Distribution Agreement

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

Signature:

_____ _____

Shana Topp Date

Biological Engineering with Chemical-Sensing Macromolecular Switches:
I. Discovery and Applications of Small-Molecule Dependent Synthetic Riboswitches
II. A Genetic Toolbox for Creating Reversible Ca^{2+}-Sensitive Biomaterials

By
Shana Topp
Doctor of Philosophy

Chemistry

_____
Justin P. Gallivan, Ph.D.
Advisor

_____
Vincent P. Conticello, Ph.D.
Committee Member

_____
David G. Lynn, Ph.D.
Committee Member

Accepted:

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

_____
Date

Biological Engineering with Chemical-Sensing Macromolecular Switches:
I. Discovery and Applications of Small-Molecule Dependent Synthetic Riboswitches
II. A Genetic Toolbox for Creating Reversible Ca^{2+}-Sensitive Biomaterials

By

Shana Topp
B.S. with Highest Honors, Emory University, 2004
M.S., Emory University, 2004

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
in Chemistry
2009

Abstract

Biological Engineering with Chemical-Sensing Macromolecular Switches:
I. Discovery and Applications of Small-Molecule Dependent Synthetic Riboswitches
II. A Genetic Toolbox for Creating Reversible Ca^{2+}-Sensitive Biomaterials

By Shana Topp

Nature has evolved the ability to precisely coordinate physiological and cellular processes in response to a variety of chemical signals. This dissertation draws inspiration from the exquisite chemical-sensing abilities of natural macromolecules toward reengineering chemical-sensing systems for applications in synthetic biology or nanotechnology.

Part I focuses on the development of efficient methods to select for synthetic riboswitches, and the use of these genetic control elements to modulate complex bacterial behavior. In Chapter 2, we demonstrate that synthetic riboswitches can be used to regulate *E. coli* chemotaxis with an exogenous ligand that wild-type cells neither recognize as a chemoattractant, nor naturally detect. The reprogrammed cells can be guided toward and precisely localized to a completely new chemical signal. Chapter 3 presents the development of a high-throughput selection to identify synthetic riboswitches by selecting for cells that exhibit ligand-dependent changes in migration on semi-solid media. We also discuss complications of this method and present potential solutions to surmount these limitations. Chapter 4 discusses studies toward overcoming our previously unproductive efforts to identify synthetic riboswitches that could repress bacterial gene expression when a small-molecule ligand is provided. These studies revealed a novel mechanism by which synthetic riboswitches may function in *E. coli* cells. In Chapter 5, we present principles to introduce synthetic riboswitches into a

diverse set of prokaryotes. For species lacking dynamic inducible promoter systems, the introduction of synthetic riboswitch technologies will facilitate previously intractable genetic and biochemical studies.

Part II focuses on our efforts to develop 'smart' materials that sense specific chemical signals in complex environments and respond with predictable changes in their mechanical properties. Toward this end, we developed a genetic toolbox of natural and engineered protein modules that can be rationally combined in many ways to create reversible self-assembling materials that vary in their composition, architecture, and mechanical properties. Using this toolbox, we produced and characterized several materials that reversibly self-assemble in the presence of calcium ions. The properties of these materials could be predicted from the dilute solution behavior of their component modules, suggesting that this toolbox may be generally useful for creating new stimuli-sensitive materials.

Biological Engineering with Chemical-Sensing Macromolecular Switches: I. Discovery and Applications of Small-Molecule Dependent Synthetic Riboswitches II. A Genetic Toolbox for Creating Reversible Ca^{2+}-Sensitive Biomaterials

By

Shana Topp
B.S. with Highest Honors, Emory University, 2004
M.S., Emory University, 2004

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
in Chemistry
2009

Acknowledgments

I would like to express my deepest gratitude to my advisor, Professor Justin Gallivan, for his invaluable guidance and encouragement. He gave me the freedom and confidence to approach problems independently, yet was always available for advice. There is little I could have hoped for in an advisor that Justin did not provide. I will be forever grateful for the indelible impact he has made on my development as a scientist.

I also would like to extend my sincere appreciation to my committee members, Professors David Lynn and Vince Conticello, whose helpful advice and support were critical to my success. I am also very grateful for the guidance of the many other professors and teachers who have cultivated my interest in science and encouraged me along the way. Among these, Professor Stefan Lutz deserves special thanks, as he has taught me so much, both in the classroom and the seminar room.

Much of the work presented in this thesis would not have been possible without the enthusiastic contributions of Professor Eric Weeks and his research group. As a result of this fruitful collaboration, I had the great pleasure to work with Vikram Prasad and Gianguido Cianci, each of whom dedicated many hours to helping me direct and interpret motion pictures starring bacteria or fluorescent particles. The Weeks Lab was a home away from home, even after our experiments were complete. I certainly will never forget Wallyball. I would also like to thank Ana West, as well as Professors James Kindt and Phil Segre, for their helpful discussions and interest in the Biomaterials Project.

Additionally, I would like to extend my gratitude to the many Departmental staff members whose important contributions make this department a better place. I would especially like to thank Ann Dasher, Donna Hudson, and the wonderful stockroom staff.

It has been my good fortune to work beside an incredible group of individuals in the Gallivan Lab. I have learned many things, both from and with these special people. Shawn Desai was my scientific hero when I joined the lab as an undergraduate; he was largely responsible for my decision to embark on graduate studies. The stresses of graduate school have been reduced significantly over the years through enjoyable and captivating conversations with Shawn, Steven Dublin, Karen Riesenburger, Sean Lynch, Colleen Reynoso, and Joy Sinha. The undergraduates and other postdocs who have passed through the lab have also shaped my graduate school experience in unique ways.

Graduate school would have been much less fun without the wonderful friends I have come to know and love. I won't list everyone here, for fear of leaving someone out, but you know who you are! I am very thankful for the support and encouragement of Brad Balthaser and Stella, my best friends. Their companionship and endless enthusiasm have made traversing the final year of graduate school much more bearable.

Last, but certainly not least, I want to thank my family for being there for me at every turn. My brother, Michael, often kept me company on the phone during those late-night trips to and from the lab. Those jokes and random stories meant a lot more to me than he probably realized. I'm pretty sure that my parents know how much I appreciate their constant support and interest, but that doesn't mean I shouldn't say it. Thank you Mom and Dad, for your never-ending devotion, wisdom, and encouragement. I dedicate this work to you.

Contents

List of Figures

List of Tables

Part I

Discovery and Applications of Small-Molecule Dependent Synthetic Riboswitches

CHAPTER 1 – Introduction

1.1 Chemical Sensing is a Fundamental Property of Life

A fundamental property of living systems is the ability to detect and respond to chemical signals in the internal and external environment. Nature has evolved countless systems in which chemical stimuli induce changes that are critical to cellular processes, such as the ability to maintain metabolic homeostasis. On a cellular level, the ability to detect and respond to deviations from ideal metabolite concentrations enables an organism to function at optimal levels, while avoiding energy expenditures associated with biochemical pathways that are not necessary at that time. How, then, do single-celled organisms such as *E. coli* achieve this feat?

In many cases, proteins have evolved that bind specific chemicals and either directly affect, or set off a signaling cascade to affect, the expression of genes related to the production of those molecules. In this way, proteins required for the synthesis of specific metabolites are not expressed when these or related metabolites are present in sufficient quantities. Such systems are critical for energy conservation, since limiting the production of unnecessary proteins saves energy that can be invested elsewhere. The *lac* operon[1] of *E. coli* is a canonical example of a metabolite-sensing, energy conservation system.[2]

The *lac* operator is used to repress or activate the expression of proteins involved in lactose metabolism and import. Specifically, the *lac* repressor binds the *lac* operator DNA sequence in the absence of lactose. This effectively prevents the transcription of genes that metabolize lactose (*lacZ*) or import lactose into the cell (*lacY*). However, when lactose concentrations build, allolactose is formed and binds the lac repressor

protein, inducing a conformational change that prevents the repressor from binding at the lac operator sequence. Transcription is then permitted, and enzymes are produced to enable the conversion of lactose into glucose, as well as to import lactose from the extracellular environment. Through a similar but opposing mechanism, the presence of glucose down-regulates transcription of the lac operon, because glucose is the preferred energy source even in the presence of lactose.

The *trp* operon[3] is another example of a metabolite-sensing, energy conservation system in *E. coli* cells. Like the *lac* operon, the *trp* operon system enables the intracellular concentration of a metabolite (tryptophan) to regulate the expression of enzymes involved in its biosynthesis. However, in addition to relying upon a tryptophan-binding protein to repress transcription initiation by binding the *trp* operator sequence in the presence of the metabolite, the *trp* operon also features a *cis*-acting regulatory component. This *cis*-acting element utilizes a 14-amino acid leader peptide, which includes two tryptophan residues, to regulate the expression of tryptophan biosynthesis genes. If tryptophan is at sufficient concentrations within the cell, an abundance of tryptophan-loaded tRNA molecules permits translation of the leader peptide, which attenuates gene expression through the formation of a transcriptional terminator sequence of the mRNA.[4] However, when tryptophan concentrations are very low, tryptophan-loaded tRNA levels are also low, causing ribosomal stalling within this leader sequence. Upon ribosomal stalling, a transcriptional anti-terminator forms, enabling the mRNA encoding the tryptophan biosynthesis genes to be transcribed. Thus, repression and attenuation mechanisms work in concert to provide an incredible level of control over tryptophan levels within a cell.

1.2 Metabolite-Sensing can be Achieved Using RNA

While elucidation of the *trp* attenuation mechanism demonstrated that the structure of an mRNA molecule may influence gene expression, it was not yet clear that RNA can also bind metabolites directly to regulate gene expression. However, over the past decade, it has become increasingly apparent that small-molecule binding RNA molecules serve to regulate a variety of biochemical pathways across all domains of life.[5] Particularly abundant in Eubacteria, these metabolite-sensing RNA switches (riboswitches) are typically found in the 5′ untranslated regions (UTR) of metabolic genes, where they activate or repress biochemical pathways involving their cognate ligands. Notably, these riboswitches perform their critical tasks without the need for protein cofactors.[6] In fact, the first riboswitches were hypothesized to exist, in part, because it had proven impossible to identify protein cofactors involved in several metabolite-regulated biochemical pathways.[7,8] For example, the discovery of an FMN riboswitch in *B. subtilis* in 2002 ended a decade-long search for a putative repressor protein that might use FMN as a cofactor.[9]

At least a dozen riboswitch classes have been discovered;[5] they respond to adenosylcobalamin (AdoCbl), thiamin pyrophosphate (TPP), lysine, glycine, flavin mononucleotide (FMN), guanine, adenine, glucosamine-6-phosphate (GlcN6P), 7-aminoethyl 7-deazaguanine (preQ1), and S-adenosylmethionine (SAM), cyclic di-GMP,[10] and S-adenosylhomocysteine (SAH).[11] Each of these classes has been found, or is predicted to exist, within the genomes of many species of bacteria, leading some to postulate that riboswitches may be remnants of ancient metabolic ribozymes.[12] Others may suggest horizontal gene transfer as a more likely explanation.[13] In either case, it is

inarguable that modern-day riboswitches are highly conserved genetic control elements that function to maintain metabolic homeostasis within cells.

Riboswitches are defined by several important features, such as the ability to bind a small molecule, regulate gene expression, and function as *cis*-acting elements. Another characteristic feature of riboswitches is that they are composed of an aptamer domain and an expression platform. The aptamer domain binds the metabolite with high affinity and specificity, while the expression platform converts the binding event into a structural rearrangement that results in a change in gene expression.[14] Riboswitches operate by a variety of mechanisms to regulate the expression of a gene or set of genes in a metabolic pathway.

Riboswitch-mediated changes in gene expression can be accomplished on a transcriptional level or on a translational level. The expression platform for a riboswitch that acts at the transcriptional level involves the ligand-dependent formation of a transcriptional terminator or anti-terminator structure. Metabolite binding can also modulate gene expression through RNA splicing.[15] Finally, riboswitches can function on a translational level by providing ligand-dependent access to the bacterial Shine-Dalgarno sequence. Figure 1.1 depicts the various mechanisms of the TPP riboswitch in different organisms.[14] It should be noted that transcriptional and translational mechanisms are observed in both Gram-negative and Gram-positive bacteria. Additionally, several Gram-positive species feature GlcN6P-binding ribozymes.[15] While cumulative data suggest that the mechanisms shown in Figure 1.1 are consistent with the expression platforms that predominate for each type of organism,[5] it is not entirely clear why Gram-negative and Gram-positive bacteria appear to favor different riboswitch mechanisms.

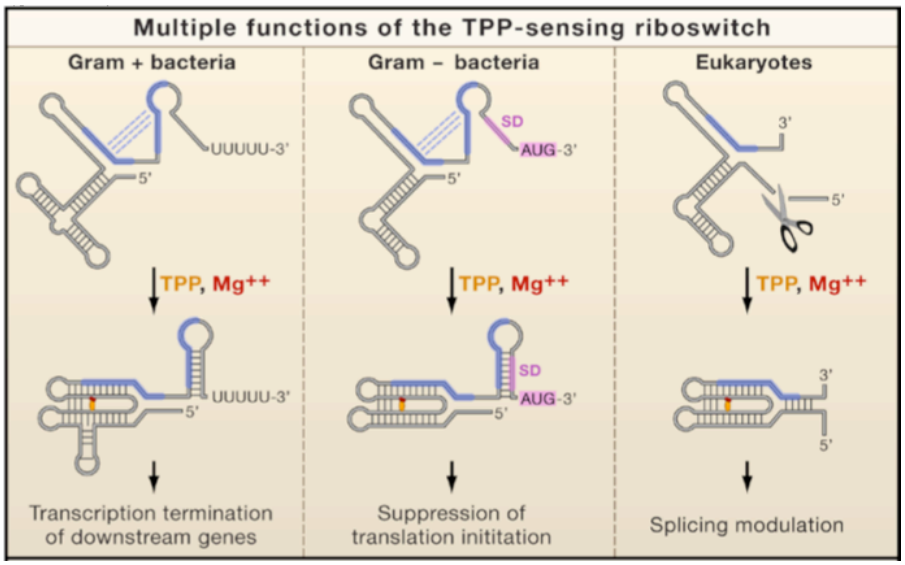

Figure 1.1: Mechanisms of Natural Riboswitches.[1] The TPP riboswitch modulates gene expression by different mechanisms in various organisms. In Gram-positive bacteria, TPP-binding causes a conformational shift that induces the formation of a transcriptional terminator, preventing expression of the downstream genes. In Gram-negative bacteria, the riboswitch functions by a translational mechanism, in which the Shine-Dalgarno sequence is unpaired in the absence of TPP and paired upon TPP-binding. The eukaryotic TPP riboswitch functions by a splicing mechanism in which ligand-binding prevents splicing of the nascent transcript.

1.3 Synthetic Riboswitches

Even before the discovery of natural riboswitches, there was broad interest in using RNA sequences to sense small molecules or proteins. In 1990, Ellington and Szostak developed an in vitro selection protocol to isolate RNA sequences that could bind specific dyes from large randomized pools.[16] They termed these ligand-binding RNA sequences "aptamers," which was derived from the Latin word "aptus," to fit. Independently, Tuerk and Gold reported a method, which they termed SELEX (**s**ystematic **e**volution of **l**igands by **ex**ponential enrichment), to isolate protein-binding RNA sequences from large pools of randomized RNA sequences.[17]

---

[1] Reprinted from **Cell**, *126*, Nudler, E., Flipping Riboswitches, 19-22, Copyright 2006, with permission from Elsevier.

Both groups recognized that RNA aptamers could be isolated starting from pools of ~10^{15} RNA variants by developing an in vitro selection procedure that eliminates sequences that do not bind the ligand of interest and enriches for high-affinity binders (Figure 1.2).[18] The first step toward isolating an RNA aptamer is to synthesize a ssDNA oligonucleotide with a randomized region of 20-60 nucleotides, which is flanked by two constant regions that serve as primer-binding sites for PCR amplification of the pool. This ssDNA pool is amplified using a forward primer that introduces a T7 promoter sequence, and a reverse primer that anneals to the 3′ primer-binding site.

Then, the RNA pool is produced by in vitro transcription using T7 RNA polymerase. The resulting pool is passed through a column to which the ligand of interest is covalently attached. Extensive washing eliminates non-specific binders, while RNA sequences that bind specifically are eluted with buffer containing free ligand. This enriched pool is then reverse transcribed to obtain a cDNA pool, which is amplified by PCR to obtain dsDNA. This cycle is then repeated 10-20 times, until high affinity binders predominate.

To identify an aptamer that discriminates between closely related molecules, it is possible to introduce a counter-selection step in which RNA bound to the column is washed with buffer containing a structurally related molecule. This counter-selection step eliminates sequences that fail to discriminate between the desired ligand and the structurally related molecule. After removing these sequences, the column is washed with the desired ligand to obtain an enriched pool of sequences that bind the ligand of interest with high specificity. In vitro selections are then continued to further enrich the pool for high-affinity binders.

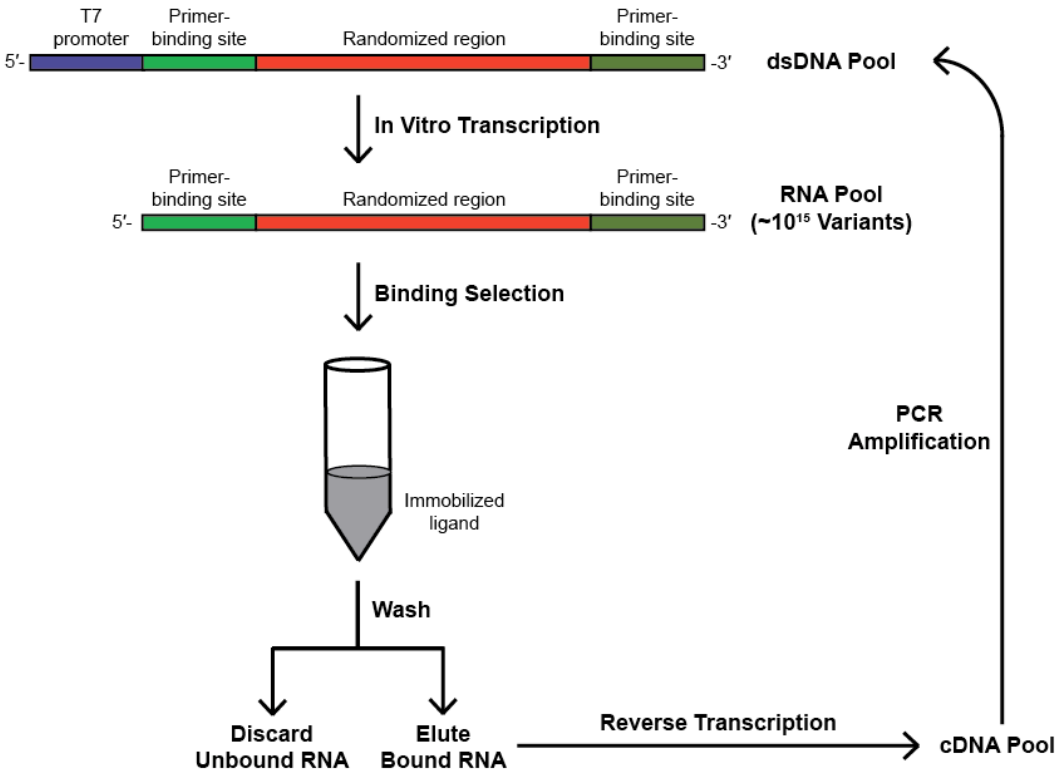

Figure 1.2: Aptamer Selection Scheme. RNA aptamers may be isolated from large pools of randomized sequences using in vitro selection. The pool is enriched for sequences that bind an immobilized ligand and are eluted with free ligand. To obtain high-affinity binders, the selected sequences are reverse transcribed, amplified, and subjected to further rounds of selection.

As the tremendous potential of RNA aptamers became apparent, further studies were initiated to select for sequences that could tightly bind new ligands with high specificity. Perhaps the best-known example is the theophylline-binding aptamer, which was isolated by SELEX in 1994.[19] This aptamer was selected to bind theophylline tightly (K_D = 320 nM) and with high specificity. The aptamer binds caffeine 10,000-fold less tightly, even though the two compounds are nearly identical in structure. This specificity was 10-times better than the performance of available antibodies, while the binding-affinity was 100-times better than had been achieved with any previously identified small-molecule binding aptamer.[19]

Even while the existence of natural riboswitches remained speculative, several groups anticipated that these aptamers could be used to regulate gene expression in various organisms. It was known that highly structured regions in the 5′ UTR of genes could cause marked reductions in gene expression in eukaryotic and prokaryotic cells.[20-23] Therefore, it was postulated that by positioning a small-molecule binding aptamer in the 5′ UTR of a gene, ligand-inducible gene expression could be accomplished in vivo. These early examples demonstrated that small-molecule binding aptamers could be incorporated within the 5′ UTR of genes in a variety of organisms to generate synthetic riboswitches that could respond to theophylline,[24-26] antibiotics,[27] or dyes.[28] Since these initial observations, our lab[29] and others[30-32] have been keenly interested in developing high-throughput screens or selections to isolate synthetic riboswitches that respond to a variety of new ligands. These synthetic riboswitches could then be used as tools in diverse applications, such as efforts to modulate gene expression in a ligand-dependent fashion, to create genetic circuits, or to reprogram complex cellular behavior.[33]

Part I of this thesis will focus on the development of high-throughput screens and selections for synthetic riboswitches that can activate or repress gene expression in *E. coli* cells or in other bacteria, as well as the use of these ligand-inducible genetic control elements to reprogram cellular behavior. These studies will employ the oft-used theophylline aptamer as proof-as-principle. Since aptamers can be selected de novo to bind essentially any ligand that is capable of interacting with RNA, we anticipate that the future holds great promise for the development of useful synthetic riboswitches that respond to a diverse array of ligands and can function by various mechanisms in a wide range of organisms.

