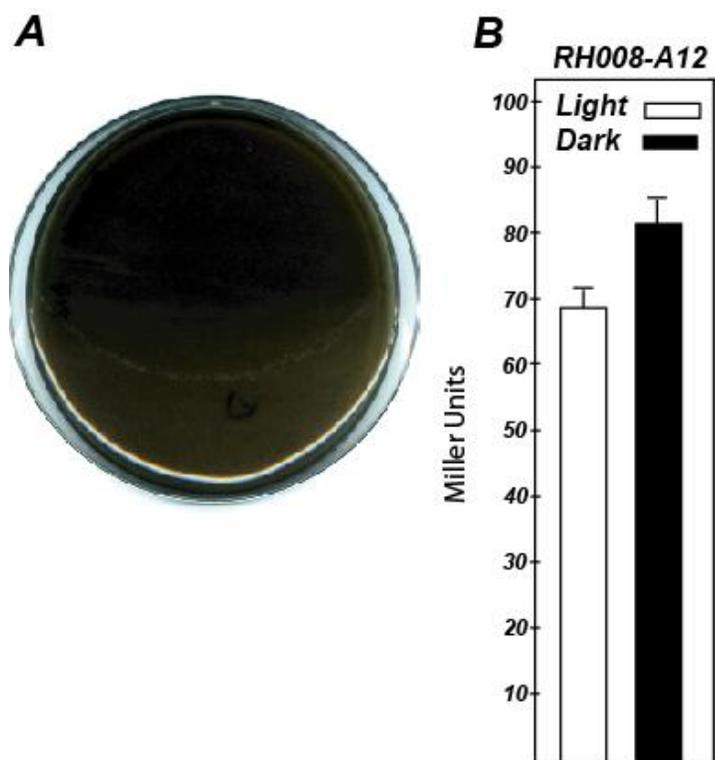
Figure 7.4 *Anomoly RH008-A12.* (A) Qualitative assay shows increased gene expression in the presence of light. "D" represents half of plate covered with foil. (B) β-galactosidase activity of clone RH008-A12 reveals no activation of gene expression upon exposure of light.

To investigate the possibility of the media playing a role in the qualitative differences we have observed, we conducted a negative control experiment with only solid media. Despite the claims of S-Gal being insensitive to light, a small but noticeable color difference was observed. When exposed to light, the media inoculated with no bacteria became slightly darker in the presence of light. The reasons for this are unknown but it is possible that the differences observed could be due to a small breakdown of S-gal or an increase in the media's concentration due to evaporation. The

temperature inside of the incubator remained constant through the experiment but it is likely that the direct light could have increased the temperature of the media.

7.3 Conclusion and Future Directions

Despite our failure to construct a photoinducible system with which to control bacterial gene expression, we have learned a great deal about some of the caveats present in this type of approach. Our opting to use *lacZ* as a reporter gene in our screening proved to cause a number of difficulties. To begin, the S-gal media used in these experiments, despite claims to the contrary, appears to be slightly affected by exposure to direct light making the qualitative detection of changes in β-galactosidase activity very difficult. Also, we were unable to obtain a strain of bacteria that could be used as a positive control to confirm that our experimental setup (e.g. light source, incubator, etc) was adequate. Furthermore, control experiments indicated that *E. coli* are also affected by direct exposure to light with either cell growth or β-galactosidase expression being markedly reduced in the presence of light. Taken together, these results suggest that a genetic selection for photoinduced changes in gene expression may be a better approach. To do this, one would only need to replace the *lacZ* gene downstream of the *ompC* promoter of the *E. coli* strain CP919 with an antibiotic resistance gene such as *cat*. Preliminary control experiments and a survey of the literature indicate chloramphenicol is insensitive to light exposure and therefore an excellent choice of antibiotic.

A complete redesign of the protein chimera should also be considered. Simply joining the photoreceptor to the histidine kinase may be a drastic oversimplification of design. A great deal of success in the engineering of hybrid proteins with switch-like

characteristics has been attributable to approaches involving domain insertions[18] where one protein is inserted into the scaffold of another. In our case, constructing a library of chimeras where the histidine kinase domain is inserted into the photoreceptor at varying positions may prove to be rewarding.

