In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing material of this type. I agree that permission to copy from, or to publish, this dissertation may be granted by the Professor under whose direction it was written, or in his absence, by the Dean of the Graduate School when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

_____

Shawn K. Desai

Engineering Control of Gene Expression in Bacteria Using RNA-Small Molecule Interactions

By

Shawn K. Desai
Doctor of Philosophy

Department of Chemistry

_____

Dr. Justin P. Gallivan
Advisor

_____

Dr. David G. Lynn
Committee Member

_____

Dr. Vincent P. Conticello
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

_____

Date

Engineering Control of Gene Expression in Bacteria Using RNA-Small Molecule Interactions

By

Shawn K. Desai
B.A., University of South Florida, 2001

Advisor: Justin P. Gallivan, Ph.D.

An Abstract of
A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

2008

Abstract

Small molecule-responsive riboswitches, which are composed entirely of RNA and control gene expression, have recently been discovered in a variety of organisms, including bacteria. These regulators have all the attributes of protein-based sensors, but are less complex and smaller than their protein counterparts. Because of their relatively simple design and broad recognition capabilities, many researchers are interested in constructing riboswitches in bacteria that respond to small molecules of their choosing. Once assembled, these engineered bacteria could be used for a variety of applications, such as sensing landmines, or for directing cancer-killing bacteria to cancerous cells.

In this thesis, we present our research directed, broadly, toward engineering small molecule-responsive riboswitches in bacteria. Chapter 2 describes our successful construction of a small molecule-responsive synthetic riboswitch in *Escherichia coli*, followed by a thorough characterization of its mechanism. After determining that it operates by activating translation, we used the riboswitch in both screens and selections for small molecules.

Chapter 3 presents the development and application of high-throughput screens and selections for synthetic riboswitches. We used these techniques to discover synthetic riboswitches in bacteria with outstanding activation ratios, and sequencing of the resulting variants allowed us to propose a potential model for their function. Several studies were undertaken to test the model, and it appears to be correct.

In Chapter 4, we explore the transferability of synthetic riboswitches between the similarly related organisms, *E. coli* and *Acinetobacter baylyi*. We chose *A. baylyi*

because it has a number of traits that *E. coli* does not have, such as natural competence, and we wished to enable the conditional control of gene expression in this potentially useful organism. We also adapted our previous high-throughput screen for use with *A. baylyi*, and used it to identify several synthetic riboswitches with large activation ratios.

Chapter 5 describes our efforts toward using a synthetic riboswitch-mediated auxotroph to functionally clone an unknown enzyme from *Coffea arabica* that catalyzes the rate-limiting step in caffeine catabolism.

Finally, Chapter 6 presents our attempts at constructing synthetic riboswitches that respond to small molecules involved in plant isoquinoline alkaloid biosynthesis.

Engineering Control of Gene Expression in Bacteria Using RNA-Small Molecule Interactions

By

Shawn K. Desai
B.A., University of South Florida, 2001

Advisor: Justin P. Gallivan, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

2008

Acknowledgements

Wow. I can't believe I'm actually writing this section. I guess I shouldn't be that surprised; I have been in graduate school for almost six years now. I have so many people to thank, but I don't have enough room for everyone. So, if you are not in here, don't get mad at me (you may be mentioned in one of the figures in this thesis, check and see).

The first person I need to thank is my advisor, Dr. Justin Gallivan. I have never gone hungry in the Gallivan lab, and I have never been denied the opportunity or equipment necessary to do any experiment. This is mainly due to Justin's prolific fundraising ability, and I am thankful to him for that. Besides the physical necessities, he has provided encouragement and has always pushed me to "think like a Ph.D." I think I finally understand what that means now, and it will serve me well the rest of my life. Thank you, Justin.

Thanks to my committee members, Dr. David Lynn and Dr. Vince Conticello. I have probably not taken advantage of their vast knowledge as much as I should have, but I am grateful that they have taken the time to attend my committee meetings and read my thesis. Thank you.

Over the years, I have worked with some wonderful people in the lab. Steven Dublin, Shana Topp, Sean Lynch, Karen Riesenburger, Joy Sinha, and Colleen Reynoso have provided me with hours of fun and friendly discussion, especially about politics. I never had to wonder if any of them would help me whenever I needed anything (anybody

want to help me move?), and I am thankful to them for that. I hope I have been as good a friend to them as they have been to me.

My Florida posse, Donald Exum and Chris Papp (it's a small posse), have been great friends throughout the years. I thank them for their friendship, and I am grateful that Donald will be my best man, and that Chris will be a groomsman in my wedding.

Special thanks to my parents, Phyllis and Kish, who have always been there for me. They have always encouraged me, and provided for my brother and I our entire lives (I still need plenty of encouragement-$, Dad). I couldn't ask for better parents.

Good luck to my brother, Kevin, who is currently studying for his Ph.D. at Florida State University. Get back to work.

Now on to the most important person I have met while in graduate school, Holly Carpenter. For those of you who don't know, Holly and I will be getting married soon, and I am thankful that I found such a wonderful person to spend my life with. She has always been supportive of me, and she is the most kind and caring person I have had the pleasure of knowing. I dedicate this thesis to Holly, the love of my life.

Contents

List of Figures

List of Tables

Chapter 1 Introduction: Natural and Synthetic Control of Gene Expression Using Small Molecules

1.1 Small Molecules Can Regulate Gene Expression

How do bacteria regulate complex, intertwined metabolic networks? From this question, a variety of seminal research questions sprang, and subsequent research based on this question provided the foundation for this thesis. This thesis is concerned with non-natural ways in which small molecules can specifically modulate gene expression in bacteria, but a thorough understanding of natural mechanisms is warranted.[1]

Jacob and Monod were the first to theorize that small molecules were capable of influencing individual genes within the bacterium *E. coli*. They were interested in a simple question: How do *E. coli* cells regulate the expression of the lactose-hydrolyzing enzyme β-galactosidase? They raised this question after observing a series of mutant *E. coli* strains that displayed either constitutive expression of β-galactosidase, or no expression of β-galactosidase.[2]

The authors believed, correctly, it would be an inefficient use of resources if *E. coli* constantly produced the metabolic enzymes needed for lactose metabolism, and they theorized that an unknown mechanism for regulating this process existed. Through a series of elegant experiments, they showed that β-galactosidase was produced when wild-type *E. coli* cells were grown with lactose as the sole carbon source.[2] What's more, when lactose was removed from the medium and replaced with an alternate carbon source, β-galactosidase production was attenuated. They isolated a genetic region (later known as the *lac*I gene and *lac* operator) that was responsible for this control. At this point, the authors hypothesized that bacteria contain regulatory mechanisms capable of turning genes on and off in response to specific environmental signals.[2] Therefore, bacteria were not disorganized, but were exquisitely controlled. Because of this leap in

understanding, Jacob and Monod were awarded the Nobel Prize for physiology or medicine in 1965.

Jacob, Monod, and others later revealed that the genes responsible for lactose metabolism were organized in an operon (Figure 1.1) that was regulated in a relatively straightforward manner.[4]

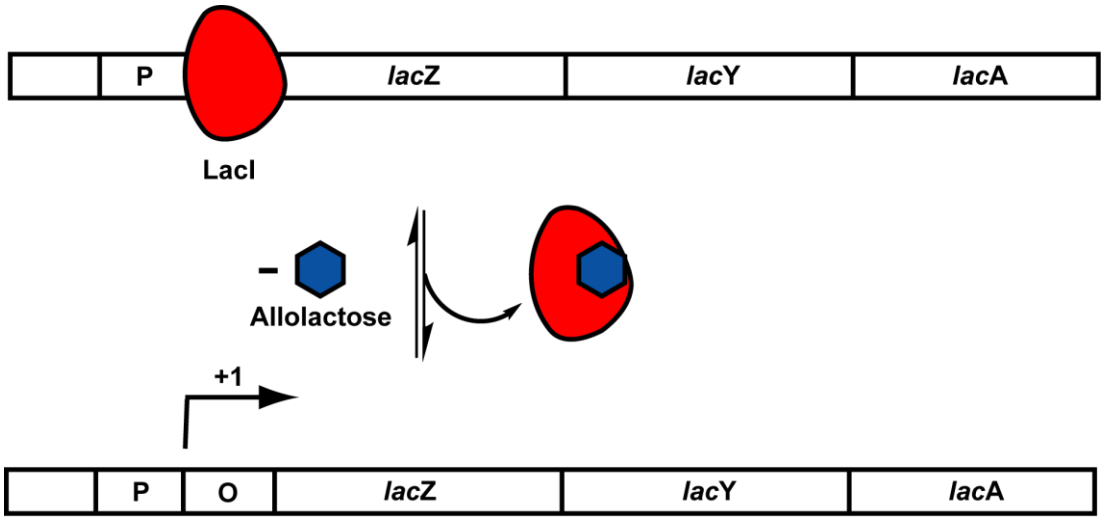

Figure 1.1: The *lac* Operon. The LacI protein binds to the *lac* operator sequence, in the absence of allolactose (formed from lactose), and prevents transcription of the downstream genes. When allolactose is present, the LacI protein binds the sugar and its affinity for the *lac* operator sequence decreases, which allows transcription. This is an abbreviated description of the system.

The LacI protein (also known as the lac repressor) is constitutively expressed from the *lac* operon. When *E. coli* cells are grown in the absence of lactose, the LacI protein binds to the *lac* operator DNA sequence. Binding of the LacI protein to the *lac* operator prevents transcription of the downstream genes. The DNA-bound LacI protein acts as a steric "roadblock" of transcription.[5]

When *E. coli* cells are grown in the presence of lactose, and the absence of glucose, the LacI protein binds to allolactose, a lactose metabolite, and releases its grip from the *lac* operator, and transcription of the downstream genes proceeds at a high level.

This process is slightly more complicated, due to the presence of the catabolite activated protein (CAP), which is involved in sensing cyclic-AMP.[4]

When *E. coli* cells are grown in the presence of lactose and glucose, transcription of the *lac* operon does not occur at a high level. This is due to the absence of the CAP-cAMP complex, which is not formed in the presence of glucose.[4] Thus, transcription of the *lac* operon occurs only when lactose is provided as the sole carbon source. It should be noted, the delineation of this pathway took many years and led to a fundamental shift in thinking about gene control.

1.1.1 The LacI Protein

In the *lac* operon, the LacI protein plays a central role. This protein is responsible for the allolactose-responsive regulation of the *lac* operon. An exploration of the details of this extraordinary small molecule sensor is presented below.

The LacI protein is composed of 360 amino acids, and forms a homotetramer. Each subunit has a molecular mass of approximately 37.5 kDa, giving a final molecular mass of approximately 150 kDa.[6] Within an *E. coli* cell, the concentration of the LacI protein is less than 0.002% of total protein.[4] This low concentration foreshadowed the high affinity of the repressor for the 17 base pair *lac* operator DNA sequence, which has been experimentally determined to have a dissociation constant of 10^{-13} M.[4] Such a high affinity is necessary to prevent dissociation of the repressor and DNA, which are at very low concentrations inside of *E. coli*. Upon binding allolactose, the affinity of the lac repressor for the *lac* operator sequence decreases by three orders of magnitude, from

approximately 10^{-13} M to 10^{-10} M.[4] At this point, RNA polymerase is free to bind to the *lac* operator sequence and initiate transcription of the downstream genes.

Over the years, researchers have identified a number of different genes in *E. coli*, and other bacteria, that are regulated by protein-small molecule interactions. Some of these pairs include: tetracycline-TetR,[7] arginine-ArgR,[8] and hypoxanthine-PurR.[9] These are a few of the known pairs, but many genes associated with the metabolism of small molecules are not regulated by protein factors, as we will discuss.

1.2 Riboswitches

1.2.1 Discovery

There was a conundrum in the field of bacterial genetics for many years; researchers were not able to identify the proteins responsible for regulating the expression of numerous small molecule biosynthetic genes, including adenosylcobalamin (Ado-Cbl), flavin mononucleotide (FMN), and thiamine pyrophosphate (TPP). It was possible that unknown proteins were expressed at such a low concentration that they were difficult to isolate. An intriguing hypothesis, but after the genomes of many bacteria were sequenced, no genes encoding potential protein regulators were found, though great effort was put forth to identify them.

Sequencing data, instead, revealed conserved regions within the 5' untranslated regions (5' UTR) of several small molecule biosynthetic genes. And these sequences were found in phylogenetically diverse species of bacteria, indicating they were most likely ancient in origin.[10] The conserved regions identified first were referred to as the *thi* box,[11] B_{12} box,[12] SAM box,[13] and *RFN* element,[10] and they were found upstream of

genes associated with thiamine pyrophosphate, Ado-Cbl, *S*-adenosylmethionine (SAM), and riboflavin metabolism, respectively. These regions were short in length, but were contained within longer 5' UTR sequences that ranged in length from approximately 100 to 300 base pairs.

Experiments indicated that the B_{12} box RNA was highly structured,[14] and mutant reporter constructs were made that unpaired a stem loop within this region.[15] Cells harboring these mutant constructs were not ligand responsive, but the loss of function could be compensated with mutations that reformed the stem loop, indicating the structure of the B_{12} box was important for physiological function.[15] Similar mutational analyses were conducted using the *thi* box, and it was shown that its structure was also important .[16]

In light of the structural evidence, several researchers theorized that Ado-Cbl bound directly to the 5' UTRs of the *cob* and *btu*B genes from *S. typhimurium* and *E. coli*, respectively, and that this binding altered gene expression.[17,18] Both of these 5' UTRs contain the B_{12} box, and in vitro transcribed RNA of the *cob* 5' UTR adopted different structures in the presence and absence of Ado-Cbl, as determined by dimethylsulfate probing.[17] Furthermore, the presence of Ado-Cbl prevented *E. coli* ribosomes from binding to the ribosome binding site (RBS) of the *btu*B gene, in vitro.[18] The inhibition of ribosome binding was strong evidence that Ado-Cbl controlled gene expression by altering RNA conformation, but this was not conclusive evidence of direct binding of Ado-Cbl to the RNA.

Further attempts were made to identify a direct interaction between Ado-Cbl and the 5' UTR of the *cob* gene. Equilibrium dialysis experiments were performed using radiolabeled Ado-[^{14}C]-Cbl and *cob* mRNA, but no shift in ligand was observed.[17]

The paper by Nahvi et al.[19] is generally credited as the first to conclusively demonstrate that the 5' UTR of the *btu*B gene from *E. coli* specifically binds Ado-Cbl, and that this binding event is responsible for changes in gene expression. The authors conducted in vitro structure probing on a 202 nucleotide RNA comprising the 5' UTR of the *btu*B gene, and they estimated an apparent K_d of 300 nM for Ado-Cbl, based on structural modulations that occurred when the RNA was incubated with increasing concentrations of the ligand.[19] No modulation of structure was observed when analogs of Ado-Cbl were used, suggesting the small molecule-RNA interaction was specific. Equilibrium dialysis using Ado-[^{3}H]-Cbl and the 202 nucleotide leader RNA was also performed, and the authors observed a shift in equilibrium of the radiolabeled compound, with a two-fold increase of radiolabeled compound into the chamber that contained RNA.[19] An equal distribution of ligand would have occurred if the RNA did not bind Ado-Cbl. It is difficult to reason why previous authors did not observe a similar effect when they conducted their equilibrium dialysis experiments, but it may have been due to preparation of their RNA, length of the RNA used, or other contaminating factors.

1.2.2 Genetic Control Mechanisms

After the initial confirmation of the Ado-Cbl riboswitch, a variety of small molecule sensitive riboswitches were discovered in both eukaryotes[20] and prokaryotes,[21] which we will focus on here. To date, the small molecules known to associate with

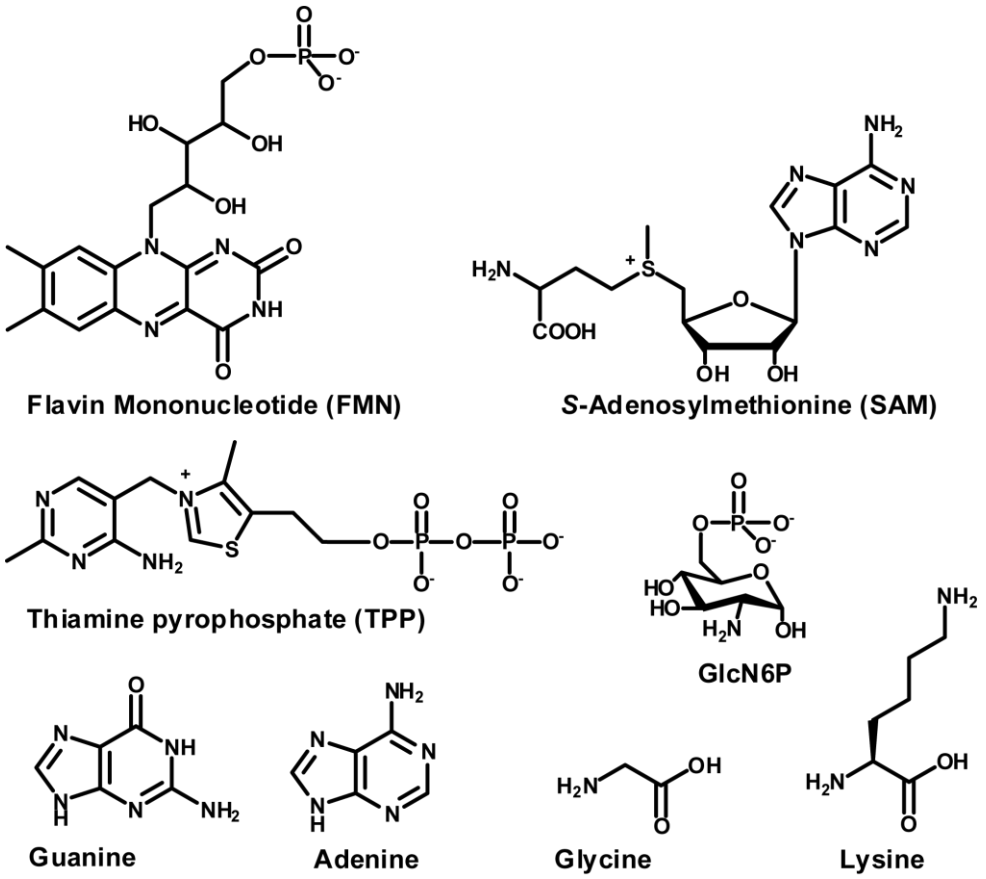

Figure 1.2: Natural Riboswitch Ligands.

riboswitches include: adenosylcobalamin,[20] adenine,[22] guanine,[22] glycine,[23] lysine,[24] flavin mononucleotide[25] (FMN), *S*-adenosylmethionine,[26] thiamine pyrophosphate,[27] glucosamine-6-phosphate[28] (GlcN6P), 2'-deoxyguanosine,[29] and 7-aminomethyl-7-deazaguanine[30] (Figure 1.2). In addition, a magnesium (Mg^{2+}) responsive riboswitch has also been identified[31]. The chemical diversity of these small molecules is impressive, and highlights the ligand recognition capability of RNA. It is also impressive to note the size ranges of the molecules listed; from Mg^{2+} (molecular weight 23.41) to glycine (molecular weight 75.07) to Ado-Cbl (molecular weight 1579.58).

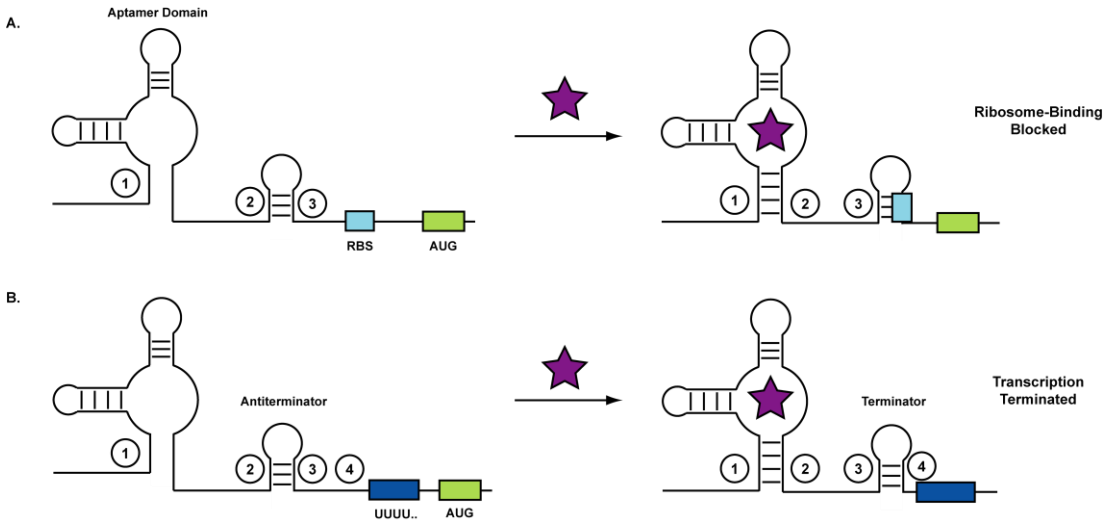

Figure 1.3: Riboswitch Control Mechanisms in Prokaryotes. Most riboswitches are composed of both an aptamer domain and an expression platform. (A) Translational control of gene expression. In the presence of ligand (star) the riboswitch pairs the ribosome-binding site, which prevents ribosome binding, and, thus, translation. (B) Transcriptional control of gene expression. In the presence of ligand, a terminator structure forms, which leads to premature truncation of the downstream mRNA. (Adapted from Gallivan 2007)

Most prokaryotic riboswitches are organized into two distinct domains: an aptamer domain and an expression platform[32] (Figure 1.3). The aptamer domain is responsible for small molecule recognition, and ligand binding drives conformational changes within the expression platform, leading to modulation of the downstream gene(s). It should be emphasized that the aptamer domains from riboswitches bind their cognate ligands with high affinity and specificity.[24,27] This discrimination is necessary in the cellular milieu, where metabolites often resemble each other in structure. Structural and mutational work with the aptamer domains of the adenine and guanine riboswitches,[3,33,34] two chemically similar molecules, demonstrates this specificity. Their aptamer domains are similar in sequence, and their crystal and NMR structures,[3,34] combined with mutational analyses,[33] revealed that a single nucleotide was responsible for discrimination amongst guanine and adenine, due to changes in hydrogen bonding with the respective small molecule.

Most prokaryotic riboswitches control the transcription or translation of their downstream gene(s),[35] and the majority of these riboswitches downregulate gene expression in response to small molecule binding[35] (Figure 1.3). Control of translation is carried out by simple means. In prokaryotes, when the RBS within an mRNA is paired, ribosome binding is blocked.[36] Simply unpairing the RBS allows the ribosome to bind the mRNA, and translation proceeds.[36] Transcription is also regulated by a simple mechanism. If a terminator hairpin is formed during synthesis of an mRNA, the RNA polymerase is displaced from the RNA, and a truncated mRNA is formed.[37] Alternately, if an anti-terminator hairpin is formed during synthesis of an mRNA, a full-length transcript is formed.[37] Some researchers theorize these two simple mechanisms of gene control predated those of protein based methods, and that they may have been first used in a primordial RNA world.[27]

Riboswitches do not always modulate transcription or translation of their downstream genes in a simple allosteric fashion. For example, the glycine riboswitch from *Vibrio cholerae* controls the glycine cleavage operon, *gcv*T, and contains two aptamer domains.[23] This riboswitch uses a cooperative mechanism to activate gene expression in response to glycine, whereby glycine binding by one aptamer domain leads to an approximately 100 to 1000 fold increase in affinity for glycine at the other aptamer domain. This method of gene control allows the cell to rapidly respond to small changes in glycine concentration. Cooperative control of gene expression with a riboswitch is reminiscent of hemoglobin's response to oxygen,[38] and illustrates that RNA is capable of functions previously associated only with proteins.

Tandem riboswitches further demonstrate the power of RNA based gene control. These riboswitches contain two aptamer domains, and each domain recognizes a different small molecule.[39] This molecular construction allows tandem riboswitches to respond as two-input Boolean NOR logic gates. In other words, only one of the small molecules needs to be present for the riboswitch to modulate gene expression. *Bacillus clausii* contains a tandem riboswitch upstream of the methionine biosynthesis gene, *met*E, and downregulates gene expression in response to either SAM or Ado-Cbl. But why are SAM and Ado-Cbl used to control a methionine biosynthesis gene? Either MetH or MetE can convert homocysteine to methionine, and methionine is a precursor to SAM. This explains why SAM regulates MetE expression, but what about Ado-Cbl? Methylcobalamin, a derivative of Ado-Cbl, is the cofactor for MetH, and MetH is more efficient than MetE. Therefore, the cell only produces the less efficient enzyme, MetE, when both Ado-Cbl and SAM are low.[39] This is a stunning example of the efficiency and complexity of Nature.

One additional class of riboswitches does not regulate either transcription or translation. The GlcN6P riboswitch, found in the 5' UTR of the GlcN6P synthetase (*glms*) gene in many species of bacteria, is both a riboswitch and a ribozyme, and acts through an autocatalytic mechanism.[28] When in the presence of GlcN6P, the ribozyme self cleaves within the 5' end of the RNA transcript, outside of the coding region, and this event leads to a decrease in gene expression. Researchers theorized that the 5' RNA cleavage resulted in a destabilized transcript in vivo.[28] This was found to be the case. In *Bacillus subtilis*, after cleavage of the *glms* transcript occurs, the transcript is rapidly degraded.[40] Further biochemical work indicated that an RNAse, specifically RNAse J1,

was responsible for the degradation of the cleaved transcript. Cleavage within the mRNA creates a new 5' end that carries a hydroxyl group in place of a phosphate group, and the hydroxyl group acts as a degradation signal for RNAse J1.[40]

1.2.3 Structural Characteristics

Crystal structures of members from the TPP,[41] guanine,[1] adenine,[3] SAM[42] and GlcN6P[43] riboswitch classes have been determined, and they further blur the line between the functions/structures usually ascribed to protein and RNA. In all of the riboswitch structures to date, binding of the metabolite is facilitated through a series of hydrogen bonds, with Mg^{2+} usually present in the binding site. There, the Mg^{2+} serves to stabilize the negatively charged phosphate groups of SAM, TPP, or GlcN6P. Before the publication of these structures, the architecture of these riboswitches was gleaned from biochemical data and presented in a series of two dimensional secondary structure pictures. However, the crystal structures of some of these riboswitches differ from those originally predicted.[27] Rather than discuss all the riboswitch crystal structures identified, we will discuss two different classes, as an illustration.

Researchers thought the TPP riboswitch contained a single small molecule binding pocket, like that of an in vitro selected aptamer, but its precise location was never determined.[27] Crystal structures of the TPP riboswitches from *E. coli* and *Arabidopsis thaliana* shed light on this problem.[44,45] Small molecule binding by the TPP riboswitch was found to occur through the use of two separate parallel helices within the riboswitch (Figure 1.4), and each helix recognizes a chemically distinct segment of TPP. One helix recognizes the pyrophosphate group, and one recognizes the terminal

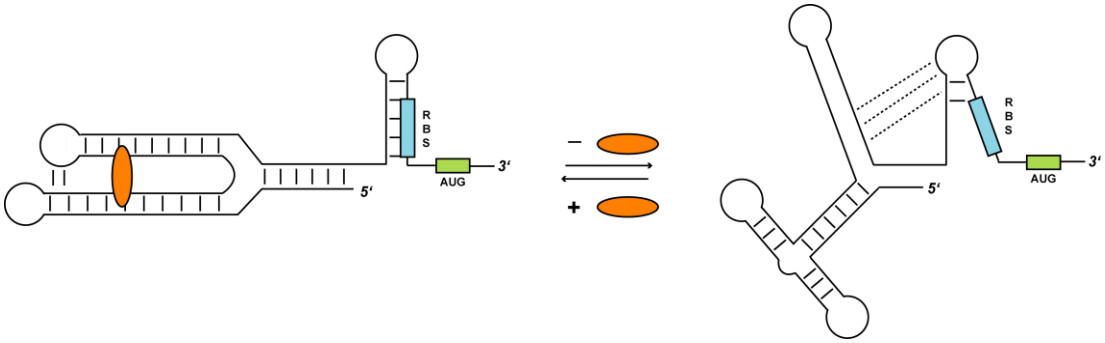

Figure 1.4: Mechanism of a Thiamine Pyrophosphate Riboswitch. The riboswitch shown is the *thiM* riboswitch from *E. coli*, which downregulates gene expression in the presence of thiamine pyrophosphate[3] (orange).

pyrimidine. Interestingly, the internal thiazole ring is not recognized by the RNA, but the thiazole ring does provide the necessary distance between the pyrophosphate and pyrimidine groups. In this way, the thiazole ring is recognized. The inability of the riboswitch to recognize the thiazole ring also explains why pyrithiamine pyrophosphate (PPP), a structurally similar molecule to TPP (Figure 1.5), acts as an antibacterial compound.[46] PPP was shown to kill bacterial cells by binding TPP riboswitches with high affinity, leading to a downregulation of TPP production, which "starves" the cell of the necessary cofactor, TPP.[46]

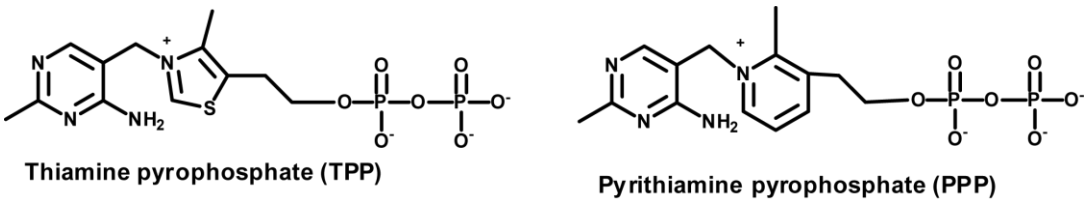

Figure 1.5: Antibacterial Analog (PPP) that Activates TPP Riboswitches.

The crystal structures of both the adenine and guanine sensing riboswitches from *Vibrio vulnificus* and *B. subtilis*, respectively, were determined recently by separate groups.[1,3] Both adopt a "tuning fork-like" structure, where the "forks" are helices, and the junction between the two helices serves as a binding pocket for the ligand

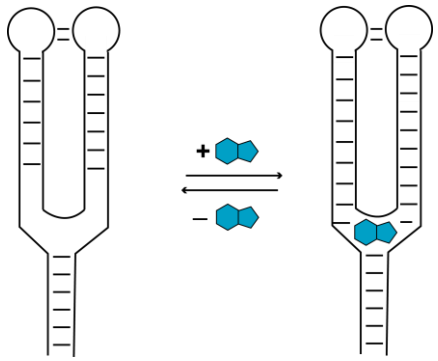

Figure 1.6: Aptamer Domain from a Guanine-Sensing Riboswitch. In the presence of adenine (blue), the aptamer domain of the guanine riboswitch, *xpt-pbu*X, from *B. subtilis* becomes ordered.[1]

(Figure 1.6). Without ligand, the binding pocket has a disordered structure. Upon ligand binding, the pocket becomes ordered, and the structure of the expression platform changes, leading to modulation of either transcription or translation. The crystal structures also showed that adenine or guanine is almost completely enveloped within the aptamer domain. This total enclosure of the ligand was unexpected. Many structures have been determined for in vitro selected small molecule-binding RNA aptamers,[47,48] and none of them sequester their ligands as thoroughly as the adenine or guanine riboswitches. Perhaps this should not be unexpected, as the ligands used for in vitro selection are often covalently attached through a linker arm to a solid support.[49] Thus, the ligands are never presented in their free form during the selections. Or, the data shows that Nature is far superior at selecting RNAs that recognize small molecules. We can only hope to approximate the capabilities of natural riboswitches in the lab.

1.3 Engineering Small Molecule Control of Gene Expression

1.3.1 Engineering Protein-Small Molecule Interactions to Control Gene Expression

Above, we reviewed a few of the natural mechanisms that cells use to control gene expression in response to small molecules. However, the small molecules that cells

are normally concerned with have physiological consequences; they are not just passive bystanders. What would happen if a cell (eukaryotic or prokaryotic) was incubated in media that contained a small molecule that was nontoxic and could not be metabolized? Most likely, nothing would occur. From a scientific and engineering perspective, this presents an interesting question that forms the basis of this thesis: *How can we alter a cell (specifically a bacterial cell) so that its gene expression will change, specifically, in response to an innocuous and orthogonal small molecule of our choosing*? Aside from being academically interesting, a number of applications in both biotechnology and medicine would benefit if this question was fully answered.

What are the cellular processes that can be altered to control gene expression in response to a specific small molecule of our choosing? For both prokaryotic and eukaryotic cells; transcription, translation, or protein splicing could be altered. For eukaryotic cells, the process of RNA splicing could also be altered. But how would the chosen small molecule modulate gene expression? Nature commonly uses protein-small molecule interactions to control gene expression, and we will focus on the manipulation of these interactions first.

The simplest approach to take advantage of protein-small molecule interactions to control gene expression is through the use of what I will refer to as "plug-and-play". In this method, existing genetic components that encode a small molecule-responsive gene regulator protein are moved from one organism to another. If all of the introduced components function within the host cell, and the small molecule is orthogonal to it, the host's gene expression will be newly responsive to the chosen small molecule. No

manipulation of the protein-small molecule interface is required with this method, but the choice of small molecule is limited.

One of the first examples of the "plug-and-play" method was achieved by the Sayler laboratory.[50] In their seminal work, they engineered *Pseudomonas fluorescens* so that it was bioluminscent in the presence of naphthalene. They constructed a plasmid that contained the bioluminescent reporter gene (*lux*) from *Vibrio fischeri* under the control of the salicylate responsive protein NahR, which was isolated from a rare strain of *P. fluorescens*. When *P. fluorescens* cells containing the plasmid were grown in the presence of naphthalene, bioluminescent light was produced in response to the production of salicylate, a catabolite of napthalene. Building on this concept, bacterial systems have been constructed that activate reporter gene expression in response to alkanes, benzene (and its derivatives), polychlorinated biphenyls, and chlorocatechols.[51] And each system took advantage of pre-existing genetic components; no alteration of the small molecule binding protein was necessary.

Separate groups have also applied the "plug-and-play" concept to whole organisms.[52,53] Both took advantage of the ecdysone responsive nuclear hormone (EcR) receptor from *Drosophilia*. Reporter genes under the control of the EcR were introduced into mice[52] and plants,[53] separately. Since ecdysone is not found in mice or plants, the authors were able to specifically activate reporter gene expression when ecdysone was provided to the organisms.

The naphthalene and ecdysone responsive systems above were transferred from prokaryote to prokaryote, or eukaryote to eukaryote, respectively. But this conservation is not always necessary. In an interesting example, the tetracycline repressor gene (*tet*)

from the *E. coli* tetracycline resistance operon was fused to a transcriptional activation domain from the herpes simplex virus.[54] The chimeric gene was introduced into mammalian cells and used to downregulate reporter gene expression in the presence of tetracycline. The system was later engineered to activate transcription in response to tetracycline.[55] This system illustrates that the genetic components used in "plug-and-play" systems need not be derived from organisms within the same kingdom. And, if additional small molecule-responsive gene regulator proteins are identified, the orthogonal "plug-and-play" system will be more widely used.

It is convenient to transfer the genetic components encoding naturally occurring small molecule-responsive gene regulator proteins from one cell, or organism, to another. But what if one wishes to control cellular gene expression in response to a small molecule for which there is no cognate gene regulator protein? This can be accomplished through the rational remodeling of a gene regulator protein, but rational design can be problematic, and has not been successfully used in many instances. As an illustration of why rational design is difficult, a single 100 amino acid polypeptide has 20^{100} possible amino acid sequences, and targeted mutagenesis strategies will only sample a minute number of these combinations. However, there have been a few examples where rational design has been successful.

Polar group-exchange has been successfully used to rationally remodel a small molecule-responsive gene regulator protein. In the polar-group exchange method, a charged amino acid group involved in binding the protein's cognate ligand is mutated to an uncharged amino acid.[56] This (sometimes) changes the small molecule binding specificity of the receptor protein. Using a crystal structure as a guide, Shi and Koh

mutated a single glutamate to an alanine, in the ligand binding pocket of a nuclear hormone receptor. This single change modified the receptor's ligand binding specificity from estradiol to a charged estradiol analog.[56] The modified nuclear hormone receptor was introduced into mammalian cells, and gene expression was newly responsive to the estradiol analog. One of the drawbacks with polar-group exchange is that it requires a crystal structure of the receptor protein, and the newly recognized small molecule cannot be substantially different in structure from the wild-type ligand. These drawbacks have limited the use of this method, and others like it (i.e. the "bump and hole" approach[57]).

The rational design of proteins is a slow process, and it is often difficult to predict how the selected mutations will affect the global structure of the modified receptor protein. Computational redesign may circumvent these problems, as it is relatively fast, can test multiple combinations of mutations, and can potentially predict the global fold of an altered protein.[58]

In an effort to take advantage of these possibilities, the Hellinga group developed a computer program that predicts how to remodel the small molecule binding pocket of gene regulator proteins.[59] In their initial work, they modified sugar-binding proteins to recognize trinitro-toluene (TNT), l-lactate, or serotonin,[60] which are chemically distinct from the wild-type ligands. Several of the modified proteins bound their designated ligand with high affinity, and one TNT-binding variant was measured to have a K_d of 2 nM.[60] The altered receptor proteins were introduced, separately, into *E. coli* and were used to control reporter gene expression in a small molecule-dependent manner, demonstrating the utility of their approach. Subsequently, the Hellinga group designed sugar binding proteins that control gene expression in response to Zn^{2+}.[61] While these

results are impressive, the Hellinga group's data should be viewed with a critical eye, in light of their recently retracted papers,[62,63] coupled with the fact that their computer programs are not publicly released.

Directed evolution is an alternative approach to computational and rational protein design, and allows the sampling of more actual variants than is possible with either method. In directed evolution, a library of mutant proteins is created in vitro, and these mutants are then screened or selected for function in vivo.[64] If the screen or selection is judiciously created, libraries of greater than 10^9 variants can be sampled.[65]

The Hillen group used directed evolution to evolve the tetracycline repressor protein (TetR) so that it recognized a new inducer molecule.[66] They created a library of mutant *tet*R genes using DNA shuffling, combined with mutagenesis, and subjected the library to four rounds of screening in *E. coli*. From their screen, they identified several variants that activated gene expression in the presence of a tetracycline analog, (4-de(dimethylamino)-6-demethyl-6-deoxy-tc), and these variants were not activated with tetracycline.[66] Sequencing revealed that their two best variants contained three mutations, and the third best variant contained five mutations, and many of the mutations were located outside of the tetracycline binding site.[66] It is doubtful rational design would have led to the numerous mutations identified, and, in fact, it may not have been possible to identify them using computational design.

Workers from the Arnold group also used directed evolution to modify the small molecule binding properties of the quorum sensing protein LuxR, which activates transcription in response to a specific homoserine lactone, with a dual positive-negative selection system.[67] The authors used *E. coli* to harbor randomized *lux*R libraries, and

identified variants that activated a chloramphenicol resistance gene in the presence of the desired small molecule; these survived. From these variants, a second selection was performed to isolate clones that did not activate gene expression in the absence of the chosen small molecule. The mutant library in the second selection was used to control the expression of a β-lactamase inhibitory protein, and it was grown in the absence of the chosen small molecule, and in the presence of carbenicillin. Variants active in the absence of the chosen small molecule were sensitive to carbenicillin, and did not survive. Using their selection strategy, the authors identified a LuxR variant that activated gene expression in response to straight chain acyl-homoserine lactones, and it was not responsive to the wild-type homoserine lactone.[67] Interestingly, a single mutation in LuxR was responsible for this effect.

In addition to the prokaryotic examples mentioned, separate groups have used directed evolution to alter the small molecule binding specificity of eukaryotic nuclear hormone receptors. In the Doyle group, a library of randomized nuclear hormone receptor genes, controlling an adenine synthesis gene, was introduced into a yeast strain deficient in adenine production.[68] Yeast containing the library were grown in the absence of adenine, and in the presence of a structural variant of the receptor's native ligand. Only nuclear hormone receptors that activated gene expression in response to the new ligand produced adenine, and, thus, survived. The authors refer to this type of system as "chemical complementation" because the cell's life is dependent on the chosen small molecule.[68] The Zhao group also modified the small molecule binding pocket of a nuclear hormone receptor, but they used a screen instead of a selection to identify

functional variants.[69] In both cases, the mutant nuclear hormone receptors activated gene expression in vivo in response to the desired ligand.

Gene regulator proteins that are responsive to small molecules serve as attractive targets for modification. They have pre-existing binding pockets, are easily manipulated using standard laboratory techniques, and are usually well studied. But in all the research surveyed above, the modified proteins recognize small molecules structurally similar to those recognized by the wild-type protein (except in the case of the Hellinga group). Therefore, with these techniques, the ability to control gene expression with a small molecule of one's choosing is limited by the small molecule binding properties of the parent protein. Is it possible to skirt this limitation?

Parent Gene Regulator Protein	**Wild-type Ligand**	**Ligand Recognized after Modification**	**Modification Method**	**Ref.**
human estrogen receptor	estradiol	carboxylate-functionalized estradiol	rational design	56
ribose-binding protein	ribose	trinitrotoluene	computation	60
glucose-binding protein	glucose	l-lactate	computation	60
arabinose-binding protein	arabinose	serotonin	computation	60
ribose-binding protein	ribose	Zn^{2+}	computation	61
tetracycline repressor	tetracycline	tetracycline analog (cmt3)	directed evolution	66
LuxR	3-oxo-hexanoyl-homoserine lactone	octanoyl-homoserine lactone	directed evolution	67
retinoid X receptor	9-*cis* retinoic acid	synthetic retinoid (LG335)	directed evolution	68
human estrogen receptor	estradiol	4,4'-dihydroxybenzil	directed evolution	60

Table 1.1: Alteration of Small Molecule-Responsive Gene Regulator Proteins. Proteins altered to recognize different ligands, and the method of modification used. These variants were used to control gene expression in vivo.

1.3.2 Engineering RNA-Small Molecule Interactions to Control Gene Expression

Riboswitches take advantage of the binding energy associated with RNA-small molecule interactions to drive changes within their expression platforms, leading to changes in gene expression.[35] And, as stated earlier, riboswitches can recognize a number of structurally diverse molecules.[35] These findings raise a simple question: Is it possible to engineer riboswitches that respond to a small molecule of our choosing? To achieve this goal, we need to construct both parts associated with a typical riboswitch: the RNA aptamer domain and the expression platform. Let us focus on the aptamer domain first.