1.4 References

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

(2) Stoebel, D.M.; Dean, A.M.; Dykhuizen, D.E. *Genetics* **2008,** *178*, 1653-1660.

(3) Monod, J.; Cohen-Bazire, G. *C. R. Hebd. Seances. Acad. Sci.* **1953,** *236*, 530-532.

(4) Lee, F.; Yanofsky, C. *Proc. Natl. Acad. Sci.* **1977,** *74*, 4365-4369.

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

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

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

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

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

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

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

(12) Soukup, J.K.; Soukup, G.A. *Curr. Opin. Struct. Biol.* **2004,** *14*, 344-349.

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

(14) Nudler, E. *Cell* **2006,** *126*, 19-22.

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

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

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

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

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

(20) Katz, L.; Burge, C.B. *Genome Res.* **2003,** *13*, 2042-2051.

(21) Meijer, H.A.; Thomas, A.A. *Biochem. J.* **2002,** *367*, 1-11.

(22) van der Velden, A.W.; Destree, O.H.J.; Voorma, H.O.; Thomas, A.A. *Int. J. Dev. Biol.* **2000,** *44*, 843-850.

(23) Xia, X.; Holcik, M. *PLoS ONE* **2009,** *4*, e4136.

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

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

(26) Bayer, T.S.; Smolke, C.D. *Nat. Biotech.* **2005,** *23*, 337-343.

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

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

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

(30) Famulok, M. *Curr. Opin. Mol. Ther.* **2005,** *7*, 137-143.

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

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

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

CHAPTER 2 – Guiding *E. coli* with Small Molecules and RNA

2.1 Introduction

Chemotactic bacteria navigate complex chemical environments by coupling sophisticated information processing capabilities to powerful molecular motors that propel the cells forward.[1-3] A general method to reprogram the ligand sensitivity of the bacterial chemo-navigation system would enable the production of cells that autonomously follow arbitrary chemical signals, such as pollutants or disease markers. Equipping bacteria that can degrade pollutants,[4] synthesize and release therapeutics,[5] or transport loads[6] with the additional ability to localize to a specific chemical signal would open new frontiers in bioremediation, drug delivery, and synthetic biology. For example, such cells could be engineered to follow specific pollutants in soil and to degrade them. Alternatively, such cells could be engineered to selectively target small-molecule signals of disease, thus providing a powerful drug-delivery system.[7] However, the complexity of the bacterial chemosensory system makes reprogramming a cell to follow a completely new chemical signal a formidable challenge.

E. coli recognize chemoattractants using five transmembrane receptor proteins, which cluster with one another[8-11] and interact with a set of well characterized cytosolic proteins to effect changes in the directional rotation of the flagellar motor.[3,12] In the absence of a chemical gradient, individual *E. coli* execute a random walk characterized by smooth runs punctuated by tumbles that often result in a change of direction.[1-3] When a cell moves up a chemoattractant gradient, the flagellar motor preferentially rotates

---

Adapted with permission from: S. Topp and J.P. Gallivan. *J. Am. Chem. Soc.*, **2007**, *129*, 6807-6811.

counterclockwise, resulting in less frequent tumbling and longer runs toward the attractant; when the chemoattractant concentration becomes constant, the cell resumes a random walk.[2] Similarly, when a cell moves down a chemoattractant gradient, the tumbling frequency increases. A defining feature of *E. coli* chemotaxis is that individual cells do not simply respond to the absolute concentration of a stimulus at a given time. Rather, individual *E. coli* cells integrate the concentration of a chemical stimulus over a 1-3 s period[2] and migrate up attractant gradients by adjusting the frequency of tumbling based on the concentration differences between these time points.

In principle, bacteria can be programmed to respond to a new chemical signal by engineering an existing chemoreceptor protein to recognize a new ligand. Although *E. coli* have only 5 chemoreceptor proteins, they perform chemotaxis toward greater than 30 compounds,[1] indicating that some chemoreceptors recognize multiple compounds. A recent effort to engineer the *tar* (aspartate) receptor[13] produced relatively modest changes in ligand specificity, consistent with the tendency of engineered proteins to display broadened rather than shifted ligand specificities.[14,15] While it may be possible to produce more dramatic changes in receptor specificity using rational design, directed evolution,[13] or computational methods,[16] such efforts are ultimately limited by structural constraints enforced by the receptor scaffolds and the need to interface with the existing signaling network.[17] Thus, engineering cells to follow a new stimulus presents a difficult molecular recognition problem.

Faced with these challenges, we sought to bypass the chemoreceptors entirely by developing an effective, generic approach to reprogram *E. coli* to detect, follow, and localize to a new chemical signal. Over 40 years of genetics and biochemistry research

has provided remarkable insight into the mechanisms of *E. coli* chemotaxis. The signal transduction pathway that converts ligand binding by the chemoreceptors into a change in the rotational direction of the flagellar motor is comprised of six chemotaxis signaling proteins (Che A, B, R, W, Y, and Z).[18] The chemoreceptors interact with CheB, CheR, and CheW, which collectively control the autophosphorylation rate of CheA, which in turn, phosphorylates the protein CheY. CheY controls the rotational direction of the flagellar motor. When CheY is phosphorylated (CheY-P), it binds to the flagellar switch protein FliM and induces the flagellum to rotate clockwise (CW), which causes the cell to tumble. Swimming is restored by the phosphatase CheZ, which dephosphorylates CheY-P and causes the flagellum to rotate CCW (Figure 2.1, *top*). *E. coli* lacking the *cheZ* gene (Δ*cheZ,* strain RP1616) cannot dephosphorylate CheY-P, tumble incessantly, and are thus non-motile (Figure 2.1, *bottom*).

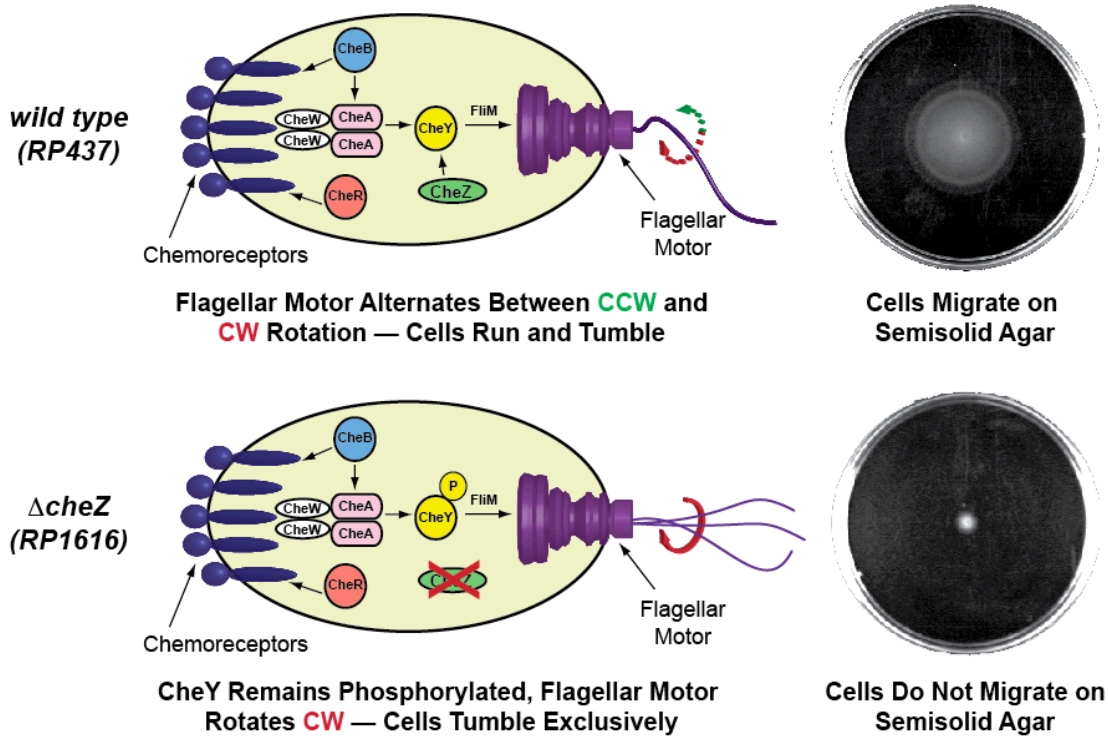

Figure 2.1: The CheZ Protein is Critical for Chemotaxis. Wild-type cells express CheZ and migrate on semisolid agar. Cells lacking CheZ tumble in place and are nonmotile. (Adapted from Ref. 18).

Previous studies have shown that inducing the expression of CheZ can restore motility in a CheZ knockout strain by reactivating wild-type chemotaxis.[19-21] We anticipated that using a ligand-inducible expression system to control the production of CheZ would enable us to guide cells toward higher concentrations of a new ligand in a process known as pseudotaxis.[22] This system would differ from classical *E. coli* chemotaxis in that cell motility toward the new ligand would be dictated by the absolute ligand concentration at a specific time, rather than the concentration differences between two points in time.

Key to these efforts is the development of a tunable ligand-inducible expression system that can precisely control the production of the CheZ protein. While a variety of bacterial expression systems are ligand-inducible, many of these, such as the *lac* repressor and the araC transcriptional regulator, use proteins to recognize ligands. Consequently, engineering these protein-based expression systems to respond to new ligands presents many of the same challenges that make engineering the chemoreceptor proteins difficult. In contrast to inducible expression systems that use proteins to recognize ligands, riboswitches[23,24] control gene expression in a ligand-dependent fashion by using RNA aptamers to recognize ligands. Using powerful in vitro selection techniques,[25-27] it is possible to generate aptamers that tightly and specifically recognize new ligands without the need for a pre-existing RNA scaffold. We,[28,29] and others,[30-32] have shown that a variety of aptamers generated by in vitro selection can be engineered into synthetic riboswitches that regulate gene expression in a ligand-dependent fashion. Here we show that a synthetic riboswitch can guide *E. coli* toward a new, non-metabolized ligand without protein engineering. These reprogrammed cells migrate

preferentially up a ligand gradient and have the unique ability to localize to a specific chemical signal, which enables precise spatial patterning. We anticipate that the ability to engineer cells to follow new chemical signals will provide new opportunities in bioremediation and synthetic biology.

2.2 Results and Discussion

2.2.1 Regulating CheZ Expression with a Theophylline-Dependent Riboswitch

Since *E. coli* lacking a single gene in the signaling pathway (strain RP1616, Δ*cheZ*) tumble incessantly and are essentially non-motile,[21] we asked whether a synthetic riboswitch could restore ligand-dependent motility to *E. coli* RP1616 by activating the translation of CheZ in response to the alkaloid theophylline (Figure 2.2). Reprogramming a cell to follow theophylline is a challenging test: It is neither chemoattractive, nor extensively metabolized, and because it is structurally dissimilar to natural chemoreceptor ligands, which are predominantly amino acids, dipeptides, and sugars, it would be difficult to reengineer a chemoreceptor to selectively bind it.

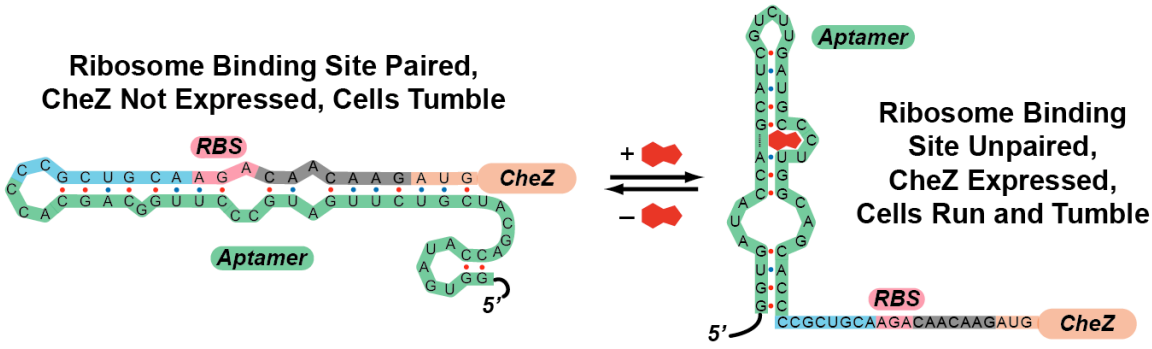

Figure 2.2: Mechanism of a Riboswitch for Theophylline-Dependent Chemotaxis. Model for how the theophylline-sensitive synthetic riboswitch controls the translation of the CheZ protein. In the absence of theophylline (left), the mRNA adopts a conformation in which the ribosome binding site is paired and translation of CheZ is inhibited. In the absence of CheZ, the protein CheY-P remains phosphorylated and the cells tumble in place. In the presence of theophylline (shown in red), the mRNA can adopt a conformation in which the ribosome binding site is exposed and CheZ is expressed, thus allowing the cells to run and tumble.

2.2.2 Ligand-Dependent Migration of *E. coli* Populations on Semi-Solid Agar

To investigate ligand-inducible motility, we introduced the *cheZ* gene under the control of a theophylline-sensitive synthetic riboswitch into *E. coli* RP1616 cells (hereafter referred to as "reprogrammed cells"), plated the reprogrammed cells onto semi-solid media containing various ligand concentrations, and measured their migration radius after 10 hours (Figure 2.3A). The distance that the cells migrate increases as a function of ligand concentration until reaching a maximum at 2 mM theophylline, after which further concentration increases lead to cell death (Figure 2.3B). When theophylline is replaced with caffeine, which is structurally similar (Figure 2.3C), but does not bind to the riboswitch[29] the reprogrammed cells are non-motile, demonstrating that motility changes are riboswitch-dependent (Figure 2.3B).

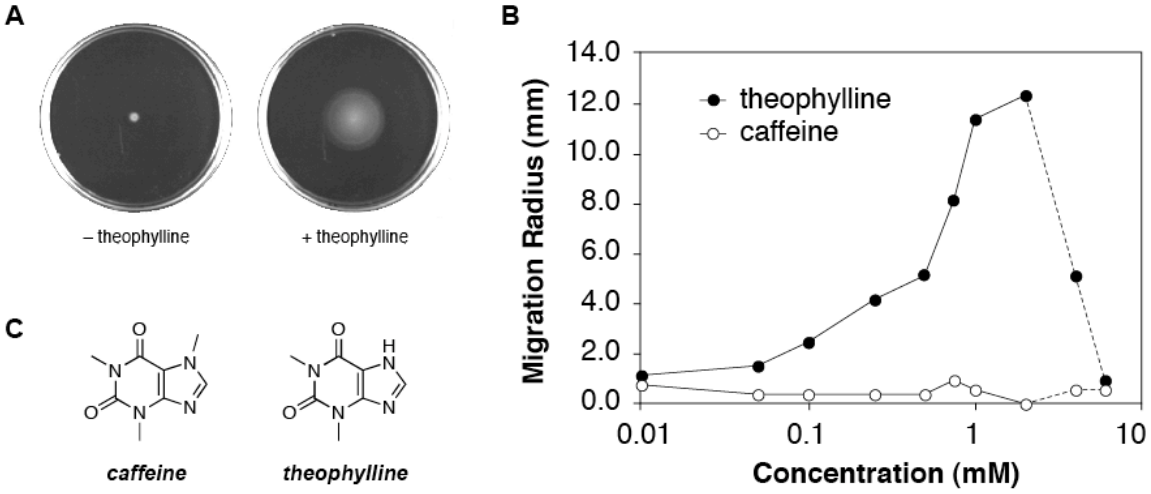

Figure 2.3: The Migration Distance of Reprogrammed Cells Increases in a Dose-Dependent Manner. (A) Migration of reprogrammed cells on semi-solid media in the absence or presence of theophylline (2 mM). (B) Migration radius of reprogrammed cells as a function of ligand concentration. All cells were grown on semi-solid media for 10 h at 37 °C; uncertainties in measurement are smaller than the symbols. (C) Caffeine and theophylline are structurally similar, with the exception that caffeine is methylated at the N7 position. The aptamer binds theophylline 10,000 times more tightly than caffeine.

2.2.3 Testing the Gradient-Sensing Abilities of Reprogrammed Cells

The increased migration of the reprogrammed cells at higher theophylline concentrations suggested that these cells might preferentially migrate up a theophylline gradient. To test this idea, we plated the reprogrammed cells at three locations on semi-solid media containing a theophylline gradient (Figure 2.4). Cells plated near the edges, where the local theophylline concentrations are relatively constant, do not bias their movement in a particular direction. In contrast, cells plated near the center of the plate experience larger changes in theophylline concentration over a short distance and bias their migration toward increasing theophylline concentrations. This behavior is consistent with the data in Figure 2.3B, in which significant changes in motility are observed in this concentration range.

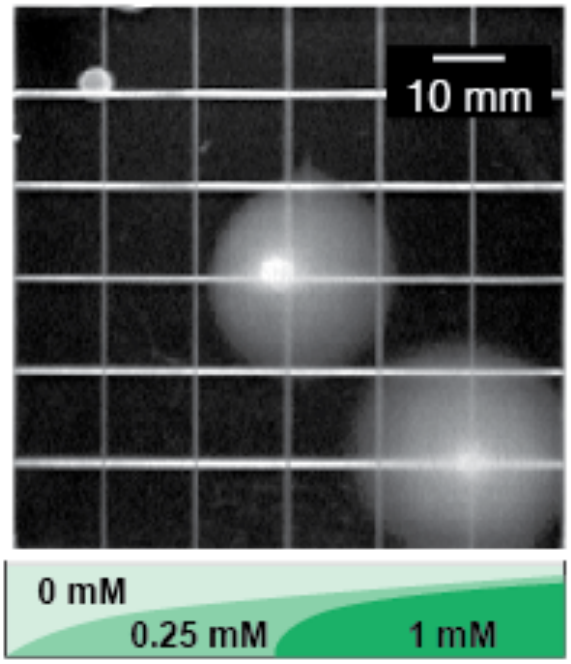

Figure 2.4: Reprogrammed Cells Migrate up a Theophylline Gradient. Reprogrammed cells were plated at three locations on semi-solid media containing a theophylline gradient. The layering scheme used to prepare the gradient and the theophylline concentrations are shown. Cells were grown for 10 h at 37 °C.

Taken together, the data show that populations of reprogrammed cells recapitulate many features of chemotactic *E. coli*, including the ability to migrate up a ligand gradient. However, unlike wild-type *E. coli*, which rarely stop moving, populations of these engineered cells do not migrate in the absence of theophylline. To investigate the mechanisms that influence cell migration and to explain the phenotypic similarities and differences between the reprogrammed cells and wild-type *E. coli*, we assessed the ability of these cells to perform chemotaxis toward a natural chemoattractant in the presence of theophylline, as well as their ability to respond to various static concentrations of theophylline in the absence of chemoattractant gradients.

To determine if theophylline restores chemotaxis toward natural chemoattractants, we plated reprogrammed and wild-type cells onto semi-solid minimal media plates that presented a gradient of the chemoattractant L-aspartate and a static theophylline concentration (0 or 2 mM). As expected, reprogrammed cells do not perform chemotaxis toward aspartate in the absence of theophylline (Figure 2.5). When theophylline is present, however, the reprogrammed cells perform chemotaxis toward aspartate in a manner similar to wild-type cells.

Taken with the data in Figure 2.3B, these results indicate that the reprogrammed cells exhibit stronger chemotaxis toward natural chemoattractants at higher static theophylline concentrations, suggesting that the theophylline concentration may serve to establish the steady-state tumbling frequency of these cells. Thus, chemo-navigation is achieved, in part, because the riboswitch acts as a molecular brake. In the absence of theophylline, the brake is engaged and the reprogrammed cells tumble in place. The addition of theophylline releases the brake in a dose-dependent fashion by inducing the

expression of CheZ, which allows the cells to navigate gradients of natural chemoattractants.

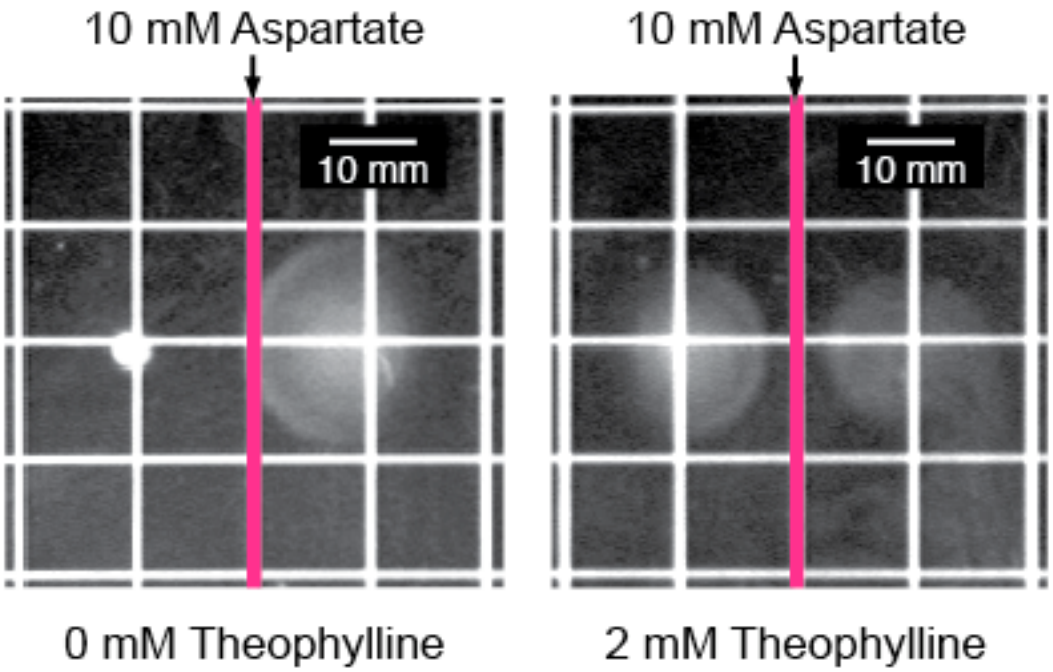

Figure 2.5: Reprogrammed Cells Exhibit Theopylline-Dependent Migration Toward Aspartate. Migration of reprogrammed (plated left of center) and wild-type cells (plated right of center) on minimal media containing a gradient of aspartate and a static concentration of theophylline (0 mM or 2 mM). Aspartate was applied to the surface of the plate along the pink line; theophylline was added to the media. Reprogrammed cells only perform chemotaxis when theophylline is present.