7.4 Experimental

General Considerations All plasmid manipulations utilized standard cloning techniques. All constructs have been verified by DNA sequencing performed by MWG. Purifications of plasmid DNA, PCR products, and enzyme digestions were performed using kits from Qiagen. o-nitrophenyl-β-D-galactopyranoside (ONPG), ampicillin chlorampneicol and kanamycin were purchased from Sigma. S-gal media was also purchased from Sigma and prepared according to manufacturer's specifications. Synthetic oligonucleotides were purchased from IDT. All plasmid experiments were performed in *E. coli* CP919 (Kan$_R$) cells obtained from Dr. Gerald Hazelbauer's lab at the University of Missouri. Plasmids pACYC(TAL,pCL) which contains the biosynthetic genes for holo-PYP formation, and pZJ137 which contains the PYP photoreceptor were obtained from Dr. John Kyndt at the University of Arizona.

Library Contruction

The photoreceptor- histidine kinase chimera was constructed using the following strategy. The photoreceptor from plasmid pZJ137 was PCR amplified using forward primer SAL 129, which anneals to the RBS of the Ppr gene and contains a KpnI

restriction site, and reverse primer SAL 130 which anneals to the photoreceptor gene approximately 1800 bases into the gene and contains an in-frame HindIII restriction site. This PCR product was digested with KpnI and HindIII, gel purified and cloned into the sites present in pSAL172 to yield pSAL676. pSAL172 is identical to pSAL8.1* aside from the strong ptac promoter being replaced with a weaker IS10 promoter. An identical strategy was used to create a construct with a photoreceptor chimera containing only the first 129 amino acids of the PYP domain with the exception of reverse primer SAL130 being replaced with primer SAL 133.

The histidine kinase from the envZ protein was amplified from *E. coli* chromosome using forward primer SAL 134, which contains a HindIII restriction site followed by 9 randomized nucleotides, and reverse primer SAL 132 which contains a SacI restriction site. The PCR product was digested with HindIII and SacI, gel purified and cloned into the pSAL676 vector cut with the same enzymes to yield a plasmid library pRH008.

Screening Experiments

Selective S-gal media was prepared with ampicillin (50 μg/mL)to maintain the chimeric photoreceptor plasmid, chloramphenicol (34 μg/mL) to maintain the pACYC(TAL,pCL) plasmid and kanamycin (10 μg/mL). IPTG was also added to the media to induce the expression of tyrosine ammonia lyase and *p*-hydroxycinnamic acid ligase. *E. coli* CP919 harboring the pACYC(TAL,pCL) plasmid were made electrocompetant and transformed with library of photoreceptor chimeras and plated on selective S-gal media and grown overnight at 37 ℃. A compact fluorescent light bulb (Ultra mini 23W, Westpointe) was used for conditions requiring light exposure. Compact fluorescents have a Color

Rendering Index (CRI) of ~80 and are known to emit blue light. CRI provides a quantitative measure of the ability of a light source to reproduce the colors of various objects faithfully in comparison with an ideal or natural light source. All Miller assays and "bacteria photography" experiments were conducted exactly as described by Levskaya *et al.*[7]

7.5 References

(1)   Young, D. D.; Deiters, A. *Org. Biomol. Chem.* **2007**, *5*, 999-1005.

(2)   Lin, W.; Albanese, C.; Pestell, R. G.; Lawrence, D. S. *Chem. Biol.* **2002**, *9*, 1347-1353.

(3)   No, D.; Yao, T. P.; Evans, R. M. *Proc. Nat. Acad. Sci. U. S. A.* **1996**, *93*, 3346-3351.

(4)   Dinan, L.; Hormann, R. E.; Fujimoto, T. *J. Comput. Aided Mol. Des.* **1999**, *13*, 185-207.

(5)   Alberto, M.; Caroline, S.; Cliff, A. H.; John, T.; Ian, J. *Plant Journal* **1999**, *19*, 97-106.

(6)   Hegemann, P. *Annu. Rev. Plant Biol.* **2008**, *59*, 167-189.

(7)   Levskaya, A.; Chevalier, A. A.; Tabor, J. J.; Simpson, Z. B.; Lavery, L. A.; Levy, M.; Davidson, E. A.; Scouras, A.; Ellington, A. D.; Marcotte, E. M.; Voigt, C. A. *Nature* **2005**, *438*, 441-2.

(8)   Gambetta, G. A.; Lagarias, J. C. *Proc. Nat. Acad. Sci. U. S. A.* **2001**, *98*,

10566-10571.