Years prior to the discovery of riboswitches, researchers developed a method known as Systematic Evolution of Ligands by EXponential enrichment (SELEX).[70-72] With SELEX, in vitro selection is used to isolate short, small molecule or protein-binding RNAs or DNAs from large random libraries[49] (Figure 1.7). And the size of the random library can contain up to 10^{15} variants,[49] which is five to six orders of magnitude larger than the largest library possible with in vivo selections.[73] The RNAs or DNAs isolated from these selections are known as aptamers, from the latin 'aptus' which means 'to fit'.[72]

Typically, in RNA aptamer selections, a DNA oligonucleotide that contains two constant sequences, flanking a randomized region of between ten and sixty nucleotides, is synthesized, and one of the constant regions contains a promoter sequence for the phage T7 RNA polymerase.[49] The randomized oligonucleotide is made double stranded through the use of PCR or reverse transcriptase, and is then transcribed in vitro. The transcribed RNA is then isolated and applied to a column containing either a small

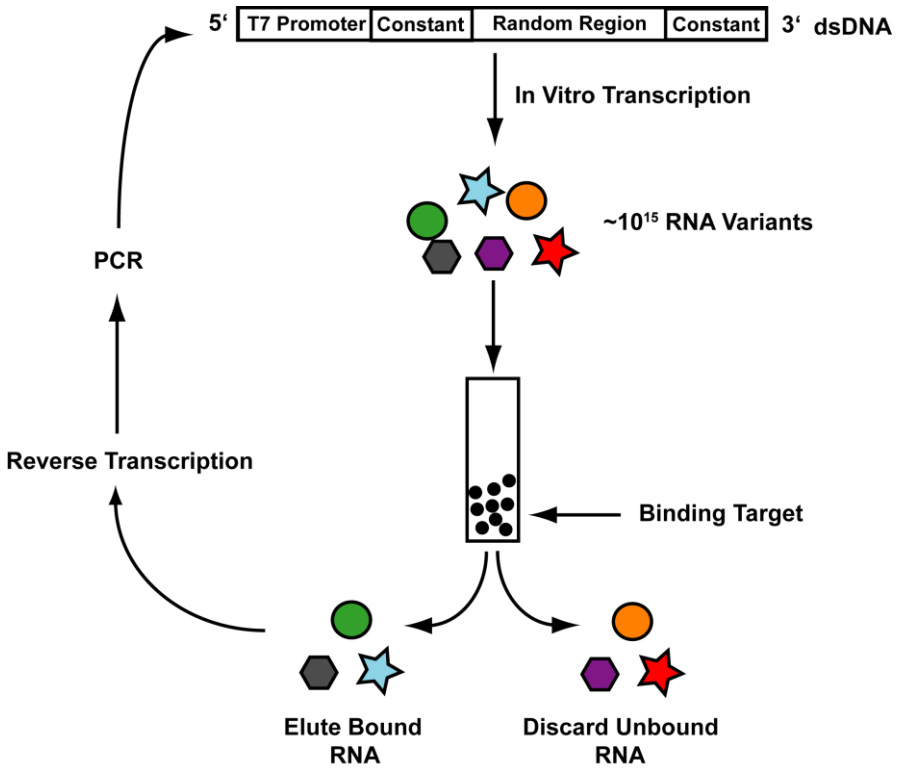

Figure 1.7: SELEX Schematic. Binding RNAs are selected from a large library, with a series of partitioning and amplification steps.

molecule bound to a solid support, or a protein. Loosely bound variants are washed away, and bound RNAs are eluted using either free small molecule or a denaturing agent, such as urea. Finally, the eluted RNA is reverse transcribed, followed by PCR amplification. These steps constitute what is referred to as a ‘round’ of SELEX,[49,70] and multiple rounds are commonly performed during a single experiment.

RNA aptamers have been raised to bind a variety of small molecules (Figure 1.8). A truncated listing of these includes: cyanocobalamin,[74] flavin mononucleotide,[75] β-nicotinamide mononucleotide,[76] arginine,[77] guanine,[78] ATP,[79] biotin,[80] dopamine,[81] tetramethylrosamine,[82] tetracycline[83] and theophylline.[84] Many of these RNA aptamers bind their cognate ligands with high affinity and specificity. In fact, one theophylline aptamer binds theophylline with a K_d of 100 nM, and has a 10,000 fold lower affinity for

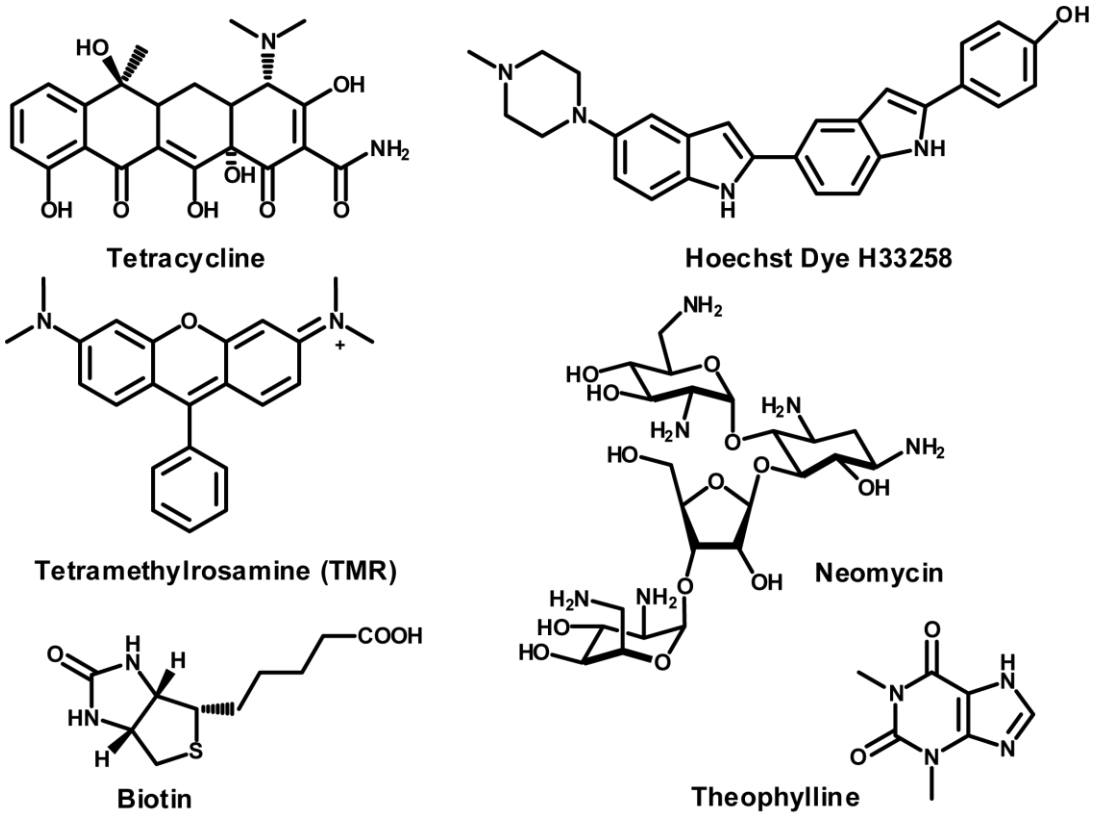

Figure 1.8: RNA Aptamer Targets. Aptamers to the ligands shown have been used to control gene expression in living organisms.

caffeine,[84] which differs from theophylline by only one methyl group. Thus, in vitro selected RNA aptamers have recognition capabilities that approach those of the aptamer domains from natural riboswitches, and, therefore, may serve as potential sensors for applications involving the in vivo control of gene expression.

Werstuck and Green were the first to use RNA-small interactions to control gene expression in a living organism.[85] In eukaryotic cells, the ribosome scans mRNAs from the 5' to 3' end until a start codon is reached, then translation begins.[86] However, an increase in secondary structure within the 5' UTR of eukaryotic genes leads to a decrease in translation, due to inhibition of ribosome scanning.[87] The researchers hypothesized that by inserting a small molecule-binding RNA aptamer into the 5' UTR of a reporter

gene, they could control gene expression in a small molecule dependent fashion. They believed the small molecule would bind to the RNA aptamer, increase its secondary structure, and inhibit translation of the downstream gene.[85]

The authors raised RNA aptamers to a cell-permeable Hoechst dye, inserted a single copy of one of the aptamers into the 5' UTR of a β-galactosidase reporter gene construct, and transfected Chinese hamster ovary cells with it.[85] Cells containing the reporter construct were grown in the presence and absence of different concentrations of the Hoechst dye, and β-galactosidase activity was subsequently measured. In the presence of the Hoechst dye, reporter gene expression decreased in a small molecule dependent fashion;[85] demonstrating that a small molecule of one's choosing can be used to control gene expression using RNA-small molecule interactions.

The approach of Werstuck and Green is superior to protein-based methods for engineering small molecule control of gene expression for a number of reasons. First, no pre-existing protein scaffold was required in their method, and, second, the ligand-binding domain was selected de novo, which cannot be done, currently, with proteins. As an aside, the Hoechst dye aptamer used by Werstuck and Green has never been used by any other group, indicating it may not function as described, but their concept has held up to scrutiny.

Building on the intial work of Werstuck and Green, several groups have used a variety of additional RNA-small molecule interactions to downregulate translation in eukaryotic systems.[88-90] Harvey et al. used a theophylline and biotin aptamer, separately, to control reporter gene expression in *Xenopus oocytes* and wheat germ extract in a small molecule dependent fashion.[89] Workers from the Suess group used a tetracycline

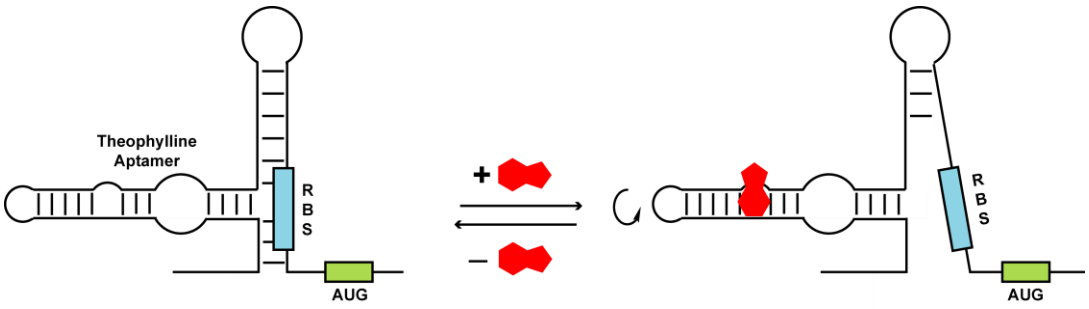

Figure 1.9: Mechanism of a Theophylline-Activated Riboswitch. Workers from the Suess group designed a theophylline-activated riboswitch in *B. subtilis* based on helix slippage. In the absence of theophylline, the RBS is paired, and gene expression is inhibited. However, when theophylline is present (red), it binds to the aptamer and causes a slippage of the paired RBS by one base, resulting in an unpaired RBS and translation begins.

aptamer to modulate translation of a reporter gene in yeast in a tetracycline dependent manner,[88] and they subsequently showed that the thermodynamics associated with the small molecule-aptamer interaction were responsible for inhibiting translation of the reporter genes;[91] just as Werstuck and Green theorized.

Other groups have modified Werstuck and Green's concept and applied it to controlling gene expression in bacteria.[92-94] As mentioned in the discussion of natural riboswitches found in bacteria, simply pairing the RBS of a gene decreases its translation, and unpairing the RBS increases translation.[35,36] The Suess group designed a theophylline-responsive riboswitch in *B. subtilis* using these principles[92] (Figure 1.9). In their work, they designed a hairpin within the 5' UTR of the *xyl*R gene (a repressor protein) that paired the RBS. They then coupled a theophylline aptamer to the hairpin, through a previously identified 'communication module' sequence.[95] They theorized that theophylline binding would alter the aptamer's structure and cause the hairpin to slip by a single nucleotide, exposing the RBS, and allowing translation.[92] Members of the Suess group introduced the designed riboswitch construct into *B. subtilis*, along with a β-galactosidase reporter construct that was downregulated by the XylR protein. The cells

were grown with increasing concentrations of theophylline, and β-galactosidase expression decreased in parallel, due to an increase in translation of the *xyl*R gene.[92] The authors achieved a maximum regulatory factor of 8.8, when the cells were grown with 6 mM theophylline. While impressive, this regulatory factor has been surpassed by synthetic riboswitches discovered in our lab.[93]

Concurrently with the Suess group, our lab constructed a theophylline-responsive riboswitch in *E. coli*.[94] We simply inserted a theophylline aptamer into the 5' UTR of a β-galactosidase reporter gene, and introduced the construct into *E. coli*. When cells containing the construct were grown in the presence of theophylline, we observed an increase in reporter gene expression. Further studies showed this increase was dose dependent, and could be used to perform genetic selections for theophylline. This work will be discussed in greater detail in Chapter 2.

In the examples above, a rational strategy was used to construct the RNA-based genetic switches; directed evolution can also be useful in this realm. Workers from the Liu lab used directed evolution to convert a previously selected tetramethylrosamine (TMR) aptamer[82] into a transcriptional activator in yeast.[96] Using common laboratory techniques, they connected the TMR aptamer to an RNA-based transcriptional activator[97] through a seven base pair randomized linker. They cloned the randomized library upstream of a histidine biosynthesis gene, and introduced the library into a histidine deficient strain of yeast. Yeast containing the library were grown in the absence of histidine, and in the presence of TMR. Only clones that activated transcription of the histidine biosynthesis gene survived.[96] From their selection, many clones survived. But, from these, only four variants were TMR responsive. Their best clone activated gene

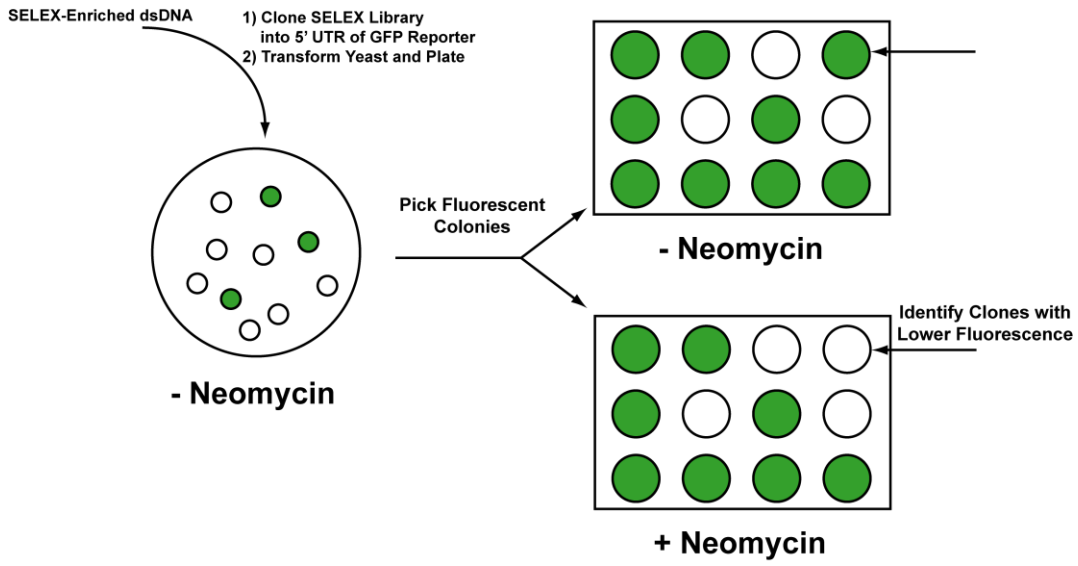

Figure 1.10: In vitro/In vivo Combination to Discover Synthetic Riboswitches. A SELEX-enriched PCR product is cloned upstream of a GFP reporter gene, transformed into yeast, and grown in the absence of neomycin. Fluorescent clones are picked and used to inoculate microtiter plates supplemented with and without neomycin. After growth, clones which display lower fluorescence in the presence of neomycin are selected. Successful variants are analyzed by sequencing.

expression 10.3 fold over the basal level, when grown with 10 μM TMR.[96] Sequencing revealed the identity of the linker nucleotides, and further tests showed these linker nucleotides were important for the TMR-dependent phenotype, demonstrating that insights could be gained from the selected sequences.

In addition to the eukaryotic example mentioned above, members of our laboratory used directed evolution to identify theophylline riboswitches with outstanding activation ratios in *E. coli*.[93] Using a simple screening method, we identified variants that had activation ratios of up to 36 fold, when grown in the presence of 1 mM theophylline. Sequencing of the variants also allowed us to propose a potential mechanism for our switches. This research will be discussed further in Chapter 3.

Combining the concepts of rational design and directed evolution, Suess and coworkers produced neomycin-sensitive riboswitches in yeast[98]. First, they performed six rounds of SELEX for the antibiotic neomycin (Figure 1.10). They cloned the

enriched library into the 5' UTR of a green fluorescent protein reporter gene, transformed it into yeast, and plated the cells in the absence of neomycin. Fluorescent colonies were used to inoculate duplicate 96-well plates containing growth media. One set of plates contained neomycin in the media, and one set did not. Thirty clones that showed decreased fluorescence in the presence of neomycin were tested further, but the tests revealed many of these clones were only mildly affected. After manipulation of their best clone, they identified a variant that exhibited a 7.5 fold decrease in fluorescence, when grown with 100 μM neomycin[98]. Their results demonstrate the power of combining in vitro and in vivo techniques, in addition to demonstrating how creativity can lead to solutions in this field of research.

1.4 Conclusion

It has been over 40 years since Jacob and Monod made their fundamental discovery that small molecules can control gene expression. Since their discovery, researchers have identified a variety of systems that Nature uses to control gene expression in a small molecule dependent fashion. And while discoveries of these natural systems were underway, many researchers pushed past the natural boundaries and successfully learned how to control gene expression in response to small molecules of their choosing. Research in this field is progressing at an exponential rate, and, surely, many additional discoveries will be published in short order.

1.5 References

(1) Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, *432*, 411-415.

(2) Jacob, F.; Monod, J. *J. Mol. Biol.* **1961**, *3*, 318-356.

(3) 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*, 1729-1741.

(4) Lewis, M. *Cr. Biol.* **2005**, *328*, 521-548.

(5) Wilson, C. J.; Zhan, H.; Swint-Kruse, L.; Matthews, K. S. *Cell. Mol. Life Sci.* **2007**, *64*, 3-16.

(6) Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P. Z. *Science* **1996**, *271*, 1247-1254.

(7) Saenger, W.; Orth, P.; Kisker, C.; Hillen, W.; Hinrichs, W. *Angew. Chem. Int. Ed.* **2000**, *39*, 2042-2052.

(8) Maas, W. K. *Microbiol. Rev.* **1994**, *58*, 631-640.

(9) Choi, K. Y.; Zalkin, H. *J. Bacteriol.* **1994**, *176*, 1767-1772.

(10) Gelfand, M. S.; Mironov, A. A.; Jomantas, J.; Kozlov, Y. I.; Perumov, D. A. *Trends Gen.* **1999**, *15*, 439-442.

(11) Miranda-Rios, J.; Morera, C.; Taboada, H.; Davalos, A.; Encarnacion, S.; Mora, J.; Soberon, M. *J. Bacteriol.* **1997**, *179*, 6887-6893.

(12) Lundrigan, M. D.; Koster, W.; Kadner, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1479-1483.

(13) Grundy, F. J.; Henkin, T. M. *Mol. Microbiol.* **1998**, *30*, 737-749.

(14) Richterdahlfors, A. A.; Andersson, D. I. *Mol. Microbiol.* **1992**, *6*, 743-749.

(15) Ravnum, S.; Andersson, D. I. *Mol. Microbiol.* **1997**, *23*, 35-42.

(16) Miranda-Rios, J.; Navarro, M.; Soberon, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9736-9741.

(17) Ravnum, S.; Andersson, D. I. *Mol. Microbiol.* **2001**, *39*, 1585-1594.

(18) Nou, X. W.; Kadner, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7190-7195.

(19) Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R. *Chem. Biol.* **2002**, *9*, 1043-1049.

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

(21) Tucker, B. J.; Breaker, R. R. *Curr. Opin. Struc. Biol.* **2005**, *15*, 342-348.

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

(23) Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R. *Science* **2004**, *306*, 275-279.

(24) Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R. *Gene Dev.* **2003**, *17*, 2688-2697.

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

(26) Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R. *Nat. Struct. Biol.* **2003**, *10*, 701-707.

(27) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952-956.

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

(29) Kim, J. N.; Roth, A.; Breaker, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16092-16097.

(30) Roth, A.; Winkler, W. C.; Regulski, E. E.; Lee, B. W. K.; Lim, J.; Jona, I.; Barrick, J. E.; Ritwik, A.; Kim, J. N.; Welz, R.; Iwata-Reuyl, D.; Breaker, R. R. *Nat. Struct. Mol. Biol.* **2007**, *14*, 308-317.

(31) Dann, C. E.; Wakeman, C. A.; Sieling, C. L.; Baker, S. C.; Irnov, I.; Winkler, W. C. *Cell* **2007**, *130*, 878-892.

(32) Winkler, W. C. *Arch. Microbiol.* **2005**, *183*, 151-159.

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

(34) Noeske, J.; Richter, C.; Grundl, M. A.; Nasiri, H. R.; Schwalbe, H.; Wohnert, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1372-1377.

(35) Mandal, M.; Breaker, R. R. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451-463.

(36) Gold, L. *Annu. Rev. Biochem.* **1988**, *57*, 199-233.

(37) Yanofsky, C. *J. Bacteriol.* **2000**, *182*, 1-8.

(38) Edelstein, S. J. *Annu. Rev. Biochem.* **1975**, *44*, 209-232.

(39) Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R. *Science* **2006**, *314*, 300-304.

(40) Collins, J. A.; Irnov, I.; Baker, S.; Winkler, W. C. *Genes Dev.* **2007**, *21*, 3356-3368.

(41) Serganov, A.; Polonskaia, A.; Phan, A. T.; Breaker, R. R.; Patel, D. J. *Nature* **2006**, *441*, 1167-1171.

(42) Montange, R. K.; Batey, R. T. *Nature* **2006**, *441*, 1172-1175.

(43) Klein, D. J.; Ferre-D'Amare, A. R. *Science* **2006**, *313*, 1752-1756.

(44) Edwards, T. E.; Ferre-D'Amare, A. R. *Structure* **2006**, *14*, 1459-1468.

(45) Thore, S.; Leibundgut, M.; Ban, N. N. *Science* **2006**, *312*, 1208-1211.

(46) Sudarsan, N.; Cohen-Chalamish, S.; Nakamura, S.; Emilsson, G. M.; Breaker, R. R. *Chem. Biol.* **2005**, *12*, 1325-1335.

(47) Flinders, J.; DeFina, S. C.; Brackett, D. M.; Baugh, C.; Wilson, C.; Dieckmann, T. *Chembiochem* **2004**, *5*, 62-72.

(48) Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J. P.; Pardi, A. *Nat. Struct. Biol.* **1997**, *4*, 644-649.

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

(50) King, J. M. H.; Digrazia, P. M.; Applegate, B.; Burlage, R.; Sanseverino, J.; Dunbar, P.; Larimer, F.; Sayler, G. S. *Science* **1990**, *249*, 778-781.

(51) Daunert, S.; Barrett, G.; Feliciano, J. S.; Shetty, R. S.; Shrestha, S.; Smith-Spencer, W. *Chem. Rev.* **2000**, *100*, 2705-2738.

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

(53) Martinez, A.; Sparks, C.; Drayton, P.; Thompson, J.; Greenland, A.; Jepson, I. *Mol. Gen. Genet.* **1999**, *261*, 546-552.

(54) Gossen, M.; Bujard, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5547-5551.

(55) Gossen, M.; Freundlieb, S.; Bender, G.; Muller, G.; Hillen, W.; Bujard, H. *Science* **1995**, *268*, 1766-1769.

(56) Shi, Y. H.; Koh, J. T. *Chem. Biol.* **2001**, *8*, 501-510.

(57) Belshaw, P. J.; Schoepfer, J. G.; Liu, K. Q.; Morrison, K. L.; Schreiber, S. L. *Angew. Chem. Int. Edit.* **1995**, *34*, 2129-2132.

(58) Baker, D. *Philos. T. Roy. Soc. B* **2006**, *361*, 459-463.

(59) Looger, L. L.; Hellinga, H. W. *J. Mol. Biol.* **2001**, *307*, 429-445.

(60) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185-190.

(61) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11255-11260.

(62) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Science* **2008**, *319*, 569-569.

(63) Allert, M.; Dwyer, M. A.; Hellinga, H. W. *J. Mol. Biol.* **2007**, *366*, 945-953.

(64) Lin, H.; Cornish, V. W. *Angew. Chem. Int. Ed.* **2002**, *41*, 4402-4425.

(65) Woycechowsky, K. J.; Hilvert, D. *Eur. J. Biochem.* **2004**, *271*, 1630-1637.

(66) Scholz, O.; Kostner, M.; Reich, M.; Gastiger, S.; Hillen, W. *J. Mol. Biol.* **2003**, *329*, 217-227.

(67) Collins, C. H.; Leadbetter, J. R.; Arnold, F. H. *Nat. Biotechnol.* **2006**, *24*, 708-712.

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

(69) Chockalingam, K.; Chen, Z. L.; Katzenellenbogen, J. A.; Zhao, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5691-5696.

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

(71) Ellington, A. D.; Szostak, J. W. *Nature* **1992**, *355*, 850-852.

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

(73) Taylor, S. V.; Kast, P.; Hilvert, D. *Angew. Chem. Int. Ed.* **2001**, *40*, 3310-3335.

(74) Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33*, 973-982.

(75) Burgstaller, P.; Famulok, M. *Angew. Chem. Int. Edit.* **1994**, *33*, 1084-1087.

(76) Lauhon, C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 1246-1257.

(77) Geiger, A.; Burgstaller, P.; vonder Eltz, H.; Roeder, A.; Famulok, M. *Nucleic Acids Res.* **1996**, *24*, 1029-1036.

(78) Kiga, D.; Futamura, Y.; Sakamoto, K.; Yokoyama, S. *Nucleic Acids Res.* **1998**, *26*, 1755-1760.

(79) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550-553.

(80) Wilson, C.; Nix, J.; Szostak, J. *Biochemistry* **1998**, *37*, 14410-14419.

(81) Mannironi, C.; DiNardo, A.; Fruscoloni, P.; Tocchini-Valentini, G. P. *Biochemistry* **1997**, *36*, 9726-9734.

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

(83) Berens, C.; Thain, A.; Schroeder, R. *Bioorgan. Med. Chem.* **2001**, *9*, 2549-2556.

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

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

(86) Kozak, M. *Gene* **1999**, *234*, 187-208.

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

(88) Suess, B.; Hanson, S.; Berens, C.; Fink, B.; Schroeder, R.; Hillen, W. *Nucleic Acids Res.* **2003**, *31*, 1853-1858.

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

(90) Grate, D.; Wilson, C. *Bioorg. Med. Chem.* **2001**, *9*, 2565-2570.

(91) Muller, M.; Weigand, J. E.; Weichenrieder, O.; Suess, B. *Nucleic Acids Res.* **2006**, *34*, 2607-2617.

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

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

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

(95) Soukup, G. A.; Breaker, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3584-3589.

(96) Buskirk, A. R.; Landrigan, A.; Liu, D. R. *Chem. Biol.* **2004**, *11*, 1157-1163.

(97) Buskirk, A. R.; Kehayova, P. D.; Landrigan, A.; Liu, D. R. *Chem. Biol.* **2003**, *10*, 533-540.

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

Chapter 2 Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch that Activates Protein Translation

2.1 Introduction

Synthetic riboswitches have a number of potential applications, including in genetic selections for small molecules. In some genetic selection experiments, the cell's life is dependent on a necessary small molecule.[1,2] The necessary small molecule could be directly or indirectly provided to the cell, where the indirect production could be the product of a biocatalyst. Therefore, genetic selections could also be used to isolate particular biocatalysts from large gene libraries made from techniques including mutagenic PCR,[3,4] DNA shuffling,[5,6] and incremental truncation.[7-9] Additionally, the gene libraries used could be from metagenomic or cDNA sources.[10] And selections would be particularly useful in identifying unknown genes that are responsible for the biosynthesis of medically relevant small molecules (such as paclitaxel or bryostatin).

Genetic selections in bacteria are particularly powerful because libraries of greater than 10^9 members can be transformed into bacteria and tested in a single experiment.[2] And, in contrast to a genetic screen, where individual members from a library must be tested individually, in a genetic selection experiment, only the surviving colonies contain the desired trait.[1] Also, screens typically have a much lower throughput than selections, and cannot sort through libraries containing more than 10^6 members.[11] Screens are often more expensive due to the instrumentation used (i.e. plate readers, robotic workstations, etc.). Genetic selection experiments in bacteria, in their simplest form, require only the cells containing the gene library, growth media, and an incubator, which are relatively low in cost. However, genetic selections are typically used to discover biocatalysts that complement strains which are auxotrophic in required metabolic pathways[12,13] (i.e. amino

acid biosynthesis, glycolysis, etc.) or provide resistance to a particular environmental toxin[14] (such as antibiotics), thus limiting their usage.

We are interested in developing a genetic selection method in bacteria that overcomes these limitations. In a perfect scenario, our method would select for the production of a small molecule of our choosing in a reaction independent manner. This would enable the identification of biocatalysts from libraries based solely on their product, without reliance on their enzymatic mechanism. And, as stated earlier, these "designer auxtrophs" could be used to sort through randomized gene libraries or heterologously cloned cDNA libraries.

A genetic selection method that can detect a desired small molecule requires a few components. One, it needs a mechanism for sensing the desired small molecule. And, two, it needs a way to modulate changes within the cell in response to small molecule sensing, where the change could be in levels of gene expression. Many groups have taken advantage of engineered protein-small molecule interactions to control gene expression.[15-18] But these types of systems, as stated in Chapter 1, are difficult to engineer, and the small molecules they can recognize are often limited by the recognition properties of the parent protein.

RNA may provide a platform for creating a sensor that binds to a small molecule of our choosing, and induces changes in gene expression. RNA aptamers have been selected in vitro that bind a variety of small molecules,[19] and, unlike protein engineering, they are not limited by the recognition capabilities of a pre-existing scaffold. And many groups have taken advantage of RNA aptamers to control gene expression in a small molecule dependent manner,[20-22] which presaged the discovery of a number of natural

small molecule responsive riboswitches that respond to a variety of small molecules.[23-26] Taken together, the evidence suggests synthetic riboswitches may be useful for small molecule dependent genetic selections.

In this Chapter, we present a synthetic riboswitch that activates protein translation in *E. coli* in response to small molecule binding, and we use this switch in both genetic screens and selections to detect the presence of theophylline. We further demonstrate that the recognition capabilities of the aptamer in vitro are retained in vivo. And we also present a variety of experiments directed towards understanding the mechanism of this synthetic riboswitch. In addition, we present data indicating that genetic selections may be used to alter the ligand recognition capabilities of a synthetic riboswitch. Genetic selections that take advantage of synthetic riboswitches may be useful for selecting biocatalysts in short order.

2.2 Results and Discussion

2.2.1 Creation of a Synthetic Riboswitch

We decided to construct a synthetic riboswitch with the previously described mTCT8-4 aptamer.[27] This aptamer binds theophylline with a K_d of 100 nM, binds the structurally similar molecule caffeine with a 10,000 fold lower affinity, and has been thoroughly characterized.[22,28] Previously, the theophylline aptamer was cloned into the 5' UTR of a reporter gene and used to downregulate translation in response to theophylline in wheat germ translation extracts.[22] The Suess group have also used the aptamer to turn on reporter gene expression in *B. subtilis* in the presence of theophylline. However, there are no examples of a synthetic riboswitch in *E. coli*. We wished to use

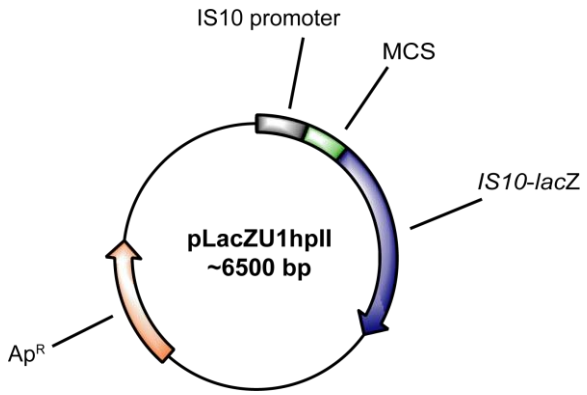

Figure 2.1: pLacZU1hpII Plasmid Map. The plasmid contains a weak IS10 promoter; a multiple cloning site (MCS); and an *IS10-lac*Z reporter gene fusion.

E. coli as a host, since it has high transformation efficiencies, well-understood genetics, and is commonly used as a host in directed evolution experiments.[29]

To create the synthetic riboswitch, we cloned the mTCT8-4 aptamer five base pairs upstream of the ribosome-binding site of the β-galactosidase reporter gene (IS10-*lac*Z) in the pLacZU1hpII plasmid[30] (Figure 2.1). This plasmid has a number of useful features: it contains a weak constitutive IS10 promoter, which is followed by a multiple cloning site in the 5' UTR, a putatively weak RBS, and a β-galactosidase gene fused with an IS10 transposase sequence.[30] These features were originally designed into the plasmid in order to measure small changes in gene expression in vivo in response to peptide-binding RNAs cloned into its 5' UTR. We reasoned this plasmid was an excellent platform to measure small changes in gene expression in response to small molecule binding.

We hypothesized that the aptamer, in the presence of theophylline, would form a tighter secondary structure, which would inhibit the ribosome from binding to the RBS. We based our hypothesis on a previous study that showed when hairpins with increasing secondary structure were cloned upstream of a RBS of a reporter gene in *E. coli*,

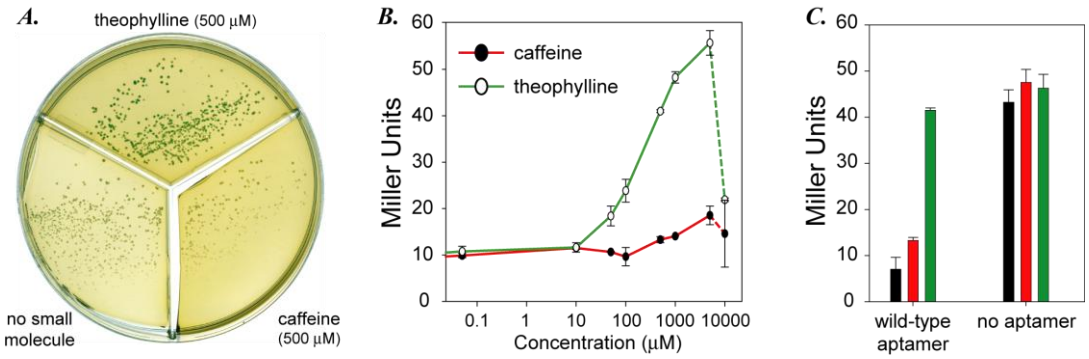

Figure 2.2: Characterization of a Synthetic Riboswitch. (A) Plate-based assay for β-galactosidase activity. The color changes only in the presence of theophylline. (B) Dose response curve for β-galactosidase activity (dashed lines indicate toxicity). (C) β-galactosidase activity in response to no small molecule (black), 500 µM caffeine (red), or 500 µM theophylline (green).

expression of the downstream gene decreased concurrently.[31] Our expectation was also in line with the measurements of natural riboswitches that operate at the translational level,[26] as well as a eukaryotic system that incorporated the theophylline aptamer.[22]

To test the activity of the synthetic riboswitch, *E. coli* harboring this construct were grown on LB agar supplemented with ampicillin, the chromogenic β-galactosidase substrate X-gal, IPTG [used to inhibit the background activity of β-galactosidase, thus improving our signal-to-noise ratio], and either no small molecule, 500 µM caffeine, or 500 µM theophylline. Cells containing the construct were grown for 14 h at 37 ^{o}C and were then incubated at 4 ^{o}C to allow hydrolysis of X-gal to occur. The plates were analyzed after 48 h and cells grown on theophylline appeared blue, whereas cells grown on media supplemented with either no small molecule or caffeine showed no color (Figure 2.2). This result suggested that theophylline was responsible for activating the expression of β-galactosidase, contrary to our hypothesis.

We used the liquid based assay method of Miller[30] to accurately quantify the theophylline-dependent increase in β-galactosidase activity we observed on the solid agar plates. Miller units for cells grown in the presence of theophylline showed a

dose- dependent increase, whereas cells grown with caffeine showed only a small increase in β-galactosidase activity. A control construct, lacking the aptamer, was also tested, and it showed no theophylline or caffeine dependence, indicating theophylline was not generically increasing β-galactosidase activity (Figure 2.2). It should be noted, the construct with no aptamer produced β-galactosidase activity similar to the construct containing the aptamer when grown with 500 μM theophylline, suggesting the theophylline aptamer does not interfere with translation.

From our dose response curve with theophylline, we estimated that an extracellular theophylline concentration of 250 μM was necessary to achieve a half-maximal response (Figure 2.2). Although we were not able to grow the cells with more than 5 mM theophylline due to toxicity (Figure 2.2), we believe our estimate is, nonetheless, accurate. However, the extracellular concentration of 250 μM is substantially higher than the K_d for the aptamer measured in vitro. One possible cause for this discrepancy arises from the permeability of theophylline. Previous studies have shown that when *E. coli* are grown with 10 μM theophylline, the internal concentration is only 7 nM,[32] a three-order of magnitude difference. These results suggest that either theophylline does not enter the cell to a great extent, or there is an active efflux mechanism, though it is difficult to prove either mechanism. If the external to internal concentration of theophylline varies linearly, we estimate that at an external concentration of 250 μM, the internal concentration is approximately 175 nM, which is near the K_d for the aptamer's theophylline binding in vitro. Therefore, our results indicate the in vitro properties of the theophylline aptamer are maintained in vivo. Additionally, we tried to increase theophylline permeability by growing cells with

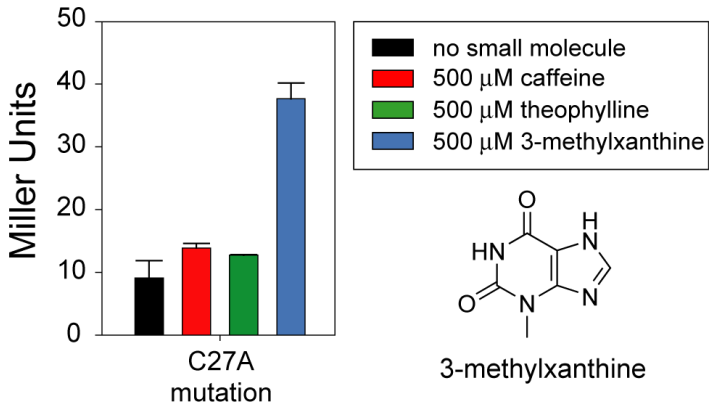

Figure 2.3: β-Galactosidase Activity of the C27A Mutant Riboswitch.

theophylline and DMSO, but no significant increase in β-galactosidase activity was observed.

We conducted further studies to investigate whether the β-galactosidase activity increase we observed was due to theophylline and not a generic response, and to determine if other in vitro properties of the aptamer transfer in vivo. We introduced a point mutation (C27A) into the aptamer sequence of our synthetic riboswitch which, in vitro, lowers theophylline affinity by approximately 10-fold, and slightly increases 3-methylxanthine binding.[33] *E. coli* cells harboring this mutant construct were grown in media supplemented with either no small molecule, 500 µM caffeine, 500 µM theophylline, or 500 µM 3-methylxanthine, and β-galactosidase activity was measured. Theophylline had only a mild effect on β-galactosidase activity, whereas 3-methylxanthine retained its response (Figure 2.3). This data indicates, again, that the in vitro properties of the aptamer are transferrable in vivo, and correlates well with the known adenine and guanine riboswitches that recognize different ligands due to a single base pair change in their aptamer domains.[34]

Our results demonstrate the sensitivity of riboswitch based assays for small molecules, and that it is possible to transfer the in vitro properties of an aptamer into an in vivo system. Additionally, the dose-dependent response of our assay indicates it may be useful for monitoring the intracellular concentrations of endogenous and non-endogenous small molecules. We believe this assay will prove useful for monitoring the production of small molecules by biocatalysts. This will enable a variety of screening and selection experiments with the goal of identifying biocatalysts from large libraries.

2.2.2 Determining the Mechanism of Action

We originally hypothesized that our synthetic riboswitch would show a theophylline dependent decrease in β-galactosidase activity. Clearly, this was not the case. In a previous report, the authors designed a theophylline activated riboswitch based on a proposed helix-slippage mechanism, where theophylline binding would cause the theophylline aptamer to twist and, thus, unpair the RBS of a reporter gene.[35] We did not have a defined strategy; we simply cloned the aptamer into the 5' UTR of the reporter gene. To investigate the mechanism, we moved the aptamer either closer or further away from the RBS (Figure 2.4). In most cases, the riboswitch was still activated with theophylline. The exception is the 0 bp spacing (not shown); it is not theophylline responsive, indicating the number of bases between the aptamer and the RBS is important. Since the riboswitch was relatively insensitive to the position of the aptamer, we decided to test whether it was controlling the transcription or translation of the reporter gene. A transcriptional mechanism would explain why the distance dependence does not affect function dramatically. Since, in transcriptional control, the synthesis of

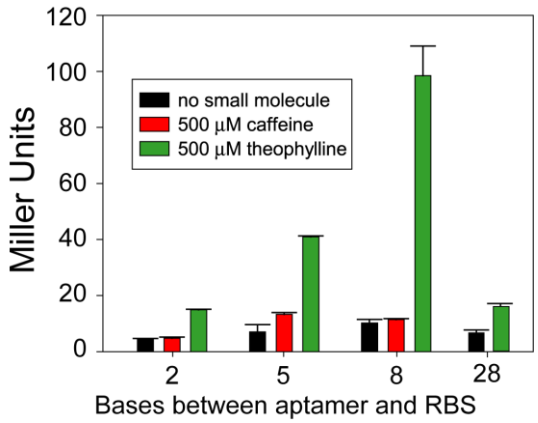

Figure 2.4: Varied Spacings Between the Aptamer and RBS. β-galactosidase activity for constructs with varied spacings when grown under different conditions. The wild-type is the five base spacing.

the mRNA is modulated. This type of mechanism would also be similar to that of the adenine riboswitch,[34] which activates transcription.

We created a transcriptional fusion between two genes to determine if the riboswitch was acting at the transcriptional or translational level.[36] The construct was made as follows: the first gene consisted of the mTCT8-4 aptamer, a RBS, and a 61-amino acid N-terminal fragment of the IS10 transposase followed by three in-frame stop codons. This gene segment was fused through a 28 bp linker sequence to a second gene that contained a RBS and a full-length wild-type *lac*Z sequence (Figure 2.5). Using this construct, if the expression of the *lac*Z gene is theophylline dependent, the riboswitch controls transcription. If, however, the riboswitch controls translation, then the transcriptional fusion would not show a theophylline dependent increase in *lac*Z expression because the first gene segment (IS10) is independent from, and not a translational fusion to, the downstream *lac*Z gene, which contains its own RBS. Our results indicate that β-galactosidase expression is not theophylline dependent, and that the riboswitch acts at the translational level (Figure 2.5).

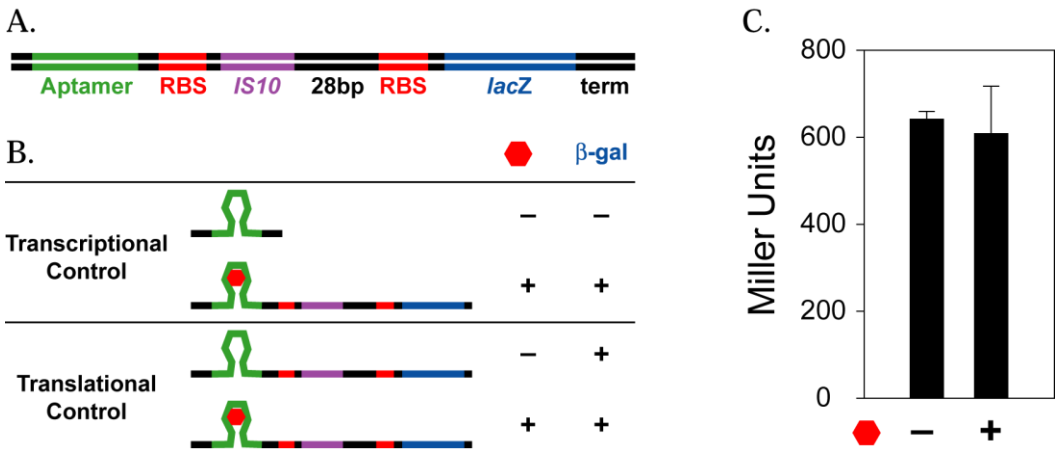

Figure 2.5: Transcriptional Fusion of a Synthetic Riboswitch. (A) Shematic of transcriptional fusion (aptamer = mTCT8-4, RBS = ribosome binding site, IS10 = 61 amino acid fragment from IS10 transposase, term = transcriptional terminator). (B) Expected outcomes for transcriptional and translational control when theophylline (red hexagon) is present or absent. (C) β-Galactosidase activity for the transcriptional fusion expressed in *E. coli*. Results show it is theophylline independent, indicating translational control.