2.2.4 Quantifying the Behavior of Reprogrammed Cells using Light Microscopy

To investigate additional mechanisms influencing taxis, we used video microscopy to observe and track the motility of individual reprogrammed cells in rich liquid media containing various static theophylline concentrations.[33,34] Since rich liquid media largely eliminates chemoattractant gradients, the effects of different theophylline concentrations can be studied independently of the gradient-induced chemotaxis that occurs on semi-solid media. The tracking data reveal that a prime determinant of the

population behavior is a sharp concentration-dependent rise in the fraction of motile cells (Figure 2.6A).

At low theophylline concentrations, most cells tumble in place, consistent with the behavior of cells lacking CheZ. At higher concentrations, a greater fraction of the population is motile, consistent with theophylline-induced expression of CheZ and a transition from a tumbling to a running phenotype. These observations help explain both the theophylline-dependent increases in motility on semi-solid media (Figure 2.3B) and the gradient-sensing behavior seen in Figure 2.4. The dose-dependent increases in migration distance on semi-solid agar are related to the increases in the fraction of motile cells determined by tracking the paths of individual cells (Figure 2.6B).

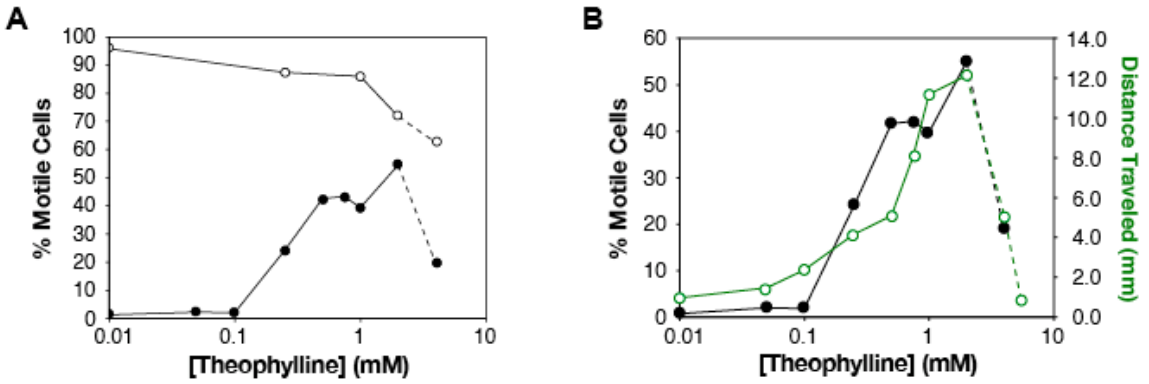

Figure 2.6: The Fraction of Motile Cells Increases in a Ligand-Dependent Manner. (A) The fraction of the population of cells that is motile for reprogrammed (solid circles) and wild-type cells (open circles) is shown as a function of the theophylline concentration. (B) When plotted on the same chart, the fraction of motile cells (left axis, black) is related to the distance traveled by the population of cells on semi-solid agar (right axis, green). For both graphs, uncertainties in measurement are smaller than the symbols.

To test whether motile cells exhibit pseudotaxis toward theophylline, we determined the average maximum run speeds (Figure 2.7A) and tumbling frequencies (Figure 2.7B) as functions of theophylline concentration for the motile populations of the

wild-type and the reprogrammed cells While motile cells from both strains showed modest theophylline-dependent changes in maximum run speed (Figure 2.7A), only the reprogrammed cells showed dramatic changes in tumbling frequency (Figure 2.7B). Whereas wild-type *E. coli* maintain a steady-state tumbling frequency of approximately 0.5 s^{-1} regardless of the absolute concentration of a chemoattractant,[19] reprogrammed cells adopt lower steady-state tumbling frequencies at higher theophylline concentrations, which allows cells to migrate up theophylline gradients via pseudotaxis.

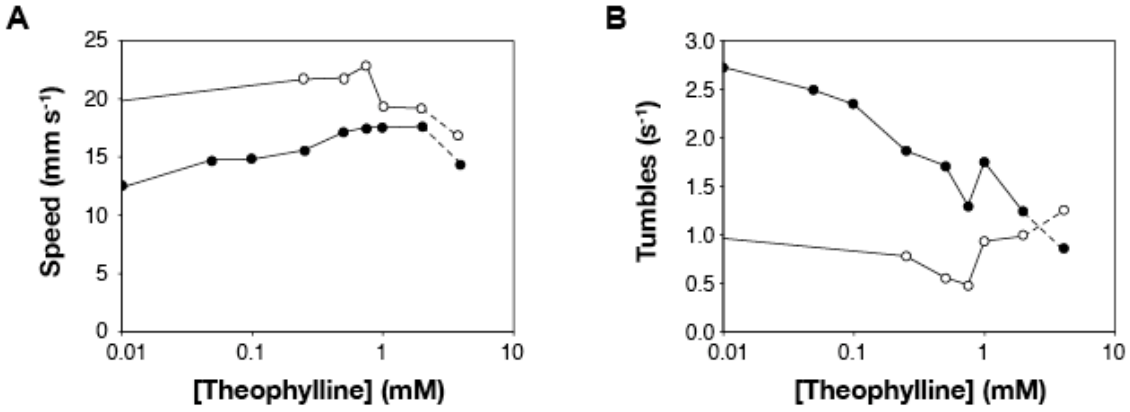

Figure 2.7: Theophylline Affects the Run Speed and Tumble Frequency of Reprogrammed *E. coli*. (A) Average run speeds of the motile fraction of the reprogrammed cells (solid circles) and wild-type cells (open circles). (B) Average tumble frequencies of the motile fraction of the reprogrammed cells (solid circles) and wild-type cells (open circles). Dashes indicate theophylline toxicity. Uncertainties in measurement are smaller than the symbols.

2.2.5 Patterning of Reprogrammed Cells on Semi-Solid Media

Since reprogrammed cells migrate farther at higher ligand concentrations and essentially stop migrating at low concentrations, we anticipated that these cells might also navigate paths containing the ligand. To test this idea, we plated cells onto semi-solid media patterned with the various compounds shown in Figure 2.8. Wild-type cells migrate radially without regard for the patterned compounds, while RP1616 cells lacking the riboswitch are essentially non-motile. In contrast, the reprogrammed cells

exclusively follow the theophylline-containing path and ignore the other paths. This behavior is possible because, unlike wild-type *E. coli*, reprogrammed cells become less motile as they move down a concentration gradient and, in the limit, are non-motile. Thus, these cells have the unique ability to localize to a specific target, which may be particularly useful for targeting cells to a specific disease site for applications in biomedicine. In the future, cells may be reprogrammed with a pair of riboswitches, such that one switch could enable cells to migrate up gradients of a target molecule, while the other could prohibit motility once very high concentrations of the target molecule are detected. In this way, cells could be directed up a ligand gradient to a site of interest, where they could be made to accumulate and perform a desired task.

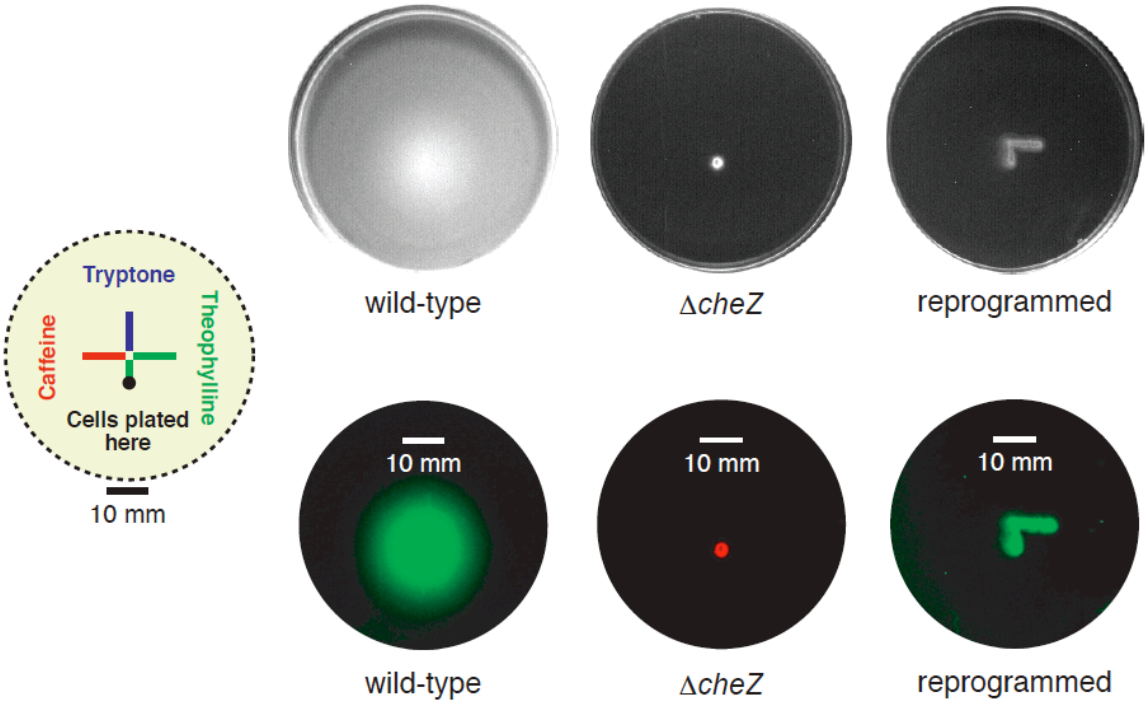

Figure 2.8: Reprogrammed Cells Migrate Along a Theophylline Path. Semi-solid media was patterned with solutions of various ligands (left). Cells were plated at the location shown and grown for 10 h at 37 °C. The wild-type cells harbor no switch, but express GFP. RP1616 (Δ*cheZ*) cells also harbor no switch, but express a red fluorescent protein. The reprogrammed cells harbor the switch and express GFP. The plates on top are shown as imaged, and the plates below have been enhanced with false color.

2.2.6 Changing the Ligand Specificity to 3-Methylxanthine

The studies described above serve as proof-of-principle that a synthetic riboswitch can be used to reprogram *E. coli* cells to detect and migrate toward new ligands that are not recognized by wild-type cells. To test whether reprogrammed cells can be reengineered to respond to another ligand, we introduced a previously reported point mutation that reduces the aptamer's affinity for theophylline and increases its affinity for the related compound 3-methylxanthine.[35] Specifically, nucleotide C27 of the mTCT8-4 aptamer interacts directly with the ligand in its binding pocket; mutation of this nucleotide (C27A) causes the aptamer to change binding specificity from theophylline to 3-methylxanthine.[*]

To test the affect of this mutation on cell motility, we measured the migration distances of reprogrammed cells harboring the mTCT8-4 aptamer or the C27A version of the aptamer at several ligand concentrations (Figure 2.9). The data match the prediction that cells harboring the mutated aptamer would no longer migrate in the presence of theophylline, but would migrate in the presence of 3-methylxanthine. However, in the absence of ligand, the migration distance of reprogrammed cells featuring the C27A version of the riboswitch is greater than that of reprogrammed cells harboring the parent riboswitch. We suspect that this difference may result from an increase in "leakiness" in the absence of ligand for the C27A version of the switch, permitting higher background expression levels of the CheZ protein. This hypothesis is supported by secondary-structure predictions of the ligand-free state of the switch, in which the RBS is less paired in the absence of theophylline for the C27A mutant riboswitch compared to the original

[*] It should be noted that nucleotide C27 of the mTCT8-4 aptamer corresponds to position C22 of the aptamer studied by Breaker et al.[35]

riboswitch. We anticipate that the C27A version could be made less leaky by engineering or selecting for riboswitches with increased pairing of the RBS sequence in the absence of theophylline. Ultimately, these results suggest that it will be relatively straightforward to engineer cells to migrate toward a variety of new ligands by applying the principles described here to reprogram *E. coli* cells with small molecules and RNA.

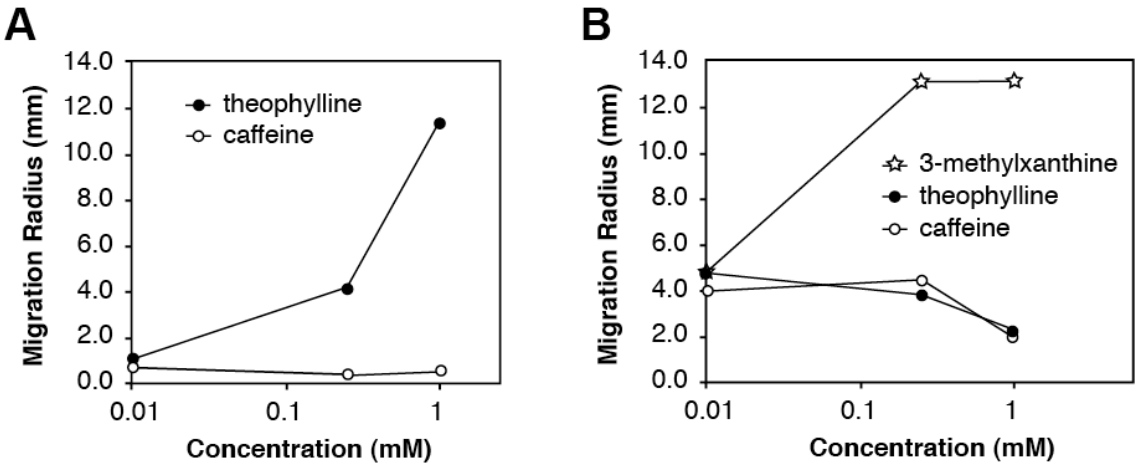

Figure 2.9: The Ligand-Specificity of Reprogrammed Cells can be Changed. (A) Reprogrammed cells expressing CheZ under the control of a riboswitch with the mTCT8-4 aptamer migrate on semi-solid agar in the presence of theophylline (closed circles) but not in the presence of caffeine (open circles). B) Cells with the C27A version of the riboswitch do not migrate far when grown in the presence of theophylline or caffeine, but become very motile when grown on 3-methylxanthine (stars). Uncertainties in measurement are smaller than the symbols.

2.3 Conclusion

The reprogrammed cells described here capture many of the features displayed by naturally chemotactic bacteria such as gradient sensing, but do so through different mechanisms. Unlike natural chemotaxis in which cells detect gradients by integrating the concentration differences between chemoattractants at two points in time, reprogrammed cells respond to theophylline gradients via pseudotaxis, in which the motility of a cell changes as a function of the local theophylline concentration. Because motility changes are mediated through protein synthesis induced by the synthetic riboswitch, these cells

respond to changes in the concentration of new ligands more slowly than wild-type cells, which respond to chemoattractants over a period of seconds by adjusting the methylation state and the degree of clustering of the chemoreceptor proteins. While the response rate of our reprogrammed cells to new ligands is slower than the natural chemotaxis response of *E. coli*, these cells show unique behaviors that are not displayed by wild-type bacteria. As an example, reprogrammed cells become less motile as the ligand concentration decreases, which allows these cells to not only detect gradients of a new ligand (Figure 2.4), but also to precisely target a patterned ligand (Figure 2.8). Finally, because the displacement of a reprogrammed cell is proportional to the ligand concentration, and cells that migrate the farthest are easily identified in large populations, we anticipate that motility-based screens using synthetic riboswitches may provide a powerful and inexpensive technique to detect the production of small molecules through biocatalysis. Studies toward this end will be the focus of Chapter 3.

2.4 Experimental

General Considerations

Synthetic oligonucleotides were purchased from Integrated DNA Technologies. Culture media was obtained from EMD Bioscience. Amino acids were purchased from Sigma. Ampicillin was purchased from Fisher. DNA polymerase, restriction enzymes, and the pUC18 cloning vector were purchased from New England Biolabs. The pGFPuv and pDsRed-Express vectors were purchased from Clontech. Purifications of plasmid DNA, PCR products, and enzymatic digestions were performed using kits from Qiagen. Plasmid manipulations were performed using *E. coli* TOP10F′ cells (Invitrogen) that

were transformed by electroporation. All new constructs were verified by DNA sequencing at the NSF-supported Center for Fundamental and Applied Molecular Evolution at Emory University. Experiments were performed with *E. coli* strains RP437 and RP1616, both of which were kindly provided by J. S. Parkinson.

Plasmid Construction

The *cheZ* gene was cloned from *E. coli* TOP10F′ cells using PCR. A cassette containing the *tac* promoter, a theophylline-sensitive riboswitch, and the *cheZ* gene was assembled using PCR and was subcloned into the *Bam*HI and *Sac*I sites of SKD1248, a derivative of pUC18 (Ap^R), to create the riboswitch-containing plasmid. A fluorescent reporter gene, GFPuv, flanked by a constitutively active promoter and a transcriptional terminator was amplified from the plasmid pGFPuv and subcloned into the *Sap*I site of the riboswitch-containing plasmid.

Dose-Dependent Migration on Semi-Solid Media

To perform the migration experiments shown in Figure 2.3, selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin, and various concentrations of theophylline) was prepared in Petri dishes (85 mm dia.). Diluted cell suspensions from mid-log-phase cultures (1.5 μL, ~2×10^5 cells/μL) were applied to the center of the plates, which were dried in air for 15 min, and incubated at 37 °C for 10 h. The migration radii were determined by measuring the diameter of the outermost ring of growth, dividing by two, and subtracting the radius of migration of non-motile RP1616 cells grown under the same conditions. The motility of the parent strain (RP437) was unaffected by

concentrations of caffeine or theophylline below 2 mM; above this concentration, motility was decreased as a result of cell death.

Gradient Response to Theophylline on Semi-Solid Media

Macroscopic motility experiments using a gradient of theophylline were performed as above, with the following modifications: Media was prepared in 100 mm square Petri dishes. Layers (15 mL) of selective 0.25% agar containing theophylline (1 mM, 0.25 mM, 0 mM) were poured in the pattern shown in Figure 2.4. Each layer was allowed to solidify for 50 min before applying the following layer. After all layers were applied, the media was allowed to equilibrate at room temperature for 3.5 h, after which diluted cell suspensions from mid-log-phase cultures (3 μL, ~2×10^5 cells/μL) were applied to the locations shown in Figure 2.4. The plates were dried in air for 15 min, and the cells were grown at 37 °C for 13.5 h.

Gradient Response to Aspartate on Semi-Solid Minimal Media

Square Petri dishes (100 mm) were filled with 50 mL of M9 minimal media containing 0.25% agar, 2% (v/v) glycerol, 50 μg/mL ampicillin, and 1 mM each of methionine, leucine, histidine, and threonine. Plates also contained a static concentration of theophylline (0 mM or 2 mM). After two hours, 40 μL of a 10 mM solution of L-aspartate dissolved in M9 salts was evenly distributed along a thin line in the center of each plate, as indicated in Figure 2.5. The plates were left at 21 °C for 3.5 h to permit diffusion of aspartate. Cells were grown to mid-log-phase in selective tryptone broth (50 μg/mL ampicillin). Cells were pelleted at 4,000 rcf and resuspended with M9 salts.

After two washes with M9 salts, diluted suspensions of cells (2.0 μL, ~2×10^5 cells/μL) were applied approximately 13 mm from the center of the plates. For each plate, the reprogrammed cells were spotted to the left of center, while wild-type cells were spotted to the right of center. Plates were dried in air for 10 min and incubated at 30 °C for 16 h.

Microscopic Behavior on a Glass Surface

To track the behavior of individual cells, cells were grown to mid-log-phase (OD_{600} = 0.55) in selective tryptone broth containing various concentrations of theophylline. A square (~1 cm/side) was drawn on a glass microscope slide using a wax pencil, the cell suspension (3 μL) was placed at the center of the square, and a glass coverslip was placed on top of the wax to seal the chamber. Samples were imaged in two dimensions using differential interference contrast microscopy (DIC) with an inverted microscope (Leica) equipped with a C.C.D. camera and a 40× objective (N.A. = 1.4). A red filter was used to improve DIC images and to protect cells from high-intensity blue light. Digital images at the surfaces of the coverslip and the slide were captured at 30 Hz and the positions of individual cells were identified using macros within the Interactive Data Language (IDL)[36]. Cells that were visibly stuck to the glass surface or were tracked for less than 1 s were discarded from further data analysis. The remaining coordinates were used to obtain paths for individual cells over the course of each video (Figure 2.10).

To differentiate motile cells from non-motile cells within the complex populations, the mean-square displacement (MSD) of each cell was determined at lag times of 0.2 s and 0.6 s. A cell was considered motile if the following three conditions were met: MSD ≥ 4.7 μm at 0.2 s; MSD ≥ 9.4 μm at 0.6 s; MSD at 0.6 s ≥ 2.5× MSD at

0.2 s. These values were determined empirically by examining the behavior of wild-type (RP437, motile) and Δ*cheZ* (RP1616, non-motile) cells.