(9)   Cai, S. J.; Inouye, M. *J. Biol. Chem.* **2002**, *277*, 24155-24161.

(10)   Utsumi, R.; Brissette, R. E.; Rampersaud, A.; Forst, S. A.; Oosawa, K.; Inouye, M. *Science* **1989**, *245*, 1246-1249.

(11)   Shimizu-Sato, S.; Huq, E.; Tepperman, J. M.; Quail, P. H. *Nat. Biotechnol.* **2002**, *20*, 1041-1044.

(12)   Sharrock, R. *Genome Biol.* **2008**, *9*, 230.

(13)   van der Horst, M. A.; Key, J.; Hellingwerf, K. J. *Trends Microbiol.* **2007**, *15*, 554-562.

(14)   Jiang, Z.; Swem, L. R.; Rushing, B. G.; Devanathan, S.; Tollin, G.; Bauer, C. E. *Science* **1999**, *285*, 406-409.

(15)   Kyndt, J. A.; Fitch, J. C.; Meyer, T. E.; Cusanovich, M. A. *Biochemistry* **2007**, *46*, 8256-8262.

(16)   Rajagopal, S.; Moffat, K. *Proc. Nat. Acad. Sci. U. S. A.* **2003**, *100*, 1649-1654.

(17)   Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. *Chem. Biol.* **2007**, *14*, 173-184.

(18)   Ostermeier, M. *Protein Eng. Des. Sel.* **2005**, *18*, 359-364.

Chapter 8:

Summary

8.1 Summary

The work described in this thesis began with notion that ligand-binding RNA molecules could be used to control gene expression in bacteria and earns its place in a Chemistry department, not by the techniques or methodology used, but by the problem it is designed to address: Chemical recognition in vivo. While the number of engineered systems described in the literature utilizing RNA-small molecule interactions to control gene expression in both prokaryotes and eukaryotes was limited, the principles driving their function were straightforward. Therefore, it was tempting to believe these ideas could be readily applied in the development of similar genetic control systems that could respond to new ligands.

Using a well-characterized theophylline aptamer, our lab had created a synthetic riboswitch capable of activating bacterial gene expression in the presence of theophylline[1]. The high-throughput screens and selections described in this thesis have assisted in identifying improved riboswitches, while the sequence data from these switches has enabled us to develop and test a mechanism for their function[2]. The relative ease with which this riboswitch was constructed supported the notion that the rapid creation of new synthetic riboswitches capable of detecting small molecules was within reach. Yet the question remains: Are we any closer to creating designer genetic control systems that can be used to control bacterial gene expression in response to a chosen small molecule?

Like most questions, the answer is both yes and no. Clearly, the major hurdle to creating new synthetic riboswitches lies in the ability to isolate RNA aptamers capable of binding small-molecule ligands. The selection of RNA aptamers is largely considered to be a "solved problem" and, at first glance, this would appear to be true. RNA aptamers are readily selected to bind macromolecular targets[3] yet small-molecule aptamers are far more elusive. Furthermore, despite the existence of a few RNA aptamers that bind exogenous targets, such as dopamine, why are there so few examples of synthetic bacterial riboswitches that respond to molecules other than theophylline? The answer to this is unclear but may be found in the SELEX experiments used to select these aptamers.

In vitro selected aptamers are selected in both chemical and physical environments that may be very different from the in vivo environments they would need to function in as part of a riboswitch. With this in mind, it may be beneficial to, at the very least, include a suitable expression platform (i.e. RBS and start codon) in the constant sequences located 3' to the randomized regions of oligonucleotides used in SELEX experiments. Doing so increases the likelihood of the ribosome binding site interacting with the RNA aptamer or small molecule in some fashion, perhaps facilitating the creation of a new switch. Some of the structural arrangements seen in the constant sequences of the erythromycin aptamer we have identified seem to support this idea. Additionally, the elucidation of the mechanism dictating the function of our theophylline riboswitch appears to support this notion and is likely to be general enough such that similar mechanisms could drive the function of new switches.

The choice of small-molecule targets will also be very important. Permeability and toxicity should be well understood for a specific target. An aptamer that binds the compound tetramethylrosamine has previously been reported yet its high toxicity severely hampers efforts to create a riboswitch that responds to it. Additionally, intracellular concentrations of a small molecule must approach that of the K_d for its aptamer counterpart. These concentrations can be difficult to determine but must be considered as an impermeable molecule will likely preclude the identification of a switch in vivo. If permeability is not an issue, then it should also be established whether or not a chosen small-molecule is metabolized by the bacteria in which the riboswitch is designed to function in. Again, this would make the chosen molecule an unsuitable target.