One possible mechanism for the observed theophylline dependence of β-galactosidase activity may be that when theophylline binds to the aptamer, the RNA is stabilized in vivo. If the theophylline-bound RNA is resistant to RNAse cleavage, or some other nuclease, it will build up within the cell, thus more transcripts will be translated simultaneously. To probe this possibility, we performed Northern blots[37] on total RNA extracted from *E. coli* cells harboring the riboswitch that were grown with either no small molecule, 1 mM caffeine, or 1 mM theophylline. Equal amounts of total RNA from the cultures were separated on a denaturing agarose gel and transferred to a nylon membrane. A radiolabeled anti-sense RNA probe, specific to the riboswitch, was blotted onto the membrane and subjected to Phosphorimaging. Our results clearly indicate (Figure 2.6) that the amount of riboswitch RNA is highest in the cells grown with theophylline. This suggests that theophylline is stabilizing the RNA in vivo, but a translated RNA is also protected from endonuclease cleavage.[38] Since we have already determined the riboswitch operates at the translational level, the observed increases in

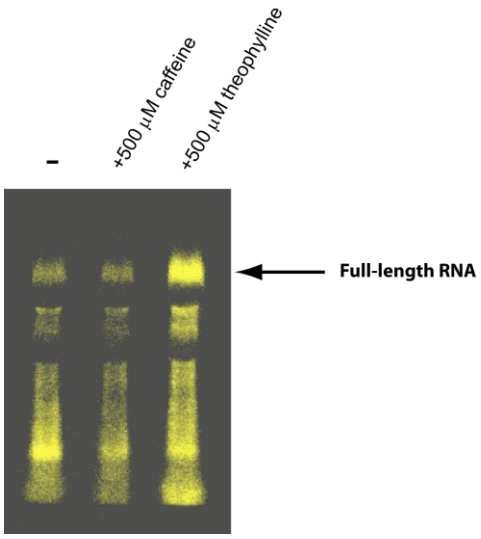

Figure 2.6: Northern Blot of Riboswitch RNA. RNA extracted from *E. coli* cells expressing the riboswitch with an 8 base pair linker was analyzed by Northern blot. When the cells were grown in the presence of theophylline, the amount of full-length RNA increased, in comparison to the full-length RNA from cells grown with no small molecule or caffeine. This is most likely due to an increase in translatable RNAs, which are slowly degraded.

riboswitch RNA are most likely due to activation of translation, and not an inhibition of endonuclease cleavage.

mFold[39] analysis of the first 190 base pairs of the transcribed riboswitch with an 8 base pair linker sequence indicates the RBS of the construct pairs with a segment of the coding region of the IS10 gene (not shown). This type of structure is reminiscent of a theophylline activated ribozyme, which was made by fusing the mTCT8-4 aptamer to the hepatitis delta virus ribozyme.[40] This mutant ribozyme was proposed to be theophylline dependent due to a ligand activated helix slippage mechanism, which could be analogous to our riboswitch. To explore the feasibility of the *mFold* structure further, we made mutations that were predicted to further pair or unpair the RBS (Figure 2.7). We created a mutant (A91U – where 1 is the first transcribed base) predicted to slightly upair the RBS, but not completely. This mutant functioned in a theophylline dependent manner in *E.coli*. A second mutant was created (C99G) that was predicted to further pair the RBS.

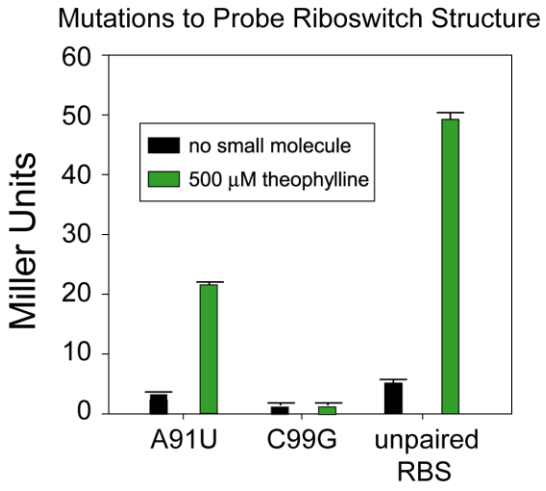

Figure 2.7: Structure Probing Mutations. Mutations were introduced into the riboswitch with an 8 base pair linker between the aptamer and RBS, based on the mFold structure. A91U was predicted to slightly unpair the RBS; C99G was predicted to pair the RBS; and the "unpaired RBS" construct was made by mutating the codons which were predicted to pair with the RBS. A91U functioned in a theophylline dependent manner, as expected. C99U was not theophylline responsive, as expected. The unpaired RBS constructed was theophylline responsive, which was not expected.

The increased secondary structure would, tentatively, increase the energy required to unpair the RBS, and we predicted this mutant would have little or no theophylline dependence. Indeed, the C99G mutation eliminated the theophylline dependence of the construct in *E. coli*. While the mutations correlated well with the *mFold* structure, we decided to make a construct that completely unpaired the RBS. If the *mFold* structure is correct, completely unpairing the RBS will eliminate theophylline dependent gene expression. To this end, the codons within the IS10 gene, opposite the RBS, were conservatively mutated to unpair the RBS and this construct was introduced into *E. coli*. Contrary to our expectations, its β-galactosidase activity was theophylline dependent, which indicates the mFold structure may not be accurate and that further studies are necessary.

We wondered what structure the riboswitch RNA adopted in vivo, and decided to probe this using in vivo dimethyl sulfate (DMS) probing.[41] In this technique, DMS is

added directly to living bacterial cells, penetrates the cells, and methylates the N^7, N^1, and N^3 positions of guanine, adenine, and cytidine, respectively. If the RNA structure is paired, DMS modification is inhibited at those positions. Total RNA from DMS modified cells is extracted, reverse transcribed using a specific radiolabeled primer, and run on a denaturing PAGE gel. DMS modification of RNA causes reverse transcriptase to pause, and the amount of pausing correlates with the extent of DMS modification, and, thus, the degree of secondary structure.[41] Specific positions are then identified by comparison to a sequencing lane of unmodified RNA. *E. coli* cells harboring the riboswitch were grown in LB supplemented with either no small molecule or 500 µM theophylline, separately. When the cells reached mid-log phase, the cells were pelleted and resuspended in buffer, then DMS was added to a final concentration of 150 mM. The total RNA was extracted after modification, subjected to reverse transcription, and run on a denaturing PAGE gel. Our results indicate that the RBS region of the riboswitch RNA is modified to the same extent when grown in the presence and absence of theophylline (Figure 2.8). This was unexpected, but may have an explanation. If only a small fraction of the riboswitch RNAs within the *E. coli* cells are turned on in the presence of theophylline, a bulk analysis method (like DMS modification) will fail to show a difference. It is also possible that the theophylline concentration used was not high enough to adequately shift the equilibrium of the RNA population in vivo for successful analysis with this technique. Further studies addressing riboswitch structure were performed, and will be presented in Chapter 3.

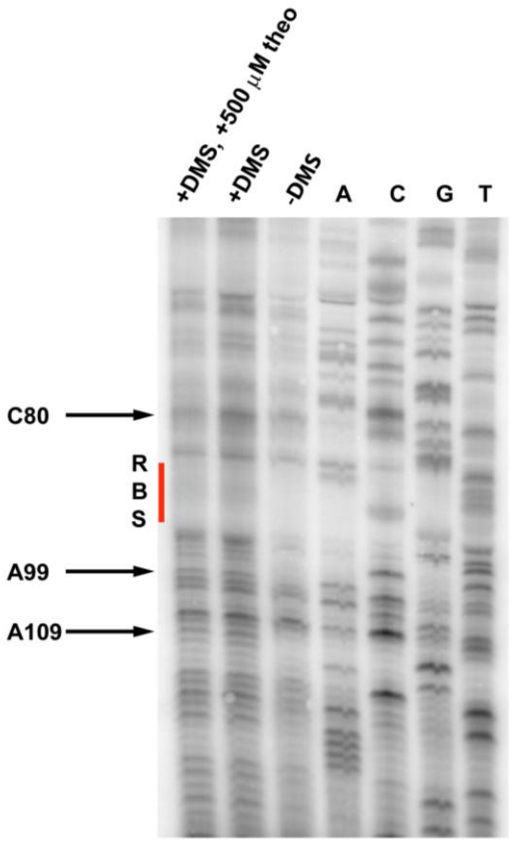

Figure 2.8: In vivo DMS Probing of Riboswitch RNA. *E. coli* cells expressing the riboswitch with an 8 bp spacer between the aptamer and RBS were grown with and without theophylline, before modification with DMS. Reverse transcription was performed on the modified RNA and run on a 10% denaturing PAGE. Clearly, the RBS region is modified to the same extent in either the presence or absence of theophylline, though a bulk analysis technique cannot determine small changes in modification. Sequencing lanes are on the right hand side of the gel.

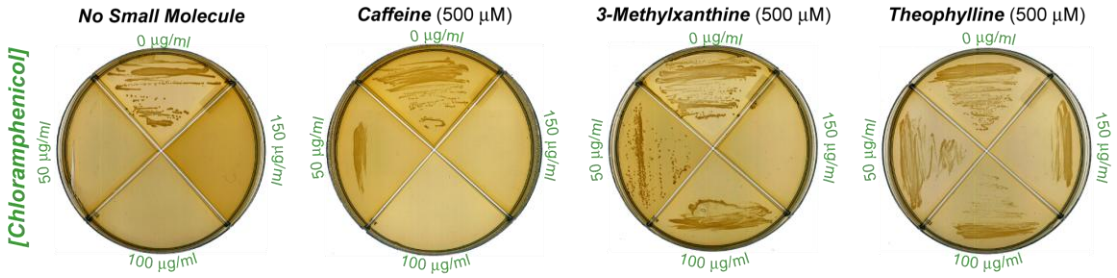

Figure 2.9: Riboswitch-Mediated Genetic Selections. *E. coli* expressing the theophylline-responsive riboswitch, upstream of the chloramphenicol resistance gene (*cat*), were grown at 37 ^{o}C for 18 h in the presence of different small molecules with varying concentrations of chloramphenicol. As expected, theophylline fully activates the riboswitch and allows growth on high concentrations of antibiotic.

2.2.3 Genetic Selections for Small Molecules Using a Synthetic Riboswitch

Our main goal in this work is to construct bacteria that are dependent on a small molecule of our choosing and to then use the newly designed bacteria for genetic selection experiments. The riboswitch presented above may be useful for this purpose, if the gene under control is changed appropriately. This scenario would allow genetic selections to be performed for the production of a desired small molecule from a large biocatalyst library, where the library could be produced in vitro or isolated from an organism, thus allowing identification of the desired biocatalyst.[42,43]

To test our hypothesis, we replaced the *lac*Z reporter gene in the riboswitch with a gene encoding the enzyme chloramphenicol acetyl transferase (*cat*), which confers resistance to chloramphenicol. And we replaced the weak IS10 promoter with the strong constitutive *tac* promoter.[44] We predicted that *E. coli* cells harboring this construct would be sensitive to chloramphenicol in the absence of theophylline, and resistant to chloramphenicol in the presence of theophylline, when translation would be activated. *E. coli* cells harboring the riboswitch-*cat* fusion gene were grown for 18 h at 37 ^{o}C on LB agar containing ampicillin and increasing concentrations of chloramphenicol (Figure 2.9). Without theophylline, the cells do not survive on chloramphenicol. With caffeine (500 μM), the cells grow slightly on 50 μg/mL chloramphenicol; this may be due to a weak interaction of caffeine with the riboswitch. Growth is observed up to a chloramphenicol concentration of 100 μg/mL, in the presence of 3-methylxanthine (500 μM). Cells grown in the presence of theophylline (500 μM) are resistant to chloramphenicol concentrations up to 150 μg/mL, as expected. These results demonstrate

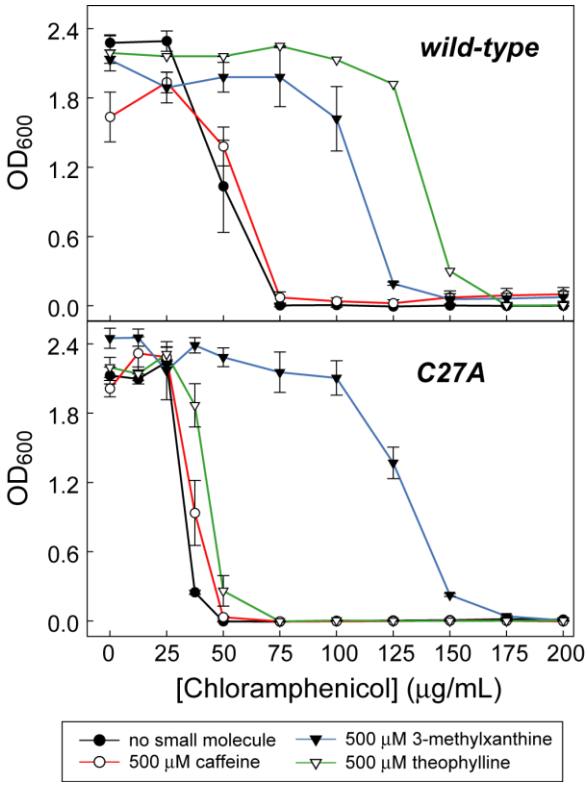

Figure 2.10: Liquid-Based Genetic Selections. *E. coli* harboring the wild-type or C27A mutant riboswitch were grown for 18 h at 37 °C in the presence of different small molecules and varying chloramphenicol concentrations. The wild-type riboswitch is most active in the presence of theophylline; the C27A riboswitch is most active in the presence of 3-methylxanthine.

that riboswitch-based genetic selections can be used to detect the presence of a small molecule by monitoring cell growth.

We also used the C27A mutant riboswitch, which recognizes 3-methylxanthine, to control the *cat* gene and performed liquid based selections with it (Figure 2.10). *E. coli* cells harboring this construct did not live at chloramphenicol concentrations above 50 µg/mL, without media supplementation. When the cells were grown with 500 µM 3-methylxanthine, they were resistant to chloramphenicol concentrations up to 150 µg/mL. As expected, the cells were not resistant to chloramphenicol concentrations above 50 µg/mL when grown with 500 µM caffeine or theophylline, since these molecules do not activate the C27A mutant riboswitch. Our results demonstrate that a

small molecule of one's choosing can be genetically selected for, using a synthetic riboswitch. The designed *E. coli* are, in essence, designer auxotrophs that require a specific, non-endogenous, small molecule to survive when grown with antibiotic. We predict these auxotrophs will be useful for isolating desired biocatalysts from large libraries, based solely on their reaction products.

2.2.4 Genetic Screens and Selections to Discover Synthetic Riboswitches in *E. coli*

The above experiments focused on developing synthetic riboswitches that recognize a particular small molecule target. But our results suggest that the power of screens and selections may be used to discover synthetic riboswitches with new small molecule recognition properties. If we create a library of mutant riboswitches in *E. coli*, it may be possible to isolate mutants with the desired small molecule recognition properties by monitoring an appropriate output (color, growth). To test our hypothesis, we asked whether we could identify a functional small molecule sensitive riboswitch against a background of mutant riboswitches that were not small molecule sensitive. Our parent riboswitch is based on the mTCT8-4 aptamer and activates translation in response to theophylline, whereas the C27A mutant does not. We decided to probe whether we could isolate the parent riboswitch against a background of theophylline insensitive C27A riboswitches by performing spiking experiments in which parent riboswitch DNA was diluted with the C27A riboswitch DNA. We decided to perform screens first, where we would monitor the expression of the *lac*Z gene. Plasmid mixtures of the parent and C27A plasmids in varying ratios (1:1, 1:10, and 1:100) were used to electroporate *E. coli*. These cells were grown for 14 h at 37 ^{o}C on LB agar supplemented with ampicillin,

X-gal, IPTG, and 1 mM theophylline. After growth, the cells were set at 4 ^{o}C for 48 h, to allow hydrolysis of the X-gal substrate. After hydrolysis, the plates containing 1:1 and 1:10 dilutions showed blue and white colonies that corresponded to their plasmid ratios, and five blue colonies were picked for sequencing. The plate from the 1:100 dilution showed only two blue colonies, and these were also picked. Sequencing revealed that all the blue colonies were the parent riboswitch, as expected. We also sequenced 12 white colonies, and 11 out of 12 were identified as the C27A mutant. Thus, the screen has a very low incidence of false positives.

Since the screening procedure worked well, we decided to perform similar experiments using a selection where we could use even larger dilutions of functional riboswitches against non-functional riboswitches. Again, we used the parent and C27A riboswitches, but we used them to control the *cat* gene. Plasmid mixtures of increasing dilutions of the parent versus the C27A mutant (1:10^3, 1:10^4, 1:10^5, and 1:10^6) were used to electroporate *E. coli*. Cells containing the plasmids were first plated at high density on LB agar supplemented with ampicillin and grown for 12 h at 37 ^{o}C. The resulting colonies were scraped off the plates into liquid LB media supplemented with ampicillin and grown at 37 ^{o}C until the cultures were saturated. Samples of cells from each culture were plated on LB agar supplemented with ampicillin (50 µg/mL), theophylline (1 mM), and chloramphenicol (100 µg/mL) and were grown at 37 ^{o}C for 18 h. Colonies were visible after this time, for all plasmid dilutions used. We picked five colonies from each plate and sequenced their plasmids. As expected, all the sequenced clones were the parent riboswitch. Therefore, genetic selection can be used to identify functional, small

molecule responsive riboswitches from a pool that contains a 10^6-fold excess of mutant riboswitches with different ligand specificities.

2.3 Conclusion

In this Chapter, we have presented the successful construction of a synthetic riboswitch in *E. coli* that activates protein translation in response to a specific small molecule. And we have shown that synthetic riboswitches are useful in both screening and selection experiments in response to a non-endogenous small molecule. Also, we demonstrated that the in vitro recognition properties of the aptamer transfer to in vivo systems.

Taken together, our results indicate that it is possible to create strains of *E. coli* that depend on a small molecule of our choosing for survival. Since these selections depend only on the presence of the small molecule, and not how it was made, synthetic riboswitch mediated screens and selections are not limited to monitoring a particular class of reaction such as bond-breaking or bond forming.[16] Therefore, this property may be useful for cloning biosynthesis genes for which the products, but not the synthetic pathways, are known. Because methods for generating aptamers are well known,[19] we anticipate that synthetic riboswitches will be constructed that respond to a variety of chemically distinct small molecules. Synthetic riboswitches should also prove useful for monitoring the production of small molecule products from directed evolution experiments.

Additionally, we have shown that *E. coli* cells harboring a synthetic riboswitch that responds to a specific ligand can be isolated from a one million fold background of

mutant synthetic riboswitches that respond to a closely related ligand. This result suggests that genetic selections may be useful in selecting synthetic riboswitches that respond to new ligands in vivo; this may potentially circumvent aptamer selections, which are time consuming.

2.4 Experimental

General Considerations

All plasmid manipulations were performed according to standard cloning techniques.[29] All plasmids were verified by sequencing. Qiagen kits were used for purification of plasmid DNA, PCR products, and enzyme digestions. Caffeine, theophylline, 3-methylxanthine, *o*-nitrophenyl-β-d-galactopyranoside, ampicillin, and chloramphenicol were purchased from Sigma. IPTG and X-gal were purchased from US Biological. Oligonucleotides were purchased from IDT. TOP10 *E. coli* cells were used in all experiments.

Plate Based Assays for β-galactosidase Activity

LB agar plates (100 mm diameter) containing ampicillin (50 μg/mL) and supplemented with either no small molecule, caffeine (500 μM), or theophylline (500 μM) were spread with 9 μL of a 20% IPTG solution and 50 μL of a 2% solution of X-gal in DMF. The parent synthetic riboswitch plasmid (pSKD177.2) was introduced into *E. coli* TOP10 cells by electroporation, and recovered in SOC media for 20 min at 37 ^{o}C. A 10 μL aliquot was spread on the LB plates, and the cells were grown for 14 h at 37 ^{o}C. The cells were then incubated at 4 ^{o}C until blue color was visible (48-96 h).

Quantitative Assays for β-galactosidase Activity

Separate colonies of *E. coli* TOP10 cells harboring either the no-aptamer (pLacZU1hpII), synthetic riboswitch (pSKD177.2), C27A synthetic riboswitch (pSKD185.1), or transcriptional fusion (pSKD323l.1) plasmid were picked from an LB agar plate containing ampicillin (50 µg/mL) and grown overnight at 37 ^{o}C in three separate culture tubes containing LB media supplemented with ampicillin (50 µg/mL). 5 µL of the overnight culture was used to inoculate 1.5 mL of LB media supplemented with ampicillin (50 µg/mL), and varying concentrations of caffeine, theophylline, or 3-methylxanthine. The cells were grown at 37 ^{o}C until an OD_{600} of 0.3-0.5 was reached. Cells were cooled on ice for 15 min, lysed and β-galactosidase activity was measured by monitoring the hydrolysis of *o*-nitrophenyl-β-d-galactopyranoside spectrophotometrically with a plate reading spectrophotometer. Assays were conducted in triplicate.

In vivo DMS Structure Probing

One colony of *E. coli* TOP10 cells harboring the synthetic riboswitch plasmid (pSKD177.2) was picked from an LB agar plate supplemented with ampicillin (50 µg/mL) and used to inoculate LB media supplemented with ampicillin (50 µg/mL) and grown overnight at 37 ^{o}C. 250 µL of the overnight culture was used to inoculate three separate LB cultures (25 mL) supplemented with ampicillin (50 µg/mL) and either no small molecule (two cultures) or theophylline (500 µM). Cells were grown with shaking at 37 ^{o}C until the OD_{600} reached 0.5, and the cells were pellet by centrifµgation at 4 ^{o}C. Cell pellets were resusupended in 5 mL TM buffer (10 mM Tris, 10 mM $MgCl_2$, pH 7.0) and 71 µL DMS (Sigma) was added to one of the cultures containing no small

molecule and the theophylline supplemented culture, and shaken at room temperature for one minute. β-mercaptoethanol (304 µL) was then added to quench the reaction. Total RNA from each cell culture was extracted with Triazole reagent according to the manufacturer (Sigma) and ethanol precipitated. Extracted RNA was resuspended in DEPC dH_2O. Separate reverse transcription reactions were setup on the extracted RNA using a 5' P^{32} labeled primer (SKD-146) and AMV reverse transcriptase (Promega) according to the manufacturer. Four separate RNA sequencing reactions were also setup on the unmodified RNA, where each reaction contained a single dideoxynucleotide, and each reaction was run as before. Fractions of the reverse transcription reactions were mixed with gel loading dye (Ambion) and loaded into separate wells of a 31 cm long, 4 mm thick gel, containing 8% acrylamide and 7 M urea. The gel was run at 1350 volts until the dye reached the end of the gel, dried, and exposed to a Phosphorimager screen (GE Healthcare) overnight. Images were acquired using a Phosphorimager and analyzed using ImageQuant (GE Healthcare).

SKD-146
RT sequencing primer for mTCT8-4 construct (54bp from RBS)
5′-GTTGAGTCGTTTTAAGTGTAATTCG

Plate-based Chloramphenicol Resistance Assays

E. coli TOP10 cells harboring the synthetic riboswitch plasmid (pSKD314) were streaked on LB agar plates containing ampicillin (50 µg/mL) and varying concentrations of chloramphenicol and either no small molecule, caffeine (500 µM), 3-methylxanthine (500 µM), or theophylline (500 µM). Cells were incubated at 37 ^{o}C for 18 h. Assays were conducted in triplicate.

Solution-based Chloramphenicol Resistance Assays

E. coli TOP10 cells harboring the parent (pSKD314) or the C27A mutant synthetic riboswitch plasmid (pSKD441.1) were picked from an LB agar plate supplemented with ampicillin (50 µg/mL) and used to inoculate LB media supplemented with ampicillin (50 µg/mL) and grown at 37 ^{o}C overnight. Aliquots of cells (10^5 cells) were used to inoculate 1 mL of LB media in a 24-well plate (Costar). The LB was supplemented with ampicillin (50 µg/mL) and varying concentrations of chloramphenicol and either no small molecule, caffeine (500 µM), 3-methylxanthine (500 µM), or theophylline (500 µM). The plate was shaken at 37 ^{o}C for 18 h, and the OD_{600} reading was taken with a plate reading spectrophotometer. All growth studies were conducted in triplicate.

Plate-based Screening for a Functional Riboswitch, "Spiking Experiment"

An LB agar plate (100 mm diameter) supplemented with ampicillin (50 µg/mL) and theophylline (1 mM) was spread with 20 µL of a 20% IPTG solution and 50 µL of a 2% solution of X-gal in DMF. The parent synthetic riboswitch plasmid (pSKD177.2) and the C27A synthetic riboswitch plasmid (pSKD185.1) were mixed in varying molar ratios (1:1, 1:10, and 1:100) and 1 µL was used to electroporate *E. coli* TOP10 cells. The cells were recovered in 400 µL SOC for 15 min at 37 ^{o}C. 1 µL of the recovered mixture was diluted to 100 µL in SOC and 50 µL was spread on the agar plates, above. Cells were grown for 14 h at 37 ^{o}C, and then incubated at 4 ^{o}C for 48-96 h. Individual blue and white colonies were grown and sequenced.

Plate-based Selection for a Functional Riboswitch, "Spiking Experiment"

The parent synthetic riboswitch (pSKD314) and the C27A mutant synthetic riboswitch (pSKD441.1), with the chloramphenicol resistance gene, were mixed and 1 µL was used to electroporate *E. coli* TOP10 cells (~5 x 10^9 cells). The molar ratios of pSKD314 to pSKD441.1 used were 1:10^3, 1:10^4, 1:10^5, and 1:10^6. Electroporated cells were recovered in 400 µL SOC for 1 h at 37 ^oC. 50 µL of the cells were spread on an LB agar plate supplemented with ampicillin (50 µg/mL) and grown at 37 ^oC for 12 h. After growth, 1 mL of LB was spread on the plates and the cells were scraped into the media. This mixture was used to inoculate 5 mL LB supplemented with ampicillin (50 µg/mL) and grown until saturation (~10^9 cells/mL) at 37 ^oC. A 1-20 µL aliquot of the saturated culture was spread onto LB agar plates supplemented with ampicillin (50 µg/mL), chloramphenicol (100 µg/mL), and theophylline (1 mM) and grown at 37 ^oC for 18 h. Five colonies from each selection plate were grown and their plasmids were sequenced.

Plasmid Construction

All constructs are derived from the previously described plasmid pLacZU1hpII, which comprises a weak IS10 promoter, followed by a ribosome-binding site, and a gene encoding a fusion between a fragment of the bacterial IS10 transposase and the *lac*Z gene. Expression of this plasmid in *E. coli* provides the "no aptamer" control in Figure 2.2.

Plasmid pSKD177.2 (wild-type aptamer)

The following strategy was used to introduce the mTCT8-4 aptamer into the 5'-UTR of the IS10-*lac*Z fusion gene in the pLacZU1hpII plasmid. pLacZU1hpII was used as a template for a PCR reaction with the forward primer SKD-57 and the reverse primer SKD-56. The PCR product was gel purified, digested with *Kpn*I and *Hind*III, gel purified, and cloned into the *Kpn*I and *Hind*III sites of pLacZU1hpII as described. The *Kpn*I site is located immediately 5' of the aptamer sequence and a 5bp spacer (*TATAA*) was included immediately 3' to the aptamer sequence before the ribosome-binding site of the IS10-*lac*Z gene.

SKD-57
5'-CCCGGTACC**GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACC***TATAA*AGACAACAAGATGTGC
GAACTCG
(**mTCT8-4 aptamer,** *5bp spacer,* gene specific sequence)

SKD-56
5'- CGACGGGATCGATCCCCCC

Plasmid pSKD185.1 (C27A mutation in the aptamer sequence)

The mutation C27A (this the 27th nucleotide of the mTCT8-4 sequence as reported in Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B., *Science* **1994,** *263*, 1425-1429, it has alternatively been referred to as C22A in Soukup, G. A.; Emilsson, G. A.; Breaker, R. R., *J. Mol. Biol.* **2000,** *298*, 623-632.) was prepared using the QuikChange method (Stratagene) using pSKD177.2 as the template and primers SKD-65 and SKD-66.

SKD-65
5'- CAGCATCGTCTTGATGCCATTGGCAGCACCTATAAAG
(C to A mutation in mTCT8-4 aptamer)

SKD-66
5'- CTTTATAGGTGCTGCCAATGGCATCAAGACGATGCTG

Plasmid pSKD314 (chloramphenicol selectable plasmid)

A cassette mutagenesis strategy was used to generate pSKD314. The chloramphenicol acetyl transferase (*cat*) gene was cloned from pBAD33 using PCR with the forward primer SKD-99 and the reverse primer SKD-94, to give product A. A separate PCR product (B) was generated using pSKD177.2 as a template with the forward primer SKD-71 and the reverse primer SKD-100. PCR products A and B contain an overlapping region and were mixed and amplified using the forward primer SKD-71 and the reverse primer SKD-94 to give PCR product C. PCR product C was digested with *Kpn*I and *Sac*I, gel purified, and cloned into the *Kpn*I and *Sac*I sites of pLacZU1hpII to afford plasmid pSKD305.1. A derivative of the Ptac1 promoter lacking the *lac* operator sequence was engineered in place of the IS10 promoter in pSKD305.1 by first generating a PCR product (D) using pSKD305.1 as a template with the forward primer SKD-125 and the reverse primer SKD-98. PCR product D was digested with *Spe*I and *Sac*I, gel purified, and cloned into the *Xba*I and *Sac*I sites of pUC18, forming plasmid pSKD314. pSKD314 thus comprised the *cat* gene fused in-frame to the 61st amino acid of IS10 throμgh a 7 amino acid linker (Trp-Pro-Gly-Ser-Pro-Ala-Ser) along with the mTCT8-4 aptamer in the 5'-UTR of the gene fusion under the control of the Ptac1 derivative promoter.

SKD-99

5′- **GGCCTGGGTCCCCTGCTAGC**GAGAAAAAAATCACTGGATATACCACCGTTG
(**7 amino acid linker,** gene specific segment)

SKD-94
5′ - GGCGCATGAGCTCTTACGCCCCGCCCTGCCACTCATCG
(gene specific segment)

SKD-71
5′- CGGCCCGGCCATAAACTGCCAGGAATTAATTTAC

SKD-100
5′– **GCTAGCAGGGGACCCAGGCCA**GATTCGTTTGATGTTATGTTTTGTTCTCGC
(**7 amino acid linker,** gene specific segment)

SKD-125
5′-TCAATCACTAGT**GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGG**GCTAACAAGTCTAGCG
AACCGCAC
(**Ptac1-derived promoter,** gene specific segment, *mutation eliminates SpeI site*)

SKD-98
5′ GTTAAATTGCCAACGCTTATTACCCAGCTCGATGC

Plasmid pSKD441.1 (C27A mutation in aptamer of chloramphenicol selectable plasmid)

A cassette mutagenesis strategy was used to make plasmid pSKD441.1. A PCR product (A) was made using pSKD314 as a template with forward primer JPG-7 and reverse primer SKD-66. A separate PCR product (B) was formed using pSKD314 as a template with forward primer SKD-65 and reverse primer JPG-8. PCR products A and B contain an overlap, were mixed, and amplified using primers JPG-7 and JPG-8, forming PCR product C. PCR product C was digested with *Kpn*I and *Sac*I and cloned into these sites in plasmid pSKD314 forming plasmid pSKD441.1.

JPG-7
5′– GCGATTAAGTTGGGTAACGCCAGGG

JPG-8
5′– GTATGTTGTGTGGAATTGTGAGCGG

Plasmid pSKD345.1 (transcriptional fusion)

The transcriptional fusion contained 3 stop codons after the 61st amino acid of the IS10 transposase in plasmid pSKD177.2 followed by a 28 bp pair spacer and then a ribosome binding site and *lac*Z gene cloned from pRS415. To make this construct, a

PCR product (E) was generated using pSKD177.2 as a template with the forward primer SKD-71 and the reverse primer SKD-134. A second PCR product (F) was made using plasmid pRS415 as a template with the forward primer SKD-133 and the reverse primer SKD-98. PCR products E and F overlap, and they were mixed and amplified using primer SKD-71 and primer SKD-98, to give PCR product G. PCR product G was gel purified, digested with *Kpn*I and *Sac*I, gel purified, and cloned into the *Kpn*I and *Sac*I sites in pLacZU1hpII, to give pSKD345.1.

SKD-133
5′- **CTGACTCCTCGAGTATAAAG**ACAACAAGATGACCATGATTACGGATTCACTGGCCGTC
(overlapping region, gene specific segment)

SKD-134
5′-
CTTTATACTCGAGGAGTCAGAGATCTCAGTTTATTATTAGATTCGTTTGATGTTATGTTTTGTTCTCGC
(overlapping region, gene specific segment)

2.5 References

(1) Woycechowsky, K. J.; Hilvert, D. *Eur. J. Biochem.* **2004**, *271*, 1630-1637.

(2) Taylor, S. V.; Kast, P.; Hilvert, D. *Angew. Chem. Int. Ed.* **2001**, *40*, 3310-3335.

(3) Cadwell, R. C.; Joyce, G. F. *PCR Methods Appl.* **1992**, *2*, 28-33.

(4) Cadwell, R. C.; Joyce, G. F. *PCR Methods Appl.* **1994**, *3*, S136-140.

(5) Stemmer, W. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10747-10751.

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

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

(8) Ostermeier, M.; Nixon, A. E.; Shim, J. H.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3562-3567.

(9) Ostermeier, M.; Shim, J. H.; Benkovic, S. J. *Nat. Biotechnol.* **1999**, *17*, 1205-1209.

(10) Brady, S. F.; Chao, C. J.; Handelsman, J.; Clardy, J. *Org. Lett.* **2001**, *3*, 1981-1984.

(11) Olsen, M.; Iverson, B.; Georgiou, G. *Curr. Opin. Biotechnol.* **2000**, *11*, 331-337.

(12) Oue, S.; Okamoto, A.; Yano, T.; Kagamiyama, H. *J. Biol. Chem.* **1999**, *274*, 2344-2349.

(13) Yano, T.; Oue, S.; Kagamiyama, H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5511-5515.

(14) Yano, T.; Kagamiyama, H. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 903-907.

(15) Azizi, B.; Chang, E. I.; Doyle, D. F. *Biochem. Biophys. Res. Commun.* **2003**, *306*, 774-780.

(16) Baker, K.; Bleczinski, C.; Lin, H.; Salazar-Jimenez, G.; Sengupta, D.; Krane, S.; Cornish, V. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16537-16542.

(17) Firestine, S. M.; Salinas, F.; Nixon, A. E.; Baker, S. J.; Benkovic, S. J. *Nat. Biotechnol.* **2000**, *18*, 544-547.

(18) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185-190.

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

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

(21) Grate, D.; Wilson, C. *Bioorg. Med. Chem.* **2001**, *9*, 2565-2570.

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

(23) Miranda-Rios, J.; Navarro, M.; Soberon, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9736-9741.

(24) Nou, X.; Kadner, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7190-7195.

(25) Nudler, E.; Mironov, A. S. *Trends Biochem. Sci.* **2004**, *29*, 11-17.

(26) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952-956.

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

(28) Zimmermann, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186-9192.

(29) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001.

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

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

(32) Koch, A. L. *J. Biol. Chem.* **1956**, *219*, 181-188.

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

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

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

(36) Beckwith, J. *Methods Enzymol.* **2000**, *326*, 3-7.

(37) Alwine, J. C.; Kemp, D. J.; Stark, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5350-5354.

(38) McCormick, J. R.; Zengel, J. M.; Lindahl, L. *J. Mol. Biol.* **1994**, *239*, 608-622.

(39) Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406-3415.

(40) Soukup, G. A.; Breaker, R. R. *Trends Biotechnol.* **1999**, *17*, 469-476.

(41) Wells, S. E.; Hughes, J. M. X.; Igel, A. H.; Ares, M. *Method Enzymol.* **2000**, *318*, 479-493.

(42) Rothman, S. C.; Voorhies, M.; Kirsch, J. F. *Protein Sci.* **2004**, *13*, 763-772.

(43) Kast, P.; Asif-Ullah, M.; Jiang, N.; Hilvert, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5043-5048.

(44) de Boer, H. A.; Comstock, L. J.; Vasser, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 21-25.

Chapter 3 High-Throughput Screens and Selections for Synthetic Riboswitches

3.1 Introduction

Riboswitches are genetic control elements composed entirely of RNA, and regulate gene expression in response to ligand-binding, without the need for proteins.[1,2] In their simplest form, riboswitches contain an aptamer domain, which recognizes the ligand, and an expression platform, which changes gene expression in response to ligand binding.[2] Riboswitches are found in a large number of prokaryotes,[3-6] and have also been found in a smaller number of eukaryotes.[7] Natural riboswitches control gene expression in response to endogenous metabolites, and several researchers have created synthetic riboswitches that respond to nonendogenous small molecules.[8-13] In theory, synthetic riboswitches can be engineered to respond to any nontoxic, cell-permeable molecule that is capable of interacting with an RNA. These synthetic riboswitches could be used to detect the production of small molecules for applications in directed evolution, or for the sensing of environmental contaminants. While aptamer selection methods are well established,[14] there are not general methods for converting these aptamers into riboswitches that function optimally in prokaryotes. To fully exploit their potential, new methods to create optimized riboswitches in bacteria must be developed.

A number of methods have been used to create synthetic riboswitches that control eukaryotic transcription or translation. In the first example, Werstuck and Green cloned an aptamer into the 5' untranslated region (5' UTR) of a eukaryotic reporter gene.[13] Ligand-binding by the RNA increased the secondary structure of the RNA, and reduced translation of the downstream gene. Other groups used a similar approach to create synthetic riboswitches in eukaryotes that decreased gene expression in response to other nonendogenous ligands.[9,10,12] In addition, workers from the Liu group used directed

evolution to discover an aptamer based RNA unit in yeast that activates transcription in response to the small molecule tetramethylrosamine.[15]

Although most natural riboswitches have been found in bacteria, it is surprising that there are few examples of synthetic riboswitches that function in bacteria. The Suess group reported a theophylline-dependent riboswitch in *B. subtilis* which activated translation using a designed helix-slippage mechanism,[11] and we reported a theophylline responsive riboswitch that activates translation in *E. coli*.[8] Our previously reported riboswitches were useful, but were deficient in a number of areas. First, in the absence of ligand, protein translation was still active. Second, our best riboswitch showed a signal to background ratio (activation ratio) of approximately eight in the presence of 500 μM theophylline. While this activation ratio was useful for performing genetic screens and selections, an increased ratio of signal to background is desirable. Finally, we conducted a number of studies to determine the mechanism of our first riboswitches, but we were not able to pinpoint it. To address these issues, we developed a number of high-throughput approaches that allow us to identify synthetic riboswitches which have very low levels of gene expression in the absence of ligand, and high levels of gene expression in the presence of ligand. These characteristics lead to synthetic riboswitches with large activation ratios. Sequencing of the evolved synthetic riboswitches allowed us to propose and test a model for synthetic riboswitch function. Our results indicate that starting from a single aptamer, a number of synthetic riboswitches can be formed, lending weight to the theory that the aptamer domains of natural riboswitches were recruited from other functions, and subsequently used as genetic control elements.[16] We believe that the high-throughput screens and selections presented here will be generally useful for

discovering synthetic riboswitches that respond to new ligands, and that these synthetic riboswitches will also have better performance characteristics.

3.2 Results and Discussion

In our previous publication, we observed that changing the length of the sequence separating the aptamer and the ribosome binding site (RBS) changed both the expression levels and dynamic range of a synthetic riboswitch.[8] However, we were not able to determine why changes in sequence led to these effects. *mFold* analysis of the initial synthetic riboswitches indicates that the bases between the aptamer and RBS are paired in the absence of ligand, and we anticipated that by randomizing these bases and screening or selecting for function, we could discover improved variants and gain insight into their function.

3.2.1 Creation of a Library of Randomized Mutants

Previously, we created synthetic riboswitches by cloning a theophylline-binding aptamer at various positions upstream of a β-galactosidase reporter gene (*IS10-lac*Z) that was controlled at the transcriptional level by a weak, constitutive IS10 promoter. To increase the signal, we replaced the IS10 promoter with the stronger Ptac1 promoter.[17] And, as expected, the *tac* promoter increased the level of β-galactosidase expression in the presence and absence of theophylline (Figure 3.1). For our selection experiments, we replaced the *lac*Z gene of this plasmid with either *ccd*B, which encodes for a toxic gyrase inhibitor protein,[18] or *cat*, which confers chloramphenicol resistance.

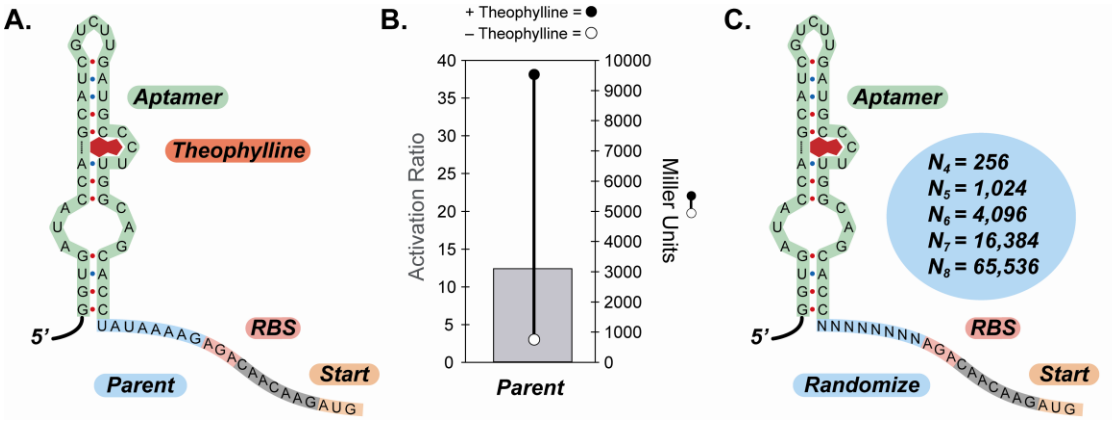

Figure 3.1: Diagram of the 5' Region of a Synthetic Riboswitch, Performance Characteristics of a Synthetic Riboswitch, and Randomized Region. (A) Sequence of parent synthetic riboswitch. (B) β-galactosidase activity of the parent riboswitch shown in (A) under the control of the *tac* promoter when expressed in *E. coli*. β-galactosidase activity was measured in the absence (open circles) and presence (closed circle) of 1 mM theophylline. The activiation ratio (green bar) is defined by dividing Miller units in the presence of theophylline by those in the absence. (C) Sequence of randomized riboswitch library. Total possible combinations for each randomization length are listed.