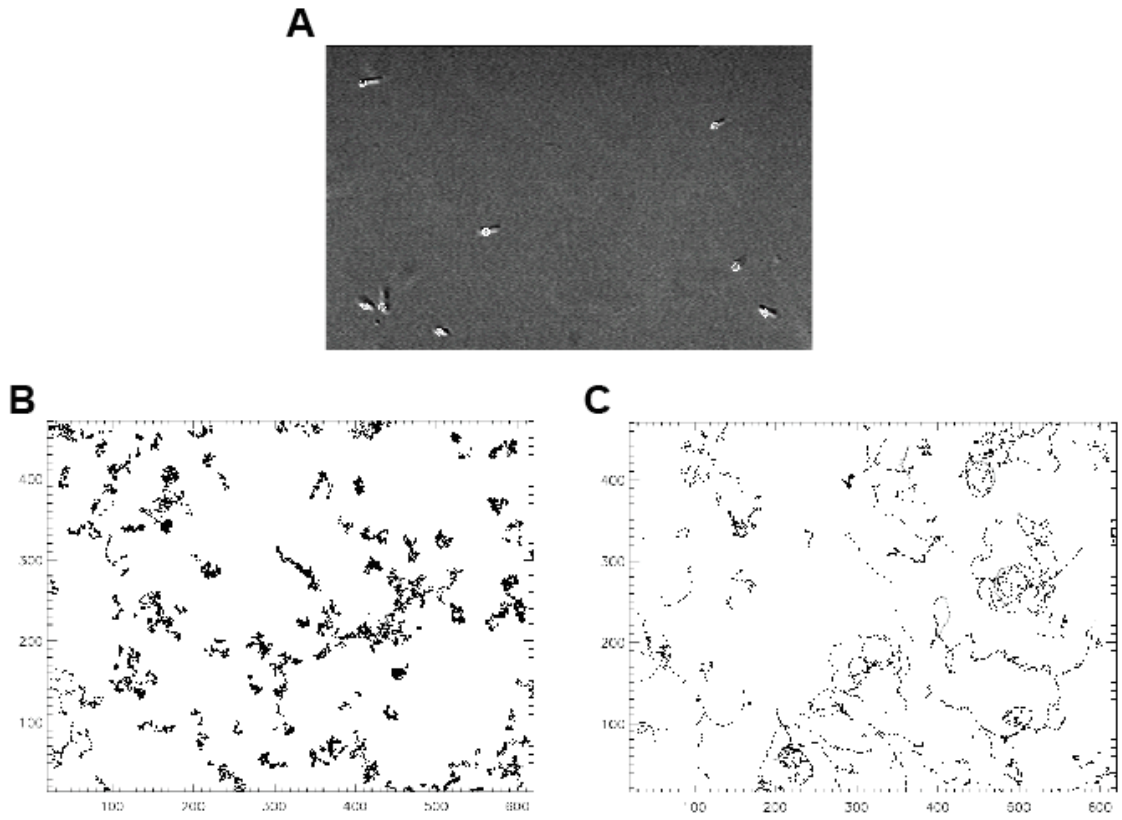

Figure 2.10: Tracking to Determine the Paths of Individual Cells. (A) A single frame of a video obtained by light microscopy shows *E. coli* cells overlaid with circles. Each circle indicates that a specific cell has been identified as a feature for tracking experiments. B) Paths of all cells identified over the course of a single movie. The reprogrammed cells were grown without theophylline. The high density of the coordinates for the cells indicate that they are nonmotile or tumbling in place. (C) Paths identified for reprogrammed cells grown in 0.75 mM theophylline. The low density of the coordinates for these cells indicates that many are motile and are able to migrate far distances at this concentration of theophylline.

Determining the fraction of cells that contribute to the overall motility of the population is complicated by the fact that motile cells appear on both the top and bottom of the chamber, while dead and non-motile cells tend to collect on the bottom of the chamber. At low theophylline concentrations, most cells on the bottom of the chamber are non-motile, while at high concentrations, a large fraction of these cells are dead. While these cells do not move in either case, their existence impacts the motility of the

population as a whole. To account for this contribution, data were collected from locations on both the top and bottom of the chamber for equal lengths of time, and the fraction of motile cells in the combined population was determined as a function of theophylline concentration for wild-type and reprogrammed cells (Figure 2.6A).

To determine the behavior of motile cells, data were collected on the top of the chamber, where cell motion was generally not obstructed by dead or non-motile cells. To minimize tracking errors, the x and y coordinates of each track were independently filtered by calculating the mean of the 3 median points of a 5 point moving window as previously described.[33] Using this filtered data, the maximum run speed of each track was calculated by averaging the top 10% instantaneous velocities in the track. A cell was considered to be tumbling when its instantaneous velocity was less than 30% of its maximum run speed, and its rate of change of direction (RCD) exceeded 5.9 rad/s for at least 2 consecutive frames. The tumbling frequency of the motile population was determined by dividing the total number of tumbling events by the total tracking duration summed over all tracks. The run speeds and tumbling frequencies are shown as a function of theophylline concentration for wild-type and reprogrammed cells (Figure 2.7).

Spatial Localization on Semi-Solid Media

To perform the spatial localization experiments shown in Figure 2.8, selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin) was prepared in Petri dishes (85 mm dia.). After solidification of the media, solutions of caffeine or theophylline (10 μM in tryptone broth), or tryptone broth alone, were applied in the pattern shown in Figure 2.8 by spotting with a micropipet (1 μL/mm), and the plates were

air-dried for 90 min. Diluted cell suspensions from mid-log-phase cultures (2 μL, ~2×10^5 cells/μL) were applied at the location shown in Figure 2.8, the plates were dried in air for 15 min, and incubated at 37 °C for 10 h. Plates were imaged on a Kodak transilluminator with NEN filters #25 (red) or #61 (green). False color (red or green) was applied using ImageJ; adjustments to brightness and contrast were performed to the whole image using Adobe Photoshop 8.0.

Changing the Ligand Specificity of Reprogrammed Cells

To test cells harboring the C27A mutant version of the riboswitch, a cassette containing the *tac* promoter, the mutated riboswitch, and the *cheZ* gene was assembled using PCR and was subcloned into the *Bam*HI and *Sac*I sites of SKD1248, a derivative of pUC18 (Ap^R). TOP10F′ cells were transformed with these ligations, and cells were plated on selective media. Single clones were grown in liquid media, and plasmids were isolated to verify the sequence of the constructs. RP1616 cells were transformed with the verified construct and were plated on selective media.

To perform the migration experiments shown in Figure 2.9B, selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin, and various concentrations of theophylline, caffeine, or 3-methylxanthine) was prepared in Petri dishes (85 mm dia.). Diluted cell suspensions from mid-log-phase cultures (1.5 μL, ~2×10^5 cells/μL) were applied to the center of the plates, which were dried in air for 15 min, and incubated at 37 °C for 10 h. The migration radii were determined by measuring the diameter of the outermost ring of growth, dividing by two, and subtracting the radius of migration of non-motile RP1616 cells grown under the same conditions.

2.5 References

(1) Adler, J. *Annu. Rev. Biochem.* **1975**, *44*, 341-356.

(2) Wadhams, G.H.; Armitage, J.P. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 1024-1037.

(3) Parkinson, J.S.; Ames, P.; Studdert, C.A. *Curr. Opin. Microbiol.* **2005**, *8*, 116-121.

(4) Ellis, L.B.; Roe, D.; Wackett, L.P. *Nucl. Acids Res.* **2006**, *34*, D517-521.

(5) Steidler, L.; Hans, W.; Schotte, L.; Neirynck, S.; Obermeier, F.; Falk, W.; Fiers, W.; Remaut, E. *Science* **2000**, *289*, 1352-1355.

(6) Weibel, D.B.; Garstecki, P.; Ryan, D.; Diluzio, W.R.; Mayer, M.; Seto, J.E.; Whitesides, G.M. *Proc. Natl. Acad. Sci.* **2005**, *102*, 11963-11967.

(7) Anderson, J.C.; Clarke, E.J.; Arkin, A.P.; Voigt, C.A. *J. Mol. Biol.* **2006**, *355*, 619-627.

(8) Maddock, J.R.; Shapiro, L. *Science* **1993**, *259*, 1717-1723.

(9) Bray, D.; Levin, M.D.; Morton-Firth, C.J. *Nature* **1998**, *393*, 85-88.

(10) Gestwicki, J.E.; Kiessling, L.L. *Nature* **2002**, *415*, 81-84.

(11) Sourjik, V.; Berg, H.C. *Nature* **2004**, *428*, 437-441.

(12) Baker, M.D.; Wolanin, P.M.; Stock, J.B. *Curr. Opin. Microbiol.* **2006**, *9*, 187-192.

(13) Derr, P.; Boder, E.; Goulian, M. *J. Mol. Biol.* **2006**, *355*, 923-932.

(14) Matsumura, I.; Ellington, A.D. *J. Mol. Biol.* **2001**, *305*, 331-339.

(15) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Mc, Q.G.S.; Roodveldt, C.; Tawfik, D.S. *Nat. Genet.* **2005**, *37*, 73-76.

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

(17) Ames, P.; Studdert, C.A.; Reiser, R.H.; Parkinson, J.S. *Proc. Natl. Acad. Sci.* **2002**, *99*, 7060-7065.

(18) Bren, A.; Eisenbach, M. *J. Bacteriol.* **2000**, *182*, 6865-6873.

(19) Parkinson, J.S.; Houts, S.E. *J. Bacteriol.* **1982**, *151*, 106-113.

(20) Kuo, S.C.; Koshland, D.E., Jr. *J. Bacteriol.* **1987**, *169*, 1307-1314.

(21) Huang, C.; Stewart, R.C. *Biochim. Biophys. Acta* **1993**, *1202*, 297-304.

(22) Ames, P.; Yu, Y.A.; Parkinson, J.S. *Mol. Microbiol.* **1996**, *19*, 737-746.

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

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

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

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

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

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

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

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

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

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

(33) Alon, U.; Camarena, L.; Surette, M.G.; Aguera y Arcas, B.; Liu, Y.; Leibler, S.; Stock, J.B. *EMBO J.* **1998**, *17*, 4238-4248.

(34) Berg, H.C.; Brown, D.A. *Nature* **1972**, *239*, 500-504.

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

(36) Crocker, J.C.; Grier, D.G. *J. Coll. Interface Sci.* **1996,** *179*, 298-310.

CHAPTER 3 – Riboswitches on the Move: A High-Throughput Selection Based on Cell Motility

3.1 Introduction

A major goal of chemical and synthetic biologists is to create ligand-dependent genetic control systems to report on cellular metabolism, to construct synthetic gene circuits, or to reprogram cellular behavior. Because designing genetic switches *de novo* is challenging, many groups have employed directed evolution strategies to achieve these goals.[1,2] Our lab recently reported a high-throughput robotic screen that identified synthetic riboswitches that displayed low background levels of gene expression in the absence of a ligand, and strongly activated gene expression in the presence of the ligand.[3] We subsequently demonstrated that these riboswitches could control bacterial motility in a ligand-dependent fashion, which was the focus of Chapter 2.[4] Because the differences in cell motility at various ligand concentrations were easy to distinguish using only a ruler, we asked whether motility differences could be the basis of a high-throughput selection to discover new synthetic riboswitches from large libraries. We envisioned that this method could equal, if not exceed, the throughput of our previously reported robotic screen, and could be performed at a fraction of the cost.

Herein, we present an inexpensive and operationally simple selection method based on cell motility that not only approaches the throughput of a genetic selection, but also provides the quantitative nature of a genetic screen. We further show that this selection quickly identifies synthetic riboswitches that display low background levels of

---

This chapter contains sections from the pre-peer reviewed version of the following article:
S. Topp and J.P. Gallivan. *ChemBioChem*, **2009**, *9*, 210-213.

gene expression in the absence of a ligand, and robust increases in the presence of a ligand. We anticipate that motility-based selections will be generally useful in the discovery of rare events from large genetic libraries.

There is good precedent for using motility to select for rare events. For decades, microbiologists have identified rare mutants by spotting cells at the center of a Petri dish containing semi-solid media and manually selecting cells that migrated abnormally.[5] Capitalizing on these successes, Goulian and coworkers recently developed a motility-based selection to discover mutant chemoreceptors that could recognize a new ligand.[6] However, selecting riboswitches presents an additional challenge. Because a switch displays two different phenotypes, depending on whether or not the ligand is present, a selection must be able to quantitatively assay both the 'on' and the 'off' states. We envisioned that a motility-based selection could be used to discover 'on' switches by selecting for mutants that don't move in the absence of the ligand, and then further selecting for mutants that move in the presence of the ligand. Of course, there are many ways that one can imagine setting up the experiment (e.g. looking for 'off' switches), but the key is to be able to rapidly and inexpensively determine the phenotype under two sets of conditions.

To compare our motility selection to a more traditional screening technique, we assayed combinatorial libraries that were similar to those previously reported by our lab,[3] which were comprised of four to eight randomized base pairs flanked by the mTCT8-4 theophylline aptamer[7] and the ß-galactosidase reporter gene. *E. coli* cells harboring these libraries were assayed using a multi-step process that used a robotic colony picker and an automated liquid handling system.[3] Although this assay identified several riboswitches

that could activate gene expression greater than 20-fold in the presence of 1 mM theophylline, it relied on expensive robotics and used large quantities of consumables. We anticipated that a motility-based selection could eliminate the need for specialized capital equipment, and identify synthetic riboswitches using only standard consumables.

3.2 Results and Discussion

To select for riboswitches using motility as the readout, we used *cheZ* as a reporter gene. CheZ plays a critical role in *E. coli* chemotaxis by dephosphorylating the CheY-P protein,[8] which binds to the flagellar motor and causes running cells to tumble. Optimal levels of CheZ are necessary for *E. coli* cells to migrate on semi-solid media. If too little CheZ is present, the level of CheY-P will increase, and the cells will tumble incessantly and not migrate.[9] If cells have excess CheZ, they will swim very smoothly and rarely tumble. Because cells that do not tumble become embedded in the semi-solid media, they cannot migrate.[10] Thus it is critical to ensure that CheZ is not overexpressed in these assays.

3.2.1 Pre-selection for RBS Sequences that Provide Optimal CheZ Levels

Because the introduction of an aptamer in the 5′ UTR of a reporter gene generally does not increase expression levels, we hypothesized that optimization of the RBS for CheZ production using a given promoter would enable us to set a maximal expression level to overcome the parabolic relationship of migration radius versus CheZ concentration for *E. coli* migrating on semi-solid media. To identify an optimal RBS sequence for CheZ production from the weak *IS10* promoter[11] before introduction of the

aptamer, we randomized nine consecutive base pairs 5′ to the *cheZ* start codon, extracted plasmid from ~15,000 pooled colonies, and transformed *E. coli* strain RP1616 (Δ*cheZ*) with the plasmid library. 600,000 cells were spotted at the center of a selective motility plate, which was incubated overnight at 30 °C. Cells located within the outer migration ring were transferred to liquid media. After growth to mid-log phase, the selected cells were spotted again on a motility plate for a second round of selection. At the end of four rounds, the outer ring of cells migrated to the edge of the plate, and DNA sequencing of the population revealed that the pool had converged on adenine-enriched sequences.

To isolate single library members and to verify that the changes in motility were plasmid-dependent, we transformed electrocompetent RP1616 cells with the pool of plasmids remaining after the fourth round of selection. Seven transformants were grown in liquid media and sequenced to determine the identity of each randomized region. Additionally, these cells were subcultured, grown to mid-log phase, and spotted on semi-solid media to identify individual clones that might exhibit motility similar to wild-type cells. Of the seven clones tested, four were approximately as motile as wild-type cells, one was somewhat less motile, and two were completely nonmotile. For the most motile clone, a mutation in the constant region before the randomized region appears to be incorporated as a part of the putative RBS for CheZ expression (Figure 3.1).

Because the introduction of an aptamer 5′ to the RBS of a reporter gene generally does not increase overall expression beyond the level achieved by the transcript lacking the aptamer, we anticipated that most cells found at the center of the plate during the subsequent riboswitch selections would be nonmotile as a consequence of low CheZ expression rather than because overexpression of CheZ caused cells to become embedded

in the media. A library in which the mTCT8-4 theophylline aptamer was followed by 8 consecutive randomized base pairs was cloned 5′ to the preselected RBS sequence. TOP10F′ cells were transformed with this ligation and plated on selective media. Plasmid DNA was extracted from the resulting pool of 10,000 clones and was used to transform the JW1870 strain (Δ*cheZ*, Keio collection).[12]

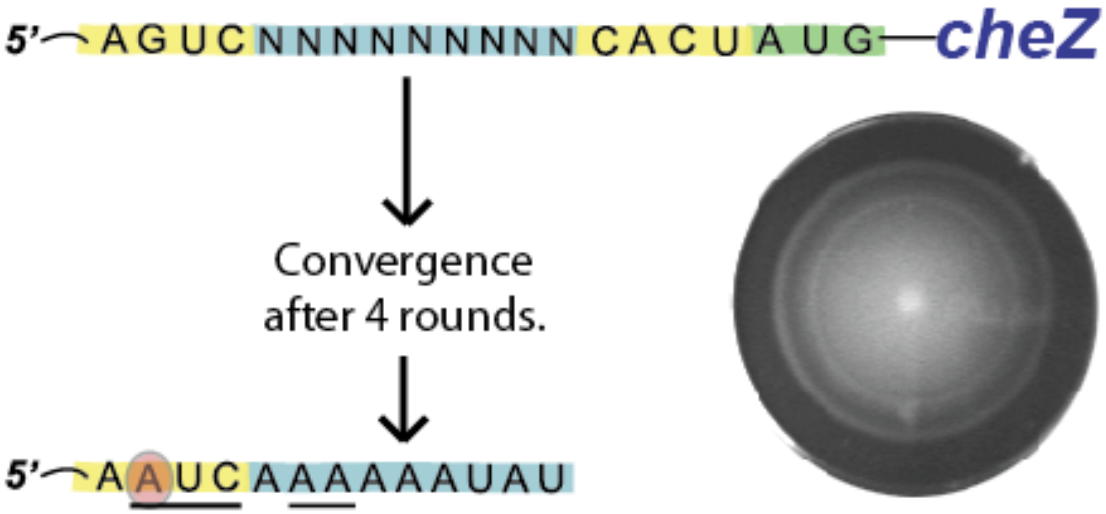

Figure 3.1: Pre-Selection for an Optimized RBS Sequence. After four rounds of selection, the most motile clone featured a purine-rich sequence in the randomized region, while a mutation from G to A in a constant region may contribute to the putative RBS sequence (underlined). The plate shows the migration of the selected clone on semi-solid media after incubation at 30 °C for 13 h.

For the first round of motility selection, 600,000 cells were spotted at the center of a plate containing semi-solid media without theophylline. Non-motile cells remaining at the center of the plate were transferred to liquid media and were spotted on plates without theophylline for a second round of negative selection. A suspension of non-motile cells selected from the center of the second selection plate was spotted on a pair of motility plates containing 0 mM or 1 mM theophylline. For the third round of selection, cells were collected from the outermost ring of the plate containing theophylline, grown in liquid media, and spotted again on plates with and without theophylline. Cells remaining

at the center of the 0 mM theophylline plate were collected for the fourth and final round, and DNA sequencing showed convergence to the sequence shown in Figure 3.2. The predicted structure[13] of this mRNA sequence suggests that the preselected RBS is unpaired both in the aptamer structure and lowest energy structures, while the randomized region contains a new putative RBS, which is paired with a portion of the aptamer in the absence of ligand. Two of five clones that were sequenced following the third round of selection were identical to the sequence that would reach convergence after the fourth round. The other three sequences selected after three rounds also featured putative RBS sequences within the N8 randomized region.

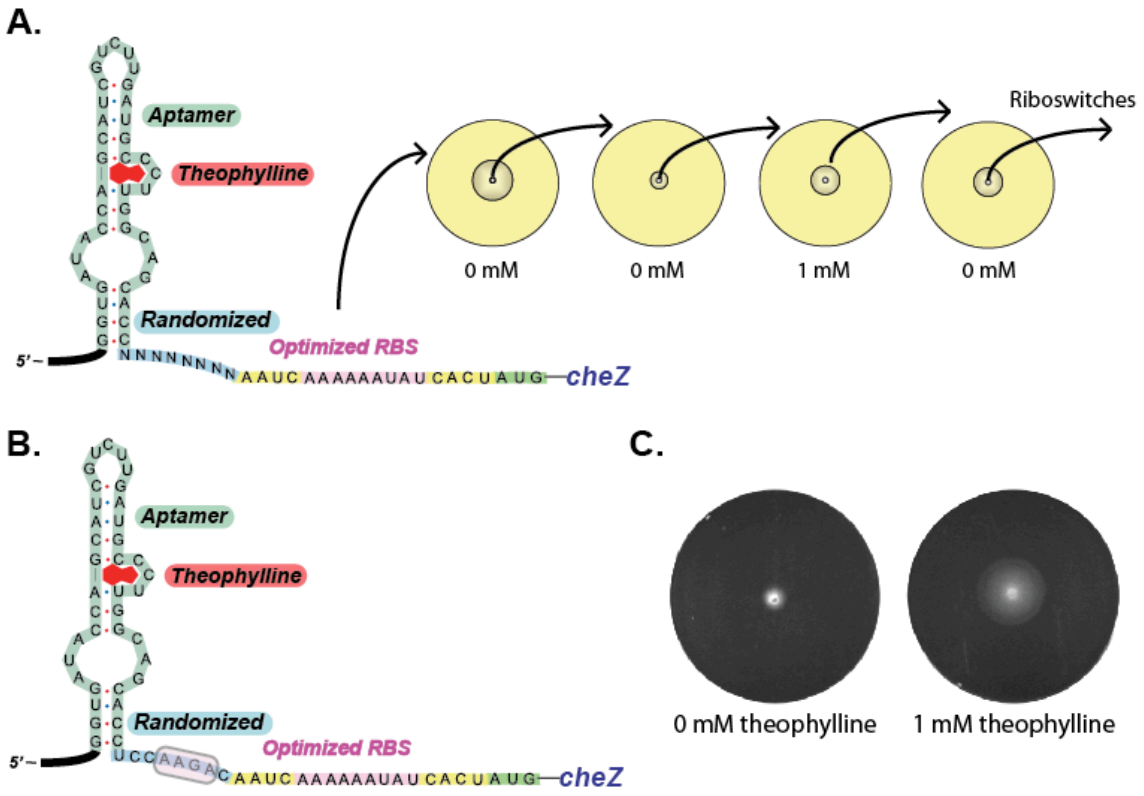

Figure 3.2: Selection for Riboswitches with an Optimized RBS Sequence. (A) A randomized region of 8 nucleotides was cloned between the theophylline aptamer and the pre-selected RBS sequence. The selection strategy entailed choosing cells that were non-motile in the absence of ligand but motile in the presence of theophylline. (B) The region that was randomized contains a putative RBS sequence, indicating that prior optimization of the RBS sequence is unnecessary. (C) The motility selection yielded riboswitches that prevent migration on semi-solid agar in the absence of theophylline, but induce migration in the presence of 1 mM theophylline.