A well-planned approach to coupling the target molecule to a solid support should also be established to improve the successes of SELEX experiments. For our SELEX experiments involving erythromycin, we were fortunate to have Dr. Sam Reyes available to assist in the coupling of erythromycin to a sepharose column. His careful consideration in choosing strategies for capping the unmodified attachment sites on the sepharose column such that functional groups likely to interact with RNA non-specifically were not generated greatly improved out chances of success. Additionally, coupling strategies that require overly alkaline or acidic conditions can racemize or degrade target molecules prior to performing any selection experiments. Choosing very mild, near neutral "click chemistry" conditions in coupling erythromycin to sepharose helped to limit the amount degradation of our target molecule as the lactone contained within an erythromycin molecule can be readily hydrolyzed under acidic or

basic conditions. These types of ligand properties do not necessarily preclude it from being a suitable target, but should be considered when considering reaction conditions and choosing experiments. A molecule that is easily hydrolysable could result in the selection of an aptamer that binds something other than the intended target.

The availability of the target ligand is also a very important. This may prove to be troublesome in cases when a switch is being created to aid in the discovery of a biocatalyst capable of producing a molecule that is difficult to construct synthetically. From the selection of an aptamer to the screen and identification of a riboswitch, there is a significant amount of small molecule required for each step. Limited availability of a small molecule can limit the number of rounds of SELEX, screening steps and RNA structural studies thus impairing the discovery and characterization of new switches.

While the results of our SELEX experiments seem to indicate that we have selected an erythromycin-binding aptamer, more experiments must be performed to confirm this. Moving forward, it will be beneficial to explore novel methods for selecting aptamers that exploit the physical and biochemical properties afforded by small molecule-RNA interactions. For example, some have utilized allosteric ribozymes that induce self-cleavage upon ligand binding[5]. Could it also be possible to exploit the reverse phenomenon displayed by self-ligating ribozymes at low temperature[6]? With this strategy, circularized RNA would be insusceptible to 5' or 3' degradation and thus more apt to "survive" enzymatic breakdown.

If given a hypothetical aptamer that binds a cell permeable, nontoxic molecule, would it be possible to create a new riboswitch? Given the complexity of natural riboswitches and the unanticipated results observed in creating our original theophylline riboswitch, it is unlikely that the rational design of a riboswitch from this hypothetical aptamer will be the best approach to creating switches that either activate or repress gene expression in the presence of its effector. As such, combinatorial approaches involving genetic screens and selections for desired functions are potentially very powerful tools in these endeavors. The work presented here describes multiple screening methods that can be employed and will likely enhance efforts to both identify and improve the function of synthetic riboswitches. The large amount of data generated from the use of these screens to improve a theophylline riboswitch has allowed us to elucidate some principles that will also aid in the semi-rational design of new synthetic riboswitches and guide future aptamer selection experiments.

Efforts to create new synthetic riboswitches will continue to be driven by the potential power that they possess. The ability to detect specific chemical signals in vivo has a number of implications in the fields of enzyme engineering and synthetic biology. Sadly, the move from aptamer to riboswitch has proven to be a far more difficult than it once appeared. The work contained in this thesis will go a long way towards the identification of and improvement of synthetic riboswitches; however, the day in which one can choose a target molecule and rapidly create a synthetic riboswitch that responds to it will likely not be realized until methods for selecting RNA aptamers are improved.

8.2 References

(1)   Desai, S. K.; Gallivan, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 13247-13254.

(2)   Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P. *Chem. Biol.* **2007**, *14*, 173-184.

(3)   Nimjee, S. M.; Rusconi, C. P.; Sullenge, B. A. *Annu. Rev. Med.* **2005**, *56*, 555-583.

(4)   Grate, D.; Wilson, C. *Proc. Nat. Acad. Sci. U.S.A.* **1999**, *96*, 6131-6136.

(5)   Soukup, G. A.; Emilsson, G. A.; Breaker, R. R. *J. Mol. Biol.* **2000***, 298,* 623-632.

(6)   Kazakov, S. A.; Balatskya, S. V.; Johnston, B. H. *RNA*. **2006**, *12*, 446-456