We used cassette-based PCR mutagenesis to create a library that contained a randomized 8 base pair region between the aptamer and RBS (Figure 3.1), and this length was chosen based on our previous observation that longer sequence lengths negatively affected synthetic riboswitch function.[8] This library was then cloned into the appropriate vector and transformed into *E. coli*.

3.2.2 Selection System for Optimally Functioning Riboswitches

We envisioned that an optimal synthetic riboswitch would have both low levels of gene expression in the absence of ligand, and high levels of gene expression in the presence of ligand. To this end, we developed a selection system that would produce riboswitches with the desired characteristics. In our system (Figure 3.2), the randomized synthetic riboswitch library is first cloned upstream of the toxic reporter gene *IS10-ccd*B, transformed into *E. coli* and grown in the absence of theophylline. Living cells should contain synthetic riboswitches that are not expressing *IS10-ccd*B and, thus, survive.

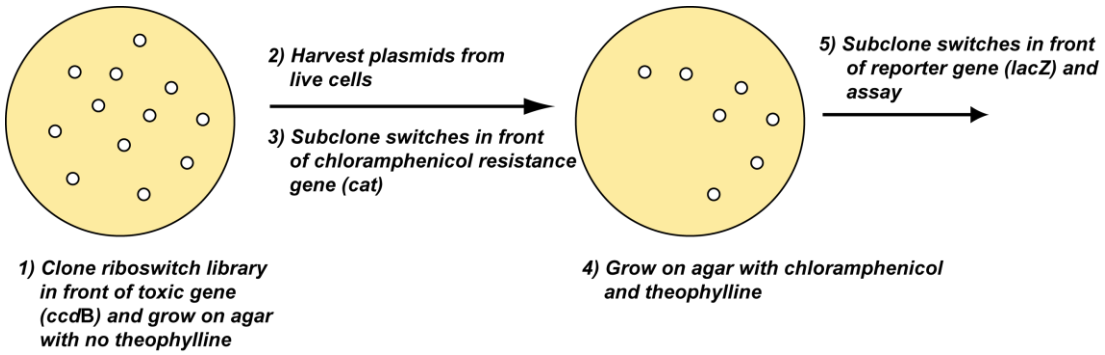

Figure 3.2: Diagram of the Selection System for Synthetic Riboswitches. Potential synthetic riboswitches are first selected for "off" in the absence of theophylline, using the toxic *ccd*B gene. Candidate riboswitches from live cells are then cloned upstream of the chloramphenicol resistance gene (*cat*) and grown on chloramphenicol and theophylline - the "on" selection. Potential riboswitches are then cloned upstream of *lac*Z and the β-galactosidase activity is assayed in the presence and absence of theophylline.

Living cells are then harvested and the plasmids are extracted. Synthetic riboswitches from the negative selection are cloned upstream of an *IS10-cat* reporter gene, transformed into *E. coli*, and grown in the presence of both chloramphenicol and theophylline. Only cells with sufficient levels of *IS10-cat* expression will be resistant to chloramphenicol and survive the positive selection. At this point, the living cells should contain synthetic riboswitches with the characteristics we desire. Synthetic riboswitches that survive the positive and negative selections can be cloned upstream of a β-galactosidase reporter gene and assayed for function.

We cloned the 8 base pair randomized riboswitch library upstream of the *IS10-ccd*B reporter gene under the control of the *tac* promoter, transformed *E. coli* cells with the library, and plated the cells on LB agar supplemented with ampicillin (50 μg/mL) and increasing concentrations of theophylline and grew the cells at 37 ^{o}C for 18 h. On the plate with no theophylline, a lawn of colonies (~10^5) was observed, but very few colonies (hundreds) grew on the plate supplemented with 1 mM theophylline, indicating that a number of synthetic riboswitches were activating expression of the toxic gene and that the reporter gene was, indeed, toxic. Colonies from the plate with no added

theophylline were harvested and the plasmids were extracted. Potential synthetic riboswitches from the negative selection plasmids were digested out and cloned upstream of a plasmid containing the *IS10-cat* reporter gene under the control of the *tac* promoter. This library was transformed into *E. coli* and approximately 10^5 cells were plated, separately, on LB agar supplemented with ampicillin (50 μg/mL), theophylline (500 μM), and varying concentrations of chloramphenicol (0, 25, 50 and 75 μg/mL). After 18 h growth at 37 ^{o}C, a lawn of colonies was observed on the plate with no chloramphenicol, but only 20-100 colonies were found on the plates containing chloramphenicol. This result suggested that the majority of the selected riboswitches do not have a high level of background expression, as expected. Two colonies from each chloramphenicol-containing plate were sequenced. Surprisingly, the plasmids isolated from the plate containing 25 μg/mL chloramphenicol were the parent riboswitch. However, the sequences of the potential riboswitches from the plates containing 50 and 75 μg/mL chloramphenicol were different from each other, and the parent.

To accurately characterize the potential riboswitches, we separately cloned the four sequences upstream of the *IS10-lacZ* reporter gene under the control of the *tac* promoter and transformed *E. coli* cells with the constructs. Using the method of Miller,[19] we assayed the β-galactosidase activity of cells harboring the constructs grown in the presence and absence of theophylline (500 μM). Unfortunately, the sequences isolated from the plate containing 75 μg/mL chloramphenicol were not responsive to theophylline. One possibility is that these riboswitches are dependent on the *ccd*B gene sequence for function, as was observed with the parent riboswitch, which is dependent on the downstream coding region for function. But this is unlikely, since both selection

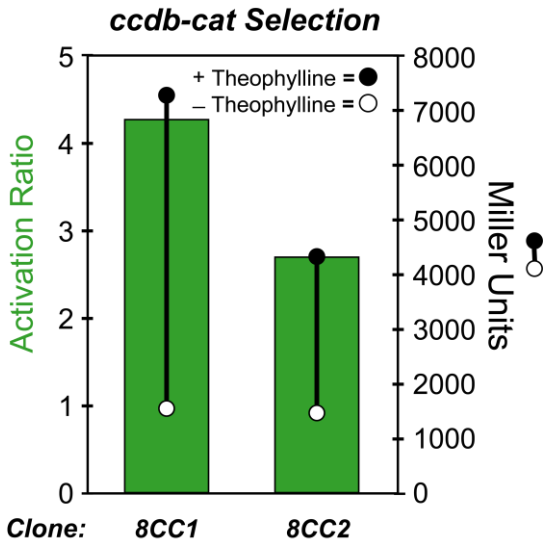

Figure 3.3: β-Galactosidase Activities for Riboswitches from *ccd*B-*cat* Selections. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 500 µM theophylline, and the standard errors of the mean are less than the circle diameters.

plasmids contain fusions with the *IS10* gene, and the putative interaction would have to occur with a gene segment over 180 base pairs downstream of the start codon, into the fused gene.

The sequences isolated from the selection plate that contained 50 µg/mL chloramphenicol were theophylline responsive, in contrast to those isolated from the plate that contained 75 µg/mL chloramphenicol (Figure 3.3). Using the method of Miller[19], we measured the β-galactosidase activity of cells harboring these constructs (8CC1 and 8CC2) after growth in the presence and absence of theophylline (500 µM), and, unfortunately, neither construct displayed an activation ratio greater than the parent riboswitch (the parent's activation ratio, when grown with 500 µM theophylline, is approximately 5). It is apparent that both riboswitches have low activations ratios because they have high levels of β-galactosidase activity in the absence of ligand.

Since we only sampled a small fraction (four) of the variants from our selections, we believed there were, potentially, other riboswitches with better characteristics

remaining. To test this, we scraped the colonies from the positive selection plates that contained 50 µg/mL chloramphenicol and harvested the plasmids. The potential synthetic riboswitches were digested out and cloned upstream of the *IS10-lac*Z reporter gene and transformed into *E. coli*. The β-galactosidase activity for 96 random colonies was assayed after growth in the presence and absence of 500 µM theophylline, and, unfortunately, no variants had an activation ratio greater than 4. While these results indicated that the selected riboswitches do not have performance characteristics superior to those of the parent's, the system does produce functioning riboswitches, which is our first indication that our randomization strategy, followed by selection, is useful.

3.2.3 High-Throughput "Stamping" Assay for Optimally Functioning Riboswitches

The selection method described above has a number of flaws. First, the process is time consuming due to the multiple cloning and cell growth steps required. All together, a single selection experiment takes approximately six days. Second, the "off" selection step is not stringent because the ccdB protein is not extremely toxic and is tolerated at moderate expression levels (not shown). Because of this, our selections produced riboswitch variants with high levels of gene expression in the absence of ligand.

To remedy these problems, we developed a novel assay that uses a single reporter gene, and involves only one cloning step and two cell growth steps. *E. coli* cells harboring the randomized riboswitch library (N8) upstream of the *IS10-lac*Z reporter gene, under the control of the *IS10* or *tac* promoter, were plated on LB agar supplemented with X-gal (a chromogenic β-galactosidase substrate) and ampicillin (50 µg/mL), in the absence of theophylline. The cells were then grown at 37 ^{o}C for 14 h.

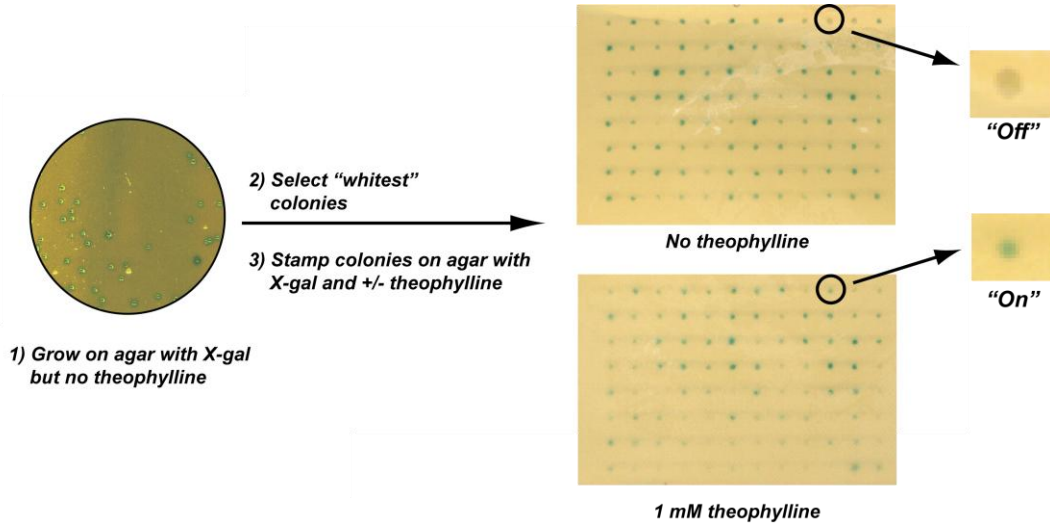

Figure 3.4: Diagram of the High-Throughput "Stamping" Assay. *E. coli* cells harboring potential riboswitches (cloned upstream of *IS10-lac*Z) are plated onto selective agar supplemented with X-gal. A colony-picking robot is used to pick the whitest colonies (low levels of β-galactosidase activity). These colonies are stamped on agar supplemented with X-gal and either no small molecule or theophylline (1 mM). After growth, colonies that are blue only in the presence of theophylline are assayed further.

Most colonies appeared blue, indicating they were expressing β-galactosidase in the absence of theophylline. However, a number of colonies appeared white, suggesting low levels of β-galactosidase expression. We hypothesized that a number of these white colonies contained riboswitches that could activate gene expression in the presence of theophylline. To test our hypothesis, we used a colony-picking robot to isolate the 96 whitest colonies from each plate of approximately 4000 colonies. The colony-picking robot has a 96 pin picking head, and the whitest colonies picked were physically "stamped" (Figure 3.4) onto LB agar supplemented with X-gal, ampicillin (50 µg/mL), and either no small molecule or theophylline (1 mM). After 14 h growth at 37 ^{o}C, the plates were left at 4 ^{o}C for 4 h to allow any background β-galactosidase activity to appear. Colonies were then visually inspected for β-galactosidase activity in the presence and absence of theophylline. Colonies that showed β-galactosidase activity (blue color)

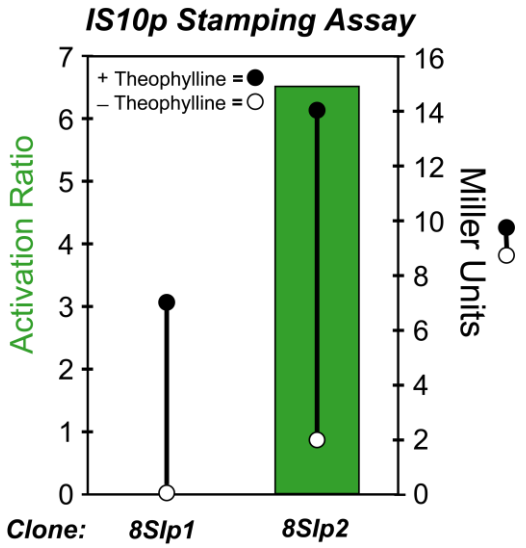

Figure 3.5: β-Galactosidase Activities for Identified Riboswitches from Stamping Assay with *IS10* Promoter. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters. The activation ratio for clone 8SIp1 is not shown because no β-galactosidase activity was measured in the absence of theophylline.

only in the presence of theophylline were assayed further using the method of Miller.[19] The results from an actual "stamping" plate are shown in Figure 3.4.

Functional synthetic riboswitches were found in both the plasmid libraries (~4000 colonies assayed from each) that were under the control of either the *IS10* or *tac* promoters, indicating that promoter strength is not critical for this method. The clones with the highest activation ratios from each promoter library are shown in Figures 3.5 and 3.6. The β-galactosidase activity for each clone in the presence and absence of theophylline (1 mM) is also shown. From these data, it is clear that the stamping assay identifies synthetic riboswitches that display low levels of β-galactosidase activity in the absence of theophylline and high levels of β-galactosidase activity in the presence of theophylline. Indeed, clone 8SIp1 has an indefinitely high activation ratio, as we could not measure any β-galactosidase activity in the absence of theophylline. To confirm this, *E. coli* cells harboring this construct were streaked on LB agar supplemented with X-gal,

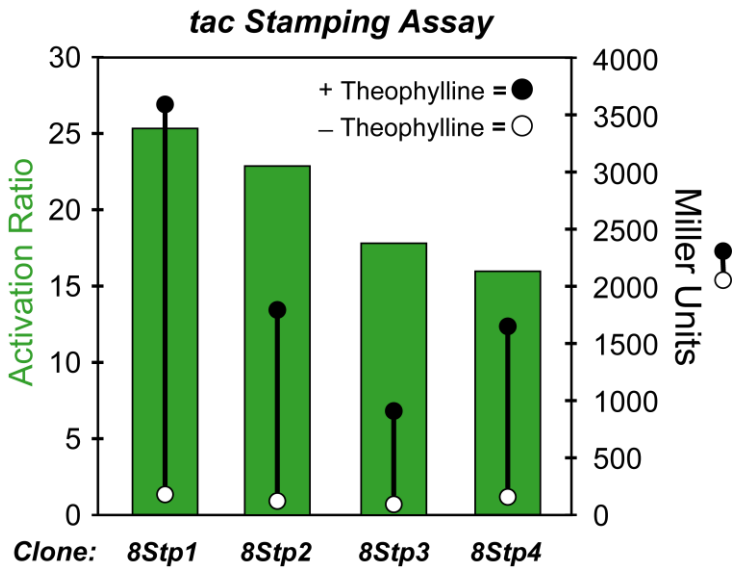

Figure 3.6: β-Galactosidase Activities for Identified Riboswitches from Stamping Assay with *tac* Promoter. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters.

but no theophylline, and grown for 14 h at 37 ^{o}C, followed by an incubation period of 96 h at 4 ^{o}C. At this point, no β-galactosidase activity was observed visually. To further probe clone 8SIp1, we inserted it downstream of the *tac* promoter and assayed its β-galactosidase activity in the presence and absence of theophylline (1 mM), and it produces an activation ratio of 36 (Figure 3.10-Clone 8.1), suggesting that promoter strength strongly influences the activation ratio. Natural riboswitches typically show regulatory factors of approximately 100 in the presence of their cognate ligand,[5] indicating that our evolved synthetic riboswitches are within this functional realm and that the stamping method is quite effective for identifying functional variants.

In a previous publication, we used a synthetic riboswitch for both screens and selections for small molecules. To use the original riboswitch in a screen, it was necessary to supplement the LB agar with IPTG to inhibit the background level of β-galactosidase activity.[8] With riboswitch 8SIp1, the addition of IPTG to the media is

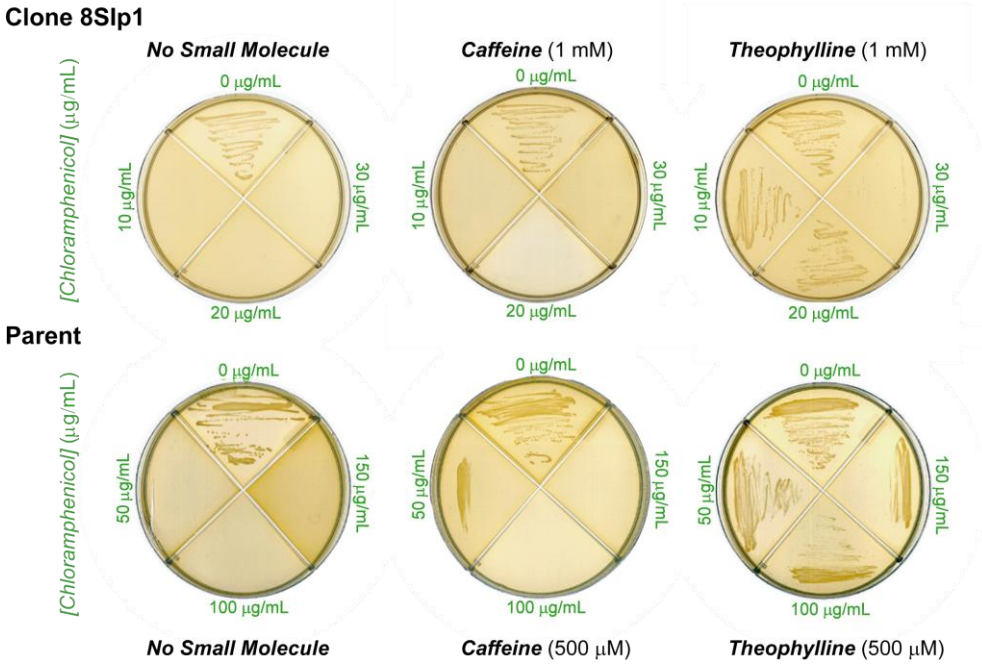

Figure 3.7: Resistance Assays with the Parent and 8SIp1 Riboswitches. *E. coli* cells harboring either the parent or 8SIp1 synthetic riboswitches, cloned upstream of the chloramphenicol resistance gene (*cat*), were grown on agar supplemented with varying chloramphenicol concentrations and small molecules. Riboswitch 8SIp1 requires lower concentrations of chloramphenicol than the parent riboswitch. Clone 8SIp1 also displays no background growth in the presence of caffeine with any concentration of chloramphenicol, unlike the parent.

not necessary because no background β-galactosidase activity appears. Therefore, clone 8ISp1 is clearly superior to the wild-type riboswitch for use in screens. We wondered if clone 8SIp1 would be equally superior in a selection experiment. The synthetic riboswitch from clone 8ISp1 was cloned downstream of the strong *tac* promoter and upstream of an *IS10-cat* fusion gene and introduced into *E. coli*. Cells harboring this construct were grown on LB agar supplemented with ampicillin (50 μg/mL) and either no small molecule, caffeine (1 mM), or theophylline (1 mM), with varying concentrations of chloramphenicol. After 18 h incubation at 37 ^{o}C, the plates were visually inspected for growth. Figure 3.7 shows these results, along with a comparison of the resistance profile of cells harboring the parent riboswitch.[8] Lower concentrations of chloramphenicol are needed for the cells harboring the synthetic riboswitch from clone 8SIp1, as it has lower

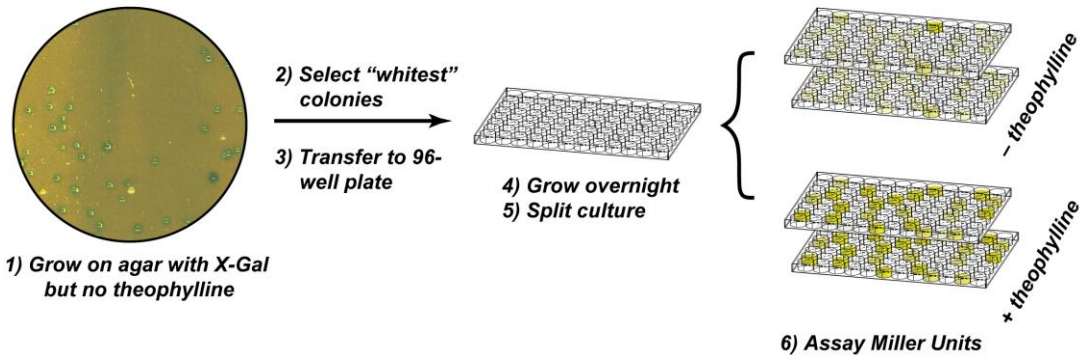

Figure 3.8: Diagram of the 96-Well High-Throughput Assay. Potential riboswitches are identified by plating cells onto agar supplemented with X-gal but no theophylline. A colony-picking robot is used to identify the whitest colonies (low levels of β-galactosidase activity) and inoculate a 96-well plate. After growth, the culture is used to inoculate 96-well plates containing media supplemented with or without theophylline. β-galactosidase activities are then measured, as described.

levels of expression in the presence and absence of theophylline. Also, no background growth is observed for cells containing the riboswitch from clone 8SIp1 on the plate containing caffeine (1 mM) and chloramphenicol (10 μg/mL). This is in contrast to the parent riboswitch, which shows growth on the plate containing caffeine (500 μM) and chloramphenicol (50 μg/mL). Because of these properties, the synthetic riboswitch from clone 8SIp1 should prove to be more sensitive in selection experiments than the parent riboswitch and suggests evolved riboswitches which respond to other ligands should be equally adept for both screens and selections.

3.2.4 Plate-Based Assay for Optimally Functioning Riboswitches

While the stamping assay above led to a number of synthetic riboswitches with optimal characteristics, it requires a visual inspection of colonies in the presence and absence of theophylline, and we desired a more quantitative approach. To this end, we adapted a high-throughput β-galactosidase activity assay,[20,21] which uses 96 well plates, for use in discovering synthetic riboswitches (Figure 3.8). As in the stamping assay,

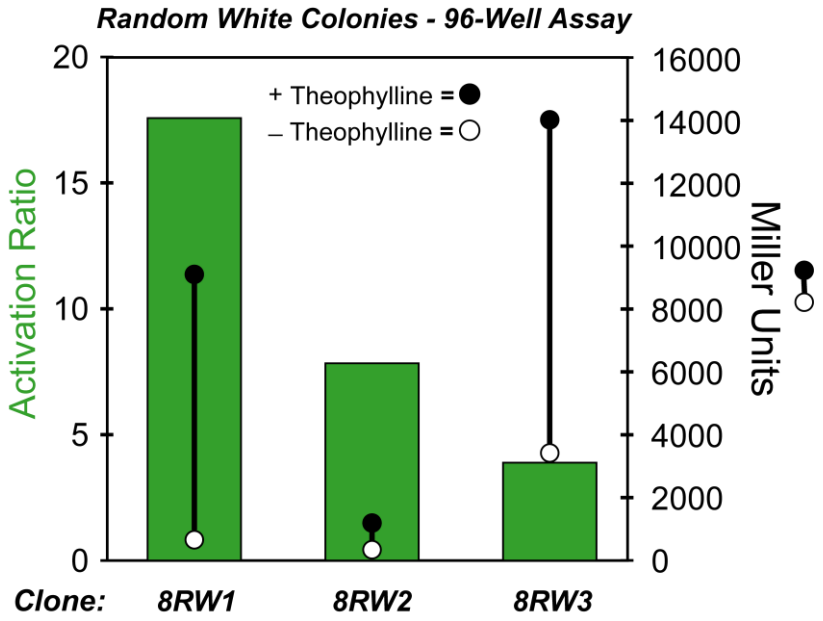

Figure 3.9: β-Galactosidase Activities for Riboswitches Identified from Random White Colonies, Followed by 96-Well β-Galactosidase Assay. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters. Since we did not screen for low β-galactosidase activity in the absence of theophylline, the Miller Units in the absence of theophylline are (sometimes) high.

E. coli cells harboring the randomized synthetic riboswitch library were grown on LB agar supplemented with ampicillin (50 μg/mL) and X-gal, but no theophylline. After growth, the 96 whitest colonies from each plate of approximately 4000 colonies were picked using a colony-picking robot. These clones were used to inoculate LB supplemented with ampicillin in 96-well plates and were grown overnight. A small fraction of the culture from each well was used to inoculate two new 96-well plates that contained LB supplemented with ampicillin (50 μg/mL) and either no small molecule or theophylline (1 mM), in duplicate. These plates were grown at 37 ^{o}C for 2-2.5 h, and the β-galactosidase activity was subsequently assayed.[20,21] We used the activation ratios, determined by dividing the Miller units for individual clones grown in the presence of theophylline by those grown in the absence, to identify functional synthetic riboswitches.

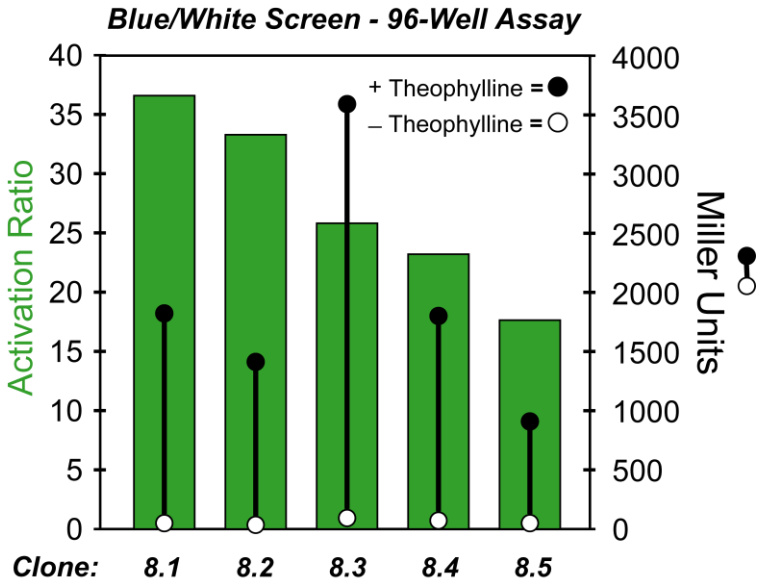

Figure 3.10: β-Galactosidase Activities for Riboswitches Identified from Whitest Colonies from Blue/White Screen, Followed by 96-Well β-Galactosidase Assay. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters. Since we screened for low β-galactosidase activity in the absence of theophylline, the Miller Units in the absence of theophylline are indeed low.

During the course of developing the assay, we encountered a number of problems. At first, we chose to lyse the cells using a standard chloroform and SDS mixture. We discovered that this combination of lysis agents, while useful in the standard Miller Units assay, did not give rise to complete lysing of the cells in a 96-well plate format. A variety of experiments were tried before we settled on using a combination of PopCulture® reagent and lysozyme. We discovered that the optimum method to lyse the cells with this reagent is to use a multichannel pipetor to pipet the cell mixture up and down five times, and to allow the cells to sit at room temperature for 10 minutes. After this treatment, the cells are completely lysed. Also, the OD_{600} of the assayed cultures must be grown within a certain range (0.3-0.5) before lysis, or the assay gives inconsistent results. We tested the modified assay using a single riboswitch clone in all the wells of a 96-well plate and, using the modifications, the assay produced accurate

data, with low standard deviations across all wells of the plate (not shown), suggesting our assay is highly reproducible.

Initially, we did not supplement the LB agar plates with X-gal. We simply picked 96 random colonies from approximately 4000 and used them in our high-throughput assay. Three functional riboswitches were found using this assay (Figure 3.9). Not surprisingly, two of the clones have high levels of β-galactosidase activity in the absence of theophylline, as we did not screen against it. A critical component of the success of this assay is our use of X-gal supplementation to identify synthetic riboswitches with low levels of β-galactosidase activity in the absence of theophylline. Several variants identified from X-gal supplemented plates are listed in Figure 3.10. It should be noted that clone 8.1 was found twice: once using the stamping assay, and once using the 96-well plate assay, suggesting that multiple methods can be used identify a single synthetic riboswitch.

3.2.5 Possible Mechanism of Action for Synthetic Riboswitch Function

In our assays we discovered a variety of variants, and we sequenced several of them. These sequences allowed us to propose a model of riboswitch function. In silico folded[22] structures of two of the clones, from the 5' start of the aptamer to the 3' end of the start codon, are shown in Figure 3.11. In both structures, the minimum energy structures on the left show that the putative RBS is paired in the absence of theophylline. When the theophylline-binding aptamer structure[23-25] is formed on the right hand side of Figure 3.11, the RBS is not paired. The difference in free energy between the minimum energy structures and the structures in which the aptamer secondary structure forms is

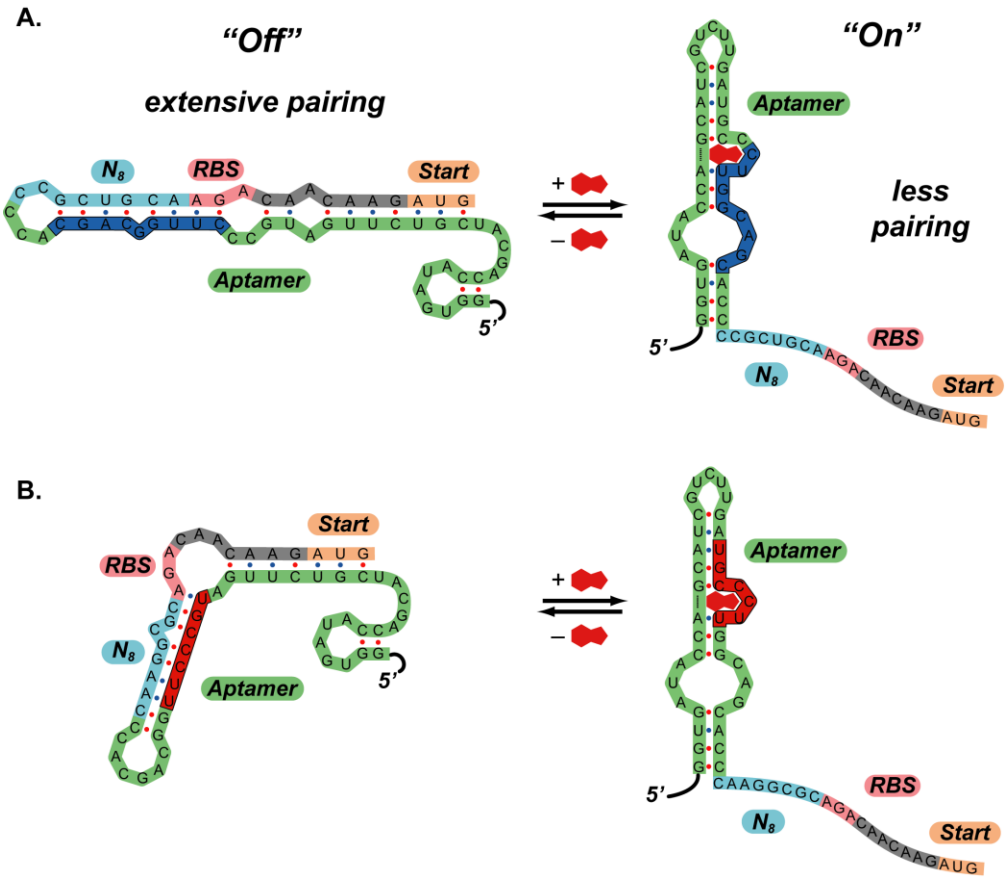

Figure 3.11: Predicted Mechanisms of Action of Synthetic Riboswitches. (A) Riboswitch RNA from clone 8.1 in the absence (left) and presence (right) of theophylline. Without theophylline (left), the ribosome binding site is paired. In the presence of theophylline (right) the ribosome binding site is unpaired. (B) Riboswitch RNA from clone 8.2, in the absence and presence of theophylline. Again, the ribosome binding site is predicted to be unpaired only in the presence of theophylline.

calculated to be less than free energy of theophylline binding to the aptamer (-9.2 kcal/mol), and suggests that thermodynamics drive the conversion of the folded structures into the unfolded structures in the presence of theophylline. Pairing of the RBS is known to inhibit translation of the downstream gene,[26] and unpairing is known to allow translation.[27] In our model, the formation of the aptamer in the presence of theophylline unpairs the RBS and increases translation.

In our previously reported riboswitch,[8] we simply inserted a theophylline aptamer upstream of a reporter gene and we did not determine whether it was dependent on the downstream gene segment, though *mFold* indicated the RBS of the riboswitch paired

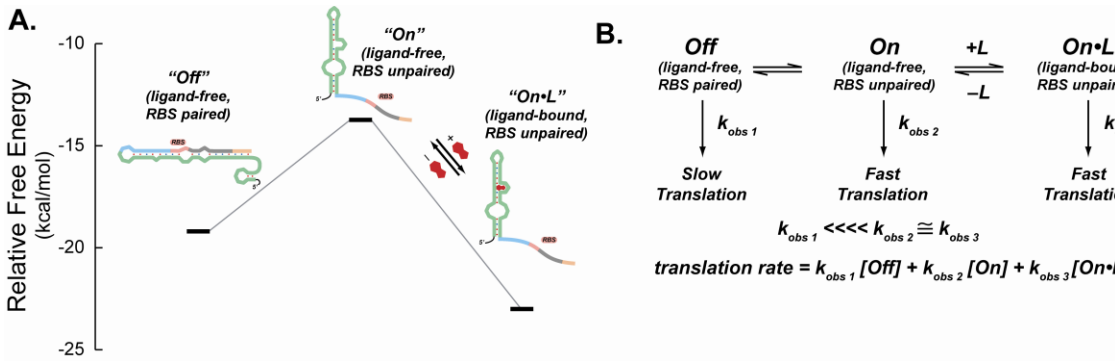

Figure 3.12: Model for Synthetic Riboswitch Function. (A) The predicted structures for riboswitch clone 8.1, as determined in silico. The experimentally determined energy for theophylline binding to the aptamer (-9.2 kcal/mol) is added to the unliganded structure where the aptamer is formed. (B) Kinetic model for synthetic riboswitch function. When the ribosome binding site is paired ("Off" state), translation is slow. The "Off" state is proposed to be in equilibrium with the "On" state, where the ribosome binding site is unpaired and translation is fast. Since this state is not highly populated, background gene expression is low. When ligand is present (L), the eqiulibrum is shifted towards the "On●L" state, and a large increase in translation occurs.

with the downstream gene. To test whether the original riboswitch was dependent on the downstream gene, we deleted the N-terminal IS10 fusion (S.A.L.). If this segment is necessary for function of the original riboswitch, the activity of the deletion mutant will not be affected with theophylline. Indeed, the Miller units of the deletion construct increased dramatically in the absence of theophylline, and its activity was not affected by theophylline, indicating the previous riboswitches were dependent on the downstream gene.

If our model is correct, the functional unit of the newly identified riboswitches is contained solely within the 5' UTR of the RNA, and the downstream sequence should not matter. To test this, we deleted the *IS10* gene segment from clone 8.1 (pSKD8.1-ΔIS10). This construct behaved similarly to clone 8.1, and its activity is theophylline responsive. These results indicate that the region between the aptamer and the start codon is responsible for the function of the identified riboswitches, and is consistent with the in silico predicted RNA secondary structures.

3.2.6 A Model for Synthetic Riboswitch Function

The data indicates that our synthetic riboswitches display low levels of translation in the absence, and high levels of translation in the presence of theophylline for two reasons. In the absence of theophylline, the RBS is paired and translation is inhibited. In the presence of theophylline, the ligand binds to the RNA, and drives the RNA into a conformation where the RBS is unpaired. And it is known that pairing of the RBS inhibits ribosome binding to the RNA.[26-28] Since translation is the slow step in protein production, the position of a pre-equilibrium between RNA structures that have the RBS paired and unpaired will factor into the rate expression for gene expression. Strongly paired sequences will likely show minimal levels of protein expression in the absence of ligand, since translation is inhibited. Even if the lower energy structures equilibrate with higher energy structures (where the RBS is unpaired), the concentration of the higher-energy structures will be low, and the overall translation rate will be minimal. The relative stabilities of these structures will determine the background levels of gene expression in the absence of theophylline. However, if theophylline is added, this model predicts that the population of mRNA structures will shift to ones where the RBS is unpaired, leading to large increases in gene expression.

Figure 3.12 shows the predicted secondary structures and free energies of the putative region responsible for riboswitch function from clone 8.1. In the minimum-energy structure, the RBS is paired. In the middle structure, the aptamer is formed and the RBS is unpaired in the absence of ligand. This structure is higher in energy than the minimal energy structure, but it is not ligand-bound. The final structure on the right shows the structure with the aptamer formed and bound to theophylline, which results in

a lowering in energy of the structure by the free energy of theophylline binding (9.2 kcal/mol).

In the absence of ligand, our equilibrium model predicts a majority of the RNA population will be in the "Off" structure and the RBS will be paired. When theophylline is added, the structure is predicted to be thermodynamically driven toward the "On●L" structure. These observations can be represented in the model shown in Figure 3.12, where the pseudofirst-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. Previous studies predicted that the translation efficiency from the "Off" structure, where the RBS is paired, would be lower in efficiency than the structures where the RBS is unpaired[26-28] ("On" and "On●L"), thus $k_{obs\ 1} <<< k_{obs\ 2} \sim k_{obs\ 3}$. Therefore, in the absence of ligand, the rate of translation would be low, and would increase in the presence of ligand.

While this model fits our experimental observations, it is difficult to determine rate constants and mRNA concentrations in vivo. Though these things are difficult to measure, we can still test certain segments of the model, such as the ability of the RNA structures to equilibrate in the presence of ligand in vitro and in vivo. Demonstrating that a riboswitch mRNA transcript can undergo a ligand-inducible shift from an untranslated state into a translated state in vivo would also provide evidence for the proposed equilibrium model and would argue against a kinetic model, where the fate (or structure) of the mRNA is decided at the moment of transcription.

In a series of in vivo experiments, Sean Lynch showed that after stopping transcription of a riboswitch mRNA within cells grown in the absence of theophylline,

the mRNA can subsequently be turned "On" by adding a high concentration of theophylline to the cells. The details of the experiments will not be presented here, but they support the equilibrium model for riboswitch function, where addition of ligand shifts the population of mRNAs from the "Off" to the "On" state. If the mechanism of the riboswitch was kinetically controlled, the mRNA equilibrium could not have been shifted post transcriptionally.

Since the in vivo data support an equilibrium model in which ligand induces conformational changes, one might expect that ligand-dependent changes could be observed through RNA footprinting studies. To test this possibility, we carried out in vivo footprinting experiments by using the SHAPE method.[29] In this method, *N*-methylisatoic anhydride (NMIA) modifies the 2'-OH of RNA, depending on the local nucleotide flexibility. If the RNA is modified at a certain position, reverse transcriptase will pause at the modified site and fall off the template, leading to a truncated product.[29] Comparison of truncated products versus a known sequencing ladder allows one to map the location and extent of modification at certain positions.

To probe this possibility, we added NMIA to cells that contained a synthetic riboswitch, which were grown with either no small molecule or 1 mM theophylline. Modified RNA was extracted from the cells and used as a template for reverse transcriptase using a radiolabeled primer specific to a region 3' of the riboswitch. The reverse transcription reactions were separated by denaturing PAGE and the banding patterns due to NMIA modification of the riboswitch mRNA from cells grown in the presence and absence of theophylline were compared. We did not observe a difference in NMIA modification of riboswitch mRNA from cells grown in the presence and

absence of theophylline, though we did observe RNA modification that was not present in an untreated control.

SHAPE is a bulk analysis technique and provides an average of how the RNAs in a given population are structured. Therefore, if only a small fraction of mRNAs inside of a cell are in a translatable state, the SHAPE method would not allow their observation since they would be diluted against a background of non-translatable mRNAs. If our hypothesis is correct, removing the aptamer from riboswitch 8.1, but keeping the RBS, will lead to a dramatic increase in translation of the reporter gene. To this end, we removed the aptamer sequence from clone 8.1 and tested its β-galactosidase activity in vivo. Indeed, the construct lacking the aptamer had approximately six times more β-galactosidase activity than the riboswitch-containing construct when grown in the presence of 1 mM theophylline. This result indicates that only ~16% of the transcripts are activated inside cells grown in the presence of 1 mM theophylline, leaving ~84% inactive. Therefore, our in vivo SHAPE experiments correlate with this result and the technique may not have allowed us to "see" the small fraction of translatable mRNAs in vivo.