To quantitate the activity of this riboswitch, we used PCR to amplify the sequence and cloned it into the 5′ UTR of a *cheZ-lacZ* translational fusion. TOP10F′ cells were transformed with this construct, which was assayed for ß-galactosidase activity when cells were grown with or without theophylline (1 mM). The assay revealed that this riboswitch activated gene expression 5-fold in the presence of 1 mM theophylline. Although this result was encouraging, we envisioned that a truly useful motility selection method should be capable of identifying riboswitches with activation ratios greater than 5-fold.

We suspected that pre-optimization of the RBS may be unnecessary, and perhaps detrimental, as the ensuing motility assays revealed that the clones that survive the selection generally have another RBS sequence incorporated in their randomized regions. Additionally, subsequent results using other selection methods revealed that the sequence landscape for synthetic riboswitches can be enlarged by assaying larger combinatorial libraries (N10 or N12) that lack a predetermined RBS.[14] Because the strength of the promoter chosen for a selection assay on semi-solid media establishes the set of sequences that have the potential to enable *E. coli* to migrate across the plate under any particular condition, we decided to create libraries with both the weak *IS10* promoter[11] and the strong *tac* promoter.[15]

3.2.2 A Library for Motility Selections

Since the strength of the promoter will ultimately dictate the maximum expression level of the *cheZ* gene, we began with two different promoters: a ‘weak’ *IS10* promoter[11] and the *tac* promoter,[15] which is 60 to 100-fold stronger. We anticipated that the motility

selections would readily reveal which promoter provided the appropriate CheZ expression level. Using cassette-based PCR mutagenesis, we constructed a library in which the mTCT8-4 theophylline aptamer[7] was followed by 10 consecutive randomized base pairs (N10) in the 5′ untranslated region of *cheZ* (Figure 3.3).

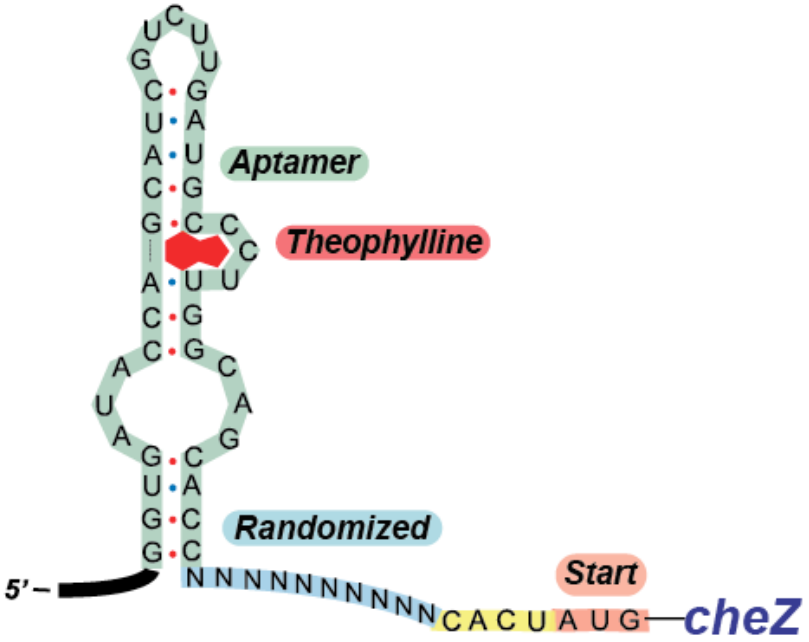

Figure 3.3: Library for Motility Selections without a Predetermined RBS. Sequence of a combinatorial library lacking a predetermined RBS sequence. A region of 10 randomized bases is positioned between the aptamer stem and the RBS of a reporter gene. The constant sequence between the randomized region and the start codon has been changed from the sequence we used in previous screens for synthetic riboswitches. This library was cloned into two plasmids to enable transcription from the strong *tac* promoter or the weak *IS10* promoter.

3.2.3 Motility Selections with Two Different Promoters

E. coli strain JW1870 (Δ*cheZ)* was transformed separately with the N10 plasmid libraries under the control of either the *IS10* or the *tac* promoter. After each transformation, ~600,000 cells were spotted on selective semi-solid media with or without theophylline and grown at 30 °C for 13 h (Figure 3.4). Cells that did not migrate in the absence of theophylline were isolated and subjected to a second round of negative

selection. After 18 h, non-motile cells on the theophylline-free plate were isolated and spotted on plates with and without theophylline. Populations of cells from both libraries displayed increased motility in the presence of theophylline, and cells were collected by scraping the region of the plates that extended beyond the outermost visible ring of cells. Cultures derived from the weaker *IS10* promoter library grew normally, and when spotted on semi-solid media in the presence or absence of theophylline (1 mM), showed clear differences in motility. In contrast, cells derived from the stronger *tac* promoter library failed to grow, suggesting that very few migrated far across the plate.

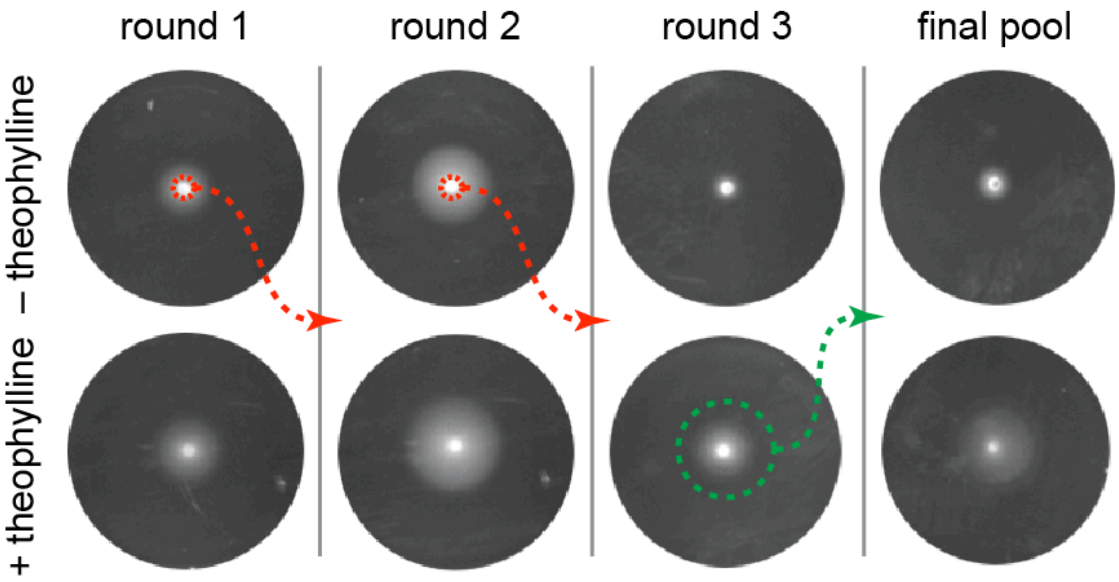

Figure 3.4: Progress of Selections for Synthetic Riboswitch Activity. Each round of selection is shown for the *IS10* promoter library. In rounds 1 and 2, cells were chosen from the center of plates without theophylline. In round 3, we picked cells from the outside of the plate containing theophylline (1 mM). The final pool shows a significant increase in motility when theophylline is present. Plates from rounds 1, 2, 3, and the final pool were incubated at 30 °C for 13, 18, 14, and 14 h, respectively. The longer incubation in round 2 was to ensure that cells had time to migrate, allowing us to pick the least motile cells.

We suspected that the selection strategy did not work well using the strong *tac* promoter because it is likely that many sequences exhibit high levels of background expression. Such over-expression of CheZ may cause the majority of the cells to become

embedded within the semi-solid media regardless of whether theophylline was present. To confirm this, we cloned the population of 5′ UTRs from each library after the second round of negative selection upstream of the ß-galactosidase reporter gene. These cells were plated onto selective media containing X-gal, but no theophylline. As expected, the majority of the cells from the weak *IS10* promoter library were white or light blue, indicating that gene expression was low in the absence of theophylline (Figure 3.5). However, the majority of cells from the *tac* promoter library were dark blue, suggesting that they were expressing substantial amounts of ß-galactosidase (and by extension, CheZ) in the absence of theophylline. Thus, many of these cells in the *tac* library were non-motile because CheZ was over-expressed and the cells became embedded in the media. This result highlights the need to control for promoter strength in motility assays.

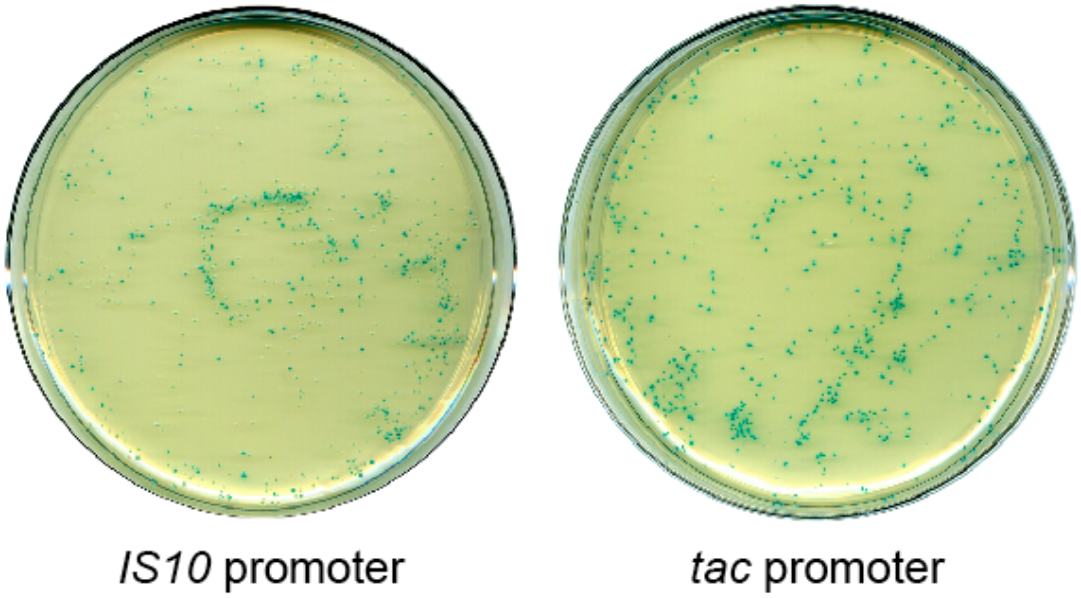

Figure 3.5: Gene Expression of *IS10* or *tac* Promoter Clones after Two Rounds of Negative Selection. Cells remaining in the pool after the second round of negative selection were tested for ß-galactosidase activity in the absence of theophylline by plating cells on media containing X-gal. When compared to the *tac* promoter pool, the *IS10* promoter pool has a greater fraction of clones that are white or light blue after two rounds of negative selection. Therefore, we may infer that many cells from the *IS10* promoter pool remain at the center of the plate because they express low levels of CheZ. In contrast, most cells from the *tac* promoter pool remain at the center of the plate because they express very high levels of the CheZ protein and become embedded in the semi-solid media.

3.2.4 Confirmation of Selected Riboswitches

Since most of the cells in the *IS10* library showed low levels of gene expression in the absence of theophylline, and significant increases in motility in the presence of theophylline, we anticipated that this library would contain functional synthetic riboswitches. To confirm this hypothesis, plasmids were extracted from the population of cells collected after the third round of selection, the DNA pool was amplified by PCR, and these sequences were cloned in the 5′ UTR of a *cheZ-lacZ* translational fusion. TOP10F′ cells were transformed with this pool obtained from the motility selection, and 95 random transformants were assayed for ß-galactosidase activity in the absence and presence of theophylline (1 mM) using a multi-channel pipettor. Of the 95 clones assayed, 12 activated gene expression more than 5-fold in the presence of 1 mM theophylline, and of these, 4 exhibited activation ratios greater than 15-fold (Figure 3.6).

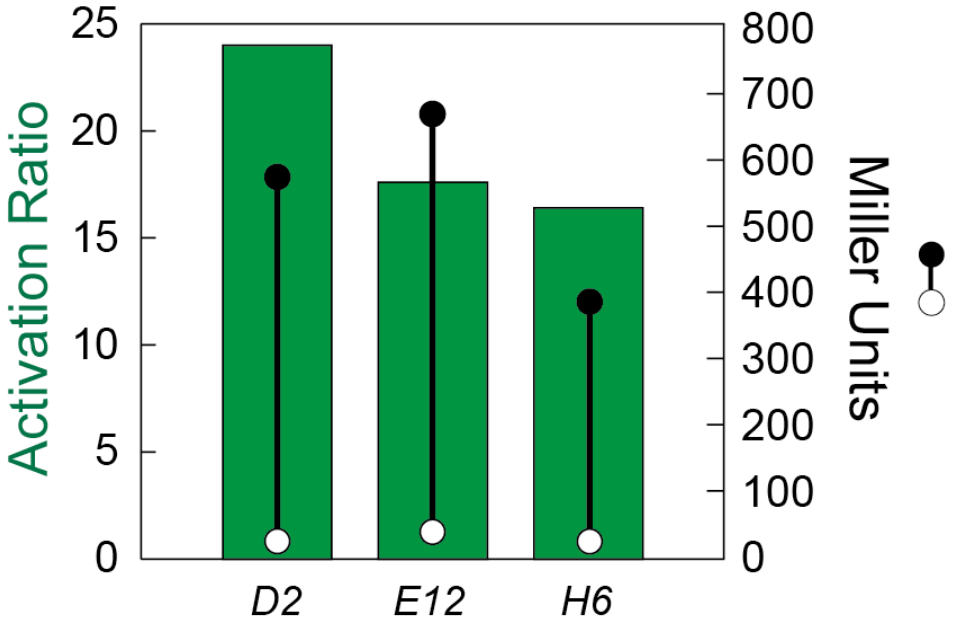

Figure 3.6: Activities of Riboswitches Identified by Motility-Based Selection. The right axis shows the ß-galactosidase activity in the absence (open circle) or presence of theophylline (1 mM, closed circle). Activities are expressed in Miller units, and the standard errors of the mean are less than the diameters of the circles. The left axis presents the activation ratios of the synthetic riboswitches (green bar), which are determined by taking the ratio of the Miller units in the presence and absence of theophylline.

Sequencing revealed that two of the 4 switches were identical in the 'N10' region (Table 3.1). Another had 11 nucleotides in this region, which may have resulted from errors in oligonucleotide synthesis, or from random mutations introduced by replication in *E. coli*; regardless of the cause, this result highlights the ability of the motility selection to discover rare events. All of the sequences could adopt folds[13] similar to riboswitches identified using our previously described screen,[3] and each contained a region of purines that could function as a ribosome binding site (Figure 3.7).

Clone	Linker Sequence	Miller U. (0 mM)	Miller U. (1 mM)	Activation Ratio
H6	UCUGAGGGCU	24	386	16
D2	UUCGAGGGCA	24	574	24
E12	CGAAGGGGCAG	38	670	18

Table 3.1: Linker Sequences of Riboswitches Discovered by Motility Selection. The linker sequences of several riboswitches are shown, with regions of similarity highlighted in red. The activation ratio for each switch is calculated by dividing its activity (measured in Miller Units) in the presence of 1 mM theophylline by its activity in the absence of theophylline. Clone H6 was isolated twice.

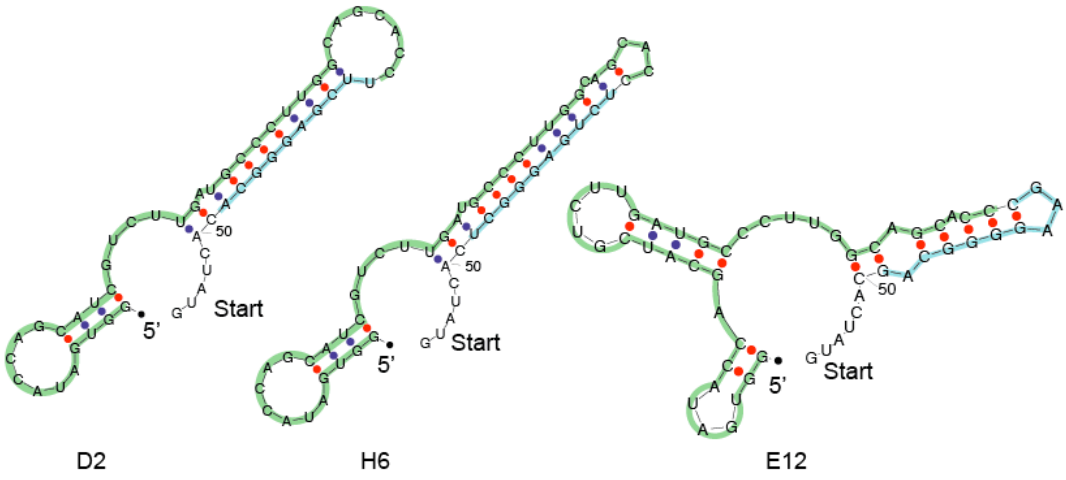

Figure 3.7: Predicted Secondary Structures of Identified Riboswitches. Predicted folds for synthetic riboswitches in the absence of the ligand. The aptamer is shaded green, and the randomized region is highlighted in blue. Each randomized region (includes RBS sequences) is extensively paired, suggesting an explanation for the low background level of gene expression in the absence of ligand.

3.2.5 Monitoring the Progression of a Library under Selective Pressure

As we considered the purine-rich nature of the sequences identified using our motility-based selection strategy, we wondered at what point in the selection process purine-rich sequences began to predominate the pool. In particular, we were interested in determining the point at which these purine-rich sequences began to resemble RBS sequences (Shine-Dalgarno sequences).[16] To address this question, we isolated single clones at random from the pool of cells remaining after rounds one, two, and three. Table 3.2 shows the DNA sequence and the number of purines for each clone.

After the first two rounds of selection, we chose cells that were non-motile on semi-solid media, which should not provide especially strong selection pressure in favor of clones that have Shine-Dalgarno sequences. Although sequences with very high complementarity to the 16S rRNA may be favored if CheZ is expressed so strongly that cells become embedded in the semi-solid agar, clones from the IS10 promoter library should not be very susceptible to such pressure. Therefore, we were not surprised to discover that, on average, these sequences include approximately three adenine residues and three guanine residues.

For the third round of selection, we chose cells that migrated far on semi-solid media containing theophylline, which provided selection pressure in favor of sequences with the ability to express CheZ under these conditions. Because the presence of an RBS generally provides enhanced gene expression over those lacking an RBS, we anticipated that the clones isolated after this positive selection round would be purine-rich and exhibit sequences with some complementarity to the 16S rRNA. Indeed, the clones selected in round three exhibit very high purine content; of the ten clones that we

sequenced, none had fewer than seven purines within the randomized region. Interestingly, of the three riboswitches featuring greater than 15-fold induction ratios, two had less than seven purines. However, the purines present in these sequences form putative Shine-Dalgarno sequences that resemble the consensus sequence.[16] This result highlights the dual requirements for optimal riboswitch function: Not only do these riboswitches need to adopt alternate conformations depending on whether a small-molecule ligand is present, but the sequence that becomes unpaired in the presence of ligand must enable the ribosome to bind for translation initiation.

Round 1	Purines	A	G
ATATTAGAGC	6	4	2
AGAGTAGTCG	7	3	4
ATTGTGATCC	4	2	2
TACAGCAAAT	6	5	1
GGACATTGCA	6	3	3
TGCGTCAGGA	6	2	4
GAACCGTTCG	5	2	3
TGGAAGCACA	7	4	3
GACTATAACA	6	5	1
AGGAGACCGG	8	3	5
Averages:	**6.1**	**3.3**	**2.8**

Round 2	Purines	A	G
GGAAGGCGCC	7	2	5
AGCGGAGAAG	9	4	5
ATGTCTGAAC	5	3	2
CTGAGGTACT	5	2	3
GAAAAATGGT	8	5	3
TTGCGAGAAT	6	3	3
GCACAAGCGT	6	3	3
CGTGATAGAG	7	3	4
GCTCGCGAAC	5	2	3
	6.4	**3.0**	**3.4**

Round 3	Purines	A	G
AGGGGAGCGC	8	2	6
CCAGGGAAAA	8	5	3
AGGGAAGTAG	9	4	5
AGGCGGGGCA	8	2	6
AGAGAGAATT	8	5	3
AACATAGGAT	7	5	2
CGCAGGCGGA	7	2	5
GACGGAAGAG	9	4	5
AAGGCGGGAG	9	3	6
AACGGAATGA	8	5	3
	8.1	**3.7**	**4.4**

Riboswitch	Purines	A	G
TTCGAGGGCA	6	2	4
CGAAGGGGCAG	9	3	6
TCTGAGGGCT	5	1	4

Table 3.2: Progression of Linker Sequences During a Motility-Based Selection. Linker sequences for random clones picked after each round of selection are shown. The number of purines for each clone is tabulated, as well as the average number of purines for each round. The linker sequences for the riboswitches shown in Figure 3.7 are presented along with the number of purines within each sequence. The bases highlighted in pink are complementary to the 16S rRNA.

3.2.6 Ligand-Dependent Motility of Cells Harboring Selected Riboswitches

The results of our motility-based selection demonstrated that the assay shows promise as a general method for efficient and inexpensive high-throughput screening. The selection identified clones harboring synthetic riboswitches with greater than 15-fold

activation ratios. Additionally, our motility-based assay selects for riboswitches featuring dynamic ranges that are conducive to cell migration on semi-solid agar. To determine whether the riboswitches selected in our assay exhibit the expected phenotypes, we cloned each into the 5′ UTR of *cheZ* and measured the migration distance of cells harboring these switches in the presence or absence of theophylline (Figure 3.8). The data show that cells harboring each switch display the expected migration phenotype.