If our equilibrium model is correct, the riboswitch RNA adopts different secondary structures in the presence and absence of theophylline. Since we could not identify these changes using the in vivo SHAPE technique, we decided to conduct in vitro structure probing of the RNA sequence comprising the 5' start of the aptamer to the 3' end of the start codon of clone 8.1. We first used T1 nuclease to probe the secondary structure of the radiolabled riboswitch RNA, after it was equilibrated in the presence and absence of 1 mM theophylline. T1 nuclease cleaves 3' to single-stranded G

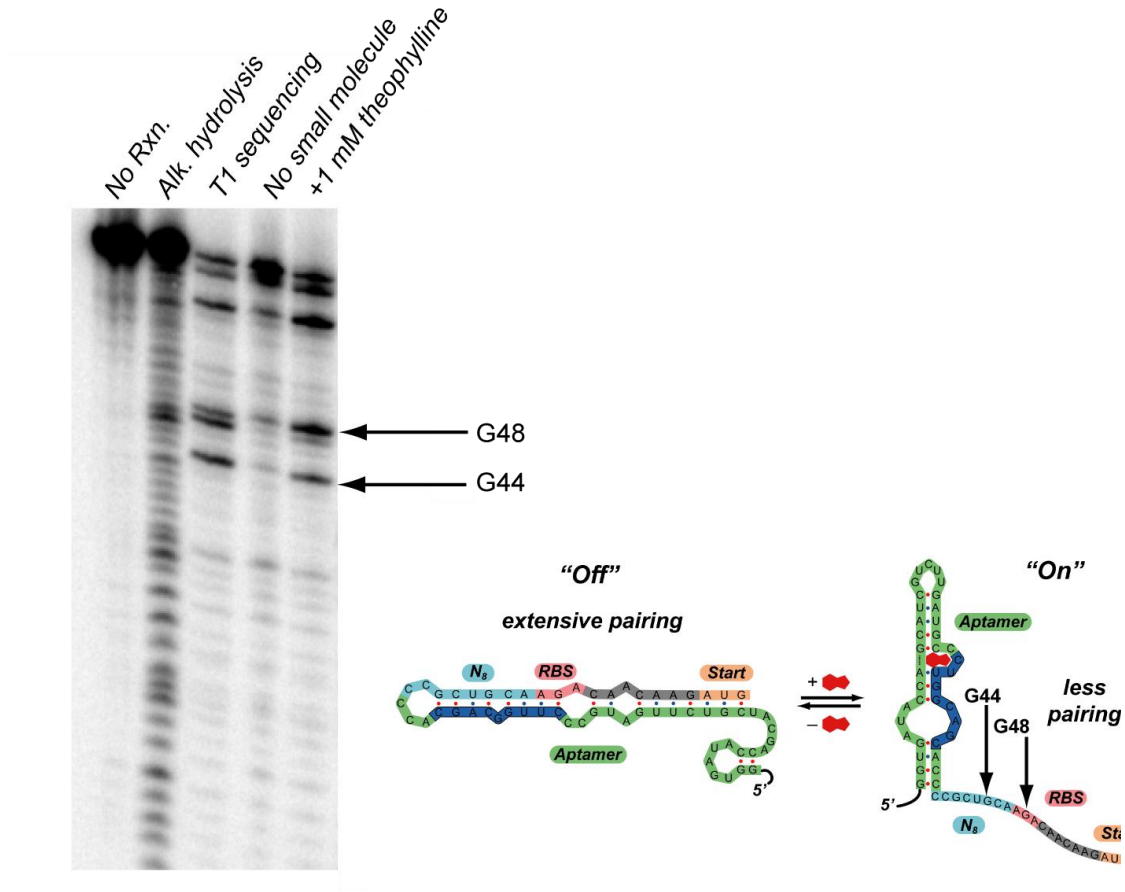

Figure 3.13: T1 Nuclease Structure Probing of Riboswitch Clone 8.1. The 10% denaturing PAGE clearly shows that the G residues within and near the ribosome-binding site are cleaved to a greater extent in the presence of theophylline. These residues are not cleaved to a great extent in the absence of theophylline, supporting the proposed equilibrium model for synthetic riboswitch function.

residues in RNA sequences,[30] and we predicted that the RBS would become unpaired in the presence of theophylline, resulting in more T1 nuclease cleavage in this region. Figure 3.13 shows the denaturing PAGE gel results of our experiment. Clearly, the RBS region is cleaved to a greater extent by T1 nuclease when equilibrated in the presence of 1 mM theophylline. And, in comparison to our in vivo SHAPE experiment where we did not know the internal cellular concentration of theophylline, the concentration of theophylline used in vitro was most likley high enough to drive a vast majority of

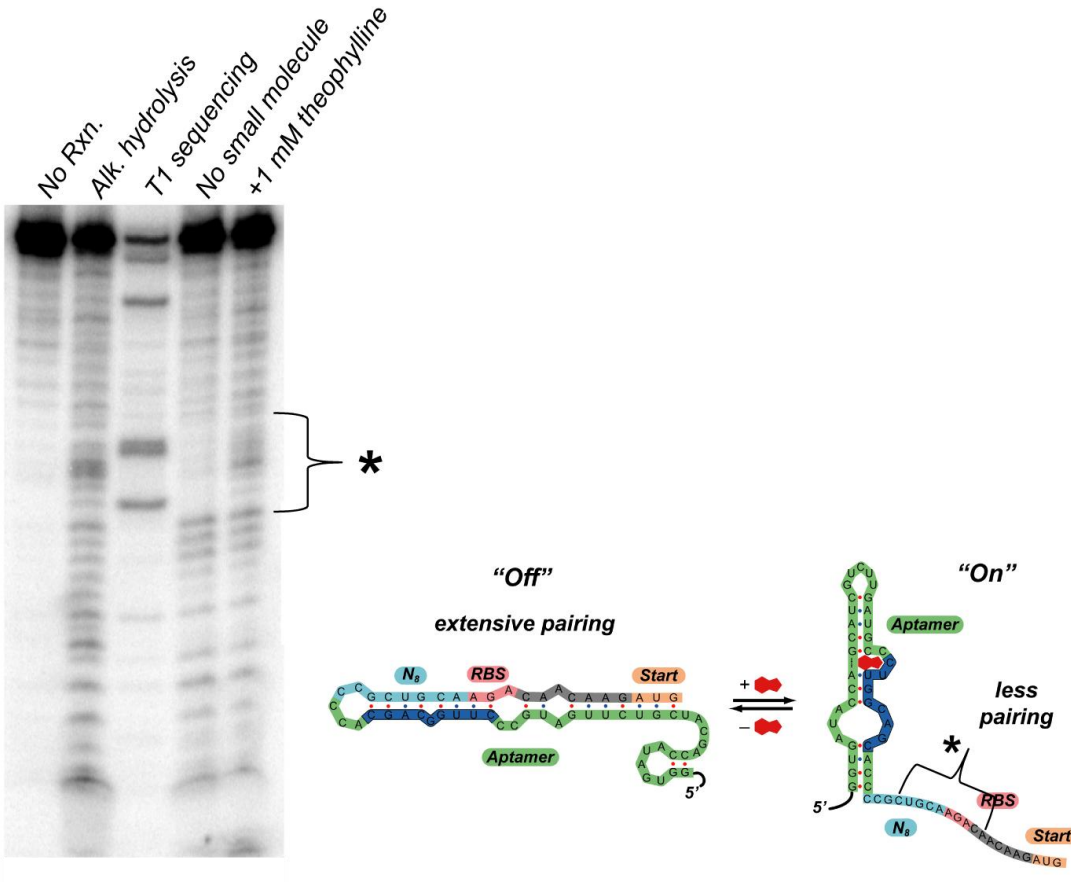

Figure 3.14: In-line Structure Probing of Riboswitch Clone 8.1. The 10% denaturing PAGE clearly shows that the residues within and near the ribosome-binding site self-cleave to a greater extent in the presence of theophylline. These residues do not self-cleave to a great extent in the absence of theophylline, supporting the proposed equilibrium model for synthetic riboswitch function.

transcripts into the ligand-bound structure (since the aptamer K_d is 100 nM), allowing us to observe the structural changes with this technique.

To support our nuclease digestion results, we also performed in-line probing on the same labeled RNA.[31] In-line probing tests the local nucleotide flexibility of each nucleotide of an RNA sequence, based on the environment of each 2' -OH. If a 2' -OH is not constrained, it participates in an intramolecular transesterification reaction and attacks the 5' phosphate group closest to it, leading to cleavage of the RNA chain.[31] After the cleavage reactions have occurred, the labeled RNA is separated on a denaturing PAGE gel next to a sequencing ladder. The relative flexibility of each nucleotide is then

analyzed based on the extent of cleavage at each particular position - more cleavage generally correlates with a more structured RNA, and less cleavage correlates with unstructured (or unpaired) RNA. Radiolabled RNA from clone 8.1 was incubated at room temperature in the presence and absence of 1 mM theophylline for 40 h to allow equilibration and intramolecular cleavage of the RNA. Figure 3.14 shows the results of our in-line probing experiment after separation by denaturing PAGE. Gratifyingly, the region containing the RBS is cleaved to a greater extent in the presence of theophylline, suggesting this region is unpaired in the presence of theophylline. Taken together, our nuclease digestion and in-line probing results suggest that the RBS regions of our riboswitches are paired in the absence of theophylline, and unpaired in the presence of theophylline. These data lend support to our equilibrium model for riboswitch function.

3.2.7 Potential Design Implications for Synthetic Riboswitches

Our current data support the equilibrium model for synthetic riboswitch function, and the model may provide insight for how to design other synthetic riboswitches. Our model suggests that it is important to balance the relative free energies of the "off state", the ligand-free "on state," and the ligand-bound "on state." If we look at the differences in energy between ligand-free "on state" and the ligand-bound "on state", it is determined by the free energy of theophylline binding to the aptamer. This puts a limit on the potential free energy state for the "off state" structures. If the free energy of the "off state" is very close to the free energy of the ligand-free "on state", there will be a large amount of background translation in the absence of ligand (due to interconversion). But, if the free energy of the "off state" is much lower than the ligand free "on state", the

presence of ligand may not result in a large proportion of ligand bound "on state" structures. Therefore, addition of ligand would not result in an increase in gene expression. These guidelines may help in the design of future synthetic riboswitches.

3.3 Conclusion

Riboswitches are genetic control elements which are composed solely of RNA and modulate gene expression in a small molecule dependent fashion.[1] Additionally, synthetic riboswitches have a number of potential applications, including in genetic screens or selections for specific small molecules.[8] Riboswitches require a ligand binding (aptamer) domain and an expression platform, which converts the ligand binding event into changes in gene expression.[2] Many methods are available to select small molecule-binding aptamers,[14] but there are few ways to convert these selected aptamers into synthetic riboswitches inside cells. In this Chapter, we presented a number of methods that identify synthetic riboswitches that have low levels of gene expression in the absence of ligand and high levels of gene expression in the presence of ligand. These synthetic riboswitches have similar characteristics to natural riboswitches (in comparison to their regulatory factors). Analysis of the sequences of these riboswitches also allowed us to propose and test a model of their function which potentially explains their mechanism of action. In total, we believe the methods presented here will be useful for creating new synthetic riboswitches which respond to a number of additional small molecules.

3.4 Experimental

General Considerations

All plasmid manipulations utilized standard cloning techniques.[32] 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), ampicillin, and chloramphenicol 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 between the mTCT4-8 theophylline aptamer and the RBS of the *IS10-LacZ*, *IS10-ccd*B, or *IS10-cat* reporter genes.

Selection System for Optimal Riboswitch Function

Library transformations of *E. coli* harboring pSKD668.5 (*ccd*B selection plasmid) were plated on LB agar supplemented with ampicillin (50 μg/mL) and grown at 37 °C for 18 h. Surviving colonies were scraped into liquid LB media (5 mL) supplemented with ampicillin (50 μg/mL) and grown with shaking at 37 °C for 6 h. A fraction (1.5 mL) of culture was used for a miniprep and the isolated plasmids were digested with *Nde*I and

*Hind*III, and the riboswitch-containing DNA fragment was separated by electrophoresis on a 1% agarose gel and purified. The resulting fragment was cloned into the *Nde*I and *Hind*III sites of pSKD314[8] and transformed into TOP10 cells. Transformations were plated on LB agar supplemented with ampicillin (50 μg/mL) and were grown at 37 °C for 14 h. The resulting lawn of colonies was scraped into liquid LB media (5 mL) supplemented with ampicillin (50 μg/mL) and grown with shaking at 37 °C for 6 h. Approximately 10^5 cells were resuspended in 100 μL liquid LB and equal aliquots were plated onto LB agar supplemented with ampicillin (50 μg/mL), theophylline (500 uM), and varying concentrations of chloramphenicol (0 μg/mL, 25 μg/mL, 50 μg/mL, and 75 μg/mL). Cells were grown at 37 °C for 18 h and single colonies were used to inoculate liquid LB media (5 mL) supplemented with ampicillin (50 μg/mL) and grown with shaking overnight at 37 °C. A fraction (1.5 mL) of each culture was used for a miniprep, and the resulting plasmids were digested with *Kpn*I and *Hind*III. The riboswitch encoding band was separated by electrophoresis on a 1% agarose gel, purified, and cloned into the *Kpn*I and *Hind*III sites of pSKD445.1. The plasmids were introduced into TOP10 cells and plated on LB agar supplemented with ampicillin (50 μg/mL) and grown at 37 °C for 14 h. Individual colonies were assayed using the method of Miller.[19]

High-throughput "Stamping" Assay

Library transformations in *E. coli* in plasmid pSKD445.1 (*tac* promoter-*IS10-lac*Z gene plasmid) or pLacZU1hpII[19] 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 96 whitest colonies from each plate were picked using a Genetix QPix2 colony picking robot and the colonies on the end of the pinheads were physically stamped on 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%) and either no small molecule or 1 mM theophylline. Plates containing the stamped colonies were grown at 37 °C for 14 h and then they were set at 4 °C for 4 h. Colonies were inspected for blue color in the presence of theophylline; and for white color in the absence of theophylline. Colonies displaying these characteristics were subcultured and subjected to assay by the method of Miller.[19]

96-well Plate-Based Screen

Library transformations pSKD445.1, harbored in *E. coli*, 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 ºC 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 (1 mM). Plates were incubated for approximately 2.5 h at 37 ºC 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:

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

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 using the method of Miller.

Resistance Assay for Clone 8SIp1

The riboswitch segment from plasmid 8SIp1 was digested with *Kpn*I and *Hind*III and separated on a 1% agarose gel. After purification of the riboswitch containing fragment, it was cloned into the *Kpn*I and *Hind*III sites in pSKD314[8] and introduced into TOP10 cells. Selection assays were performed as reported previously.[8]

***In vivo* NMIA Modification Reaction**

An overnight culture harboring clone 8.2 (400 µL) was used to inoculate two 50 mL cultures of LB/ampicillin (50 µg/mL) supplemented with either no small molecule or theophylline (1 mM). Cultures were shaken at 37ºC until the OD_{600} was 0.4 and were split into two 10 mL volumes in 50 mL Falcon tubes. A solution of NMIA in DMSO (1 mL, 300 mM NMIA, ~30 mM final concentration) was added to one tube and DMSO alone (1 mL) was added to the other. The tubes were shaken at 37 ºC for 45 min. (~5 hydrolysis half lives for NMIA). Cells were pelleted by centrifugation after NMIA modification. Total RNA was extracted from the cells by the addition of H_2O (1.5 mL) and phenol (1.5 mL, pH 4.3) followed by vortexing and heating (55 ºC, 5 min). The mixture was centrifuged at 13000 g for 10 min. at room temp. The aqueous layer was

removed, extracted with phenol (1.5 mL, pH 4.3) and centrifuged. The aqueous layer was separated and the RNA was precipitated with ethanol. Total RNA was resuspended in 50 µL of DEPC-treated H_2O.

Reverse transcription reactions were setup using total RNA as a template. NMIA modified or unmodified total RNA (2.5 µg) was added to a tube with 3 µL folding buffer (333 mM NaCl; 333 mM HEPES, pH 8.0; 33.3 $MgCl_2$) with 3 µL (~0.2 pmol) 5'-[^{32}P] labeled primer SKD-432 (which anneals 25 bp 3' of the start codon) and DEPC-H_2O to obtain a total volume of 13 µL. Tubes were heated to 65 ºC for 5 min. then set at 35 ºC for 20 min. Reverse transcription buffer (6 µL; 167 mM Tris, pH 8.3; 250 mM KCl; 10 mM $MgCl_2$; 1.67 mM each dNTP; 16.7 mM DTT) was added to each tube. Tubes were heated to 48 ºC for 1 min. followed by addition of 1 µL Superscript III (Invitrogen). Tubes were set at 48 ºC for 2 min. and the reactions were stopped by the addition of NaOH (1 µL, 4 M) followed by heating (95 ºC, 5 min). Reactions were mixed with 29 µL of neutralizing gel loading solution (29 µL; 4:25 (v/v) 1 M unbuffered Tris-HCl and stop dye [formamide, 0.5x TBE, 50 mM EDTA 0.01% (w/v) bromophenol blue].

An in vitro transcribed RNA corresponding to the riboswitch-containing region of clone 8.2 (~4 pmol) was used as a sequencing standard as described.[29] The reverse transcription reactions and sequencing reactions were separated using denaturing gel electrophoresis (7 M urea; 8% (29:1) acrylamide: bisacrylamide), (0.4 mm x 30 cm x 32 cm) and were imaged using a phosphorimager.

Reactivity profiles for the (+) and (-) NMIA reactions for total RNA from cultures grown in the presence and absence of 1 mM theophylline were compared using a pixel

intensity plot generated using NIH Image that was plotted in relation to the sequencing ladder.

Nuclease Probing

Primers SKD-580 and SKD-581 (clone 8.1) were amplified using PCR (20 cycles) and purified using a minelute kit (Qiagen). A fraction of the PCR product was used in a T7 Flash transcription kit reaction (Epicentre), separated using 10% denaturing PAGE, and eluted overnight in dH_2O. Eluted RNA was ethanol precipitated, dephosphorylated using calf intestinal phosphatase according to the manufacturer (New England Biolabs) and purified using a minelute kit. The dephosphorylated RNA was subsequently phosphorylated using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (10 µCi, MP Biomedical) according to the manufacturer, followed by purification with a minelute kit.

Radiolabeled RNA (100,000 counts per minute (cpm)) was added to separate tubes containing 1X RNA structure probing buffer (Ambion), 6 mM $MgCl_2$ and either no small molecule or 1 mM theophylline in a final volume of 9 µL. After mixing, the tubes were heated to 70 °C for 2 min and then cooled to 4 °C for 2 min. 1 µL T1 nuclease (Ambion) was added to each tube, mixed, and set at room temperature for 15 min. The reactions were quenched by adding 10 µL of stop buffer (95% formamide, bromophenol blue, 50 mM EDTA) and 3-10 µL fractions were separated by 10% denaturing PAGE. After electrophoresis, the gel was dried, exposed to a Phosphorimager screen (GE Healthcare) imaged, and analyzed with ImageJ software (National Institutes of Health).

T1 nuclease sequencing and alkaline hydrolysis reactions were setup using radiolabled RNA (100,000 cpm) according to the manufacturer (Ambion). These were quenched as before.

SKD-580
5′-TTC*TAATACGACTCACTATA***GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACC**
(*T7 promoter*, **mTCT8-4 aptamer**)

SKD-581
5′-CATCTTGTTGTCTTGCAGCGG**GGTGCTGCCAAGGGCATCAAGACGATGCTGGTATCACC**

In-line Probing

Primers SKD580 and SKD581 (clone 8.1) were amplified using PCR (20 cycles) and purified using a minelute kit (Qiagen). A fraction of the PCR product was used in a T7 Flash transcription kit reaction (Epicentre), separated using 10% denaturing PAGE, and eluted overnight in dH_2O. Eluted RNA was precipitated and dephosphorylated using calf intestinal phosphatase according to the manufacturer (New England Biolabs) and purified using a minelute kit. The dephosphorylated RNA was subsequently phosphorylated using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (10 µCi, MP Biomedical) according to the manufacturer, followed by purification with a minelute kit.

Radiolabeled RNA (100,000 cpm) was added to separate tubes containing in-line probing buffer (50 mM Tris-HCl [pH 8.5], 20 mM $MgCl_2$), and either no small molecule or 1 mM theophylline in a 10 µL final volume. After mixing, the reactions were set at 25 °C for 40 h. Reactions were quenched by addition of 10 µL stop buffer (Ambion) and 3-10 µL fractions were separated by electrophoresis on a 10% denaturing PAGE. The gel was dried, exposed to a Phosphorimager screen, and imaged as before.

T1 nuclease sequencing and alkaline hydrolysis reactions were setup using radiolabled RNA (100,000 cpm) according to the manufacturer (Ambion). These were quenched as before.

Clone	Identification Method	Linker Sequence	Miller Units (0 mM)	Miller Units (1.0 mM)	Activation Ratio
8CC1	*ccd*B/*cat* selection	CGCAGCTC	1697	7219*	4.2
8CC2	*ccd*B/*cat* selection	CCGATGGC	1537	4179*	2.7
8SIp1	Stamping	CCGCTGCA	N.D.	7.1	N.A.
8SIp2	Stamping	CCTGCCGC	2.1	14.0	6.5
8Stp1	Stamping		141	3611	25.6
8Stp2	Stamping		78	1785	22.8
8Stp3	Stamping		51	910	17.8
8Stp4	Stamping		103	1657	16.1
8RW1	Random white/ 96-well plate-based assay		531	9144	17.2
8RW2	Random white/ 96-well plate-based assay		105	810	7.7
8RW3	Random white/ 96-well plate-based assay		3452	14029	4.1
8.1	Blue-white screening/ 96-well plate-based assay	CCGCTGCA	51	1820	36
8.2	Blue-white screening/ 96-well plate-based assay	CAAGGCGC	43	1420	33
8.3	Blue-white screening/ 96-well plate-based assay	AGCTGCAA	141	3611	26
8.4	Blue-white screening/ 96-well plate-based assay	CAGCTGTC	78	1785	23
8.5	Blue-white screening/ 96-well plate-based assay	CGCTGTCC	51	910	18

Table 3.1: Synthetic Riboswitch Characteristics. Clones found and the identification method used are shown, as well as the linker sequence (unfortunately, we did not sequence all clones). The Miller Units in the presence and absence of theophylline, as well as the resulting activation ratios are also shown. (* = 500 µM theophylline)

3.5 References

(1) Nudler, E.; Mironov, A. S. *Trends Biochem. Sci.* **2004**, *29*, 11-17.

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

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

(4) Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R. *Genes Dev.* **2003**, *17*, 2688-2697.

(5) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952-956.

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

(7) Thore, S.; Leibundgut, M.; Ban, N. *Science* **2006**, *312*, 1208-1211.

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

(9) Grate, D.; Wilson, C. *Bioorg. Med. Chem.* **2001**, *9*, 2565-2570.

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

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

(12) Suess, B.; Hanson, S.; Berens, C.; Fink, B.; Schroeder, R.; Hillen, W. *Nucleic Acids Res.* **2003**, *31*, 1853-1858.

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

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

(15) Buskirk, A. R.; Landrigan, A.; Liu, D. R. *Chem. Biol.* **2004**, *11*, 1157-1163.

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

(17) de Boer, H. A.; Comstock, L. J.; Vasser, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 21-25.

(18) Bernard, P.; Gabant, P.; Bahassi, E. M.; Couturier, M. *Gene* **1994**, *148*, 71-74.

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

(20) Bignon, C.; Daniel, N.; Djiane, J. *Biotechniques* **1993**, *15*, 243-246.

(21) Griffith, K. L.; Wolf, R. E., Jr. *Biochem. Biophys. Res. Commun.* **2002**, *290*, 397-402.

(22) Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406-3415.

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

(24) Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J. P.; Pardi, A. *Nat. Struct. Biol.* **1997**, *4*, 644-649.

(25) Zimmermann, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186-9192.

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

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

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

(29) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223-4231.

(30) Knapp, G. *Methods Enzymol.* **1989**, *180*, 192-212.

(31) Soukup, G. A.; Breaker, R. R. *RNA* **1999**, *5*, 1308-1325.

(32) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001.

Chapter 4 Engineering Ligand-Activated Genetic Control Elements in *Acinetobacter baylyi* ADP1

4.1 Introduction

Model organisms provide a wealth of information through experimentation that is, often, readily transferrable to other similarly related organisms.[1] A partial listing of these includes: *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and *Tetrahymena thermophila*.[1] These organisms were chosen, mainly, because they are easily manipulated using standard laboratory techniques; they have relatively simple genetics; and they grow rapidly.[2] For prokaryotes, the model organism of choice for greater than 60 years has been, and continues to be, *Escherichia coli* - which is a Gram-negative γ-proteobacterium - because it too is readily manipulated using standard laboratory techniques, grows rapidly, and has well understood genetics.[3] Even though *E. coli* has a number of desirable features, it may not be the best prokaryotic model organism. For one, it is not naturally competent; therefore, it must be made competent.[4] And it has no natural recombination capability; thus, to modify the *E. coli* chromosome, heterologous recombinases must be introduced and expressed.[5] Because of these traits, some researchers desire an *E. coli* substitute.

One organism, *Acinetobacter baylyi* ADP1, may prove to be a viable, and superior, substitute for *E. coli*. Like *E. coli*, its genome sequence is known;[6] it is a Gram-negative γ-proteobacterium;[6] it grows optimally at standard laboratory temperatures (30 ^{o}C to 37 ^{o}C);[2] it can harbor plasmid DNA;[7] and it can metabolize a variety of carbon sources.[6,8] However, unlike *E. coli*, *A. baylyi* is naturally competent and can take up DNA from the environment when it is in the mid-log phase of growth.[8,9] The origins and survival advantage of this trait are unclear, but they are certainly desirable for researchers. In addition, *A. baylyi* harbors recombinases and can carry out the homology-

directed recombination of linear DNA, which can be harnessed by researchers to easily add or delete large segments of DNA into or out of the *A. baylyi* chromosome.[2,10] And, recently, a complete library of single gene deletions of *A. baylyi* was constructed,[11] which comprises a powerful toolkit for the study of *A. baylyi*.

Although *A. baylyi* has a number of advantages over *E. coli*, its use is not widespread. Perhaps this is due to the relatively modest number of tools (outside of the knockout collection) that have been developed for its study. For example, there are limited methods available for the conditional control of gene expression in *A. baylyi*, and the commonly used IPTG-inducible *lac* repressor system does not function within the organism.[12] It is not known why this occurs, but it is probably due to misfolding of the LacI protein. Other systems, such as the TetR-tetracycline system,[13] may function within the organism, but this has not been reported. Other gene control systems for use in *A. baylyi* are worth exploring.

Synthetic riboswitches may circumvent the problems associated with the protein-based systems mentioned above. They are composed entirely of RNA and function using simple mechanisms,[14] unlike protein-based gene control systems. Additionally, natural riboswitches are found in a vast number of bacteria species,[15] including *A. baylyi*,[16] suggesting synthetic riboswitches would also function within the organism.

Herein, we demonstrate that synthetic riboswitches, originally optimized in *E. coli*, function within *A. baylyi* and their responses are presented. In an effort to identify superior switches, we also adapted our previous high-throughput screen[17] for use with *A. baylyi*. From our screening, we identified several functioning variants and our best variant has an activation ratio of 51, which is better than any of the riboswitches

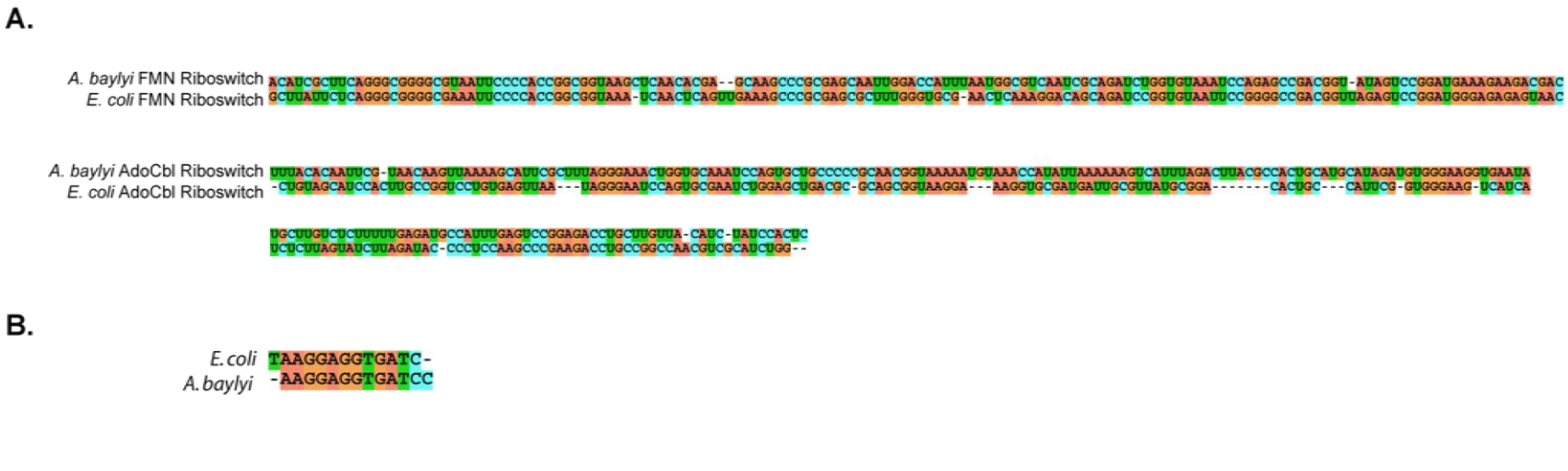

Figure 4.1: Sequence Alignments. (A) Alignment of flavin mononucleotide (FMN) and adenosylcobalamin (AdoCbl) riboswitches from *A. baylyi* and *E. coli*. (B) Alignment of 3 ' termini of 16S rRNA from *A. baylyi* and *E. coli*. The sequences shown are the reverse complements. Their similarity indicates they recognize similar ribosome binding sites.

optimized in *E. coli* that were subsequently expressed in *A. baylyi*. Our results suggest that the properties of synthetic riboswitches are sometimes different when transferred between even closely related organisms and that the axiom "you get what you select for"[18] holds true.

4.2 Results and Discussion

We reasoned that synthetic riboswitches which function in *E. coli* would also function within *A. baylyi*, and we based this on three facts. First, *A. baylyi* and *E. coli* are closely related, since they are both γ-proteobacteria.[3,6] Second, two different classes of natural riboswitches (thiamine pyrophosphate and flavin mononucleotide) are found in both organisms,[16] and their sequence alignments show they are similar (Figure 4.1). And, third, the 3' termini of the 16S rRNA from both organisms are similar, and this region of rRNA is responsible for recognizing ribosome binding sites (Figure 4.1). Taken together, the similarities between these organisms suggest that synthetic riboswitches optimized in *E. coli* should also function within *A. baylyi*.

4.2.1 Introduction of Synthetic Riboswitches from *E. coli* into *A. baylyi*

Before we introduced synthetic riboswitches into *A. baylyi*, we had to decide on an appropriate strategy. Initially, we believed that integration of a reporter gene, under the control of a riboswitch, into *A. baylyi*'s chromosome would be the simplest method to achieve our goals. We chose this strategy and integrated a *gus*A reporter gene, under the control of a theophylline-activated riboswitch, into the chromosome of *A. baylyi*. While the construct functioned in a theophylline-dependent manner, the resulting

β-glucuronidase activity at 1 mM was extremely low (between 1 to 30 Miller Units), which made characterization of the system difficult. This was most likely due to the fact that only a single copy of the reporter gene was contained within each cell, which resulted in low expression levels.

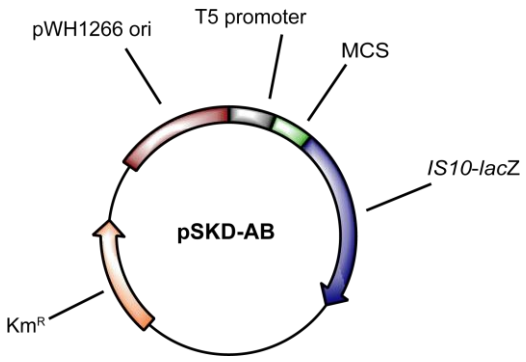

Figure 4.2: Plasmid Diagram of pSKD-AB. The plasmid contains a T5 promoter, multiple cloning site (MCS), and an *IS10-lac*Z reporter gene. The pWH1266 origin allows it to replicate in either *E. coli* or *A. baylyi*.

In addition to chromosome insertions, *A. baylyi* can also harbor plasmids, and we reasoned that if we introduced a riboswitch-controlled reporter gene into an *A. baylyi*-compatible plasmid, the multiple copy number of the plasmid would result in a higher level of reporter gene output inside the cells. To test this, we introduced an *IS10-lac*Z reporter gene into an *E. coli-A.baylyi* shuttle plasmid that contains an origin of replication that functions in both organisms[7] (Figure 4.2). The reporter gene was placed under the control of a constitutive T5 promoter,[19] and restriction sites for cloning synthetic riboswitches into the 5' untranslated region (5' UTR) of the reporter gene were also added, resulting in pSKD-AB.

In our lab, several theophylline-activated synthetic riboswitches with outstanding activation ratios have been isolated from large libraries harbored in *E. coli* using fluorescence activated cell sorting (Sean Lynch). These riboswitches have several

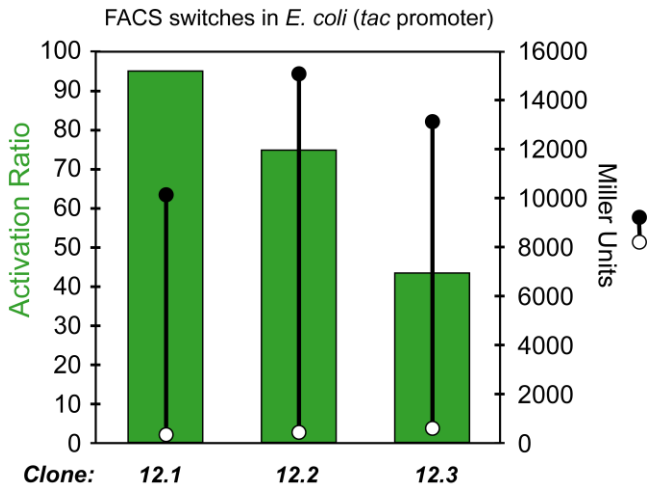

Figure 4.3: β-Galactosidase Activities for FACS Riboswitches Expressed in *E. coli*. The riboswitches, discovered using FACS, are contained in a pUC-based vector, and are under the control of the *tac* promoter. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters. Measured at 37 ^{o}C.

desirable characteristics (Figure 4.3). First, they have very low levels of gene expression in the absence of ligand. Second, they have very high levels of gene expression in the presence of ligand. For example, one of these switches has an activation ratio of 94 in *E. coli*, which is calculated by dividing the reporter gene expression level of cells grown in the presence of 1 mM theophylline by that of cells grown in the absence of ligand. Because of their ligand response characteristics, we subcloned three of the FACS switches into pSKD-AB, separately, and introduced them into *A. baylyi*.

The Miller Units[20] of these constructs are shown in Figure 4.4. In comparison to the characteristics of the riboswitches when expressed in *E. coli*, their expression levels in *A. baylyi* are much lower in the presence of ligand; and this leads to lowered activation ratios. Although the riboswitches expressed in *A. baylyi* do not have high activation ratios, they are theophylline responsive, which indicates that our original hypothesis on the transferability of synthetic riboswitches is correct.

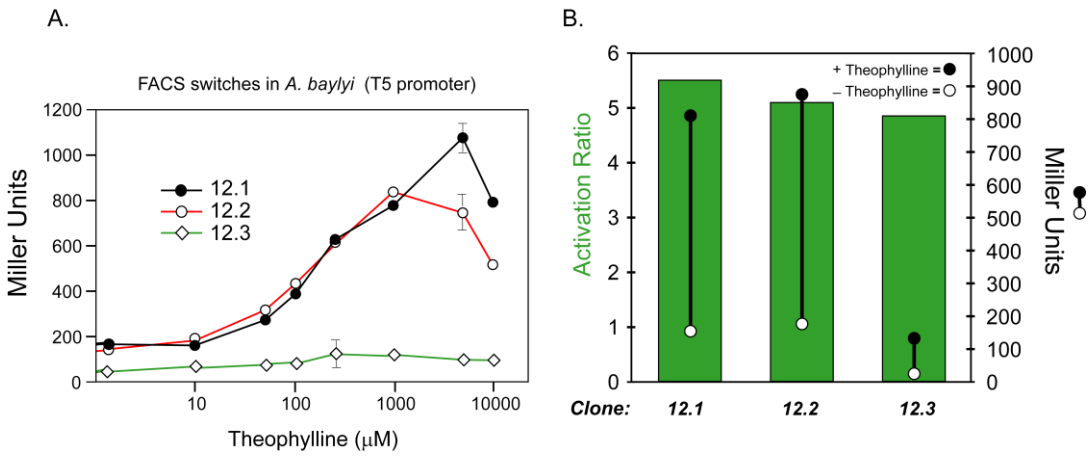

Figure 4.4: β-Galactosidase Activities for FACS Riboswitches Expressed in *A. baylyi* (T5 promoter). (A) Dose response curves for the FACS riboswitches cloned into pSKD-AB, under the control of the T5 promoter, in *A. baylyi*. Measured at 37 °C. (B) Activities corresponding to the dose response curves are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters.

Since the FACS riboswitches did not have high levels of expression in the presence of ligand when expressed in *A. baylyi*, we attempted to optimize this parameter. The FACS riboswitches were originally under the control of the constitutive *tac* promoter in a pUC-based vector[17] in *E. coli*, and we believed the T5 promoter in pSKD-AB was potentially responsible for the lackluster function of the riboswitches when expressed in *A. baylyi*. To test this, we replaced the T5 promoter in pSKD-AB with the *tac* promoter,[21] and subcloned each of the FACS riboswitches into the resulting vector, separately, and introduced them into *A. baylyi*. Figure 4.5 shows the resulting Miller Units for these constructs. While the absolute expression levels increased in the presence of ligand, the expression levels in the absence of ligand also increased; and this resulted in riboswitches with, again, low activation ratios.

Changing the promoter in pSKD-AB did not improve the activation ratios of the FACS riboswitches when expressed in *A. baylyi*, and we wondered if some other element of the plasmid was potentially responsible for this. If the *A. baylyi* origin of replication

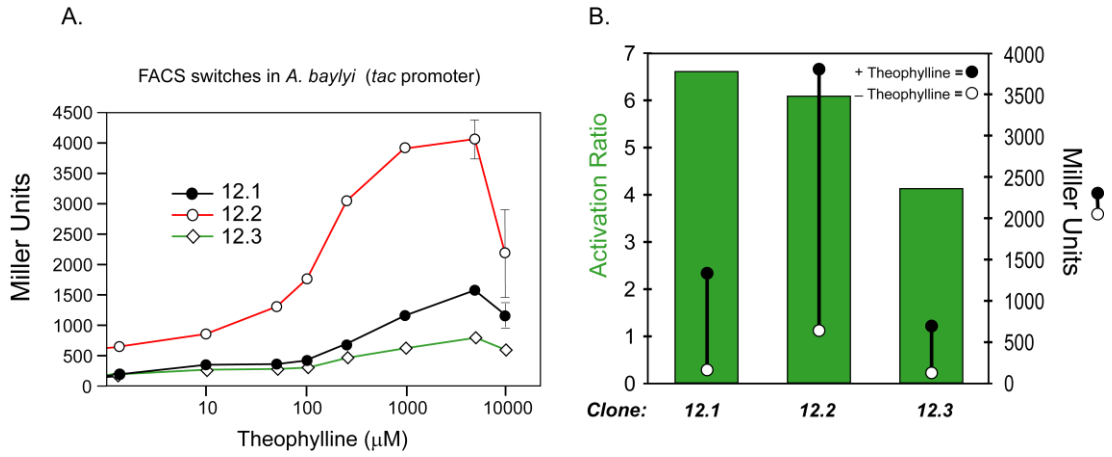

Figure 4.5: β-Galactosidase Activities for FACS Riboswitches Expressed in *A. baylyi* (*tac* promoter). (A) Dose response curves for the FACS riboswitches cloned into pSKD-AB, under the control of the *tac* promoter, in *A. baylyi*. Measured at 37 °C. (B) Activities corresponding to the dose response curves are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters.

sequence contains a promoter, it may affect the downstream construct by transcribing additional bases onto the 5' UTR, which may affect the function of the riboswitches. A previous study of the *A. baylyi* origin of replication in pSKD-AB found that it contains internal promoter sequences.[7] However, transcription does not continue outside of the origin sequence, due to the presence of transcriptional termination sequences at its 5' and 3' ends.[7] Therefore, transcription of our riboswitch constructs most likely occurs from the single promoter sequence we introduced into pSKD-AB.

pSKD-AB is a shuttle plasmid that can be harbored by either *A. baylyi* or *E. coli*, and we wished to know whether the FACS riboswitches in pSKD-AB, under the control of the *tac* promoter, would function when expressed in *E. coli*. *E. coli* cells were transformed with riboswitch constructs in pSKD-AB, and the Miller Units were measured. Figure 4.6 shows the results of these experiments and, clearly, the function of the synthetic riboswitches in pSKD-AB, when expressed in *E. coli*, is different from the original pUC-based vector constructs shown in Figure 4.3. The Miller Units of the

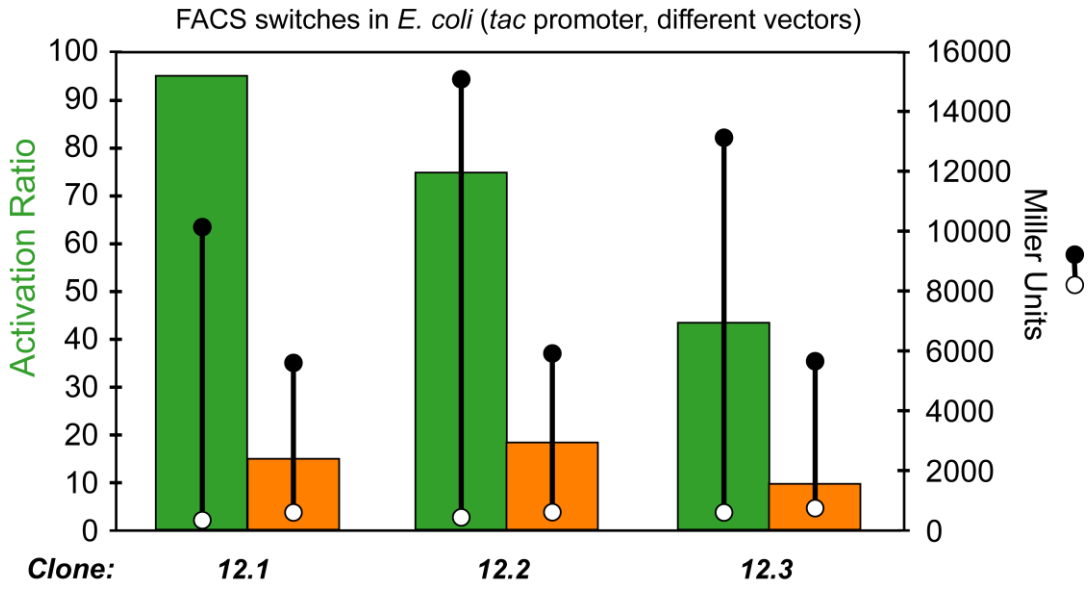

Figure 4.6: β-Galactosidase Activities for FACS Riboswitches Expressed in *E. coli* (different vectors). Activities of the FACS riboswitches in pSKD-AB (orange bars) or pUC-based vector (green bars), under the control of the *tac* promoter. Activities are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters. Measured at 37 °C.

constructs grown in the presence of ligand are not as large, and the Miller Units in the absence of ligand are larger than the pUC-based vectors, which leads to activation ratios below 18. Thus, pSKD-AB affects the FACS riboswitches through some unknown mechanism in *E. coli*, and, potentially, in *A. baylyi*.

4.2.2 Screening for Synthetic Riboswitches in *A. baylyi*

The FACS riboswitches do not function optimally when cloned into pSKD-AB and we decided to screen for synthetic riboswitches in *A. baylyi* that have more desirable properties. We adapted our previous high-throughput screen[17] for use with *A. baylyi* and used this to screen a randomized library that contained an N11 region, including the ribosome binding site, between the theophylline aptamer and the start codon of

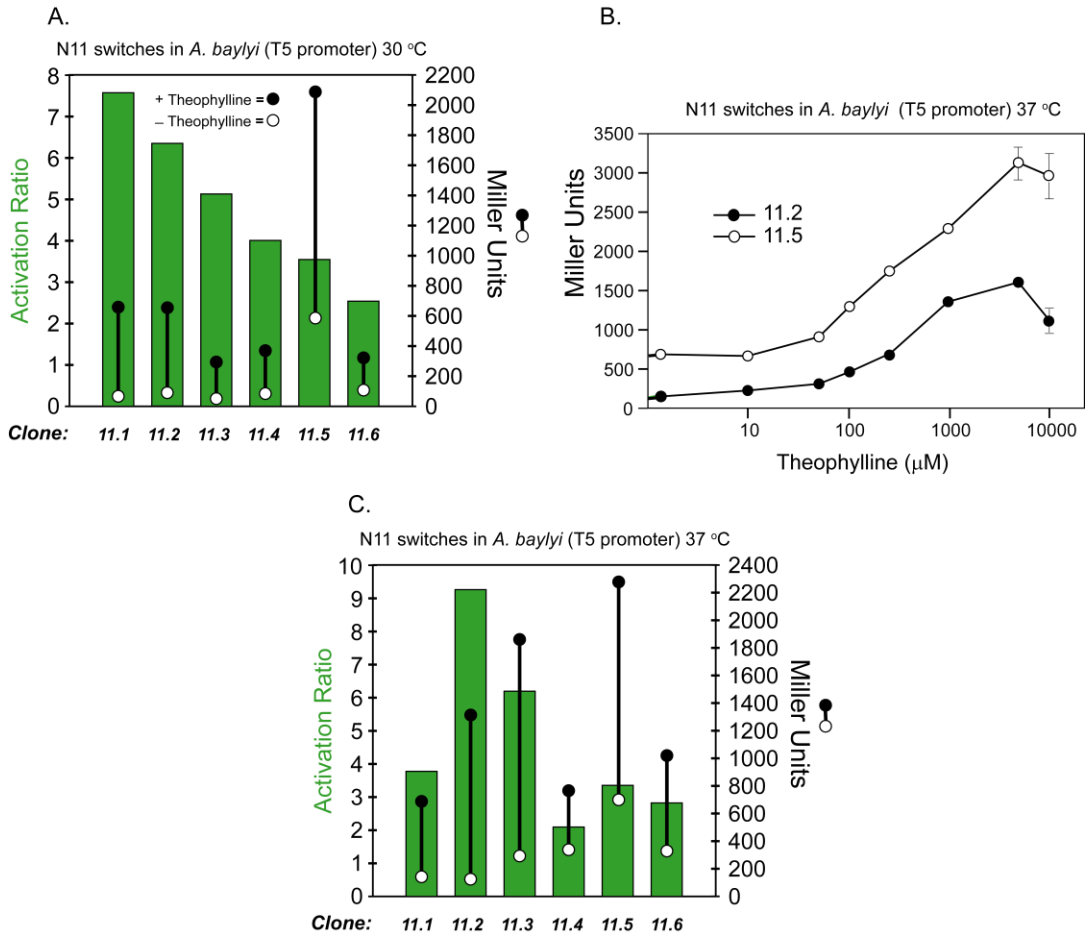

Figure 4.7: β-Galactosidase Activities for Riboswitches Identified from the High-Throughput Screen, Expressed in *A. baylyi* (T5 promoter). (A) Activities (Miller Units) for riboswitch variants identified from screening in the absence (open circles) and presence (closed circles) of 1 mM theophylline. Measured at 30 °C. Standard errors are less than the circle diameters. (B) Dose response curves for two riboswitch variants identified from screening. Measured at 37 °C. (C) Activities (Miller Units) for riboswitch variants identified from screening in the absence (open circles) and presence (closed circles) of 1 mM theophylline. Measured at 37 °C. Standard errors of the mean are less than the circle diameters.

*IS10-lac*Z. We screened approximately 20,000 *A. baylyi* colonies, and identified 6 variants with activation ratios above 2, which we characterized further. Figure 4.7 shows the Miller Units for the chosen variants when grown at 30 °C or 37 °C, and the dose response plots for two of our best constructs, grown at 37 °C, are also shown (these are the only assays for individual clones that were conducted at 30 °C, because the *A. baylyi* cells were grown at 30 °C during the screening procedure). Unfortunately, none of the variants showed activation ratios greater than 10, at either temperature.

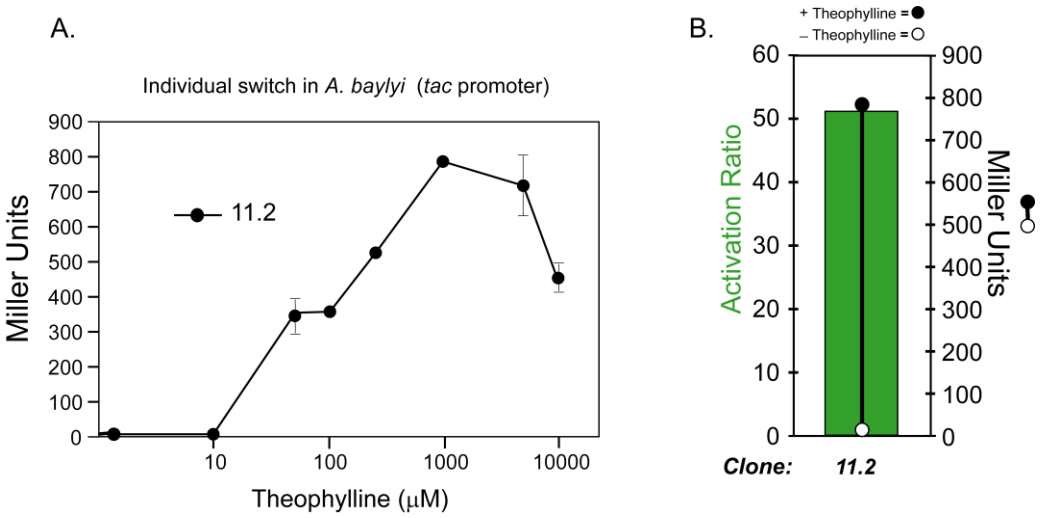

Figure 4.8: β-Galactosidase Activities for Riboswitch Clone 11.2 Identified from the High-throughput Screen, Expressed in *A. baylyi* (*tac* promoter). (A) Dose response curve for riboswitch 11.2 cloned into pSKD-AB, under the control of the *tac* promoter, in *A. baylyi*. Measured at 37 °C. (B) Activities corresponding to the dose response curve are expressed in Miller Units in the absence (open circles) and presence (closed circles) of 1 mM theophylline, and the standard errors of the mean are less than the circle diameters.

Clone Name	Linker Sequence
12.1	CTGCTAAGGTAA
12.2	CTTAGAGGCTGT
12.3	CTATGAGGCTGT
11.1	CAATCATGGGG
11.2	CTGAGAAGGGG
11.3	CAGCGCAAGGT
11.4	CCCACGCGGGG
11.5	CAAAACAGGGG
11.6	CCGGACGGGGC

Table 4.1: N11 Linker Sequences of Synthetic Riboswitches Identified from High-Throughput Screen in *A. baylyi*. The G-rich segment of the N11 linkers suggests it is the optimal ribosome binding site, since we did not specifically include one. Linker sequences of the three FACS riboswitches (12.1-12.3) are also shown.

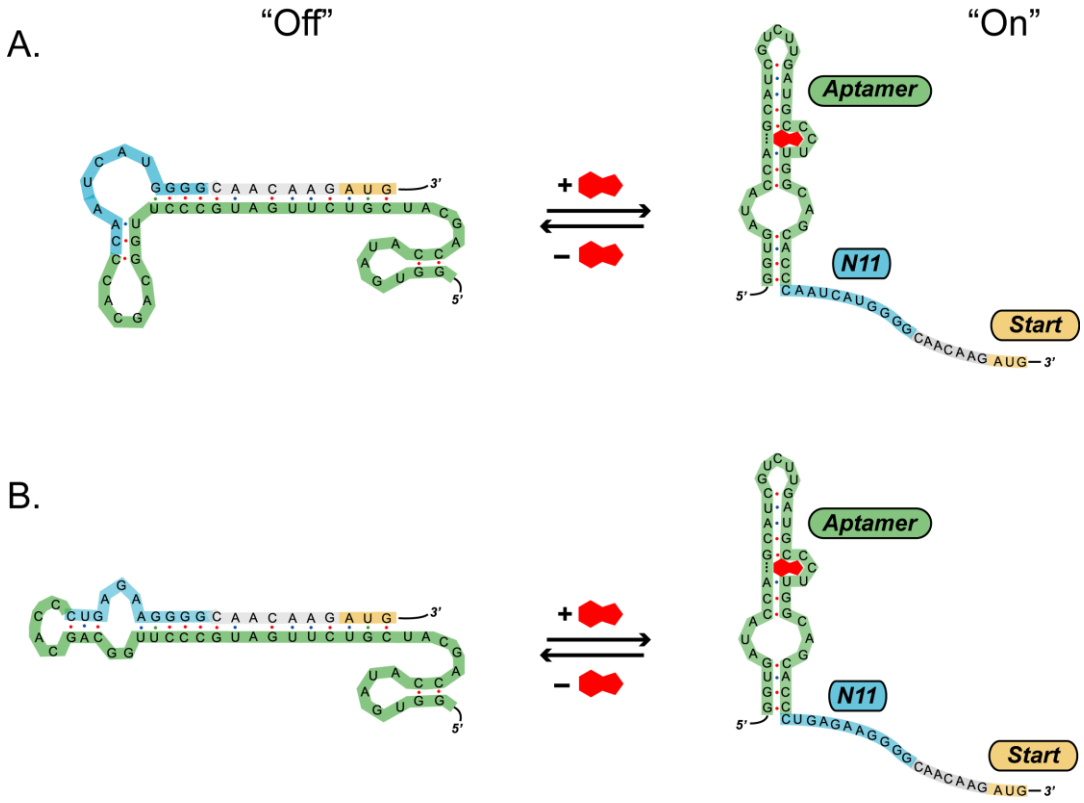

Figure 4.9: Predicted Mechanism of Action for Synthetic Riboswitches from *A. baylyi*. (A) Riboswitch RNA from Clone 11.1 in the absence (left) and presence (right) of theophylline (red). Without theophylline (left), the ribsome binding site is paired. In the presence of theophylline (right), the ribosome binding site is unpaired. (B) Riboswitch RNA from Clone 11.2 in the absence (left) and presence (right) of theophylline. The mechanism is predicted to be the same as for Clone 11.1.

Sequencing of the 6 riboswitch variants revealed that they contain a relatively G-rich segment, and this region probably serves as the ribosome binding site (Table 4.1), which suggests this sequence is optimal for *A. baylyi* expression. We also performed *mFold* analysis[22] of the sequences, (from the 5' end of the aptamer to the 3' end of the start codon) and our results show that they all fold into a hairpin structure that pairs the ribosome binding site in the absence of ligand (Figure 4.9), and this pairing is, presumably, disrupted when the aptamer structure forms in the presence of theophylline. We were pleased to observe this result, as these structures conform to our earlier reported equilibrium model for riboswitch function.[17]

To further optimize our newly identified riboswitch variants, we subcloned one of our best, 11.2, into pSKD-AB, under the control of the *tac* promoter, and introduced it into *A. baylyi*. The Miller Units for this construct and the corresponding dose response plot is shown in Figure 4.8. Gratifyingly, the activation ratio for the variant improved dramatically, and displays an activation ratio of 51 when grown with 1 mM theophylline – this is the highest activation ratio we observed in *A. baylyi*. It is interesting to note that simply changing the T5 promoter to the *tac* promoter led to a dramatic change in the function of the riboswitch variant when expressed in *A. baylyi*; this information should prove useful when engineering synthetic riboswitches in other organisms.

4.3 Conclusion

Synthetic riboswitches hold great promise for controlling gene expression in response to non-endogenous small molecules in both prokaryotes and eukaryotes.[23] Their relatively simple RNA components (an aptamer domain and expression platform) are easily obtainable. For example, powerful methods exist for selecting aptamers to the small molecule of one's choosing,[24] and recent research has presented methods to convert these aptamers into synthetic riboswitches in both prokaryotes[17] and eukaryotes.[25] Additionally, the described conversion techniques are uncomplicated, and relatively inexpensive. Therefore, the use of synthetic riboswitches for a variety of applications will, most likely, find widespread use.

One area which remains unexplored is the transferability of synthetic riboswitches between similar organisms. In theory, this should be possible because synthetic riboswitches are composed entirely of RNA and, therefore, do not depend on the host

organism's proteins.[26] We believed that synthetic riboswitches, originally optimized in *E. coli*, would function in the closely related bacterium, *A. baylyi*. We chose *A. baylyi* because it has a number of desirable properties that *E. coli* does not have, such as natural competence,[9] and we wished to enable the conditional control of gene expression in this organism, which has not been previously reported.

We found that our previously identified synthetic riboswitches (from *E. coli*) do function when expressed in *A. baylyi*, though their original ligand responses were not retained; their activation ratios were much lower in *A. baylyi*. In an attempt to remedy this problem, we adapted our previous high-throughput screen[17] for use with *A. baylyi* and used this to identify several functioning synthetic riboswitches. One of these variants displayed an activation ratio of 51 in *A. baylyi*, which was our greatest success. Taken together, transferring synthetic riboswitches between closely related organisms or screening for synthetic riboswitches directly in an organism produces an equivalent result: functioning synthetic riboswitches.

4.4 Experimental

General Considerations

All plasmid manipulations utilized standard cloning techniques. All constructs have been 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), ampicillin, and kanamycin 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) or *A. baylyi* ADP1 (ATCC) cultured in media obtained from EMD Bioscience.

Transformation of *A. baylyi*

Riboswitch constructs, cloned into pSKD-AB, were used to electroporate TOP10 cells and grown using standard procedures. A single colony was picked and grown in LB media (5 mL) supplemented with kanamycin (50 µg/mL) and shaken 37 ^{o}C overnight. A fraction of the overnight culture (1.5 mL) was used for a miniprep, resulting in a plasmid DNA concentration of approximately 100 ng/µL. A single *A. baylyi* colony was picked from an LB agar plate and used to inoculate liquid LB (5 mL) and shaken at 30 ^{o}C overnight. A fraction of the overnight culture (60 µL) was used to inoculate fresh liquid LB (900 µL) and shaken (225 rpm) at 30 ^{o}C for 2 h. At this point, 5 µL of plasmid DNA (500 µg) was added to the culture and its growth was continued at 30 ^{o}C for 2.5 h, with shaking. A fraction of this culture (100 µL) was plated on LB agar, supplemented with kanamycin (50 µg/mL), and the cells were incubated at 30 ^{o}C for 18 h. Individual colonies were picked for analysis.

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 between the mTCT4-8 theophylline aptamer and the RBS of the *IS10-LacZ* reporter gene.

96-well Plate-Based Screen

Library transformations in pSKD-AB, harbored in *A. baylyi*, were plated on large (241 mm x 241 mm) bioassay trays from Nalgene containing LB agar (300 mL) supplemented with kanamycin (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 ~2,000 colonies/plate. Cells were grown for 14 h at 30 °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 kanamycin (50 μg/mL). The plate was incubated overnight at 30 ºC 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 kanamycin (50 μg/mL). The second set of plates contained 200 μL of LB supplemented with both kanamycin (50 μg/mL) and theophylline (1 mM). Plates were incubated for approximately 2.5 hrs at 30 ºC 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 using the method of Miller.[20]

Plasmid pSKD-AB (*E. coli*-*A. baylyi* shuttle plasmid with *lacZ* reporter gene)

An acceptor plasmid was constructed by assembling overlapping PCR oligos to create a multiple cloning site as follows: *Eco*RI-*Not*I-*Xba*I-*Sph*I-RBS-*Nco*I-*Eco*RV-*Hind*III-*Spe*I-*Not*I-*Pst*I. The assembled oligos were digested with *Dra*III and *Afl*III and cloned into those sites in pSL1180 (GE Healthcare), resulting in IMBB. The TEM-1 antibiotic resistance marker in IMBB was replaced with the kanamycin resistance gene from pET28B using *Bsp*HI, resulting in IMBB-kanR. The pWH1266 origin of

replication was PCR amplified (using oligos also encoding *Sph*I and *Hind*III sites) from the original plasmid (ATCC) and subcloned into IMBB-kanR using *Sph*I and *Hind*III. The resulting plasmid was digested with *Hind*III, blunted with T4 DNA polymerase, and self-ligated, resulting in IMBB-kanR-pWH1266 ori-del *Hind*III (Ichiro Matsumura).

The *IS10-lac*Z gene, under the control of the T5 promoter, was created by amplifying pSKD445.1 using primers SKD579 and SKD574, resulting in PCR product A. Product A was digested with *Xba*I and *Pst*I, and cloned into IMBB-kanR-pWH1266 ori-del *Hind*III which was digested with *Spe*I and *Pst*I, resulting in pSKD-AB containing the T5 promoter.

The *IS10-lac*Z gene, under the control of the *tac* promoter, was amplified from pSKD445.1 using primers SKD621 and SKD574, resulting in PCR product B. Product B was digested with *Xba*I and *Pst*I, and cloned into IMBB-kanR-pWH1266 ori-del *Hind*III which was digested with *Spe*I and *Pst*I, resulting in pSKD-AB containing the *tac* promoter.

SKD-579
5′- GGAATTCGCGGCCGCT**TCTAGA**GAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTAT
AATAGATTCAATTGTGAGCGGATAACAATTTCACACA*GCTAACAAGTCTAGCGAACCGCA*
(**XbaI**, T5 promoter, *anneals to vector*)

SKD-574
5′- **CTGCAG**CGGCCGCTACTAGTA*TTATTTTTGACACCAGACCAACTGGTAATGG*
(**PstI**, *anneals to vector*)

SKD-621
5′- GGAATTCGCGGCCGCT**TCTAGA***GGAGCTGTTGACAATTAATCATCGGCTCG*
(**XbaI**, *anneals to vector*)

4.5 References

(1) Hedges, S. B. *Nat. Rev. Genet.* **2002**, *3*, 838-849.

(2) Metzgar, D.; Bacher, J. M.; Pezo, V.; Reader, J.; Doring, V.; Schimmel, P.; Marliere, P.; de Crecy-Lagard, V. *Nucleic Acids Res.* **2004**, *32*, 5780-5790.

(3) Neidhardt, F. C. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*; American Society for Microbiology: Washington, D.C., 1987.

(4) Palchevskiy, V.; Finkel, S. E. *J. Bacteriol.* **2006**, *188*, 3902-3910.

(5) Datsenko, K. A.; Wanner, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6640-6645.

(6) Barbe, V.; Vallenet, D.; Fonknechten, N.; Kreimeyer, A.; Oztas, S.; Labarre, L.; Cruveiller, S.; Robert, C.; Duprat, S.; Wincker, P.; Ornston, L. N.; Weissenbach, J.; Marliere, P.; Cohen, G. N.; Medigue, C. *Nucleic Acids Res.* **2004**, *32*, 5766-5779.

(7) Hunger, M.; Schmucker, R.; Kishan, V.; Hillen, W. *Gene* **1990**, *87*, 45-51.

(8) Young, D. M.; Parke, D.; Ornston, L. N. *Annu. Rev. Microbiol.* **2005**, *59*, 519-551.

(9) Vaneechoutte, M.; Young, D. M.; Ornston, L. N.; De Baere, T.; Nemec, A.; Van Der Reijden, T.; Carr, E.; Tjernberg, I.; Dijkshoorn, L. *Appl. Environ. Microbiol.* **2006**, *72*, 932-936.

(10) Elbahloul, Y.; Steinbuchel, A. *Appl. Environ. Microbiol.* **2006**, *72*, 1410-1419.

(11) de Berardinis, V.; Vallenet, D.; Castelli, V.; Besnard, M.; Pinet, A.; Cruaud, C.; Samair, S.; Lechaplais, C.; Gyapay, G.; Richez, C.; Durot, M.; Kreimeyer, A.; Le

Fevre, F.; Schachter, V.; Pezo, V.; Doring, V.; Scarpelli, C.; Medigue, C.; Cohen, G. N.; Marliere, P.; Salanoubat, M.; Weissenbach, J. *Mol. Syst. Biol.* **2008**, *4*, 174.

(12) Matsumura, I., Emory University, Atlanta, GA. Personal communication, 2007.

(13) Saenger, W.; Orth, P.; Kisker, C.; Hillen, W.; Hinrichs, W. *Angew. Chem. Intl. Ed.* **2000**, *39*, 2042-2052.

(14) Gallivan, J. P. *Curr. Opin. Chem. Biol.* **2007**, *11*, 612-619.

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

(16) Barrick, J. E.; Breaker, R. R. *Genome Biol.* **2007**, *8*, R239.

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

(18) Arnold, F. H. *Accounts Chem. Res.* **1998**, *31*, 125-131.

(19) Gentz, R.; Langner, A.; Chang, A. C. Y.; Cohen, S. N.; Bujard, H. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4936-4940.

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

(21) de Boer, H. A.; Comstock, L. J.; Vasser, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 21-25.

(22) Mathews, D. H.; Turner, D. H.; Zuker, M. *Curr. Protoc. Nucleic. Acid Chem.* **2007**, *11*, 11.12.11-11.12.17.

(23) Suess, B.; Weigand, J. E. *RNA Biol.* **2008**, *5*.

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

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

(26) Winkler, W. C. *Arch. Microbiol.* **2005**, *183*, 151-159.

Chapter 5 Studies Toward Functional Cloning of a Putative Caffeine N-7 Demethylase from *Coffea arabica*

5.1 Introduction

Coffee is the caffeine-containing beverage of choice for a majority of the world's population, and approximately 90% of adults in North America drink it daily in some form.[1] Mainly, coffee is consumed because it contains caffeine, which acts as a stimulant of the central nervous system, wards off sleepiness, and allows people to function for prolonged periods of time with little rest.[4] While these effects are highly desirable, caffeine can also produce a number of undesirable side effects in some people, such as heart palpitations, increased blood pressure, and insomnia.[4,5] As such, approaches to decaffeinate coffee have been, and are being, developed for those who enjoy the taste of coffee, but not the side effects of caffeine.[6] Decaffeinated coffee also serves as an economically important commodity, with sales of approximately 4 billion dollars within the United States in 1999, and demand for it is steadily increasing.[7] In combination, health implications and economic benefits provide a large incentive for research into decaffeinated coffee.

Two of the main approaches currently employed to produce decaffeinated coffee are ethylacetate and supercritical CO_2 extraction.[6] In the ethylacetate extraction process, coffee beans are soaked in a water solution that contains ethylacetate, and the extraction solution is subsequently removed, leaving decaffeinated beans. Supercritical CO_2 is also used as a solvent for the extraction of caffeine from coffee beans, but it is more expensive than ethylacetate extraction because of the need for specialized equipment, and it is less widely used. Both processes have the same drawbacks. First, they extract out a number of additional organic compounds from coffee beans, and many of these compounds give coffee its unique flavor. Because of this, synthetic flavorings must be added to the coffee

beans after the extraction process, otherwise the beverage produced from them will have an unpleasant or bland taste. And the flavorings, in large quantities, may not be healthy for consumption.[7] Second, the extraction processes and synthetic flavorings serve to increase the cost of decaffeinated coffee by an estimated 50 cents per kilogram in comparison to the price of unprocessed coffee beans.[8] Clearly, additional approaches for producing decaffeinated coffee, without the need for extraction or additional synthetic flavorings, are warranted.

One possible solution is to breed (or engineer) a coffee plant that produces beans low in caffeine content. With this type of coffee bean, there would be no need for a chemical extraction process; additional synthetic flavorings would not be necessary; and a lower priced decaffeinated coffee, free of synthetic flavorings, would be produced. Unfortunately, the two main varieties of coffee plants, *Coffea arabica* and *Coffea canephora*, which account for greater than 90% of all coffee production worldwide,[9] have complicated genetics, and a traditional program to breed variants of either species that produce beans with lower caffeine content would take approximately 20 years.[7] While there are several known species of coffee plants that produce beans with low caffeine content, including *Coffea eugenioides* and *Coffea bengalensis*, coffee brewed from their beans tastes bitter and unpleasant.[7] Furthermore, a breeding program to cross *C. eugenioides* with *C. arabica* to form a plant that produces pleasant tasting beans low in caffeine content would not be feasible due to their differing genetics.[9] Genetic engineering may be a more viable (and less time consuming) route.

To this end, the Sano group has used RNA interference to downregulate expression of a methyltransferase involved in caffeine biosynthesis in *C. canephora*.[10]

Workers from this group designed an antisense RNA specific to the 3' untranslated region of the gene encoding theobromine synthase, and infected coffee plants with a construct that constitutively produced the RNA. In the transgenic plant leaves, the caffeine content was reduced by 50-70% in comparison to a control plant.[10] Since the beans were not analyzed, it is difficult to determine whether the genetically engineered plant produces a naturally decaffeinated bean, and the taste of coffee produced from the transgenic beans is also unknown. If the transgenic beans are not low in caffeine and do not have a pleasant taste, the knockdown strategy will not be commercially successful. Additionally, the RNA interference strategy works by inhibiting the expression of an enzyme involved in the biosynthesis of caffeine, which could lead to a high concentration of precursor metabolites inside the plant. These higher metabolite concentrations may not be safe for consumption.

The RNA interference strategy (above) focused on blocking the formation of caffeine, but enhancing the rate of caffeine degradation in coffee plants would, in theory, also produce plants with a lower caffeine content,[7,11] and plants modified in this fashion would probably not harbor an increased concentration of potentially harmful precursor metabolites. The catabolism of caffeine in *C. arabica* has been studied, and its caffeine content was found to be high due to a slow rate of caffeine breakdown to theophylline by a putative caffeine N-7 demethylase[11] (7-NDM) (Figure 5.1). After theophylline is formed in *C. arabica*, it is rapidly metabolized to ammonia and carbon dioxide. Thus, the putative 7-NDM acts as a roadblock in the process of caffeine breakdown, causing caffeine to build-up within the plant. *C. eugenioides*, which is related to *C. arabica*, has a low caffeine content due to a fast rate of conversion of caffeine to theophylline by a

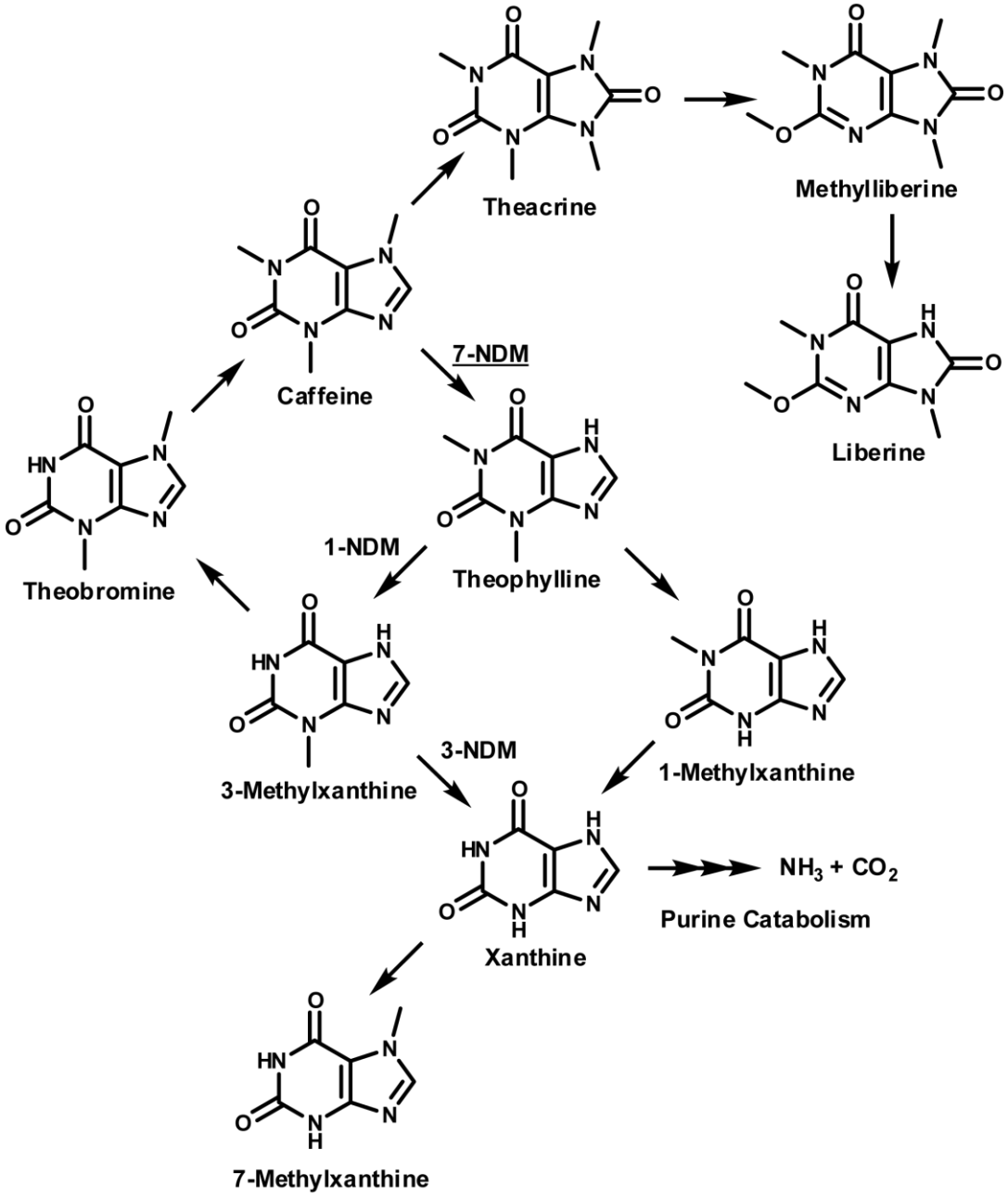

Figure 5.1: Catabolic Pathway of Caffeine in *C. arabica*. The major degradation route is proposed to occur through the labeled enzymes (7-NDM, 1-NDM, and 3-NDM)**.** The caffeine content in *C. arabica* is high due to the limited activity of 7-NDM (underlined). (NDM = *N*-demethylase)

putative 7-NDM,[11] which supports the observation that the putative 7-NDM from *C. arabica* is indeed responsible for the plant's high caffeine content. Therefore, increasing the rate of conversion of caffeine to theophylline in *C. arabica* may lead to a plant that produces beans that are both low in caffeine and good tasting. If we could isolate the putative 7-NDM from *C. arabica*, increase its catalytic rate, and reintroduce it into the plant, good-tasting, naturally decaffeinated coffee may be within reach.

The putative 7-NDM from *C. arabica* could be isolated and identified using a number of approaches, including fractionation. With fractionation, the contents of the plant's cells would be separated using column chromatography, and each fraction would be tested for enzymatic activity.[12] If the enzyme-containing fraction was identified, the sequence of the protein could then be determined using Edman degradation. But the enzymatic test for caffeine 7-NDM activity involves radioactivity and is time-consuming,[13] suggesting other identification methods may be better suited. A bioinformatics approach would be the simplest method to identify the target gene; unfortunately, the genome sequence of *C. arabica* is unpublished.

Functional cloning is another potential, and promising, method for isolating the putative 7-NDM. In functional cloning, a library of genes is introduced into a strain of bacteria (or other organism) that is deficient in a required metabolic pathway, such as amino acid biosynthesis or glycolysis.[14,15] Cells containing a gene(s) which complements the auxotrophy will, thus, survive, and the identity of the complementing gene can then be identified using DNA sequencing. If we constructed a strain of bacteria that requires the presence of theophylline to live, we could introduce a cDNA library from *C. arabica* into the bacteria and feed the cells caffeine. Cells containing the 7-NDM would

demethylate caffeine, produce theophylline, and live. The gene could then be identified using DNA sequencing. Unfortunately, as stated above, auxotrophic strains of bacteria normally have deficiencies in primary metabolic pathways, and there are few solutions to create auxotrophs outside these parameters.[15]

Synthetic riboswitch-mediated auxotrophs may circumvent these limitations. In theory, a synthetic riboswitch-mediated auxotroph can be constructed to respond to any nontoxic, cell-permeable small molecule which can interact with RNA.[2,3] Previously, we reported using a synthetic riboswitch in both genetic screens and selections for the small molecule theophylline, and these strains were not responsive to caffeine - an important characteristic for our purposes.[3] Here, we present our attempts to functionally clone the putative caffeine 7-NDM from *C. arabica* using a designer auxotroph that requires theophylline for survival. We designed a number of strategies, and, unfortunately, they were not successful. But we are hopeful that this work will provide the foundation for others who wish to conduct research in this area.

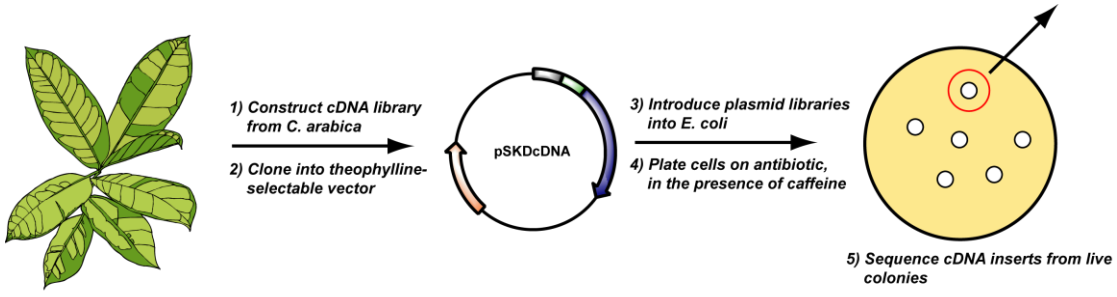

Figure 5.2: Functional Cloning Strategy for the Putative Caffeine 7-NDM from *C. arabica*. A cDNA library is created from *C. arabica*, cloned into a theophylline-selectable vector, and introduced into *E. coli*. These cells are grown in the presence of antibiotic and caffeine. Living cells will, in theory, contain a caffeine 7-NDM, which converts caffeine to theophylline.

5.2 Results and Discussion

Figure 5.2 shows the general scheme we used for functional cloning of the putative 7-NDM from *C. arabica*. We started by constructing a cDNA library from *C. arabica*, and then cloned this library into a theophylline-selectable cDNA cloning vector. After cloning, we introduced the cDNA library into *E. coli*, plated the cells on selectable media in the presence of caffeine, and allowed the cells to grow. Living cells were tested further to verify their phenotype, and their cDNA inserts were identified by sequencing.

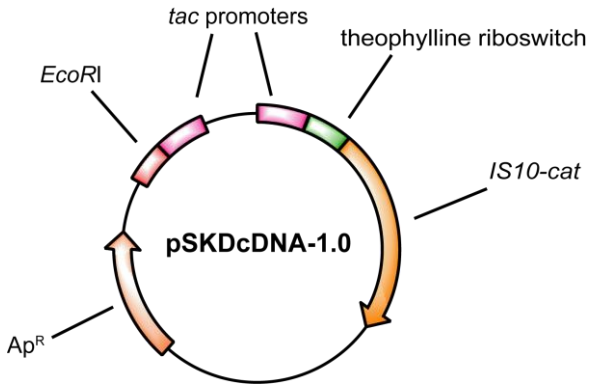

Figure 5.3: Theophylline-selectable cDNA Cloning Vector. When introduced into *E. coli*, the vector provides resistance to chloramphenicol, only in the presence of theophylline. The vector also contains a single *Eco*RI site for cloning of PCR-amplified cDNA, under the control of a constitutive t*ac* promoter. In pSKDcDNA-1.0, the theophylline riboswitch has 8 base pairs between the ribosome-binding site and the aptamer, as reported.[3]

5.2.1 Creation of a Theophylline-Selectable cDNA Cloning Vector and a *C. arabica* cDNA Library

We envisioned that our cDNA cloning vector would: be theophylline-selectable; contain a multiple cloning site for insertion of the cDNA library; and have a constitutive promoter for transcription of the cloned cDNA library. Figure 5.3 shows a diagram of this vector (pSKDcDNA-1.0) that contains a chloramphenicol resistance fusion gene

(*IS10-cat*) under the control of a theophylline-responsive riboswitch with an 8 base pair linker between the aptamer and the ribosome binding site, as reported in our previous publication.[3] Additionally, we added a transcriptional terminator at the 3' end of the *IS10-cat* coding region. The vector also has a cloning site that contains an *Eco*RI restriction site, and a constitutive *tac* promoter for transcription of the cloned cDNA, followed by a transcriptional terminator. When introduced into *E. coli*, pSKDcDNA-1.0 displays no growth after 18 h at 37 ^{o}C when plated on LB agar supplemented with ampicillin (to maintain the plasmid), chloramphenicol (50 μg/mL), and caffeine (1 mM). This is an improvement upon our previously reported theophylline riboswitch with a 5 base pair linker, which does display some growth under these conditions.[3] This observation is important, as we wished to minimize false positives from the selections due to background growth.

Since our cDNA cloning vector contained an *Eco*RI site, we needed to construct the *C. arabica* cDNA library such that both the 5' and 3' ends of the resulting DNA product contained *Eco*RI sites for cloning. To accomplish this, total RNA was extracted from the young leaves (where caffeine 7-NDM activity was measured to be highest)[13] of a *C. arabica* plant, which was subsequently reverse transcribed using a poly-T primer (Figure 5.4). Terminal transferase was used to add adenosine residues to the 3' end of the first strand cDNA product, and a poly-T primer, containing an internal *Eco*RI site, was used to PCR amplify the modified first strand cDNA, resulting in double stranded DNA with *Eco*RI sites at both termini. Figure 5.5 shows an agarose gel of the resulting PCR product and it does not appear as a single band, but rather as a smeared product, suggesting the cDNA library contains genes of varying lengths.

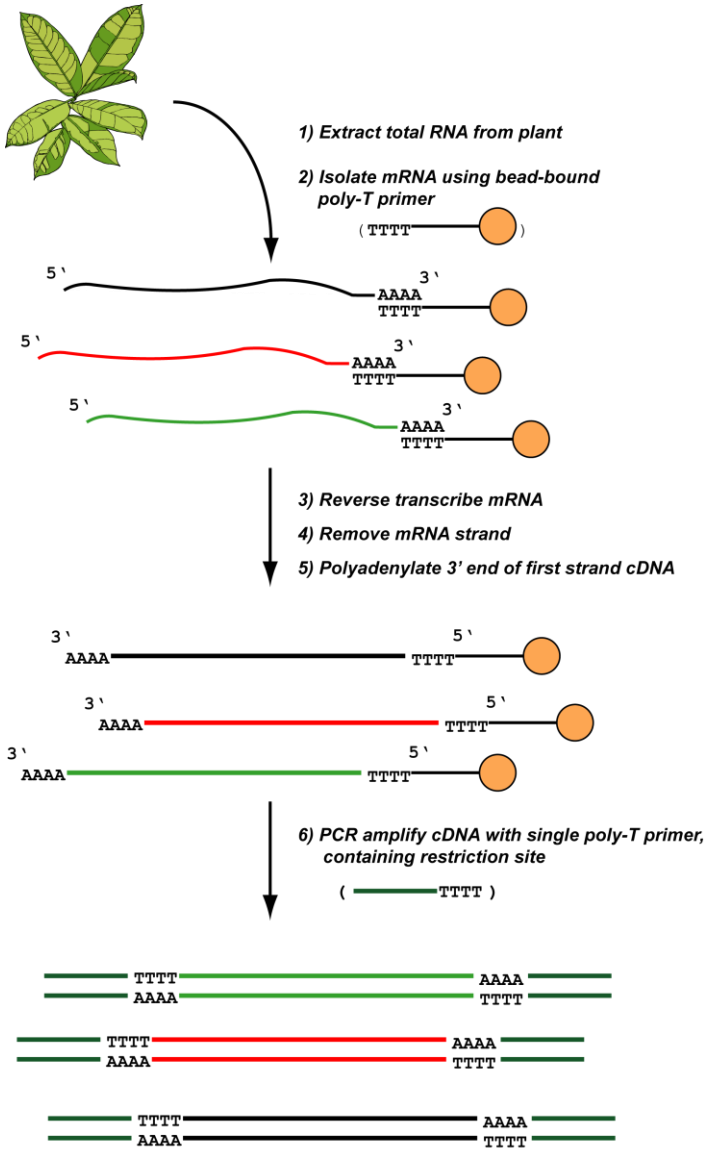

Figure 5.4: cDNA Synthesis Scheme. Total RNA is extracted from *C. arabica* and is converted into a double-stranded DNA product with terminal restriction sites.

A fraction of the cDNA PCR product was subjected to digestion with *Eco*RI and cloned into the *Eco*RI site of pSKDcDNA-1.0. The plasmid library was introduced into *E. coli* and plated on LB agar supplemented with ampicillin, resulting in a large number of colonies (~10^6). After 14 h growth at 37 ^{o}C, colony PCR was used to determine if cDNA inserts were present inside of the cloned vectors. Indeed, several of the colonies selected contained inserts of varying lengths (not shown), suggesting the cloning step was

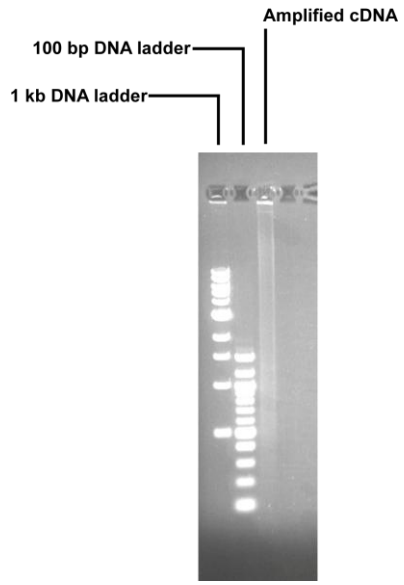

Figure 5.5: PCR-Amplified *C. arabica* cDNA on an Agarose Gel (1%). PCR amplification of the *C. arabica* cDNA library with a poly-T primer, containing an *Eco*RI site, produced a smeared product, indicating the library contained genes of various sizes.

successful and that the inserts were of various genes (sequencing also revealed several inserts were different).

5.2.2 Selections for a Putative Caffeine 7-NDM from *C. arabica*

Once we were confident that our cDNA library contained a variety of genes, we proceeded to subject the library to selection. *E. coli* cells (~10^6) harboring the cDNA-containing plasmid library were plated on LB agar supplemented with ampicillin, chloramphenicol (50 µg/mL) and either no small molecule or caffeine (1 mM, this concentration does not affect cell growth), where the plate without caffeine served as a control for the background growth of the library. The plates were set at 37 ^{o}C for 48 h, and subsequently analyzed. After this period, both plates showed six colonies. Each colony from the caffeine-containing plate was individually restreaked on LB agar, supplemented as before, in the presence and absence of caffeine, and grown as before.

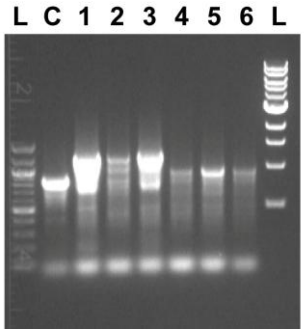

Figure 5.6: PCR Reactions of Colonies that Survived Selection on Caffeine. The agarose gel (1%) contains a 100 bp ladder (left lane), 1 kb ladder (right lane) and a vector only control (C). Colonies 1-3 contain cDNA inserts, while colonies 4-6 do not. The cDNA inserts were found to be non-coding.

Three of the colonies did not grow under either condition. But, surprisingly, three of the colonies grew both in the presence and absence of caffeine. To identify the inserts associated with the observed resistance, all six colonies from the initial plate containing caffeine were used as templates in separate colony PCR reactions, and three of the colonies (those which grew in the presence and absence of caffeine) were found to contain inserts (Figure 5.6), which were sequenced. Sequencing revealed that two of the colonies (the third sequence failed) contained, in fact, the same 390 base pair sequence. Analysis of the insert indicated it was non-coding, but was derived from the *C. arabica* plant. Its origin was determined by a BLAST analysis, and it is similar to a sequence from a plant, *Arabidopsis thaliana*. One possible resistance mechanism is that the non-coding insert interferes with the folding of the theophylline riboswitch, causing it to constitutively activate *IS10-cat* expression, but alignment of the non-coding RNA with the sequence of the theophylline riboswitch did not reveal any sequence complementarity, though a potential tertiary interaction between the two RNAs would not have been found through sequence alignment. A variety of studies would have been possible, but since the insert was non-coding we decided against further investigation.