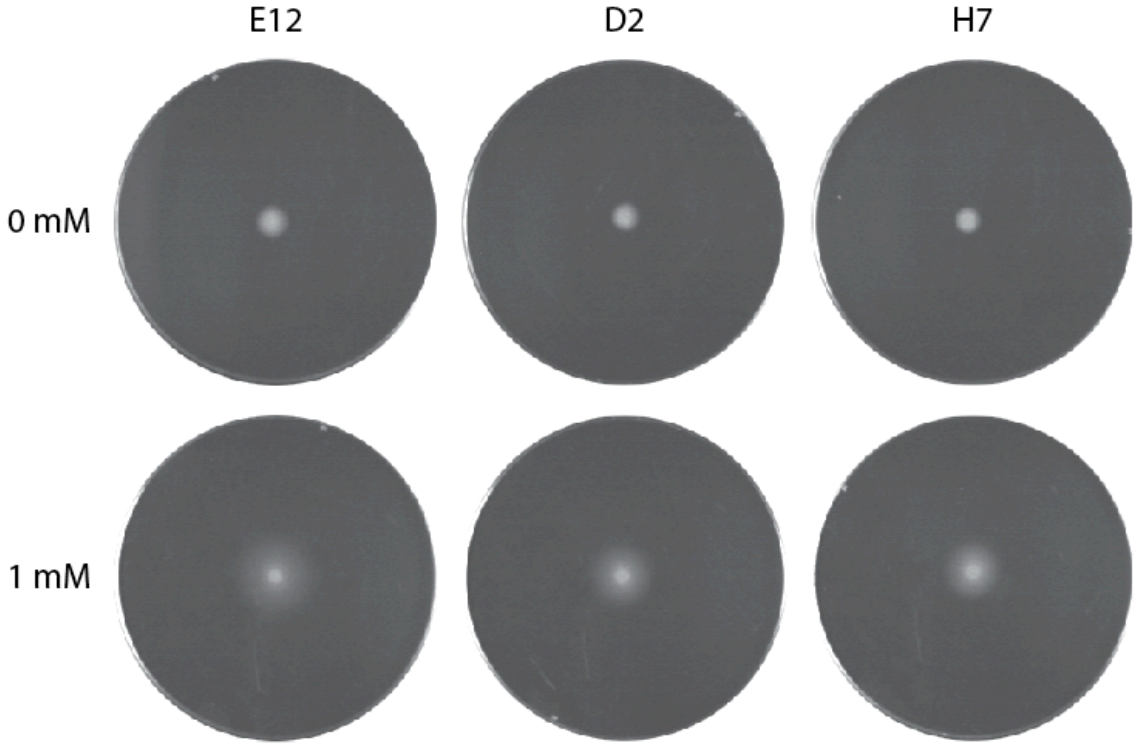

Figure 3.8: Migration of Cells Harboring New Riboswitches. *E. coli* cells (JW1870) harboring riboswitches identified using a motility-based selection strategy were spotted on semi-solid media and were grown at 30 °C for 13 h. All three riboswitches produce a non-motile phenotype when cells are grown without theophylline. The riboswitches promote migration across the semi-solid agar when cells are grown in media containing 1 mM theophylline, but the migration distances vary depending upon which riboswitch is harbored by the cells.

3.2.7 A Correlation Between Migration Distance and β-Galactosidase Activity

Intrigued by the data presented in Section 3.2.6, we were interested in determining whether the motility of cells spotted on semi-solid agar might correspond to

the absolute levels of gene expression as measured by ß-galactosidase activity (Miller Units). If this were the case, our motility selection, which can be performed on millions of clones in parallel using semi-solid agar, could be used in place of the ß-galactosidase assay, which requires significantly more resources to perform. To address this question, we cloned several riboswitches that were discovered by robotic screening into the 5′ UTR of *cheZ* and performed motility assays in the presence and absence of theophylline. We correlated the ß-galactosidase expression in the presence and absence of theophylline with the migration on semi-solid agar for these previously described clones,[3] as well as for the newly-discovered synthetic riboswitches (Figure 3.9). Regardless of the method by which the riboswitch was identified, the results correlate well, suggesting that motility on semi-solid agar is an inexpensive and effective proxy for high-throughput assays for ß-galactosidase expression that require expensive reagents and equipment.

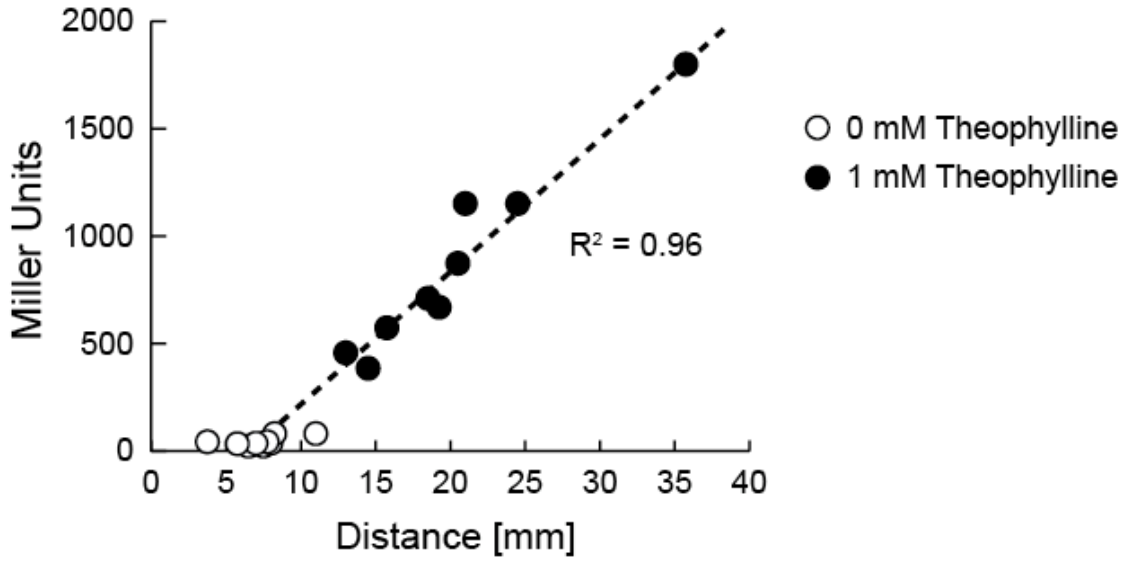

Figure 3.9: Cell Motility Correlates with Enzymatic Activity. ß-galactosidase activity (in Miller units) vs. cell motility (distance migrated on semi-solid agar) for synthetic riboswitches. Measurements in the absence of theophylline are shown as open circles; those in the presence of theophylline (1 mM) are shown as closed circles. There is a strong correlation between migration distance and ß-galactosidase activity. In the absence of theophylline, a riboswitch identified in this assay should feature < 50 Miller Units. In the presence of ligand, it should exhibit between ~ 300 and ~ 3000 Miller Units, permitting significant migration away from the center of the plate. Uncertainties in measurement are smaller than the symbols.

3.3 Conclusion

In summary, we have developed a rapid, efficient, and inexpensive method to discover synthetic riboswitches. This method is capable of performing both positive and negative selections on very large libraries (>10^5 clones) and requires no robotics, only standard consumables, and a minimal amount of human intervention. The riboswitches identified using this assay have low background levels of gene expression in the absence of ligand and display good activation ratios. We expect that this method will be useful for discovering riboswitches starting from a variety of RNA aptamers, and that the resulting synthetic riboswitches will be generally useful for controlling bacterial gene expression. Furthermore, we suggest that current genetic screens that detect changes in the production of proteins such as GFP or ß-galactosidase may be adapted to use cell motility as an inexpensive high-throughput readout.

3.4 Experimental

General Considerations

Synthetic oligonucleotides were purchased from Integrated DNA Technologies. Culture media was obtained from EMD Bioscience. Ampicillin was purchased from Fisher. Theophylline was obtained from Sigma. DNA polymerase, restriction enzymes, and the pUC18 cloning vector were purchased from New England Biolabs. Purifications of plasmid DNA, PCR products, and enzymatic digestions were performed using kits from Qiagen. All new plasmids were verified by DNA sequencing provided by Agencourt, MWG Biotech, or the Center for Fundamental and Applied Molecular Evolution at Emory University.

Construction of Libraries for Motility Selections

To construct the libraries shown in Figures 3.1, 3.2A and 3.3, cassette-based PCR mutagenesis was used to introduce the indicated sequences in the 5′ untranslated region of the *cheZ* gene. The full-length PCR products were subcloned into the *Kpn*I and *Sac*I sites of a vector derived from pUC18 (Ap^R). Plasmid manipulations were performed with *E. coli* TOP10F′ cells (Invitrogen) that were transformed by electroporation.

Selection for an Optimized RBS Sequence

Motility plates for optimizing the RBS sequence contained 25 mL selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin) in Petri dishes (85 mm dia.). RP1616 *E. coli* cells were transformed with the plasmid library isolated from TOP10F′ cells. After growing transformed RP1616 cells in selective media, diluted cell suspensions from mid-log-phase cultures (3 μL, ~6×10^5 cells) were applied to the center of the plates, which were dried in air for 15 min, and incubated at 30 °C. For four consecutive rounds, cells were selected from the outer migration ring, as described in Section 3.2.1. The selected cells were suspended in liquid media (tryptone broth, 50 μg/mL ampicillin) and grown to mid-log-phase to begin the next round of selection.

Motility-Based Riboswitch Selection with the Optimized RBS

Motility-based selections for synthetic riboswitches were performed with *E. coli* strain JW1870.[12] Motility plates containing 25 mL selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin, and 0 mM or 1 mM theophylline) were prepared in Petri dishes (85 mm dia.). To perform the assay presented in Figure 3.2A, the plasmid

library extracted from TOP10F′ cells was used to transform JW1870 cells, which were grown to mid-log phase before ~ 600,000 cells were spotted on motility plates without theophylline. Non-motile cells were collected from the center of the plate, and these cells were cultured to perform a second round of negative selection. For the third round of selection, cells were spotted on plates with and without theophylline (1 mM), and we selected cells that migrated far across the plate containing theophylline. For the fourth and final round of selection, these cells were spotted again on motility plates with and without theophylline, and cells that did not migrate from the center of the plate without theophylline were collected. Plasmid was extracted from these cells, and sequencing revealed that the pool had converged on the sequence shown in Figure 3.2B. This single clone was spotted on motility plates with and without theophylline to observe the phenotype of cells harboring this switch (Figure 3.2C)

Motility Selections Without a Pre-Selected RBS

Motility experiments were performed with *E. coli* strain JW1870. For all motility plates, selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin, and 0 mM or 1 mM theophylline) was prepared in Petri dishes (85 mm dia.). Diluted cell suspensions from mid-log-phase cultures (3 μL, ~6×10^5 cells) were applied to the center of the plates, which were then dried in air for 15 min, and incubated at 30 °C. Selected cells were suspended in liquid media (tryptone broth, 50 μg/mL ampicillin) and grown to mid-log-phase to begin the next round of selection.

To begin the selection, JW1870 cells were transformed with both plasmid libraries described in Figure 3.3, and ~600,000 cells from each library were spotted at the

center of motility plates with and without theophylline. After 13 hours, non-motile cells remaining at the center of the plates lacking theophylline were suspended in liquid media, grown to mid-log phase, and spotted on plates with and without theophylline. After 18 hours, cells that remained at the center of the plates without theophylline were transferred to liquid media and again spotted on motility plates with and without theophylline. After 14 hours, cells were collected from the region beyond the visible migration edge of the 1 mM theophylline plates. A diluted suspension of the final pool derived from the *IS10* promoter library was spotted on motility plates to confirm the theophylline-dependent migration of this population.

Construction of *cheZ-lacZ* Translational Fusions

Plasmid DNA was extracted from cell cultures grown after the second and third rounds of selection. These pools were amplified with forward primer JPG-007 (which anneals 5′ to the aptamer) and reverse primer ST-278 (which anneals to the 3′ end of *cheZ*, removes the stop codon, adds a GGSSAA linker, and introduces a *Hind*III restriction site to construct a translation fusion to *lacZ*). The pools of PCR products were subcloned into the *Kpn*I and *Hind*III sites of SAL172 (*IS10* promoter) or SAL109 (*tac* promoter), both of which contain *lacZ* and are derivatives of pUC18.[3]

β-galactosidase Promoter Comparison

TOP10F′ cells were transformed with the *cheZ-lacZ* pools constructed for both the *IS10* and *tac* promoter populations after two rounds of motility selection. Cells were

grown at 37 °C on LB/agar (25 mL) supplemented with ampicillin (50 μg/mL) and X-Gal (2.1 mg dissolved in 83.3 μL dimethyl formamide).

β-galactosidase Assay to Confirm Activity

TOP10F′ cells were transformed with the *cheZ-lacZ* pool that was constructed after the final round of motility selection. Cells were grown at 37 °C on LB/agar supplemented with ampicillin (50 μg/mL). 95 clones were picked by hand and were assayed for ß-galactosidase activity in the absence and presence of theophylline (1 mM) using a multi-channel pipettor as previously described.[3] Clones with activation ratios greater than 15 were subcultured and assayed in triplicate, as previously described.[3] Miller Units were calculated by the following formula:

Miller Units = $OD_{420}/(OD_{600}$ x hydrolysis time x [volume of cell lysate/total volume])

Motility Assay to Correlate Migration to Miller Units

Forward primer JPG-007 and reverse primer SKD-056 were used to clone several riboswitches (which were discovered previously by robotic assay)[3] into the 5′ UTR of a pUC18 derivative containing the *cheZ* gene. JW1870 (Δ*cheZ*) cells were transformed with these riboswitches. Cells were grown to mid-log phase in tryptone broth supplemented with ampicillin (50 μg/mL), and diluted suspensions (~600,000 cells) were spotted at the center of motility plates with or without theophylline (1 mM). Plates were imaged after 13 hours incubation at 30 °C. The migration radii were measured, and these values were plotted against the corresponding Miller Units for each riboswitch.

3.5 References

(1) Yokobayashi, Y.; Weiss, R.; Arnold, F.H. *Proc. Natl. Acad. Sci.* **2002,** *99*, 16587-16591.

(2) Chockalingam, K.; Chen, Z.; Katzenellenbogen, J.A. *Proc. Natl. Acad. Sci.* **2005,** *102*, 5691-5696.

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

(4) Topp, S.; Gallivan, J.P. *J. Am. Chem. Soc.* **2007,** *129*, 6807-6811.

(5) Adler, J. *Annu. Rev. Biochem.* **1975,** *44*, 341-356.

(6) Derr, P.; Boder, E.; Goulian, M. *J. Mol. Biol.* **2006,** *355*, 923-932.

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

(8) Zhao, R.; Collins, E.J.; Bourret, R.B.; Silversmith, R.E. *Nat. Struct. Biol.* **2002,** *9*, 570-575.

(9) Wolfe, A.J.; Conley, M.P.; Kramer, T.J.; Berg, H.C. *J. Bacteriol.* **1987,** *169*, 1878-1885.

(10) Wolfe, A.J.; Berg, H.C. *Proc. Natl. Acad. Sci.* **1989,** *86*, 6973-6977.

(11) Jain, C.; Kleckner, N. *Mol. Microbiol.* **1993,** *9*, 233-247.

(12) 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*, 2006.0008.

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

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

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

(16) Shine, J.; Dalgarno, L. *Nature* **1975,** *254*, 34-38.

CHAPTER 4 – Switching the Switch: An Unexpected Mechanism for Small-Molecule Dependent Translational Repression

4.1 Introduction

For nearly a decade, riboswitches have been described as genetic control elements located in an untranslated region (UTR) of mRNA that modulate transcription or translation upon binding a specific ligand or metabolite.[1] This perspective has frequently guided bioinformatics efforts to identify natural riboswitches,[2] as well as experimental efforts to develop synthetic riboswitches.[3,4] Previous work has revealed that bacterial riboswitches that act on a translational level often regulate gene expression through ligand-dependent structural changes in the 5′ UTR that provide or limit access to the RBS (defined explicitly as the region to which the 16S rRNA is complementary, or the Shine-Dalgarno sequence[5]) of an RNA transcript.[1] This insight has inspired efforts to convert synthetic aptamers into riboswitches through rational engineering[3,6,7] or genetic screens and selections.[8-10] Here, we report that synthetic riboswitches that modulate translation in bacteria can be located within the coding region of a gene, and that opportunities for developing new riboswitches may be enhanced by placing fewer restrictions on the expected mechanisms of RNA-based genetic regulation.

We have previously reported several high-throughput screening methods that can be used to identify synthetic riboswitches that modulate gene expression in *E. coli* cells,[8,9] and we have used the resulting switches to control cell behavior in a ligand-dependent fashion.[11] Additionally, the screens revealed that our switches activated gene expression by providing access to the RBS in the presence of ligand, and suggested that

the ligand is more likely to activate, rather than to repress, gene expression for the average library member. However, a given aptamer should theoretically be useful for developing riboswitches that either activate or repress gene expression. For example, the *Bacillus subtilis xpt* and *ydhl* riboswitches,[12] as well as the *Vibrio vulnificus add* riboswitch,[13] employ aptamers that bind their purine ligands in nearly identical binding pockets, yet they modulate gene expression by transcriptional termination, transcriptional anti-termination, and translational activation, respectively.

Because it should be possible to construct riboswitches that enhance or repress gene expression from a single aptamer, yet our previous work failed to identify synthetic riboswitches that repress gene expression in the presence of the ligand, we sought to develop libraries that would allow us to identify such switches. Achieving this goal would provide further insight into the mechanisms by which synthetic riboswitches can be used to regulate gene expression, and would enhance our knowledge of riboswitch construction. Moreover, there will be increased flexibility in efforts to reprogram cell behavior with RNA[14] if a single aptamer can be used to develop switches that either activate or repress gene expression.

4.2 Results and Discussion

We have shown in several recent studies that theophylline-sensitive riboswitches can be isolated from genetic libraries in which a fully randomized region of 8-12 bases is positioned between the mTCT8-4 theophylline aptamer[15] and the start codon of a reporter gene.[8,9] These switches activated gene expression by conformational changes of the RNA, from a structure in which the Shine-Dalgarno sequence (RBS) was paired in the

absence of ligand to one in which it was unpaired in the presence of theophylline.[9] Since many sequences in these libraries showed theophylline-dependent gene activation, but none exhibited significant (>5-fold) gene repression in the presence of theophylline, we hypothesized that the combinatorial library might be rationally redesigned to increase the frequency of sequences that repress gene expression in the presence of ligand.

We tested several approaches to redesign these genetic libraries and to screen for library members that function as theophylline-dependent repressors. Our studies revealed that the specific sequence positioned immediately 5′ to the start codon may provide a bias toward riboswitches that activate gene expression. Ultimately, we identified riboswitches that could repress gene expression by an unintended, yet interesting mechanism. These studies may provide general principles that could be applied in the future to develop straightforward methods to convert new aptamers into synthetic riboswitches.

4.2.1 Designing a Combinatorial Library for Ligand-Dependent Repression

To bias our combinatorial library toward riboswitches that could repress gene expression in the theophylline-bound state, we introduced two major changes that distinguish our new strategy from our previously reported library design (Figure 4.1A). First, we positioned the RBS immediately downstream of the aptamer stem, with the hypothesis that proximity to the highly structured, ligand-bound aptamer would prevent translation. We anticipated that the RBS could be made more accessible for translation in the theophylline-free state if the 5′ portion of the aptamer were to interact more favorably with an upstream sequence rather than with the 3′ portion of the aptamer. To this end, the second modification to our previous libraries involved positioning a twelve-nucleotide

randomized region immediately 5′ to the aptamer, rather than after the aptamer (Figure 4.1B).[8,9] We anticipated that in the absence of theophylline, some of the library members would have sequences in which the nucleotides in the randomized region could pair with the 5′ portion of the aptamer, disfavoring the aptamer conformation, and therefore making the RBS more accessible for translation.

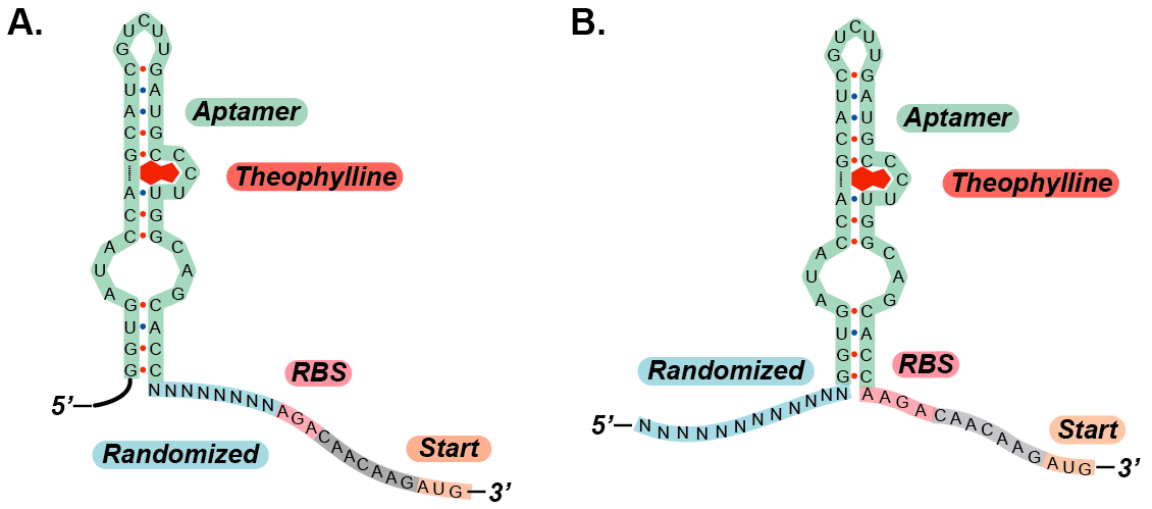

Figure 4.1: Sequences of Previously Screened and Redesigned Genetic Libraries. (A) Sequence of a previously screened library, which featured a randomized region positioned between the aptamer stem and the RBS of a reporter gene.[9] (B) Sequence of a modified library, in which the RBS is located immediately after the aptamer stem and the randomized region is position 5′ to the aptamer.