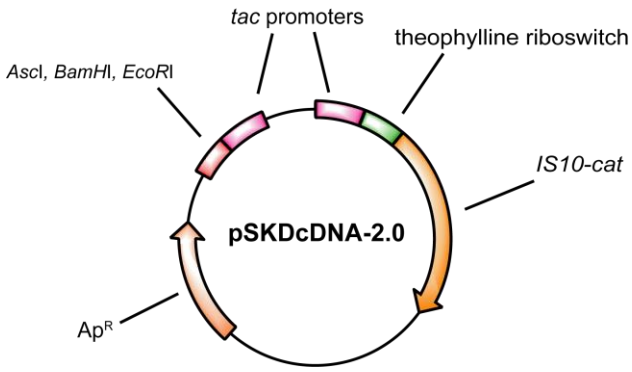

Figure 5.7: Theophylline-Selectable cDNA Cloning Vector. When introduced into *E. coli*, the vector provides resistance to chloramphenicol, only in the presence of theophylline. The vector also contains a multiple-cloning site for insertion of PCR-amplified cDNA, under the control of a constitutive *tac* promoter. In pSKDcDNA-2.0, the theophylline riboswitch is that from Clone 8.1.[2]

At this point, we believed that the selection failed for a number of reasons. First, the concentration of chloramphenicol used may have been too high, which could have interfered with the activity of the putative caffeine 7-NDM. Second, the digestion of the cDNA library with *Eco*RI may have cleaved the coding region of the putative caffeine 7-NDM, and it would, thus, not be represented in the plasmid library. To remedy these problems we modified pSKDcDNA-1.0 by replacing its theophylline riboswitch with the theophylline riboswitch from Clone 8.1[2] (Figure 5.7), which requires a lower concentration of chloramphenicol when used in selections. And we added two additional restriction sites, *Asc*I and *Bam*HI, to the cDNA multiple cloning site, that would allow us to clone the putative caffeine 7-NDM in its full-length form, assuming these restriction sites are not found within the putative gene.

As before, we PCR amplified the *C. arabica* cDNA library with primers such that the resulting products contained terminal *Asc*I or *Bam*HI restriction sites. All three PCR amplified cDNA libraries (*Asc*I, *Bam*HI, and *Eco*RI) were digested with a single enzyme, cloned into their respective sites in pSKDcDNA-2.0, and introduced into *E. coli*. Cells harboring these libraries were plated on LB agar supplemented with ampicillin and, after

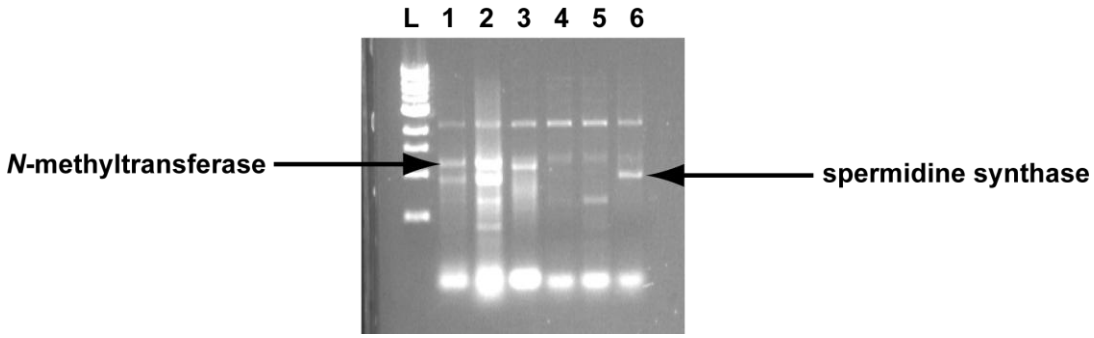

Figure 5.8: PCR Reactions with Plasmid-Cloned *C. arabica* cDNA Libraries. Plasmid-cloned cDNA was used as a template in separate PCR reactions with two sets of primers, specific for separate *C. arabica* genes, and were analyzed on an agarose gel (1%). The enzymes used for cloning were *Eco*RI (lanes 1,4), *Bam*HI (lanes 2,5) and *Asc*I (lanes 3,6); lane L is a 1 kb ladder. PCR reactions in lanes 1-3 were amplified using primers specific for an *N*-methyltransferase (~1100 bp), and it is present in all three cloned libraries. PCR reactions in lanes 4-6 were amplified using primers specific for a spermidine synthase gene (~1000 bp), and it is only present in the library cloned using *Asc*I.

growth, were scraped into liquid LB media and grown, separately, and the plasmids from a small aliquot of each culture were isolated. We used two sets of primers, specific for separate known *C. arabica* genes, in a PCR reaction, with each plasmid library serving as a template, separately. We reasoned that if the cloned cDNA libraries accurately represented the plant's genetic diversity, the two known genes should be present in each. Figure 5.8 shows an agarose gel of these PCRs. All three libraries contain the approximately 1100 base pair *N*-methyltransferase gene, but only the *Asc*I digested library contains the approximately 1000 base pair spermidine synthase gene, suggesting that it is wise to use multiple restriction sites when cloning an unknown gene from a cDNA library.

E. coli cells (~10^6) harboring each library were plated, separately, on LB agar supplemented with ampicillin, chloramphenicol (20 µg/mL) and either no small molecule or caffeine (1 mM), in duplicate. Each set was then grown under two different conditions, 30 ^oC or 37 ^oC, for a period of 10 days. During the course of growth, several colonies from each caffeine-containing plate were used in colony PCR reactions to

identify any inserts they might contain. Interestingly, these colony PCRs failed to show the presence of any cDNA insert, and restreaking of these colonies indicated they were false positives, which did not grow under the original selection conditions.

Theoretically, if the putative caffeine 7-NDM is a cytochrome P450 oxidase, which demethylates caffeine by oxidizing the N-7 methyl group of caffeine, it may not function properly within *E. coli* due to a number of factors, including folding and translation problems.[16] Other groups have successfully expressed plant cytochrome P450s within *E. coli*, but they required co-expression of a P450 plant reductase partner for activity.[16-18] In view of previous studies, we subcloned a cytochrome P450 reductase from *Candida tropicalis* (ctCPR)[19] into the *Nco*I and *Afl*II sites of an IPTG-inducible variant of the pCDF-1b vector (Figure 5.9), and introduced it into *E. coli.* We tested the toxicity of ctCPR by plating *E. coli* cells harboring the plasmid on LB agar supplemented with spectinomycin (to maintain the plasmid) and varying concentrations of IPTG (0, 10, 50, 100 and 1000 µM). After 14 h growth at 37 ^{o}C, cells harboring the plasmid did not grow on concentrations of IPTG above 50 µM, suggesting that expression of the ctCPR gene is toxic at high levels of expression.

After deciphering the toxicity levels of ctCPR expression, we made cells that harbored the ctCPR plasmid electrocompetent, and transformed them with the separate cDNA library plasmids. These cells were first propagated on LB agar supplemented with ampicillin and spectinomycin (to maintain both plasmids), scraped into liquid LB media, and grown at 37 ^{o}C until saturation. Aliquots (~10^{6}) of cells from each library were plated on LB agar supplemented with ampicillin, spectinomycin, IPTG (20 µM), chloramphenicol (20 µg/mL) and either no small molecule or caffeine (1 mM), in

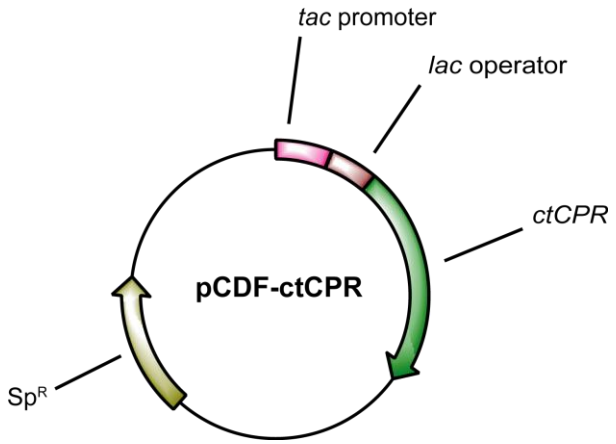

Figure 5.9: Plasmid Map of pCDF-ctCPR. Plasmid pCDF-ctCPR contains a P450 reductase gene (ctCPR) from *C. tropicalis*, under the control of an IPTG-inducible *tac* promoter. It is also confers spectinomycin resistance, and can be coexpressed in *E. coli* with any pUC-based plasmid.

duplicate. One set of plates was grown at 30 ^{o}C, and one set was grown at 37 ^{o}C, for a period of 10 days. After the growth period, hundreds of colonies were observed on the caffeine supplemented plates, and we decided to test if these cells were indeed dependent on caffeine for survival. As such, 96 colonies from each plate (30 ^{o}C or 37 ^{o}C growth) were used to inoculate separate microtiter plates containing liquid LB media supplemented with ampicillin and spectinomycin, and grown at 30 ^{o}C or 37 ^{o}C until saturation. A fraction of the saturated culture in each well was used to inoculate separate microtiter plates containg LB media supplemented with ampicillin, spectinomycin, IPTG (20 µM), chloramphenicol (20 µg/mL), and either no small molecule or caffeine (1 mM). These plates were grown at 30 ^{o}C or 37 ^{o}C for a period of 18 h and the OD_{600} absorbance for each well was measured. Interestingly, the selected colonies grew both in the presence and absence of caffeine, suggesting the ctCPR gene imparted chloramphenicol resistance, in some manner, but we did not investigate this further.

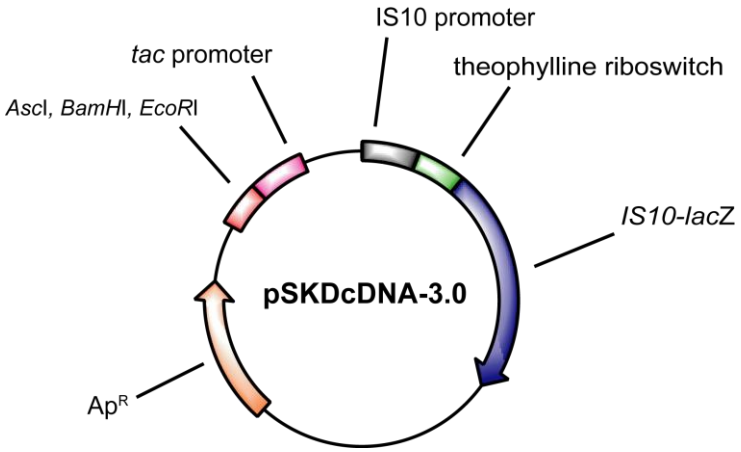

Figure 5.10: Theophylline-Screenable cDNA Cloning Vector. Similar to pSKDcDNA-2.0, pSKDcDNA-3.0 contains a cDNA multiple cloning site under control of the constitutive *tac* promoter, but it has an *IS10-lac*Z gene under the control of the IS10 promoter and a theophylline riboswitch (Clone 8.1).[2]

5.2.3 Screening for a Putative Caffeine 7-NDM from *C. arabica*

Since our genetic selection strategy did not produce the desired results, we decided to turn to a less stringent method for identifying the putative caffeine 7-NDM: genetic screens. Other groups have used screens to functionally clone gene products from large DNA libraries based on small molecule production,[20,21] and we reasoned we could screen for the enzymatic production of theophylline from caffeine. We created a theophylline-screenable cDNA cloning vector (Figure 5.9) that contains a cDNA multiple cloning site which encodes three restriction sites (*Asc*I, *Bam*HI, and *Eco*RI), and the *IS10-lac*Z reporter gene under the control of the weak IS10 promoter and riboswitch Clone 8.1.[2] *E. coli* cells harboring pSKDcDNA-3.0 display no β-galactosidase activity after three days growth on LB agar supplemented with ampicillin, X-gal and either no small molecule or caffeine (1 mM), but readily display activity when grown for the same period in the presence of theophylline (1 mM), suggesting we can use this plasmid to screen for the enzymatic conversion of caffeine to theophylline. Our three previous cDNA libraries (*Asc*I, *BamH*I, and *EcoR*I) were digested, cloned into their respective

sites in pSKDcDNA-3.0, and introduced into *E. coli*. After verifying the cDNA libraries were diverse, cells (~10^5) harboring the plasmid libraries were plated on LB agar supplemented with ampicillin, X-gal and either no small molecule or caffeine (1 mM). The cells were grown at 37 ^{o}C for 14 h and then set at 4 ^{o}C for 4 h to allow any putative demethylase activity to proceed. We observed several (28) blue colonies, in total, on the plates supplemented with caffeine. To test the observed phenotypes, we restreaked the blue colonies onto LB agar supplemented with ampicillin, X-gal and either no small molecule or caffeine (1 mM), and grew the cells as before. Unfortunately, the clones appeared blue in color in the presence and absence of caffeine, indicating their cDNA inserts were potentially interfering with the function of the synthetic riboswitch, as we had observed previously. We could have proceeded with further screening experiments by co-expressing the ctCPR plasmid with the cDNA libraries, but we decided against this.

5.3 Conclusion

Designer auxtrophic bacteria strains, based on synthetic riboswitches, are a potentially powerful tool for use in the functional cloning of unknown small molecule biosynthesis genes.[3] These types of strains can, in principle, be constructed to respond to any nontoxic, cell-permeable molecule that can interact with an RNA; the sources of the DNA libraries used can range from cDNA isolated from individual plants and animals, to metagenomic DNA isolated from environmental sources, to a mutagenized gene library; the auxotrophs depend only on the formation of the product, not on its method of production; and they are not limited to required metabolic pathways.[2,3] These traits serve

to make selections based on designer auxotrophs particularly powerful for use in cloning unknown genes, and in many other applications.

In this Chapter, we presented our attempts at functionally cloning a putative caffeine 7-NDM from *C. arabica*. The slow rate of this enzyme is responsible for the high levels of caffeine within *C. arabica*[11,13] and, it has been theorized, isolation and reintroduction of an evolved variant of this gene into *C. arabica* would lead to a naturally decaffeinated coffee plant that produces good tasting beans.[11] We used a variety of synthetic riboswitch-mediated genetic screens and selections to try to isolate the 7-NDM from *C. arabica* cDNA libraries, but our efforts were not successful. A variety of factors were potentially responsible for this, with the most likely scenario being that the enzyme did not function when expressed in *E. coli*, as has been found for other plant enzymes when expressed in *E. coli;*[16] and it is quite possible our functional cloning strategy would work in a eukaryote, such as yeast. While our attempts were unsuccessful, we envision other projects along these lines will succeed, if the small molecule target, cDNA library, and host organism are judiciously chosen.

5.4 Experimental

General Considerations

All plasmid manipulations were performed according to standard cloning techniques.[22] All plasmids were verified by sequencing. Qiagen kits were used for purification of plasmid DNA, PCR products, and enzyme digestions. Caffeine, theophylline, ampicillin, and chloramphenicol were purchased from Sigma. Spectinomycin was purchased from MP Biomedical. IPTG and X-gal were purchased

from US Biological. Oligonucleotides were purchased from IDT. TOP10 or JM109[23] *E. coli* cells were used in all experiments.

Construction of *C. arabica* cDNA Libraries

Young leaves (1 g) from a *C. arabica* plant were removed and ground under liquid nitrogen using a mortar and pestle. Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen), ethanol precipitated, and quantified by agarose gel. Poly-A tailed mRNA was isolated using a poly-T primer, which was bound to a magnetic bead (Promega), according to the manufacturer. Reverse transcription reactions on the mRNA were setup using a fraction of the bead-bound mRNA and M-MuLV reverse transcriptase (NEB) according to the manufacturer. Adenosine residues were then added to the 3' end of the first-strand cDNA by incubating a fraction of the cDNA with terminal transferase (NEB), according to the manufacturer. The cDNA library was amplified with PCR (30 cycles) using a poly-T primer (SKD-177). A fraction of this PCR product was used in subsequent PCR reactions with either primer SKD-225 or SKD-239, to add the appropriate terminal restriction sites (*EcoR*I, *BamH*I or *Asc*I). After PCR amplification, the PCRs were ethanol precipitated, and a fraction was digested with the appropriate restriction enzyme. Restriction digests were ethanol precipitated and ligated into the appropriate sites in the theophylline-selectable or theophylline-screenable cDNA cloning vectors (pSKDcDNA-1.0, pSKDcDNA-2.0, or pSKDcDNA-3.0).

SKD-177
5′- TTGCATTGACGTCGACTGAAGGAGTTTTTTTTTTTTTT
(poly-T primer)

SKD-225
5′- GCGCTACTCG***GAATTC***TTGCATTGACGTCGACTGAAGGAG
(***EcoR*I site**)

SKD-239
5′- GAATCTCCGC***GGATCC***GGCGCGCCTTGCATTGACGTCGACTGAAGGAG
(*Asc*I site, ***BamH*I site**)

Selection Procedure for cDNA Libraries and Verification of Library Diversity

TOP10 cells were electroporated with the appropriate cDNA library plasmids, plated on LB agar supplemented with ampicillin (50 µg/mL), and allowed to grow at 37 ^{o}C for 14 h. The total number of colonies from these transformations was approximately 10^6. Single colonies were analyzed for the presence of an insert with 30 cycles of colony PCR, using primers SKD-56 and SKD-178.[3] All colonies from the transformation plates were then scraped into liquid LB media and grown with shaking at 37 ^{o}C until saturation. A small fraction of each culture was used for a miniprep and the plasmid pools were used in two sets of PCR reactions with either primer set HKS-17 and HKS-18, or HKS-25 and HKS-26. These PCR reactions were used to determine if specific *C. arabica* genes were present in the cDNA libraries, after cloning into their respective plasmids. Aliquots of cells (~10^6) from the saturated cultures were diluted up to 1 mL in SOB and spread on LB agar plates supplemented with ampicillin (50 µg/mL), chloramphenicol (20 µg/mL for pSKDcDNA-2.0 or 50 µg/mL for pSKDcDNA-1.0) and either no small molecule or caffeine (1 mM). The plates were allowed to grow at either 30 ^{o}C or 37 ^{o}C for a period of up to 10 days; a wet paper towel was also placed on the lid of each plate to prevent the agar from drying. Caffeine supplemented plates were analyzed for the presence of colonies, and these were tested for caffeine dependence by growing on either solid LB agar or liquid LB, supplemented as before with chloramphenicol, and in the presence and absence of caffeine (1 mM). These clones

were then grown at 30 ^{o}C or 37 ^{o}C, for up to 10 days, and their dependence on caffeine was noted.

HKS-17
5′- GGGGGTACCATGGAGCTCCAAGAAGTCCTGC
(*C. arabica* N-methyltransferase, gene specific)
HKS-18
5′- CCCGTCGACTTACACGTCTGACTTCTCTGG
(*C. arabica* N-methyltransferase, gene specific)

HKS-25
5′- GGGGGTACCATGGCTGATGCCGTGAATAGCAATAG
(*C. arabica* spermidine synthase, gene specific)

HKS-26
5′- CCCTCTAGATCAATTTGCTTTTGAATCGATGACC
(*C. arabica* spermidine synthase, gene specific)

Selection Procedure for cDNA Libraries with Co-expression of ctCPR Gene

JM109 cells harboring pCDF-ctCPR were electroporated with the appropriate cDNA library plasmids, plated on LB agar supplemented with ampicillin (50 µg/mL) and spectinomycin (50 µg/mL), and allowed to grow at 37 ^{o}C for 14 h. Total colonies from these transformations numbered approximately 10^{6}. Single colonies were analyzed for the presence of an insert with 30 cycles of colony PCR, using primers SKD-56 and SKD-178. All colonies from the transformation plates were then scraped into liquid LB media and grown with shaking at 37 ^{o}C until saturation. Aliquots of cells (~10^{6}) from the saturated cultures were diluted up to 1 mL in SOB and spread on LB agar plates supplemented with ampicillin (50 µg/mL), spectinomycin (50 µg/mL), IPTG (20 µM), chloramphenicol (20 µg/mL for pSKDcDNA-2.0) and either no small molecule or caffeine (1 mM). The plates were allowed to grow at either 30 ^{o}C or 37 ^{o}C for a period of up to 10 days; a wet paper towel was also placed on the lid of each plate to prevent the agar from drying. Caffeine supplemented plates were analyzed for the presence of colonies, and these were tested for caffeine dependence by growing on either solid LB

agar or liquid LB, supplemented as before with chloramphenicol, and in the presence and absence of caffeine (1 mM). These clones were then grown at 30 ^{o}C or 37 ^{o}C, for up to 10 days, and their dependence on caffeine was noted.

Screening Procedure for cDNA libraries

TOP10 cells were electroporated with the appropriate cDNA library plasmids, plated (~10^4 cells) on LB agar supplemented with ampicillin (50 µg/mL), X-gal (final concentration 0.008%) and either no small molecule or caffeine (1 mM), and grown at 37 ^{o}C for 14 h. At this point, the plates were incubated at 4 ^{o}C for 4 h and the presence of blue colored colonies on the caffeine-supplemented plates was noted. Blue colonies were tested for caffeine dependence by restreaking on LB agar supplemented as before, in the presence and absence of caffeine (1 mM), and grown as before. The caffeine dependence for each clone was noted.

pSKDcDNA-1.0 (theophylline-selectable (wild-type 8 base pair switch) cDNA cloning vector)

A cassette strategy was used to create the vector pSKDcDNA-1.0 (also known as pSKD850.2). Three separate PCRs (A, B, and C) were made and assembled. PCR product A was made using primers SKD-125 and SKD-163 with template SKD186.1,[3] to amplify out the riboswitch-containing fragment which has a Ptac1 promoter variant without the *lac* operator. The chloramphenicol resistance gene fusion (*IS10-cat*) was amplified from plasmid pSKD314[3] with primers SKD-164 and SKD-99, resulting in PCR product B. A transcriptional terminator fragment was amplified from plasmid pMT416[24] with primers SKD-160 and SKD-161, resulting in PCR product C. PCR products A, B,

and C were mixed and amplified using primers SKD-125 and SKD-161. The resulting PCR product was digested with *Xba*I and *Sac*I, cloned into those sites in a pUC18[23] variant missing the *Hind*III site, introduced into *E. coli*, and the plasmid was isolated (pSKD626.4). The *Eco*RI site in pSKD626.4 was removed using Quickchange mutagenesis (Strategene) with primers SKD-181 and SKD-182, introduced into *E. coli*, and the plasmid was isolated (pSKD796). A PCR product (D), containing a Ptac1 derived promoter lacking the *lac* operator, an *EcoR*I site, and a transcriptional terminator, was made using primers SKD-174 and SKD-210 with template pMT416.[24] PCR D was digested with *Nde*I, cloned into that site in pSKD796, introduced into *E. coli*, and the resulting plasmid (pSKDcDNA-1.0) was isolated.

SKD-125
5′- TCAATCACTAGT**GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGG**GCTAAC*A*AGTCTAGC
GAACCGCAC
(**Ptac1-derived promoter,** gene specific segment, *mutation eliminates SpeI site*)

SKD-163
5′- GCTAGCAGGGGACCCAGGCCAGGGATCGATCCCCCCAAGCTTG

SKD-164
5′- GCAGGAAGCCGTTTTTTTCGTTACGCCCCGCCCTGCC

SKD-99
5′- GGCCTGGGTCCCCTGCTAGCGAGAAAAAAATCACTGGATATACCACCGTTG

SKD-160
5′- CGAAAAAAACGGCTTCCTGCGGAG

SKD-161
5′- TGATTGCGAGGAGCTCGCCGGAAAGTGAAATTGACCGATCAGAGTTTG

SKD-181
5′- CACTTTCCGGCGAGCTCCAATTCGTAATCATGGTCATAG

SKD-182
5′- CTATGACCATGATTACGAATTGGAGCTCGCCGGAAAGTG

SKD-174
5′- GGCTAGCTGACCATATGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGGAATTC
CGAAAAAAACGGCTTCCTGCGGAGGCC

SKD-210
5′- GACGCAGCATCGACATATGGTCAGCTAGCGGCGCCGCCGG

pSKDcDNA-2.0 (theophylline-selectable (Clone 8.1 switch) cDNA cloning vector)

Clone 8.1, containing the theophylline-activated riboswitch reported previously,[2] was digested with *Kpn*I and *Hind*III, and the riboswitch-containing fragment was isolated, cloned into those sites in pSKDcDNA-1.0, and introduced into *E. coli*. The resulting plasmid, pSKDcDNA-2.0 (also known as pSKD1078.1), was then isolated.

pSKDcDNA-3.0 (theophylline-screenable (Clone 8.1 switch) cDNA cloning vector)

The promoter-containing fragment from pSKD1078.1 was isolated, after digestion with *Kpn*I and *Sac*I. The *IS10-lac*Z gene fragment was isolated from pSKD445.1,[2] after digestion with *Kpn*I and *Sac*I. Both isolated fragments were ligated together and introduced into *E. coli*, resulting in plasmid pCLG. A cassette strategy was used to add the IS10 promoter to the resulting vector. A PCR product (E) was made using primers SKD-178 and SKD-283 with template pSKD1078.1. A PCR product (F) was made using primers SKD-282 and SKD-56 with template Clone 8.1.[2] PCRs E and F were mixed and amplified using primers SKD-178 and SKD-56, resulting in PCR G. PCR G was digested with *BamH*I and *Hind*III and cloned into those sites in plasmid pCLG, resulting in pSKDcDNA-3.0 (also known as pSKD1248.2).

SKD-283
5′- CATTAGGGGATTCATCAGCCCTGGCGTTACCCAACTTAATCGC

SKD-282
5′- CTGATGAATCCCCTAATGATTTTGGTAAAAATCATTAAG

pCDF-ctCPR (pCDF-1b with IPTG-inducible ctCPR gene)

A linker region containing a Ptac1 variant promoter, *lac* operator, ribosome binding site, and *Nco*I and *Afl*II sites, was constructed by PCR and inserted into pCDF-1b (Novagen). A PCR product (G) was made by amplifying pCDF-1b using primers SKD-557 and SKD-554. PCR G was digested with *Afl*III and *Asc*I and cloned into the *Nco*I and *Asc*I sites in pCDF-1b, resulting in plasmid pSKD2980.1. The cytochrome P450 reductase gene from *C. tropicalis*[19] was subcloned from pDUET-ctCPR[16] using *Nco*I and *Afl*II, and cloned into those sites in pSKD2980.1, producing plasmid pCDF-ctCPR (also known as pSKD3027.1).

SKD-557
CCCACATGT*GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGG***AATTGTGAGCGGATAACAATT**
CCCCTGTAGAAATAATTTTGTTTAAC
(*AflIII site*, *tac promoter*, ***lac* operator**)

SKD-554
GGTGGCGCGCCTCTTAAGTGTACACGAGCTCCCAATTG

5.5 References

(1) Lovett, R. *New Scientist* **2005**, *187*, 38-41.

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

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

(4) Chou, T. M.; Benowitz, N. L. *Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol.* **1994**, *109*, 173-189.

(5) Nurminen, M. L.; Niittynen, L.; Korpela, R.; Vapaatalo, H. *Eur. J. Clin. Nutr.* **1999**, *53*, 831-839.

(6) Heilmann, W. In *Coffee: Recent Developments*; Clark, R. J., Vitzthum, O. G., Eds.; Wiley-Blackwell: 2001, p 108-124.

(7) Ashihara, H.; Crozier, A. *Trends Plant Sci.* **2001**, *6*, 407-413.

(8) Hein, L.; Gatzweiler, F. *Ecol. Econ.* **2006**, *60*, 176-185.

(9) Silvarolla, M. B.; Mazzafera, P.; Fazuoli, L. C. *Nature* **2004**, *429*, 826-826.

(10) Ogita, S.; Uefuji, H.; Yamaguchi, Y.; Koizumi, N.; Sano, H. *Nature* **2004**, *430*, 417-417.

(11) Ashihara, H.; Crozier, A. *J. Agr. Food Chem.* **1999**, *47*, 3425-3431.

(12) Kato, M.; Mizuno, K.; Fujimura, T.; Iwama, M.; Irie, M.; Crozier, A.; Ashihara, H. *Plant Physiol.* **1999**, *120*, 579-586.

(13) Ashihara, H.; Monteiro, A. M.; Moritz, T.; Gillies, F. M.; Crozier, A. *Planta* **1996**, *198*, 334-339.

(14) Woycechowsky, K. J.; Hilvert, D. *Eur. J. Biochem.* **2004**, *271*, 1630-1637.

(15) Taylor, S. V.; Kast, P.; Hilvert, D. *Angew. Chem. Int. Edit.* **2001**, *40*, 3311-3335.

(16) Chang, M. C. Y.; Eachus, R. A.; Trieu, W.; Ro, D. K.; Keasling, J. D. *Nat. Chem. Biol.* **2007**, *3*, 274-277.

(17) Haudenschild, C.; Schalk, M.; Karp, F.; Croteau, R. *Arch. Biochem. Biophys.* **2000**, *379*, 127-136.

(18) Leonard, E.; Yan, Y. J.; Koffas, M. A. G. *Metab. Eng.* **2006**, *8*, 172-181.

(19) Sutter, T. R.; Sanglard, D.; Loper, J. C. *J. Biol. Chem.* **1990**, *265*, 16428-16436.

(20) Brady, S. F.; Chao, C. J.; Handelsman, J.; Clardy, J. *Org. Lett.* **2001**, *3*, 1981-1984.

(21) Uchiyama, T.; Abe, T.; Ikemura, T.; Watanabe, K. *Nat. Biotechnol.* **2005**, *23*, 88-93.

(22) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001.

(23) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985**, *33*, 103-119.

(24) Hartley, R. W. *J. Mol. Biol.* **1988**, *202*, 913-915.

Chapter 6 Studies Toward Engineering Synthetic Riboswitches that Respond to (*S*)-Coclaurine and 2-Methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline

6.1 Introduction

Engineered microbes hold great potential for the mass production of medicinally important small molecules, such as artemesinin[3] for treating malaria or paclitaxel[4] for treating cancer, in a low cost fashion. And, in contrast to traditional synthetic organic approaches for constructing molecules, microbial biosynthesis does not require toxic solvents or metals,[5] and the starting material for a number of these transformations is glucose,[6,7] which is readily available. While microbial biosynthesis and traditional organic synthesis are different in many respects, they also have many similarities; the multi-step synthesis of a natural product is planned and then conducted in a linear manner, which is analogous to the process of genetically engineering a microbe with heterologous genetic pathways where, after implementation, the product of one enzymatic reaction serves as the substrate for another.[8] However, engineering a microbe to synthesize a particular small molecule, especially a natural product, can be difficult.

If one wishes to microbially synthesize a small molecule natural product from an orthogonal organism (such as a plant) using a microbe, the genes that carry out the synthesis must be isolated, and they must function within the heterologous host. And, oftentimes, the expression levels (or activities) of genes in a heterologous biosynthetic pathway in bacteria must be optimized because one enzyme (out of many) may not be sufficiently expressed, which prevents efficient production of the desired product.[9] The optimization of gene expression levels for this purpose can be difficult if the final small molecule product has no visible signal (i.e. color) or provides no survival advantage for the cell (i.e. amino acids or glucose), and, therefore, it must be monitored with a number of other techniques, including gas chromatography or mass spectroscopy-neither of

which are high-throughput or cost effective.[10,11] As such, researchers are actively developing methods to monitor small molecule biosynthesis in vivo which are cheap and easy (high-throughput).

One potential sensing platform that may achieve these goals involves synthetic riboswitches, which can, in principle, be constructed to monitor levels of small molecules produced in bacteria; as long as an RNA aptamer can be raised to the target molecule[12,13] (issues of permeability are not necessarily important in this application, since the molecule's synthesis is carried out in vivo). Synthetic riboswitches have been constructed to respond to a number of structurally distinct small molecules in eukaroyotes (tetramethylrosamine,[14] tetracycline,[15] and neomycin[16]); and in prokaryotes (theophylline[17] and 3-methylxanthine[12]), which highlights their recognition capability. In addition, they are often highly selective for their cognate ligand. One theophylline-activated riboswitch in *E. coli* responds to theophylline but not caffeine,[12] which differs from theophylline by only a single methyl group. This suggests synthetic riboswitches may be capable of monitoring enzymatic transformations in vivo, where the substrate and product are often structurally similar.

The heterologous biosynthesis of alkaloids in bacteria serves as an ideal pathway to monitor using synthetic riboswitches, since RNA aptamers have been raised to several plant alkaloids,[18,19] and several alkaloid biosynthesis genes have been successfully expressed in bacteria.[1] Figure 6.1 shows the biosynthetic pathways for a variety of isoquinoline alkaloids found in plants, including the medicinally important analgesics morphine and codeine; and the enzymes which have been successfully expressed in *E. coli* are shown in bold.[1] Notably, a number of important alkaloids are derived from

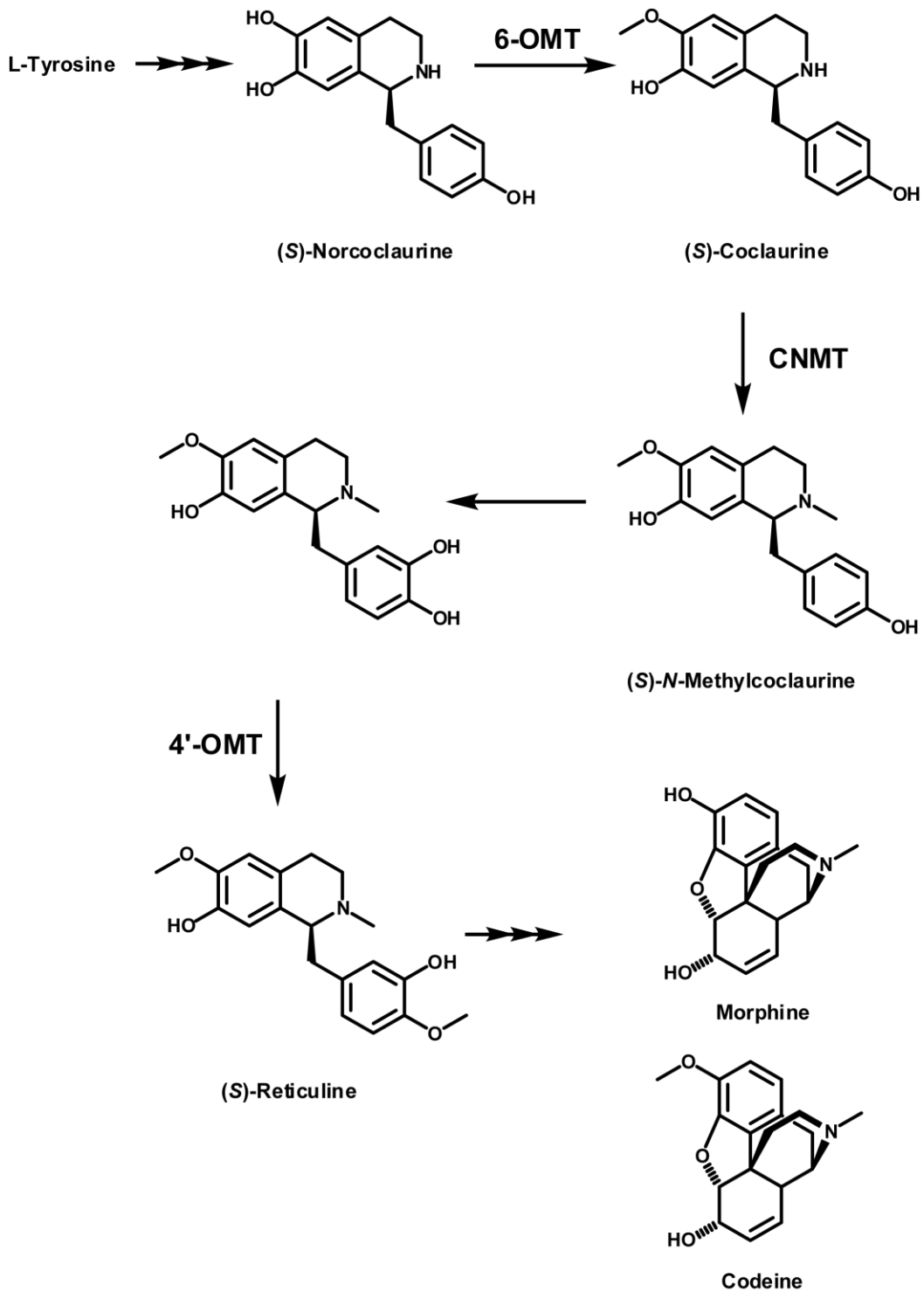

Figure 6.1: Biosynthetic Pathway for a Variety of Isoquinoline Alkaloids in Plants. 6-OMT, CNMT, and 4'-OMT from *C. japonica* have been expressed in *E. coli*.[1] Morphine and codeine are produced in *Papaver somniferum* (opium poppy).

(*S*)-reticuline,[1] and there are, to date, no simple and reasonable methods to produce it in gram quantities.[20] Biosynthesizing (*S*)-reticuline with bacteria may potentially circumvent these limitations and, if large enough quantities can be made, the resulting (*S*)-reticuline could be used in the synthesis of medicinally important alkaloids and their analogs.[20] Clearly, research in this area is important.

Herein we present our attempts at constructing synthetic riboswitches to monitor the products of two separate enzymatic reactions from the (*S*)-reticuline biosynthetic pathway, namely: (*S*)-adenosyl-l-methionine: norcoclaurine 6-O-methyltransferase (6-OMT); and (*S*)-adenosyl-l-methionine: coclaurine *N*-methyltransferase (CNMT) from *Coptis japonica*. Both of these enzymes have been functionally expressed in *E. coli,*[1] and CNMT has been used for the bioconversion of 6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline (DMTHIQ) to *N*-methyl-6,7-dimethoxy-1,2,3,4-tetrhydroisoquinoline (*N*-DMTHIQ) in *E. coli*.[2] We reasoned that if we could successfully monitor these reactions in vivo with riboswitches, we could then use our system to screen or select for enzyme variants with greater activity, and, eventually, create a bacterial strain which produces a high yield of (*S*)-reticuline. While we were not successful, our research led to improvements in our aptamer selection process, and we also designed a new method for creating synthetic riboswitches from a library of potential small molecule-binding aptamers.

6.2 Results and Discussion

In a previous publication, we reported the development and application of a high-throughput screen for identifying theophylline responsive synthetic riboswitches in

E. coli, and we reasoned the screen could be used to convert other small molecule-binding RNA aptamers into functional synthetic riboswitches.[13]

To monitor the 6-OMT reaction, presented above, we desired RNA aptamers that recognize (*S*)-coclaurine, but not (*S*)-norcoclaurine; and for the CNMT reaction, we desired RNA aptamers that recognize *N*-DMHTIQ, but not DMHTIQ (Figure 6.2).

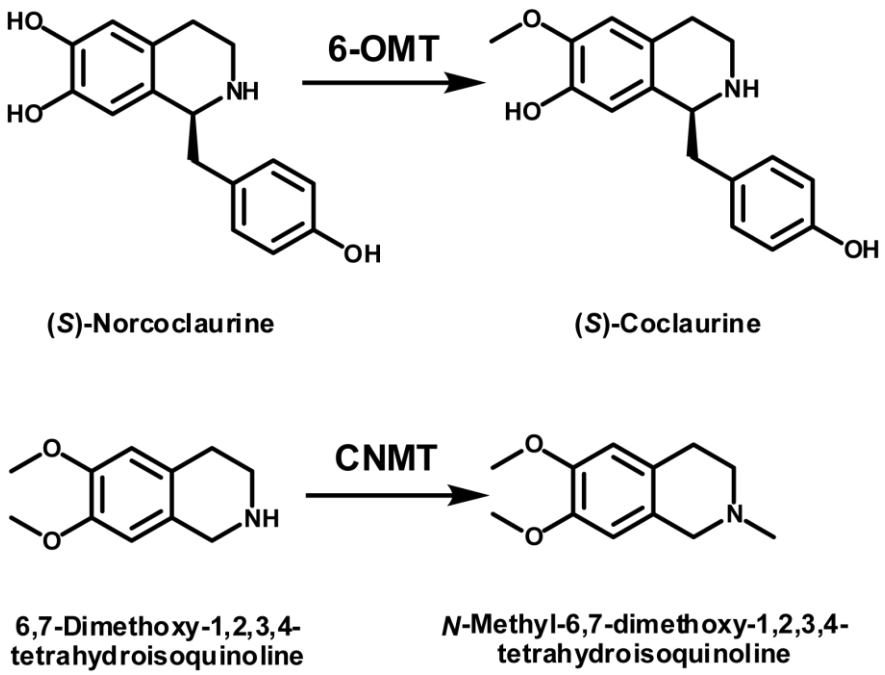

Figure 6.2: Enzymatic Reactions of 6-OMT and CNMT. Both enzymes have been expressed in *E. coli*, and the

6.2.1 Optimization of SELEX Procedure

We encountered two problems when we began our aptamer selections. First, the primers we initially chose formed a variety of PCR products, instead of a single product. Second, our RNA yields from in vitro transcription were very low. After some trial and error, we solved the primer problem by combining primers from two previously reported publications,[18,21] which were subsequently used to amplify a randomized oligo containing an N40 region. And, after several attempts with multiple systems, we were able to

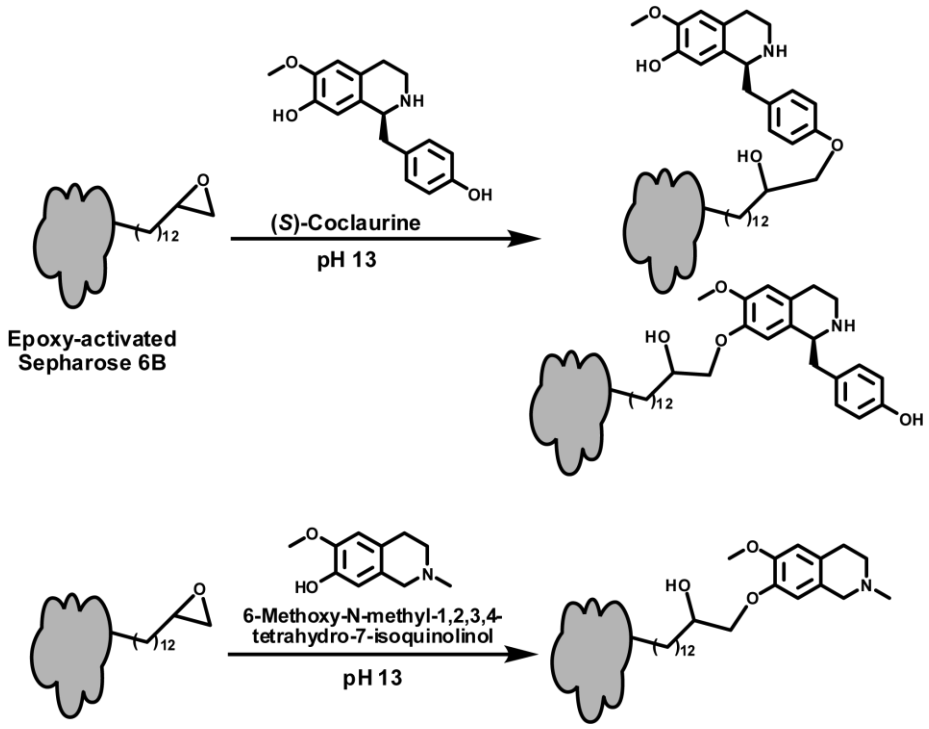

Figure 6.3: Coupling of Small Molecules Used for Selections. Both molecules were coupled to epoxide groups on Sepharose 6B.

consistently obtain RNA in high yield and purity by using the Ampliscribe™ T7-Flash in vitro transcription kit, which also decreased our RNA synthesis time from twelve hours to less than one. These improvements allowed us to carry out a complete round of SELEX in a twelve hour period (one round per day).