To identify library members that function as genetic repressors, we transformed *E. coli* cells with the library shown in Figure 4.1B, and plated 5,000 transformants on selective media containing X-gal. Because these cells were grown in the absence of theophylline and we were interested in identifying clones that would exhibit high gene expression in the theophylline-free state, we selected colonies that appeared dark blue under these conditions. To screen for riboswitches that could repress gene expression in the presence of theophylline, these clones were assayed in 96-well format, as previously described.[9]

Surprisingly, no clone showed even 2-fold lower gene expression when cells were grown in the presence of theophylline, compared to the corresponding expression level when cells were grown without theophylline. However, more than 10% of the clones exhibited 5-fold or greater activation of gene expression when cells were grown in 1 mM theophylline. To explore why this library remained biased toward sequences that activate gene expression, we sequenced several clones that exhibited at least 5-fold increases in gene expression in the presence of 1 mM theophylline. Figure 4.2A depicts the predicted secondary structures for one of these clones in the absence and presence of theophylline. Interestingly, the constant sequence (CAACAAG) positioned between the RBS and the start codon may create a potential bias toward riboswitches that provide repression in the ligand-free state and activation in the theophylline-bound state.

The other clones that exhibited greater than 5-fold induction feature similar predicted structures, indicating that the RBS is paired in the ligand-free state and unpaired in the theophylline-bound conformation. The data suggest that access to the RBS is not sufficiently obstructed by formation of the aptamer structure to prevent translation in the theophylline-bound conformation. Therefore, these results provided good impetus to design and screen a new library to overcome this bias. Section 4.2.2 will address this goal. However, before moving forward, we observed that it might be possible to improve the riboswitch shown in Figure 4.2A by increasing pairing of the RBS sequence with the aptamer in the ligand-free state.

We noted that insertion of 2 guanine nucleotides before the final adenine of the RBS sequence shown in Figure 4.2A might increase pairing between the RBS and the aptamer in the absence of ligand (Figure 4.2B). Moreover, this 2-nucleotide addition

would increase the complementarity of the RBS sequence to the 16S rRNA, potentially enabling higher activation levels when cells are grown in the presence of theophylline. Construction of this riboswitch sequence, and subsequent testing in *E. coli* cells confirmed our hypothesis. Background levels of gene expression in the absence of ligand were extremely low (< 10 Miller Units), while addition of 1 mM theophylline provided activities well over 1,000 Miller Units. This result suggests that semi-rational design may be useful in future efforts to develop synthetic riboswitches using new aptamers.

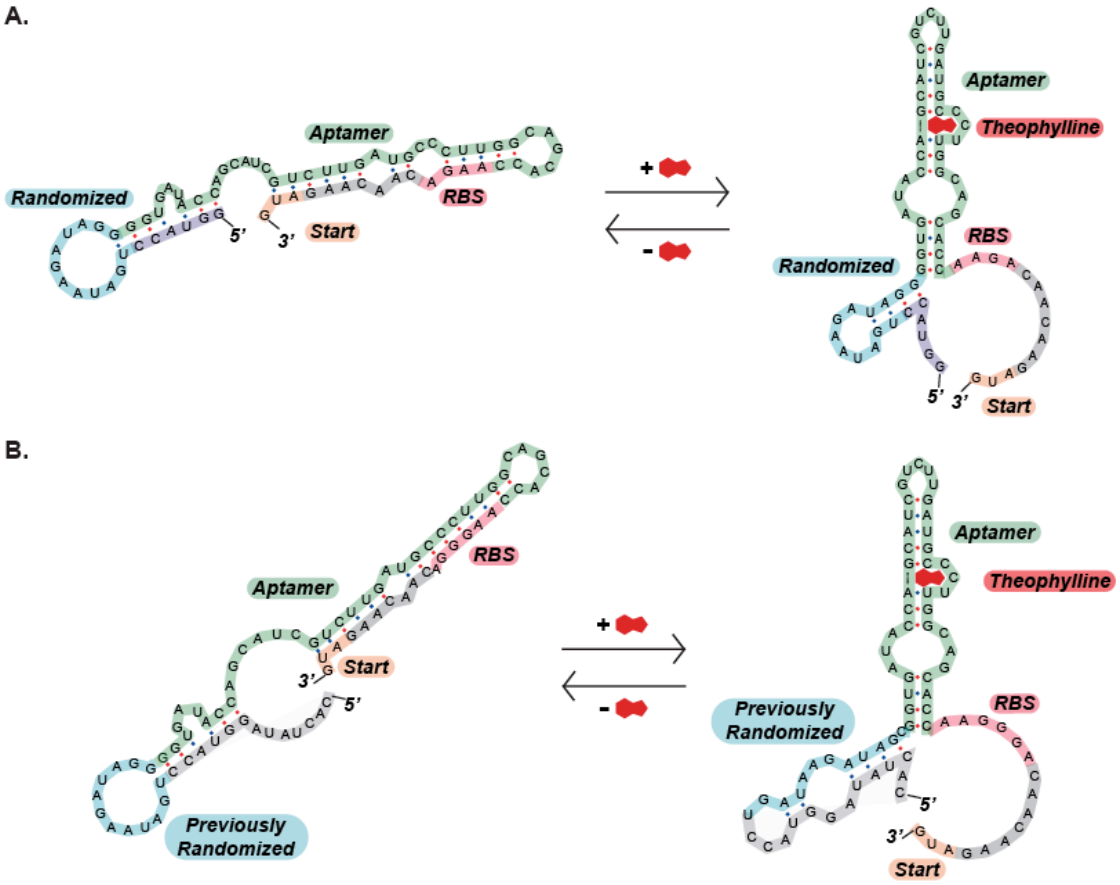

Figure 4.2: Predicted Mechanism of Action for Riboswitches that Remain Biased Toward Activation. (A) Predicted secondary structures for a clone selected from the library show that the RBS (pink) is paired in the absence of ligand (left) and unpaired in the presence of theophylline (right). These models suggest that the "CAACAAG" sequence (gray) positioned between the RBS and start codon (orange) may provide an unintended bias toward sequences that are repressed in the absence of theophylline. (B) Predicted secondary structures for a semi-rationally designed riboswitch that may activate gene expression up to 140-fold when cells are grown in the presence of theophylline.

4.2.2 A Redesigned Library Positions the RBS within the Aptamer Stem

With the realization that a library featuring the RBS immediately after the aptamer could retain a potential bias for switches that activate expression upon addition of ligand, we decided to modify our library design in such a way that the RBS would necessarily be paired in the presence of ligand. To accomplish this goal, the RBS was incorporated within the aptamer structure in a region that is expected to be paired when theophylline is bound. Specifically, we observed that swapping two pairs of bases within the stem of the mTCT8-4 aptamer would position a potential RBS within the 3′ stem of the aptamer (Figure 4.3). Because the structure and function of the aptamer have been well studied, it was apparent that the base pair swap in the aptamer stem would not perturb ligand binding significantly.[16] Lacking full appreciation for the potential role of the "CAACAAG" constant region at this stage, we did not remove this sequence. As we will see, this constant sequence did not thwart our ability to identify genetic repressors.

In addition to incorporating an RBS sequence within the aptamer stem, we positioned an 8-nucleotide randomized region immediately 5′ to the aptamer (Figure 4.3). We anticipated that in the absence of theophylline, some of the library members would have sequences in which the nucleotides in the randomized region could pair with the 5′ portion of the aptamer, causing the aptamer stem to become unpaired, and therefore making the RBS more accessible for translation. We hypothesized that these same sequences might exhibit significantly lower gene expression in the presence of ligand, since the RBS would become less accessible in the theophylline-bound conformation. To test this hypothesis, we used a simple screen to determine if the redesigned library contained riboswitches that repress gene expression in the presence of theophylline.

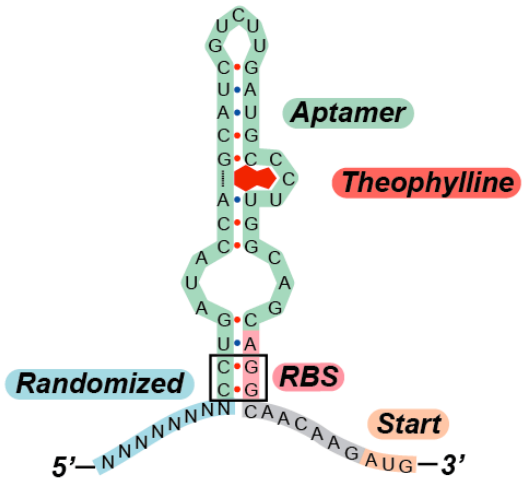

Figure 4.3: A Combinatorial Library with an RBS in the Aptamer Stem. A modified library contains a randomized region of eight bases (blue) positioned immediately 5′ to the aptamer (green). Two pairs of bases were swapped (boxed) in the aptamer stem to position the RBS (pink) within the stem.

To select for sequences with high gene expression in the absence of ligand, we cloned the library into the 5′-untranslated region of the *cat* reporter gene, which confers resistance to the antibiotic chloramphenicol, and selected clones that grew without theophylline on media containing the antibiotic. Plasmid DNA was extracted from the pool of cells that survived the chloramphenicol selection, and the region containing the randomized bases and aptamer was subcloned in the 5′-untranslated region of the *lacZ* reporter gene. To screen for clones with low gene expression in the presence of ligand, *E. coli* were transformed with this pool and plated on selective media containing X-gal and 1 mM theophylline. 94 light blue colonies were picked and grown in liquid culture to compare gene expression in the presence and absence of theophylline. These clones no longer exhibited a bias toward theophylline-dependent gene activation, and several repressed gene expression approximately 5-fold when grown in 1 mM theophylline.

Plasmid DNA was extracted from these clones for sequencing analysis, which revealed that each of the riboswitches had a different N8 sequence. To our surprise, all

three had an in-frame ATG sequence at the same position within the N8 region. It appeared that the coincidental presence of an AGG sequence several bases 5′ to the N8 library may act as an RBS to provide an unintended opportunity for start codons to be selected in the randomized region. Thus, each of the theophylline-dependent repressors has two potential translational start sites, with one positioned 5′ to the aptamer and the other positioned 3′ to the aptamer (Figure 4.4).

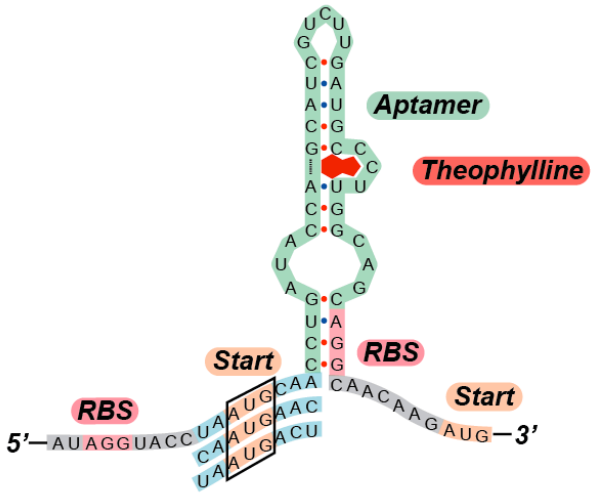

Figure 4.4: Three Clones with Potential Start Codons 5′ to Aptamer. Three 5-fold repressors were identified in the screen. The unintentional RBS (pink) is positioned 5′ to the selected sequences (blue). The old and new (boxed) start codons are shown in orange.

4.2.3 Removal of the ATG codon 3′ to the Aptamer Improves Gene Repression

Because three unique clones contained an in-frame ATG sequence 5′ to the aptamer, we hypothesized that the initiation codon 3′ to the aptamer may not play a role in repression, and that removing it may improve the performance of the synthetic riboswitches by decreasing background gene expression in the ligand-bound state. To test this idea using the three switches, we replaced the start codon following the aptamer with a 32-member NNY library, where Y (pyrimidine) was used to exclude purines from the third position to eliminate the possibility of start and stop codons in the sequence

(Figure 4.5A). In parallel, we tested whether simultaneous removal of the potentially biasing "CAACAAG" constant region may improve ligand-dependent repression (Figure 4.5B). These libraries were screened independently for improved repression.

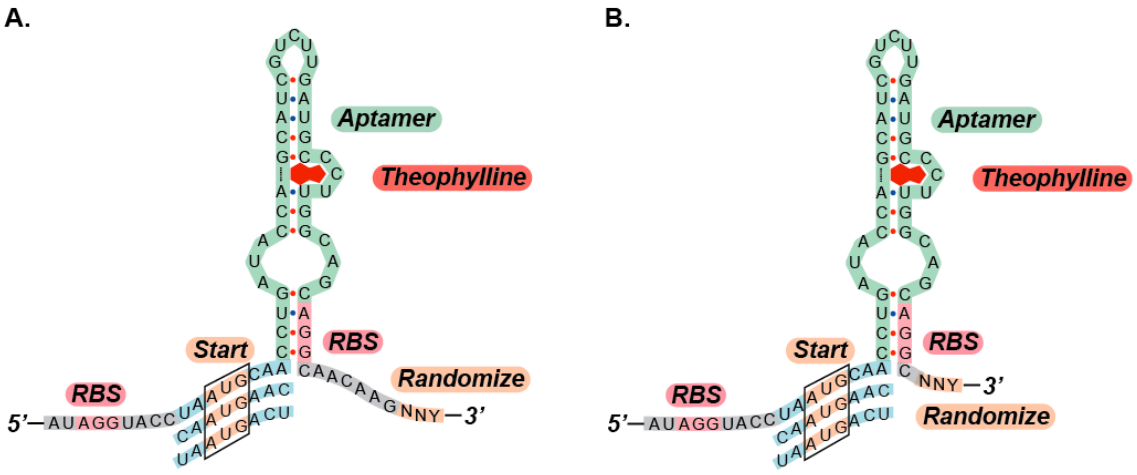

Figure 4.5: Genetic Libraries to Remove the Original Start Codon. (A) For each 5-fold repressor, the AUG after the aptamer was mutated to 'NNY' to prohibit start and stop codons. (B) Six bases of the 'CAACAAG' constant region were deleted, while the 'NNY' library was simultaneously introduced.

We assayed 23 random clones from the libraries shown in Figure 4.5A and 95 random clones from the libraries shown in Figure 4.5B. The data obtained through these screens demonstrated that the activation ratios of the riboswitches could be improved by mutating the ATG codon located 3′ to the aptamer (Table 4.1). Generally, the level of gene expression in the presence of theophylline was reduced, while the gene expression in the absence of theophylline remained relatively unchanged, indicating that the in-frame ATG before the aptamer is likely the site of translation initiation for these riboswitches.

For some clones, however, we observed that gene expression remained unchanged when cells were grown in the presence of theophylline. We sequenced several of these clones in an effort to understand why some of these library members failed to repress gene expression. The sequencing data revealed that clones that always exhibited high

gene expression often had mutations that are known to compromise the ability of the aptamer to bind theophylline (Table 4.1).

Name	5' Sequence	Aptamer Sequence	3' Sequence	Switch?
4450	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------TTC	Improved
4451	CAATGAAC	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------CGC	Improved
4452	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------TAT	Improved
4453	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------CAT	Improved
4454	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------TAC	Improved
4455	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	CAACAAGGAC	Improved
4456	TAATGCAA	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	CAACAAGATT	Improved
4457	TAATGCAA	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	CAACAAGTAT	Improved
4459	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	CAACAAGTTT	Improved
4568	TAATGACT	CCTGATACTAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------TAC	Always on
4569	TAATGCAA	CGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	CAACAAGAGC	Always on
4573	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGG	C------CCT	Always off
4575	TAATGACT	CCTGATACCAGCATCGTCTTGATGCCTTTGGCAGCAGG	CAACAAGTTT	Always on

Table 4.1: Clones Derived from Parental 5-Fold Repressors. Each 5′ region features the sequence from one of the three parental repressors and has an in-frame start codon (green). Mutations within the aptamer sequence are indicated in blue. The rightmost column indicates whether the clone functions as a repressor.

Notably, clones 4454 and 4568 have identical sequences in the regions 5′ and 3′ to the aptamer sequence, while clone 4568 has a mutation that destabilizes the aptamer (C9T). Thus, while clone 4454 represses gene expression better than its 5-fold parent switch, clone 4568 exhibits very high gene expression even in the presence of ligand. Similarly, clones 4569 and 4575 also fail to repress gene expression in the presence of ligand due to mutations that destabilize the aptamer (4569) or prevent theophylline from binding (4575). Clone 4573 represents a less common case, where expression is low regardless of whether theophylline is present. A comparison between the predicted secondary structures for this sequence versus those of functional repressors suggests that the identity of the triplet replacing the 3′ start codon may influence the accessibility of the region 5′ to the aptamer, particularly in the absence of theophylline.

The most significant improvement arising from one of the parent 5-fold repressors was for a specific sequence (Clone 4457) that enabled 27-fold repression of gene expression in *E. coli* (Figure 4.6). Half-maximal repression is observed at an extracellular theophylline concentration of approximately 300 μM. Previous work suggests that the corresponding intracellular concentration of theophylline can be estimated to be 210 nM,[17] which falls within the margin of error for the reported K_D of the aptamer (310 ± 130 nM) as determined by in vitro studies.[15]

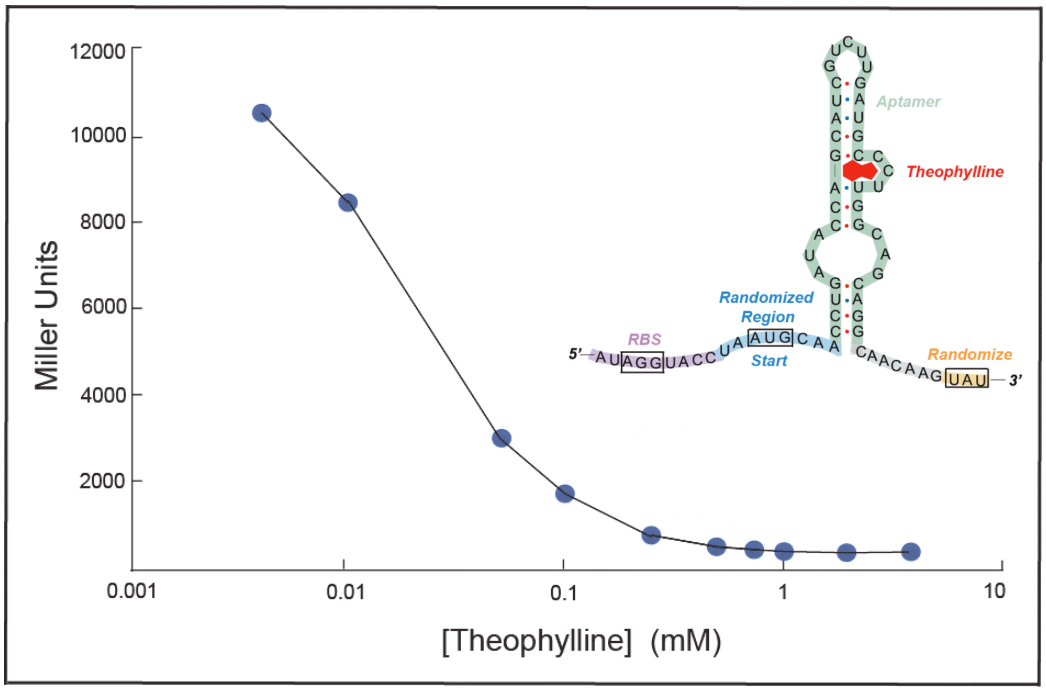

Figure 4.6: Dose Response Profile of a 27-Fold Repressor. Cells harboring the riboswitch (inset) were grown in media containing varying concentrations of theophylline. Cells grown in the absence of theophylline (0 mM) expressed very high levels of β-galactosidase, and grew very slowly. Half-maximal repression was observed at approximately 300 μM. Error bars are smaller than the size of the symbols.

To verify the theophylline-dependence of the new repressor, we assayed its performance in *E. coli* cells that were grown at various concentrations of caffeine, which is structurally similar to theophylline but binds the aptamer 10,000-fold less tightly.[15] As

expected, gene expression remained unrepressed for all caffeine concentrations tested (0 mM – 2 mM). To further test the ligand-dependence of the riboswitch, we introduced a previously verified point mutation (C27A) to the aptamer that significantly reduces its affinity for theophylline but permits binding of 3-methylxanthine.[3,6,7] The mutated repressor exhibited high levels of gene expression when cells were grown in the absence of ligand or in the presence of 1 mM theophylline, while gene expression was reduced to one tenth of the unrepressed levels when cells were grown in the presence of 1 mM 3-methylxanthine (11,000; 11,000; 910 Miller Units, respectively). These results demonstrate that the repressor modulates gene expression through specific interactions between theophylline and the RNA.

4.2.4 The Presence and Position of the 5′ ATG Is Critical for Riboswitch Function

To determine whether the ATG codon 5′ to the aptamer acts as the initiation codon, we mutated the sequence to NTG for our best repressor and assayed each of the four possible clones for gene expression when cells were grown in 0 mM and 1 mM theophylline. The data show that the switch is nonfunctional for all codons other than ATG, as gene expression in the absence of ligand is abolished to below the level of the repressed parent switch (Figure 4.7). This result provided support toward the conclusion that the ATG 5′ to the aptamer is the source of translation initiation, and hinted that our initial design was superseded by the plasticity of biological systems presented with genetic variation and exposed to selective pressures.

To further verify that the start codon 5′ to the aptamer is the site of translation initiation for our best repressor, we introduced a different point mutation to convert the

triplet immediately following the ATG to a stop codon. We suspected that gene expression levels for this construct, both in the presence and absence of theophylline, would be similar to the expression levels of the mutants lacking a start codon 5′ to the aptamer. Indeed, this point mutant exhibited low expression levels regardless of whether theophylline was present in the growth media, providing additional support for our hypothesis (Figure 4.7).