6.2.2 Selections for (*S*)-Coclaurine Aptamers

In a typical selection experiment for small molecule-binding RNA aptamers, the target small molecule is covalently attached to a solid support, such as agarose.[22] To this end, (*S*)-coclaurine was covalently attached to epoxy-activated Sepharose 6B through the hydroxyl groups of (*S*)-coclaurine (Figure 6.3), resulting in a column-bound concentration of approximately 833 µM. While (*S*)-coclaurine can couple to the epoxide groups on the column through either of its hydroxyl groups, this should not present a

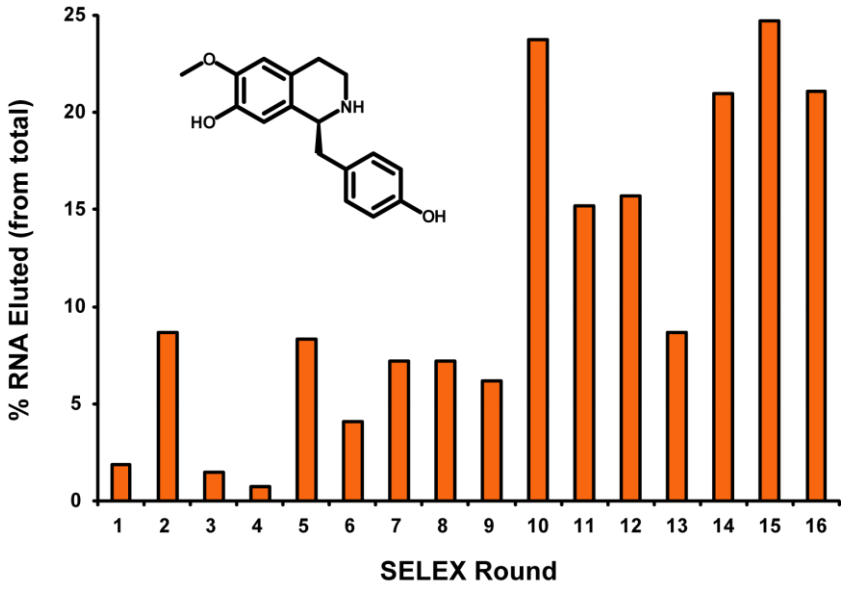

Figure 6.4: Percentage of Total RNA Eluted with (*S*)-Coclaurine (5 mM). Mutagenic PCR was performed during rounds 8 and 9; a counter selection for (*S*)-norcoclaurine was performed at round 13, prior to elution with (*S*)-coclaurine; and the wash volume was increased after round 10.

problem, as other groups have successfully selected aptamers using heterologously-coupled small molecules, such as the antibiotic streptomycin, which was also coupled to epoxy-activated Sepharose 6B through a multitude of hydroxyl groups.[23]

After (*S*)-coclaurine was coupled to the column, we performed a total of 16 rounds of selection to obtain RNA aptamers. The percentage of RNA eluted with 5 mM (*S*)-coclaurine during each round is shown in Figure 6.4. Mutagenic PCR was conducted during rounds 8 and 9, to add diversity to the pool and increase the binding percentage; a counter selection for (*S*)-norcoclaurine was performed during round 13, since we desired an aptamer which discriminates between (*S*)-coclaurine and (*S*)-norcoclaurine; and the wash volume (prior to elution) was increased from 6 column volumes to 10 column volumes after round 10, to increase the stringency of the selection. After round 15, 24% of the total RNA applied to the column was eluted with the free (*S*)-coclaurine. This was

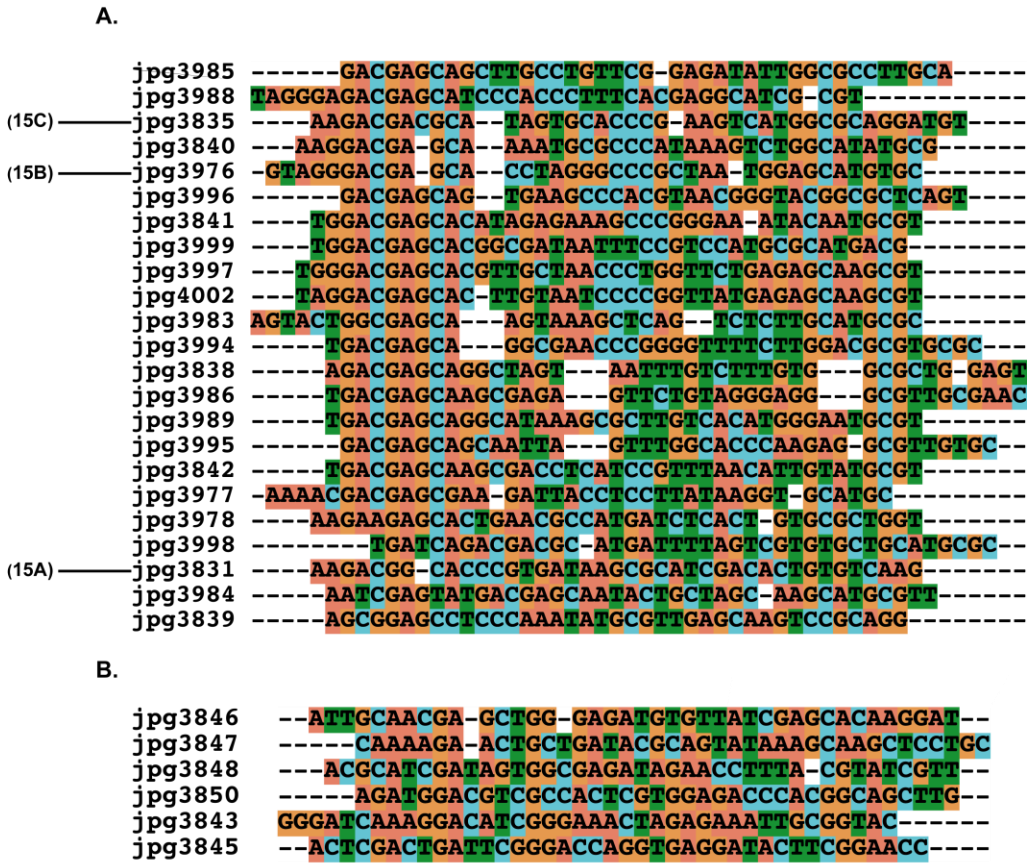

Figure 6.5: Alignment of N40 Regions from Selections. (A) N40 regions from clones after 15 rounds of selection for (*S*)-coclaurine. Sequence jpg3831 appeared four times and jpg3978 appeared twice (from 27 total sequences). The motif "ACGAGCA" appears several times. Sequences annotated as 15A, 15B, and 15C were analyzed in vitro. (B) N40 regions from clones after 15 rounds of selection for *N*-methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline. No conserved motifs appeared.

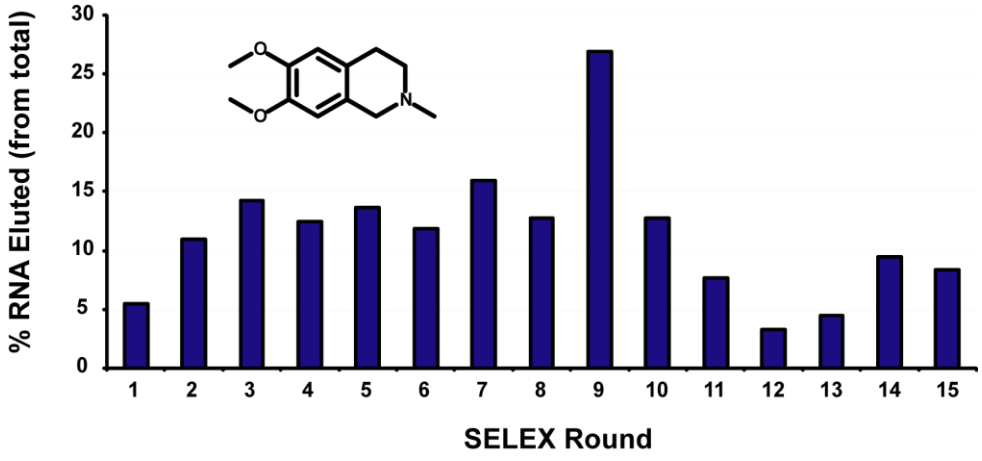

Figure 6.6: Percentage of RNA Eluted with *N*-methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline (5 mM). RNA percentage (from total) eluted during each SELEX round. Mutagenic PCR was performed during rounds 8 and 9; a counter selection with 6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline was performed at round 12; and the wash volume, before elution, was increased after round 10.

a large improvement from the 2% eluted after round 1, suggesting the selection experiment was successful.

To analyze individual members from the selection, we cloned the PCR product after round 15 into a vector and sequenced 27 random clones. One sequence appeared four times, and another appeared two times. Additionally, alignment of the resulting sequences (Figure 6.5) showed a potential conserved motif of "ACGAGCA" in several of the sequences. Combined, the appearance of a single sequence multiple times and the potential conserved motif indicated the selection procedure successfully enriched the population, but further biochemical analysis was necessary at this point to confirm the ability of the selected sequences to bind (*S*)-coclaurine.

6.2.3 Selections for *N*-methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline Aptamers

A derivative of *N*-DMHTIQ, with a free hydroxyl group, was coupled to epoxy-activated Sepharose 6B under basic conditions (Figure 6.3), resulting in a homogeneously-coupled matrix with a column-bound concentration of approximately 833 µM. After coupling, we performed a total of 15 SELEX rounds; the percentage RNA eluted with 5 mM *N*-DMHTIQ during each round is shown in Figure 6.6. A mutagenic PCR amplification was performed during rounds 8 and 9, to add diversity to the pool; a counter selection for DMHTIQ was performed during round 12, since we desired aptamers which recognize *N*-DMHTIQ but not DMHTIQ; and the wash volume was increased from 6 to 10 column volumes after round 10, to increase the stringency of the selection. Clearly, the percentage of RNA eluted did not consistently increase after the

first round, suggesting the selection was not successful (the measurement at round 9 was, most likely, an aberration). Perhaps this is not surprising, as *N*-DMHTIQ is relatively small and contains few functional groups for RNA to interact with.

Though the data indicated the selection was not successful, we cloned the PCR product after round 15 into a vector and sequenced six random clones (Figure 6.5) to verify our results. No conserved sequences were found, which was not unexpected. We decided against further sequencing or testing of these sequences, in light of our data.

6.2.4 Biochemical Analysis of Putative (*S*)-Coclaurine-Binding RNA Aptamers

Sequencing of several random clones from the (*S*)-coclaurine selections revealed that a single sequence (clone 15A) appeared four times, from a total of 27 sequenced clones, which suggested it was a potential (*S*)-coclaurine binder. To test this possibility, RNA comprising the full-length 15A RNA sequence, which included the constant regions, was transcribed in vitro with radiolabeled UTP, and the resulting RNA was isolated. This RNA was applied to both a capped Sepharose 6B column, with no (*S*)-coclaurine, and a Sepharose 6B column which was coupled to (*S*)-coclaurine. After incubation of the RNA, the columns were washed with 5 column volumes of buffer; the wash fractions and column material were then subjected to scintillation counting. For the uncoupled column, only 19% of the total RNA remained on the column after the wash. In contrast, for the (*S*)-coclaurine-coupled column, 55% of the total RNA remained on the column after the wash, indicating clone 15A binds (*S*)-coclaurine and only weakly associates with the uncoupled column material.

In the above experiment, the full-length RNA molecule isolated from the selection experiment was used, and we wondered if the N40 region, alone, would bind (*S*)-coclaurine. To this end, radiolabeled RNA, comprising the N40 region from clone 15A, was prepared using in vitro transcription. After purification, the RNA was applied to an (*S*)-coclaurine-coupled Sepharose 6B column, as before, and then washed with 5 column volumes of buffer. All fractions were subjected to scintillation counting; only 25% of the total RNA remained on the column, suggesting the constant regions are involved in the function of the RNA. This result is not surprising; the Szostak group has found that the constant regions of selected aptamers are sometimes involved in the functionality of the RNA.[24]

Since clone 15A appeared to bind the (*S*)-coclaurine-coupled column, we decided to confirm this observation using in vitro structure probing. The full-length RNA of clones 15A, 15B, and 15C (we decided to test two additional clones in parallel) was prepared in vitro and 5' end-labeled with ^{32}P. Each labeled RNA was subjected to in-line probing[25] and T1 nuclease cleavage,[26] separately, in the presence and absence of varying concentrations of (*S*)-coclaurine, followed by separation with denaturing PAGE. Analysis of the gels (not shown) indicated there was no change in the structure of the RNAs in the presence or absence of (*S*)-coclaurine. A maxiumum concentration of 1 mM (*S*)-coclaurine was used in the experiments, and this concentration should have been sufficient to alter the equilibrium of the putative aptamers in vitro (assuming their dissociation constants are less than 1 mM), but it is possible that these RNAs do not adopt different structures when in the presence of ligand. For example, the glucosamine-6-phosphate responsive riboswitch does not adopt a different tertiary

structure in the presence of ligand;[27] rather, its binding pocket is always in the proper binding conformation, and our RNA aptamers may function in a similar fashion.

We also attempted to obtain a binding constant for clone 15A using isothermal titration calorimetry, but our data was inconclusive. Unfortunately, this technique is only useful for RNAs that possess low nanomolar dissociation constants for their ligand,[28] and clone 15A may not bind (*S*)-coclaurine within that range.

6.2.5 Converting a Single Putative (*S*)-Coclaurine-Binding Aptamer into a Synthetic Riboswitch

Since the column-binding experiments indicated clone 15A binds (*S*)-coclaurine, we decided to try and convert the putative aptamer into a functioning synthetic riboswitch using our previously reported high-through put screen.[13] Clone 15A (full-length form) was cloned upstream of the *IS10-lac*Z fusion reporter gene, with a randomized region of 10 base pairs (including the ribosome-binding site) between the aptamer and start codon. This library was introduced into *E. coli* and plated on LB agar supplemented with ampicillin (to maintain the plasmid), but no theophylline, resulting in approximately 8000 total colonies, distributed amongst two separate plates. The 96 whitest colonies on each plate were picked using a robotic colony picker and their β-galactosidase activities were assayed after growth in the presence and absence of 100 μM (*S*)-coclaurine, as reported previously.[13] Unfortunately, none of the clones tested were responsive to (*S*)-coclaurine, but only approximately 8000 colonies were analyzed, and, if functional clones are rare, a screen or selection with a higher throughput may be necessary to identify them.

Topp and Gallivan have developed a high-throughput selection for synthetic riboswitches based on cell motility,[29] and we decided to use this to identify a potential (*S*)-coclaurine riboswitch from our library, since it can assay greater than 200,000 variants in a single experiment. Clone 15A was inserted upstream of the *IS10-che*Z chemotaxis gene, under the control of either the *IS10* or *tac* promoters, along with a 10 base pair randomized linker between the aptamer and start codon, and the libraries were introduced into a CheZ deficient strain of *E. coli*. Approximately 200,000 individuals from each promoter library were subjected to two successive "off" motility selections (cells which do not migrate in the absence of ligand because of low *che*Z expression) on low percentage agar, in the absence of ligand, followed by a single "on" motility selection (cells which migrate in the presence of ligand because of higher *che*Z expression) in the presence of 100 μM (*S*)-coclaurine (this concentration does not affect cell motility). After the "on" selection, the plasmids were isolated from the cells and the riboswitch-containing segments were digested out and cloned upstream of an *IS10-lac*Z reporter gene. Cells harboring the selected switches (96 colonies from both the *IS10* and *tac* promoter libraries) were assayed for β-galactosidase activity in the presence and absence of (*S*)-coclaurine, and, again, no (*S*)-coclaurine responsive riboswitches were found. The permeability of (*S*)-cocolaurine to *E. coli* is not known, and, if it is low, our screens and selections may never work. Alternatively, clone 15A may not be the appropriate RNA to use for conversion to a synthetic riboswitch, and the possibility remains that other sequences from the (*S*)-coclaurine aptamer selection are better suited for this task. However, this would require cloning individual sequences upstream of various reporter genes for use with the high-throughput screen or motility selection, as

before, which is time consuming. What is needed is a method for screening or selecting synthetic riboswitches from a pool of aptamer sequences, in a *single* experiment.

6.2.6 Identifying a Synthetic Riboswitch from a Pool of Potential Aptamers

As mentioned before, it is time consuming to clone a single aptamer upstream of a reporter gene and then screen or select for a functioning riboswitch, especially if there are a large number of potential sequences from which to choose, and the chosen sequence may never function as a synthetic riboswitch for a number of reasons, which would require testing additional sequences. We desired a better strategy. As such, we developed a method for cloning a pool of potential aptamers from a selection experiment upstream of a reporter gene, such that between the potential aptamer sequences and start codon there would be a randomized region. Figure 6.7 shows a diagram of the strategy. The initial PCR amplifies the DNA pool of sequences from an aptamer selection experiment, and this PCR also adds a randomized region of 10 base pairs between the aptamer and the start codon of the reporter gene. The resulting product is then assembled with an *IS10* gene fragment using cassette PCR, cloned upstream of an appropriate reporter gene and, after introduction into *E. coli*, a screen or selection is performed to identify functioning synthetic riboswitches.

6.2.7 Identifying an (*S*)-Coclaurine Riboswitch Using a Novel Strategy

We used the strategy presented above to clone multiple sequences from the (*S*)-coclaurine aptamer selection upstream of the *IS10-lac*Z, *IS10-che*Z, and *IS10-dsred express* reporter genes, separately. The constructs were introduced into *E. coli* and four

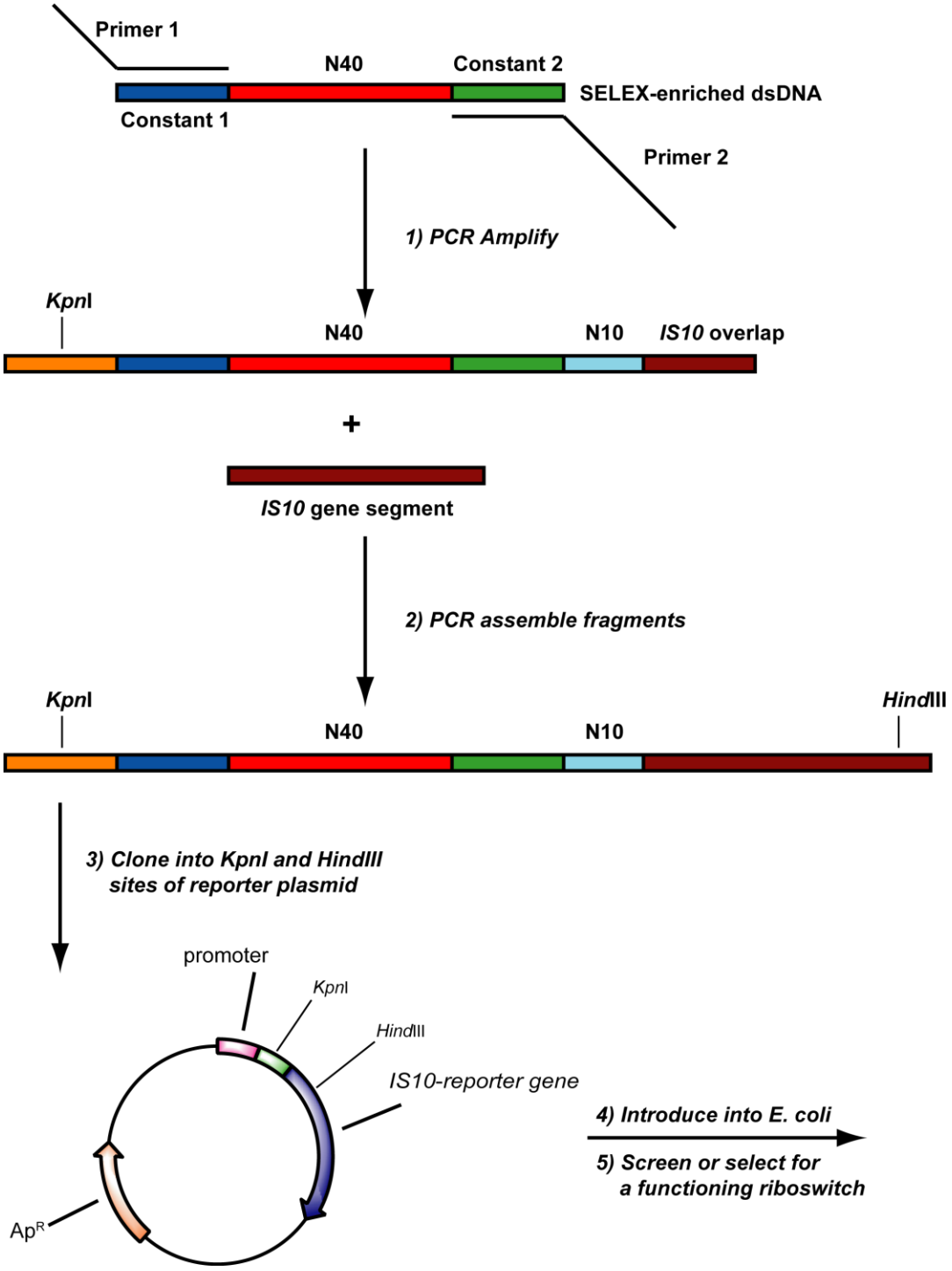

Figure 6.7: Strategy for Cloning Multiple Aptamers with Multiple Randomized Regions Upstream of a Reporter Gene. An enriched double-stranded DNA product from SELEX is amplified with two specific primers, which add a unique 5' cut site and a 3' N10 region with an overlap to the *IS10* gene fragment. This PCR product is assembled with an *IS10* gene fragment, cut with *Kpn*I and *Hind*III, and cloned into those sites in an appropriate reporter plasmid. After introduction into *E. coli*, a screen or selection is performed to identify functioning synthetic riboswitches.

clones were sequenced; all clones contained a different potential aptamer sequence and randomized region, indicating the cloning strategy was successful.

Using our high-throughput screen, we attempted to identify a synthetic riboswitch from sequences cloned upstream of *IS10-lac*Z. Approximately 8000 colonies were analyzed using our high-throughput screen, and, unfortunately, no functioning synthetic riboswitches were found. We decided to turn to a more high-throughput approach.

Libraries under the control of either the *IS10* or *tac* promoter, upstream of the *IS10-che*Z gene, were introduced into a CheZ deficient strain of *E. coli* in preparation for motility based selections. Approximately 200,000 cells from each promoter library were subjected to two "off" motility selections in the absence of ligand, and then to a single "on" selection in the presence of 100 μM (*S*)-coclaurine. Motile cells from the "on" selection were grown and the plasmids from the cells were isolated. As described earlier, the riboswitch-containing segments from the plasmids were digested out, cloned upstream of an *IS10-lac*Z reporter gene, and introduced into *E. coli*. The β-galactosidase activity for several (96) colonies from each promoter library was assayed in the presence and absence of (*S*)-coclaurine. Again, no functioning synthetic riboswitches were found.

One problem with the motility selection technique is that we could not use a concentration of (*S*)-coclaurine above 100 μM, due to inhibition of cell motility above this concentration. If the concentration of (*S*)-coclaurine required to activate a putative synthetic riboswitch from our library is greater than 100 μM, the motility selection technique will not successfully identify them. To circumvent this concentration limitation, we decided to use fluorescence activated cell sorting (FACS), which does not have a stringent a limitation on small molecule concentration, to identify a functioning

synthetic riboswitch. With FACS, cells harboring a library of potential riboswitches, upstream of a fluorescent reporter gene (we used *dsred express*), are sorted for low fluorescence and then high fluorescence after growth in the absence and presence of ligand, respectively. Other members of our lab (Sean Lynch) have used FACS to successfully identify theophylline-activated synthetic riboswitches from a large library harbored in *E. coli*, and we reasoned the technique could also identify (*S*)-coclaurine responsive riboswitches.

Cells harboring the library of potential (*S*)-coclaurine riboswitches, upstream of *IS10-dsred express*, were grown in the absence of ligand, and a fraction of the cell culture (~600,000 cells) was sorted using flow cytometry. Cells displaying low levels of fluorescence ("off") were collected. These cells were grown without ligand, and approximately 600,000 of the enriched cells were sorted; those displaying the "off" phenotype were collected again. At this point, the collected cells were grown with varying concentrations of (*S*)-coclaurine (1 mM, 2 mM, or 3 mM), and an aliquot (~600,000 cells) from each ligand concentration was sorted, and, fortunately, the high concentrations of (*S*)-coclaurine used did not affect the sorting procedure. Cells displaying high levels of fluorescence ("on") were collected, grown, and their plasmids were isolated. The riboswitch-containing segments from the plasmids were isolated, cloned upstream of an *IS10-lac*Z reporter gene, and introduced into *E. coli*. The β-galactosidase activity for several clones (96) from each "on" sorted library was assayed in the presence and absence of (*S*)-coclaurine (the concentration of (*S*)-coclaurine used in the assays was the same as that used during sorting), and, as in our previous experiments, no functioning synthetic riboswitches were found. It is difficult to pinpoint why the

sorting strategy was not successful, but the reasons are probably related to the permeability of (*S*)-coclaurine and the binding affinity of the aptamer pool.

6.3 Conclusion

The metabolic engineering of microbes for the production of small molecules is a promising field, but a number of problems remain unsolved. For one, it is difficult to monitor the intracellular concentrations of small molecules (data which can be used to optimize the biosynthetic pathway), unless the small molecule product(s) has some sort of phenotype or physiological significance.[10] And this limitation requires that researchers monitor intracellular metabolite concentrations with expensive techniques, such as gas chromatography.

Synthetic riboswitches may offer a solution; they can, in principle, be made to respond to any nontoxic, cell permeable small molecule which can interact with an RNA.[13] Thus, synthetic riboswitches can be engineered to recognize an almost limitless number of compounds. And, since they are replicated by the microbe, their cost, after construction, is relatively low.

In this Chapter, we presented our attempts at constructing synthetic riboswitches that recognize the products of two separate enzymatic reactions in the (*S*)-reticuline biosynthetic pathway, when expressed in *E. coli*. We chose to monitor these enzymatic reactions because (*S*)-reticuline is a critical intermediate in the biosynthetic pathway for a number of medicinally important alkaloids,[1] including morphine, and we reasoned that if we could monitor the in vivo production of the intermediates, we could then use our system to screen or select for improved enzyme variants.

While our attempts at constructing the desired synthetic riboswitches were unsuccessful, we improved our aptamer selection procedure: we can now carry out a complete round of SELEX in a single day. Also, we developed a novel strategy to screen or select for functioning synthetic riboswitches from a pool of potential aptamer sequences in a single experiment, and this cloning technique has been successfully used to identify an atrazine-responsive riboswitch in our lab (Joy Sinha). Combined, these improvements will prove useful for other researchers.

6.4 Experimental

General Considerations

All plasmid manipulations were performed according to standard cloning techniques.[30] All plasmids were verified by sequencing. Qiagen kits were used for purification of plasmid DNA, PCR products, and enzyme digestions. Ampicillin was purchased from Sigma. X-gal was purchased from US Biological. *N*-methyl-6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline and 6,7-dimethoxy-1,2,3,4-tetrahydroisoquinoline were obtained from Aldrich. Oligonucleotides were purchased from IDT. TOP10, JM109[31] or JW1870[32] *E. coli* cells were used in all experiments.

Aptamer Selections-General Strategy

The doubled stranded DNA template was prepared by 12 cycles of PCR in a reaction (100 µL) with template SKD-537 (20 pmole) and the outer primers SKD-493 and SKD-535 (300 pmole each). After amplification, a fraction of the product was analyzed on an agarose gel, and the remainder was purified using a minelute kit (Qiagen).

A fraction of the eluted DNA (in DEPC dH_2O) was used in a 20 µL Ampliscribe T7 Flash in vitro transcription kit (Epicentre) reaction, supplemented with 10 µCi of [α-^{32}P]UTP (MP Biomedical). The transcription reaction was incubated at 37 ^{o}C for 30 min. and then DNAseI (1 µL) was added for 15 min. at 37 ^{o}C. Following digestion, the reaction was purified by 10% denaturing PAGE, the RNA band was visualized by UV and cut out of the gel. The gel band was chopped into pieces, placed into a 50 mL conical, along with 5 mL DEPC dH_2O, and shaken overnight at 37 ^{o}C to promote elution of the RNA. After elution, the RNA was ethanol precipitated and quantitated using standard techniques.[30]

For each round of selection, a 200 µL bed volume of small molecule-bound agarose was pipeted into a 2 mL disposable plastic column (Pierce) and allowed to settle at room temperature for 30 min. At this point, the column was washed with 2 mL of binding buffer B1 (250 mM NaCl, 20 mM Tris-HCl, 5 mM $MgCl_2$, pH 7.4) to equilibrate it. An aliquot of the radiolabeled RNA pool (300 pmole) was brought to 300 µL in buffer B1 and heated to 70 ^{o}C for 2 min., cooled to 4 ^{o}C for 2 min. and then applied to the column. After a 30 min. room temp. incubation, the RNA-containing column was washed with 1.2 or 2 mL of buffer B1, depending on the desired stringency, and the wash was collected for scintillation counting. Small molecule-binding RNAs were eluted from the column with 600 µL of a 5 mM solution of free small molecule dissolved in buffer B1, and collected. A fraction of the eluted RNA (20 µL) was saved for scintillation counting, and the remaining RNA was ethanol precipitated. The percentage of eluted RNA was calculated as the percentage of the RNA eluted with free small molecule from the total amount of RNA initially applied to the column.

Following precipitation, the eluted RNA was resuspended in 20 µL DEPC dH_2O and 10 µL of this was used in a reverse transcription reaction with primer SKD-535 (20 pmole) and Superscript III reverse transcriptase (Invitrogen), and the reaction was carried out according to the manufacturer. An aliquot (10 µL) of the reverse transcribed RNA was amplified with 12-20 cycles of PCR in a reaction (100 µL) with primers SKD-493 and SKD-535 (100 pmole each). Following amplification, a fraction of the PCR product was analyzed on an agarose gel, and the remainder was purified using a minelute kit. Transcription of the DNA was carried out as before, in preparation for the next round of selection.

PCR-amplified DNA, from various rounds of selection, was cloned into a TA vector (Promega) and introduced into *E. coli*. Insert-containing clones were grown and their plasmids were sent for sequencing.

SKD-537
5′- GGGACAGGGCTAGC-N40-CTGCAGGTCGACGCATGCGCCG
(N40 randomized template)

SKD-493
5′- TTC**TAATACGACTCACTATA**GGGACAGGGCTAGC
(**T7 promoter sequence,** anneals to SKD537)

SKD-535
5′- CGGCGCATGCGTCGACCTGCAG
(anneals to SKD537)

Coupling of (*S*)-coclaurine and 6-methoxy-N-methyl-1,2,3,4-tetrahydro-7-isoquinolinol to Epoxy-activated Agarose

(*S*)-coclaurine was synthesized using the procedure of Polniaszek and Kaufman[33] by S. Reyes. 6-methoxy-N-methyl-1,2,3,4-tetrahydro-7-isoquinolinol was synthesized according to the previously published procedure[34] by S. Reyes.

Epoxy-activated Sepharose 6B (2 g, GE Healthcare) was swollen in DEPC dH_2O, added to a fritted glass filter, and washed with 400 mL of DEPC dH_2O over the course of one hour. The resulting slurry was placed into a 50 mL conical and a solution of 2.5 mM (*S*)-coclaurine or 6-methoxy-N-methyl-1,2,3,4-tetrahydro-7-isoquinolinol in a high pH buffer (0.05 Na_2HPO_4, pH 13) was added to a final volume of 15 mL, and the slurry was gently rocked for 18 h at 37 ^oC. After coupling, the mixture was added to a fritted glass filter and washed with three alternating cycles of 3 mL of low pH buffer (0.1 M NaOAc, 0.5 M NaCl, pH 4) and then 3 mL of high pH buffer (0.1 M Tris-HCl, 0.5 M NaCl, pH 8). After washing, the column material was resuspended in buffer (10 mM Tris, pH 8). A previous report has estimated that approximately a third of the hydroxyl-containing small molecules couple to an epoxy-activated column at high pH,[23] and we used this data to estimate that our column-bound concentration was approximately 833 µM.

In-line Probing

Primers SKD-493 and SKD-535 were used to amplify single sequences, cloned in a TA vector, with PCR (20 cycles) and subsequently purified using a minelute kit (Qiagen). A fraction of the PCR product was used in a T7 Flash transcription kit reaction (Epicentre), separated using 10% denaturing PAGE, and eluted overnight in dH_2O. Eluted RNA was precipitated and dephosphorylated using calf intestinal phosphatase according to the manufacturer (New England Biolabs) and purified using a minelute kit. The dephosphorylated RNA was subsequently phosphorylated using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (MP Biomedical, 10 µCi) according to the manufacturer, followed by purification with a minelute kit.

Radiolabeled RNA (100,000 counts per minute (cpm)) was added to separate tubes containing in-line probing buffer (50 mM Tris-HCl [pH 8.5], 20 mM $MgCl_2$), and either no small molecule or up to 1 mM (S)-coclaurine in a 10 µL final volume. After mixing, the reactions were set at 25 °C for 40 h. Reactions were quenched by addition of 10 µL stop buffer and 3-10 µL fractions were separated by electrophoresis on a 10% denaturing PAGE. The gel was dried, exposed to a Phosphorimager screen (GE Healthcare), imaged on a Phosphorimager, and analyzed with ImageJ software (NIH)..

T1 nuclease sequencing and alkaline hydrolysis reactions were setup using radiolabled RNA (100,000 cpm) according to the manufacturer (Ambion). These were quenched as before.

Nuclease Probing

Primers SKD-493 and SKD-535 were used to amplify single sequences cloned in a TA vector using PCR (20 cycles) and purified using a minelute kit (Qiagen). A fraction of the PCR product was used in a T7 Flash transcription kit reaction (Epicentre), separated using 10% denaturing PAGE, and eluted overnight in dH_2O. Eluted RNA was ethanol precipitated, dephosphorylated using calf intestinal phosphatase according to the manufacturer (New England Biolabs) and purified using a minelute kit. The dephosphorylated RNA was subsequently phosphorylated using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (MP Biomedical, 10 µCi) according to the manufacturer, followed by purification with a minelute kit.

Radiolabeled RNA (100,000 cpm) was added to separate tubes containing 1X RNA structure probing buffer (Ambion), 6 mM $MgCl_2$ and either no small molecule or up to 1 mM (S)-coclaurine in a final volume of 9 µL. After mixing, the tubes were heated to 70 °C for 2 min and then cooled to 4 °C for 2 min. 1 µL T1 nuclease (Ambion) was added to each tube, mixed, and set at room temperature for 15 min. The reactions were quenched by adding 10 µL of stop buffer (95% formamide, bromophenol blue, 50 mM EDTA) and 3-10 µL fractions were separated by 10% denaturing PAGE. After electrophoresis, the gel was dried, exposed to a Phosphorimager screen (GE Healthcare) imaged, and analyzed with ImageJ software (National Institutes of Health).

T1 nuclease sequencing and alkaline hydrolysis reactions were setup using radiolabled RNA (100,00 cpm) according to the manufacturer (Ambion). These were quenched as before.

Procedure to Clone Multiple Aptamers with Multiple Randomized Regions Upstream of Various Reporter Genes

The insert containing a 5' *Kpn*I site was made using a cassette PCR strategy. A PCR product (A) was made using a fraction of the PCR product from the aptamer selections (above) as a template with primers JS-20 and SKD-550. A second PCR product (B) was made using the *IS10-lac*Z reporter gene plasmid (pSKD445.1)[13] as a template with primers SKD-551 and SKD-56. Products A and B were combined and PCR amplified in a reaction with primers JS-20 and SKD-56, resulting in PCR product C. PCR product C was digested with *Kpn*I and *Hind*III and cloned into those sites in an *IS10-lac*Z (pSKD445.1) or the *IS10-che*Z containing plasmids,[29] with either the *IS10* or *tac* promoter, depending on the screening or selection procedure used. The libraries in

pSKD445.1 were introduced into TOP10 *E. coli* cells; the libraries cloned into the *IS10-che*Z containing plasmids were introduced into JW1870 *E. coli* cells.

The insert containing a 5' *Asc*I insert was also made using a cassette PCR strategy. A PCR product (D) was made using a fraction of the PCR product from the aptamer selections (above) as a template with primers SKD-517 and SKD-550. A second PCR product (E) was made using the *IS10-dsred express* reporter gene plasmid (pSKD3021.1) as a template with primers SKD-551 and SAL-61. Products D and E were combined and PCR amplified in a reaction with primers SKD-517 and SAL-61, resulting in PCR product F. Product F was digested with *Asc*I and *Nco*I, cloned into those sites in pSKD3021.1, and introduced into TOP10 cells.

JS20
5′- CCCC**GGTACC**GGGACAGGGCTAGC
(***Kpn*I site,** binds to SELEX template)

SKD550
5′- *CGAGTTCGCACAT*CTTGTTGNNNNNNNNNNGCGTCGACCTGCAG
(O*verlap with IS10 gene,* anneals to SELEX template)

SKD551
5′- CAACAAGATGTGCGAACTCG
(Anneals to IS10 gene)

SKD517
5′- CCCC**GGCGCGCC**GGGACAGGGCTAGC
(***Asc*I site,** binds to SELEX template)

SAL61
5′- CCCTTGGTCACCTTCAGCTTGGCGGTCTGGG
(Anneals 3′ to *Nco*I site in *dsred express*)

High-throughput β-galactosidase Assay of Riboswitch Libraries

TOP10 cells containing the potential (*S*)-coclaurine riboswitch libraries in pSKD445.1 were screened using the previously reported technique.[13] The concentration of (*S*)-coclaurine used in the plate-based assays was either 100 μM (for clones identified

from the motility selection or for clones which were not prescreened); or 1 mM, 2 mM or 3 mM (for clones identified from flow cytometry).

Motility Based Selections of Riboswitch Libraries

JW1870 cells harboring the potential (*S*)-coclaurine riboswitch libraries, upstream of the *IS10-cheZ* gene, were subjected to motility based selections as reported previously.[29] The concentration of (*S*)-coclaurine used in the positive selection plates was 100 µM.

Fluorescence Activated Cell Sorting of Riboswitch Libraries

TOP10 cells harboring the potential (*S*)-coclaurine riboswitch libraries, upstream of *IS10-dsred express*, were grown overnight in liquid LB (5 mL) supplemented with ampicillin (50 µg/mL) with shaking at 37 ^{o}C. An aliquot (100 µL) of the saturated culture was used to inoculate a fresh culture of liquid LB (5 mL) supplemented with ampicillin (50 µg/mL) and either no small molecule or (*S*)-coclaurine (1 mM, 2 mM, or 3 mM). The cultures were grown at 37 ^{o}C with shaking until the OD_{600} reached 0.6, then an aliquot (500 µL) was pelleted by centrifugation and resuspended in 1-4 mL of phosphate buffered saline (137 mM NaCl, 10 mM K_2HPO_4, 2.7 mM KCl, pH 7.4).

Cells were sorted with a Becton Dickinson FACSVantage SEflow cytometer using an Innova70 spectrum laser tuned to 568 nm for excitation. Fluorescence was detected through a 630/22 band pass filter. Depending on the desired outcome, either the cells displaying low fluorescence or the cells displaying high fluorescence were collected into liquid LB supplemented with ampicillin (50 µg/mL). Collected cells were grown

overnight with shaking at 37 ^{o}C, and were either archived or used to start a fresh culture for further sorting experiments

Construction of pSKD3021.1 (fluorescent reporter plasmid)

An *Asc*I site was added immediately 3' to the *Kpn*I site in plasmid pSAL6.1 by cassette PCR, resulting in plasmid pSKD3021.1. The description of the pSAL6.1 plasmid is below.

Plasmid pSAL6.1

(Ptac1 promoter-wild-type 8bp switch-*IS10 dsred express*)

This plasmid is derived from our previously reported synthetic riboswitch.[12] A cassette mutagenesis strategy was used to generate pSAL006. A PCR product (A) was generated by usingpSKD445.1[13] 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 (Clontech) 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 *Nde*I and *Afl*III. Plasmid pSKD850.2 (Chapter 2) was digested with *NdeI* and *Afl*III, dephosphorylated and gel purified. Digested PCR product C was then cloned into the digested pSKD850.2 to yield pSAL6.1.

SKD-163
5'- GCTAGCAGGGGACCCAGGCCAGGGATCGATCCCCCCAAGCTTG
(Anneals to IS10 gene)

SAL-002
5′- *GGCCTGGGTCCCCTGCTAGC*ACCATGATTACGCCAAGCTTGCATGCCTGCAG
(*Overlap with SKD-163*, anneals to vector)

SAL-003
5′- CGCGAGAA*CATGT*CTACAGGAACAGGTGGTGGCGGCCCTCGGC
(*AflIII site*, anneals to vector)

6.5 References

(1) Morishige, T.; Tsujita, T.; Yamada, Y.; Sato, F. *J. Biol. Chem.* **2000**, *275*, 23398-23405.

(2) Morishige, T.; Choi, K. B.; Sato, F. *Biosci. Biotechnol. Biochem.* **2004**, *68*, 939-941.

(3) Hale, V.; Keasling, J. D.; Renninger, N.; Diagana, T. T. *Am. J. Trop. Med. Hyg.* **2007**, *77*, 198-202.

(4) Walker, K.; Croteau, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13591-13596.

(5) Li, W.; Xie, D.; Frost, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2874-2882.

(6) Guo, J.; Frost, J. W. *Org. Lett.* **2004**, *6*, 1585-1588.

(7) Yi, J.; Draths, K. M.; Li, K.; Frost, J. W. *Biotechnol. Prog.* **2003**, *19*, 1450-1459.

(8) Chang, M. C.; Keasling, J. D. *Nat. Chem. Biol.* **2006**, *2*, 674-681.

(9) Pfleger, B. F.; Pitera, D. J.; Smolke, C. D.; Keasling, J. D. *Nat. Biotechnol.* **2006**, *24*, 1027-1032.

(10) Brady, S. F.; Chao, C. J.; Handelsman, J.; Clardy, J. *Org. Lett.* **2001**, *3*, 1981-1984.

(11) Hyun, C. G.; Bililign, T.; Liao, J.; Thorson, J. S. *Chembiochem* **2003**, *4*, 114-117.

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

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

(14) Buskirk, A. R.; Landrigan, A.; Liu, D. R. *Chem. Biol.* **2004**, *11*, 1157-1163.

(15) Suess, B.; Hanson, S.; Berens, C.; Fink, B.; Schroeder, R.; Hillen, W. *Nucleic Acids Res.* **2003**, *31*, 1853-1858.

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

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

(18) Win, M. N.; Klein, J. S.; Smolke, C. D. *Nucleic Acids Res.* **2006**, *34*, 5670-5682.

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

(20) Page, J. E. *Trends Biotechnol.* **2005**, *23*, 331-333.

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

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

(23) Wallace, S. T.; Schroeder, R. *RNA* **1998**, *4*, 112-123.

(24) Plummer, K. A.; Carothers, J. M.; Yoshimura, M.; Szostak, J. W.; Verdine, G. L. *Nucleic Acids Res.* **2005**, *33*, 5602-5610.

(25) Soukup, G. A.; Breaker, R. R. *RNA* **1999**, *5*, 1308-1325.

(26) Knapp, G. *Methods Enzymol.* **1989**, *180*, 192-212.

(27) Klein, D. J.; Ferre-D'Amare, A. R. *Science* **2006**, *313*, 1752-1756.

(28) Muller, M.; Weigand, J. E.; Weichenrieder, O.; Suess, B. *Nucleic Acids Res.* **2006**, *34*, 2607-2617.

(29) Topp, S.; Gallivan, J. P. *Chembiochem* **2008**, *9*, 210-213.

(30) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001.

(31) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985**, *33*, 103-119.

(32) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. *Mol. Syst. Biol.* **2006**, *2*, 2006-2008.

(33) Polniaszek, R. P.; Kaufman, C. R. *J. Am. Chem. Soc.* **1989**, *111*, 4859-4863.

(34) Kametani, T.; Yagi, H.; Kaneda, S. *Chem. Pharm. Bull.* **1966**, *14*, 974-980.