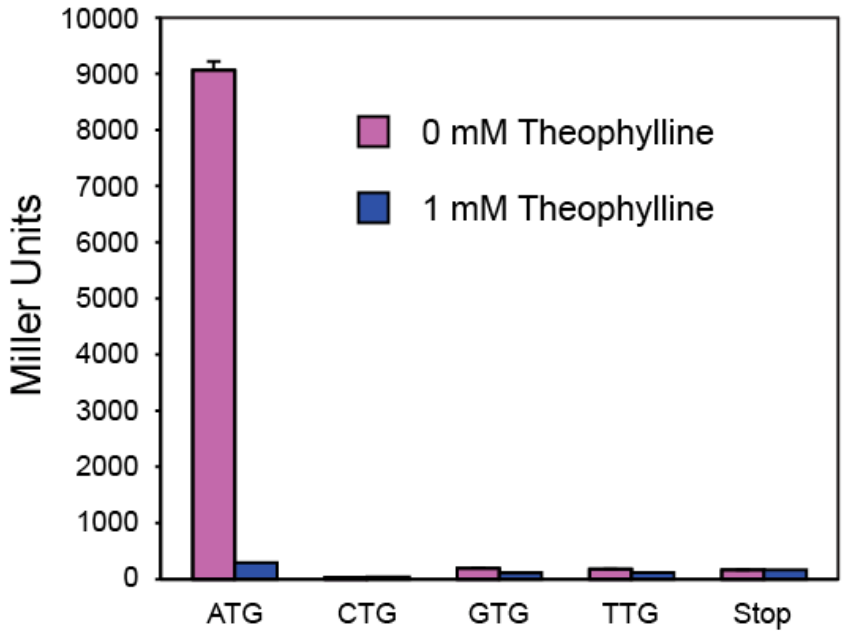

Figure 4.7: Mutational Studies Reveal that the 5′ Start Codon Is Critical. A single point mutation to remove the start codon 5′ to the aptamer abolishes gene expression to below the level of the repressed parent switch. Additionally, a different point mutant, for which the codon immediately following the ATG was mutated to a stop codon, exhibits low expression levels regardless of whether theophylline is present.

As shown in Figure 4.4, each of the 5-fold repressors that we initially identified featured an in-frame start codon positioned three bases from the beginning of the aptamer. Because the library design theoretically permitted an in-frame start codon to be positioned immediately before the aptamer, we wondered whether the specific position of

the start codon might be important for optimal repression. To test whether the riboswitch would function equally well if the start codon were positioned immediately before the aptamer, we removed the "CAA" codon immediately preceding the aptamer for the 27-fold switch shown in the inset of Figure 4.6 and assayed for β-galactosidase activity in the presence and absence of ligand (Figure 4.8A).

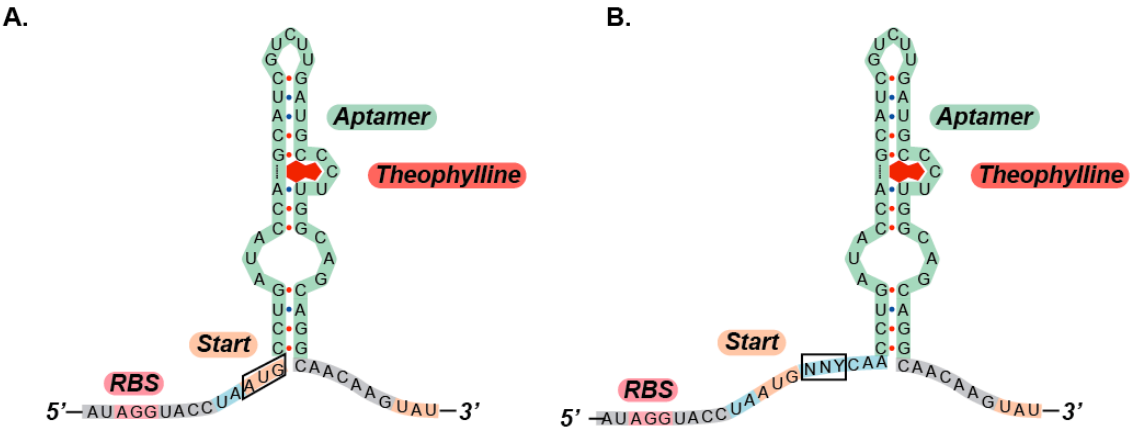

Figure 4.8: Sequence of Constructs to Test the Role of Start Codon Position. (A) Deletion of the "CAA" codon enabled the start codon (orange, boxed) to be positioned immediately 5' to the aptamer stem (green). This sequence provided ~200 MU in the presence and absence of theophylline. (B) Addition of an "NNY" library (boxed) between the start codon (orange) and "CAA" codon enabled us to position the start codon farther from the aptamer stem (green). Clones from this library repressed gene expression approximately 6-fold in the presence of 1 mM theophylline.

Gene expression remains low even in the absence of theophylline (~200 Miller Units, with 0 mM or 1 mM ligand), suggesting that the position of the start codon with respect to the aptamer stem is important. This result suggests that our "off" switches may function by sterically obstructing translation initiation when the aptamer binds theophylline. However, further studies would need to be performed to evaluate the contributions of other possible mechanisms.

To gain further insight into the importance of start codon position relative to the

aptamer, we were interested in determining how the riboswitch might function if the start codon were moved one position farther from the aptamer stem. Therefore, we designed and tested a small library of constructs in which an "NNY" codon was inserted after the start codon of our best repressor, as shown in Figure 4.8B. The ability of these modified switches to repress gene expression was diminished from 27-fold to approximately 6-fold (13,600 MU without ligand to 2,200 MU with 1 mM theophylline).

In the experiments described above, the position of the start codon relative to the aptamer stem was varied, while the distance between the RBS and start codon was kept constant. Consequently, the distance between the RBS and the 5′ end of the aptamer was varied simultaneously in these experiments. Because the positions of the start codon and RBS were simultaneously varied relative to the aptamer, the mechanism of repression is unclear. For example, it is possible that theophylline induces a conformational change that prevents the 16S rRNA from binding at the Shine-Dalgarno sequence (referred to as the RBS). An alternative hypothesis postulates that the 16S rRNA is able to bind the Shine-Dalgarno sequence in the presence of theophylline, but the ribosome is unable to bind completely due to steric obstruction by the ligand-bound aptamer structure. Finally, we might consider the option that translation initiates at the start codon and proceeds in the absence of ligand, but is thwarted by the highly structured ligand-bound aptamer.

Notably, our initial library enforced a constant distance between the RBS and aptamer stem, yet failed to reveal repressors in which the start codon was positioned immediately 5′ to the aptamer. Because the spacing between the RBS and the start codon for either case would fall within the ideal range for *E. coli*, we may infer that significant small-molecule dependent repression may not occur if the start codon is positioned

immediately 5′ to the aptamer. These results provide support for the idea that close proximity of the start codon to the aptamer may prevent translation initiation in the highly structured, ligand-bound state.

4.2.5 Mass Spectrometry Shows that the Riboswitch Is Translated

Although the loss of function data for various mutants provided strong genetic support that translation initiates 5′ to the aptamer, we chose to use mass spectrometry methods to obtain direct evidence that our best "off" riboswitch was located within the translated region. We cloned the 27-fold repressor upstream of a gene encoding the green fluorescent protein (GFP) with a C-terminal polyhistidine tag. To facilitate modular cloning of our riboswitches, each reporter gene is preceded by and translationally fused to a short (~70 amino acids), nonfunctional portion of the IS10 transposase gene.[18] Because this portion of the IS10 transposase protein contains a known post-translational cleavage site,[18] we removed the cleavage site to obtain the full length protein for mass spectrometry analysis.

As expected, when cells expressing this construct were grown with and without theophylline, only the cells grown in the absence of ligand were fluorescent. To determine whether the identity of the expressed protein corresponded to the amino acid sequence that would be anticipated if translation were initiated at the ATG before the aptamer, we used Ni^{2+}-affinity chromatography to purify the histidine-tagged fluorescent protein and performed an in-gel trypsin digest (Figure 4.9),[19] followed by tandem mass spectrometry of the resulting peptides (LC-MS/MS).[20] The expected N-terminal peptide was identified through this analysis (Table 4.2), confirming that the riboswitch is translated in the unrepressed state of the switch as a fusion to the reporter gene. This

result verified the previous conclusions that the riboswitch is part of the coding region rather than within the 5′ UTR.

While the presence of a post-translational cleavage site between the aptamer and the reporter gene for our repressors was an artifact of previous work, perhaps it would be useful to engineer such a site immediately before the gene of interest for future efforts to create novel, small-molecule dependent repressors. Taking such a step might expand the potential number of genes that could be modulated by this mechanism beyond those that tolerate translational fusions. Moreover, the amino acid sequence encoded by a given aptamer may not always produce a stable, folded protein. Therefore, encoding for post-translational removal of the leading peptide by introducing a protease cleavage site may provide enhanced stability to the protein of interest.

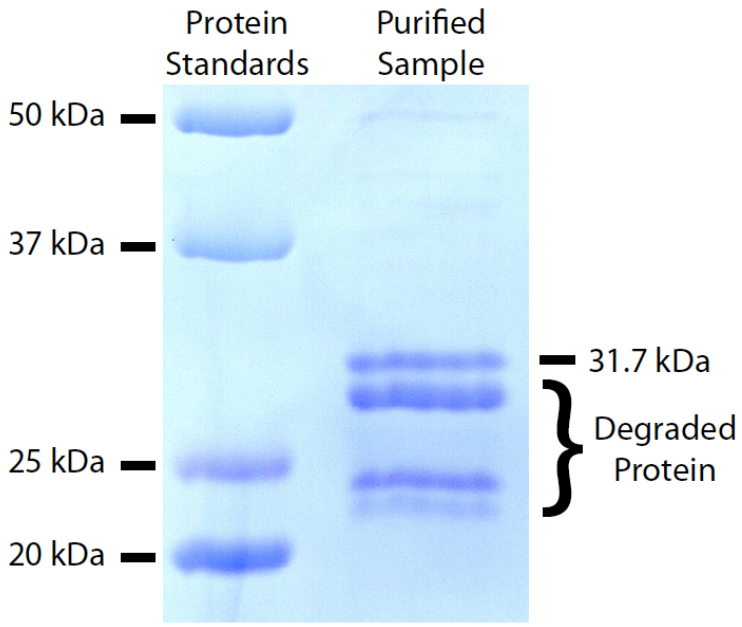

Figure 4.9: Purified Protein Is the Expected Size. 10% SDS-PAGE analysis reveals the expected band at 31.7 kDa. To avoid contamination, the band was excised from a duplicate gel, trypsinized in-gel, and analyzed by LC-MS/MS. The lower bands are likely to be degradation products of the full-length protein.

Amino Acid Sequence of Full Length, Purified Protein:

MQPDTSIVLM PLAAGNKYCE LDILHDSLYK LGGIDPAGSK GEELFTGVVP ILVELDGDVN GHKFSVSGEG EGDATYGKLT
LKFICTTGKL PVPWPTLVTT FSYGVQCFSR YPDHMKRHDF FKSAMPEGYV QERTISFKDD GNYKTRAEVK FEGDTLVNRI
ELKGIDFKED GNILGHKLEY NYNSHNVYIT ADKQKNGIKA NFKIRHNIED GSVQLADHYQ QNTPIGDGPV LLPDNHYLST
QSALSKDPNE KRDHMVLLEF VTAAGITHGM DELYKGAHHH HHH

Sequence of identified tryptic peptide	Position
R.AEVKFEGDTLVNR.I	147-159
R.DHM*VLLEFVTAAGITHGM*DELYK.G	253-275
R.DHM*VLLEFVTAAGITHGMDELYK.G	253-275
K.EDGNILGHK.L	169-177
K.FEGDTLVNR.I	151-159
K.FSVSGEGEGDATYGK.L	64-78
K.GAHHHHHH.-	276-283
K.GEELFTGVVPILVELDGDVNGHK.F	41-63
K.GIDFKEDGNILGHK.L	164-177
R.HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK.D	206-246
K.LEYNYNSHNVYITADK.Q	178-193
K.LEYNYNSHNVYITADKQK.N	178-195
K.LGGIDPAGSK.G	31-40
-.M*QPDTSIVLM*PLAAGNK.Y	1-17
-.M*QPDTSIVLMPLAAGNK.Y	1-17
-.MQPDTSIVLM*PLAAGNK.Y	1-17
-.MQPDTSIVLMPLAAGNK.Y	1-17
K.SAM*PEGYVQER.T	123-133
K.SAMPEGYVQER.T	123-133
R.TISFKDDGNYK.T	134-144
K.YC#ELDILHDSLYK.L	18-30
K.YCELDILHDSLYK.L	18-30

Table 4.2: Tryptic Fragments Identified by LC-MS/MS. The anticipated full-length sequence of the purified protein is shown (purple regions indicate that the peptide fragment was identified be mass spectrometry). Tryptic fragments identified by LC-MS/MS analysis are listed, along with the position of each peptide within the full-length protein. The N-terminal fragments (yellow) confirm that translation initiation occurs at the ATG preceding the aptamer.

4.2.6 RNA Structure Probing Studies of the 27-Fold Repressor

We demonstrated in Section 4.2.4 that the presence and optimal positioning of the 5′ start codon are important factors that can affect the performance of our RNA-based repressor. The data suggested that these repressors may sterically hinder translation initiation when the aptamer binds theophylline, while providing access for translation initiation in the ligand-free state. We decided to test this hypothesis using a well established, RNA structure probing method known as SHAPE (**S**elective 2′-**H**ydroxyl **A**cylation analyzed by **P**rimer **E**xtension).[21]

This in vitro technique can provide structural information about an RNA transcript by probing the local accessibility of each nucleotide using an electrophile such as N-methylisatoic anhydride (NMIA).[21] If a nucleotide is unconstrained, its 2′-hydroxyl group is able to adopt a more accessible conformation, thus unveiling the hydroxyl as a nucleophile for the NMIA reagent (Figure 4.10). In contrast, nucleotides that are constrained due to secondary or tertiary interactions are less likely to access the necessary conformation for reaction with NMIA.

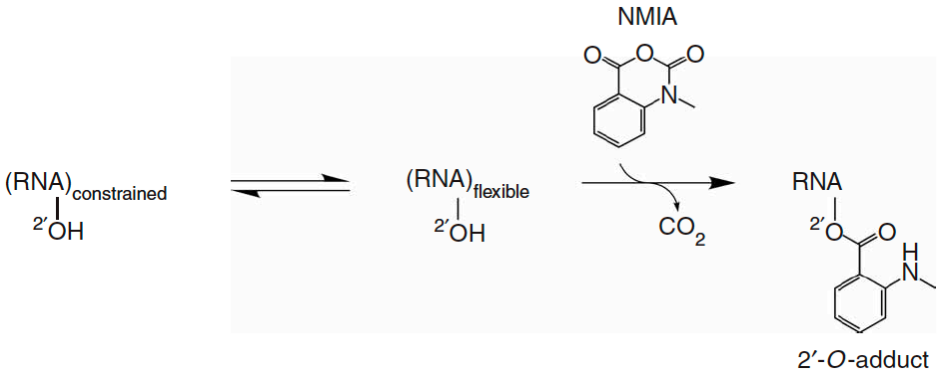

Figure 4.10: RNA Structure Can Be Probed with NMIA. Constrained nucleotides do not react with NMIA, while flexible nucleotides react with NMIA to form stable 2′-*O*-adducts.[21] These adducts prevent reverse transcription from proceeding past nucleotides that have reacted with NMIA.
(Reprinted by permission from Macmillan Publishers Ltd: Nature Protocols, KA Wilkinson, EJ Merino, and KM Weeks, 1(3): 1610-1616, copyright 2006.)

After partial modification with NMIA, the RNA is reverse transcribed using a 5′-radiolabeled DNA primer. Reverse transcription cannot proceed beyond nucleotides that have reacted with NMIA to form 2′-*O*-adducts.[21] Therefore, the accessibility of each nucleotide can be quantified by analyzing the size and intensity of the cDNA bands after separation by denaturing PAGE. We performed SHAPE analysis to investigate whether theophylline-induced changes in RNA structure may contribute to the mechanism by which our 27-fold repressor modulates gene expression.

Using our best repressor as a template, we generated a PCR product that incorporated a T7 promoter to drive transcription and a 3′ primer binding site that could be used for reverse transcription. This template was used to produce a 133-nucleotide RNA transcript, beginning with the RBS sequence at the 5′ end and extending to approximately 65 nucleotides after the aptamer. Gel purified transcripts were reacted with NMIA in the presence or absence theophylline by the method of Weeks et al.[21] Reverse transcription with 5′-radiolabeled primers provided cDNA of various lengths, which were separated by 8% denaturing PAGE. We also performed sequencing reactions, which enabled each cDNA band to be attributed to NMIA modification of a specific nucleotide. Figure 4.11 shows the PAGE analysis, as well as the specific regions within the riboswitch sequence that exhibit theophylline-dependent structural changes.

As anticipated, the portion of the aptamer comprising the theophylline-binding pocket (region 5) becomes significantly less accessible in the presence of ligand.[15] Small decreases in accessibility are also seen in regions 3 and 6, indicating that these nucleotides become more structured in the theophylline-bound state. We also observed significant changes in the accessibility of nucleotides within region 2, at the 5′ end of the

aptamer. This region shows distinct modification in the absence of theophylline, but nearly no modification in the presence of theophylline. Because the region surrounding the start codon (1) appears to be unstructured regardless of whether theophylline is present, but the nearby 5′ portion of the aptamer is less structured in the absence of theophylline, it is possible that the theophylline-stabilized aptamer conformation may sterically obstruct ribosomal binding.

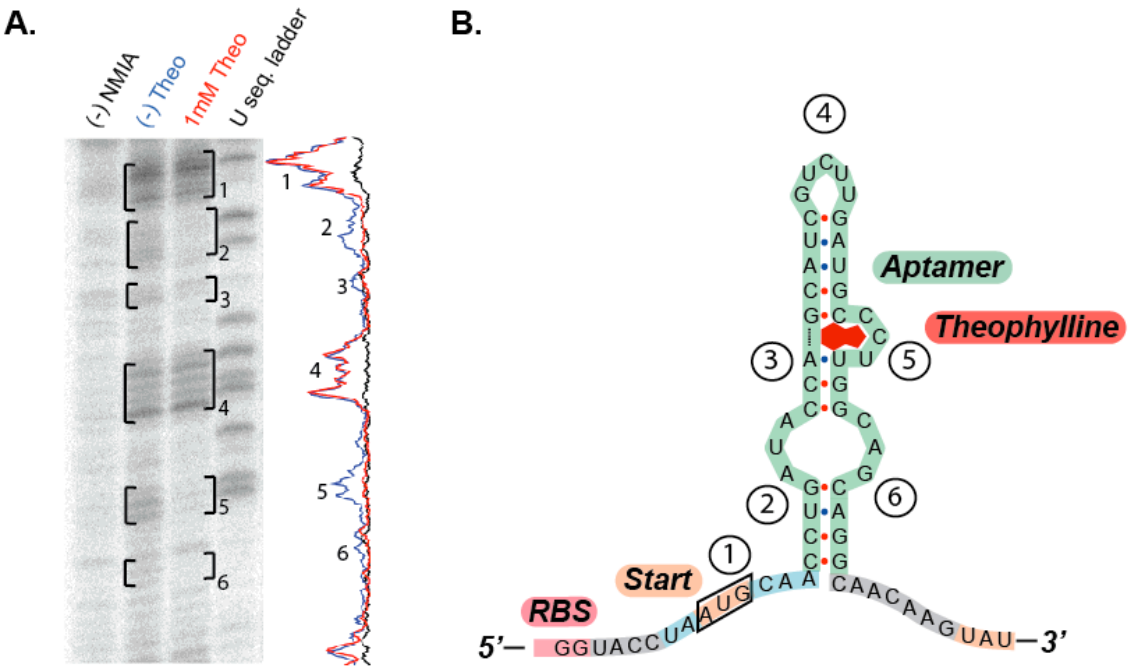

Figure 4.11: SHAPE Analysis Reveals Theophylline-Induced Structural Changes. (A) The band intensities of samples from SHAPE analysis (8% denaturing PAGE) suggest that several regions remain highly unstructured regardless of whether theophylline is present (1 and 4). Two regions are substantially more accessible in the absence of theophylline (2 and 5), while two regions are slightly more accessible in the absence of theophylline (3 and 6). (B) Sequence of the 27-fold repressor. Regions 2, 3, 5, and 6 become less accessible in the presence of theophylline. Regions 1 and 4 remain accessible under both conditions. The data suggest that the aptamer structure may form even in the absence of theophylline, but that the 5′ portion of the aptamer is less structured in the absence of theophylline.

Before moving forward, we must pause to note the limitations of these structure-probing experiments. Firstly, these studies are accompanied by the inherent qualification

that an experiment performed in vitro may not fully represent the complexities of the biological processes within a living cell. For example, in vitro SHAPE experiments bear important differences such as substantially slower mRNA degradation rates, a lack of RNA-associated proteins, and an inability to probe kinetically favored structures because the RNA is folded after in vitro transcription is complete. These factors affect the performance of our riboswitches during in vivo gene expression, yet may not be reflected in the context of an in vitro structure-probing study.[22]

Another factor that must be considered while interpreting our data is that NMIA is similar in structure to theophylline (Figure 4.12). Although the aptamer displays such high specificity that it binds theophylline 10,000 times more tightly than it binds caffeine,[15] the concentration of NMIA used in this experiment (13 mM) is five orders of magnitude greater than the aptamer's K_D for theophylline (~100 nM). If NMIA is able to bind and stabilize the aptamer conformation of the RNA transcript for our 27-fold repressor before the reagent is hydrolyzed, it might appear that little conformational switching occurs in the presence of theophylline.

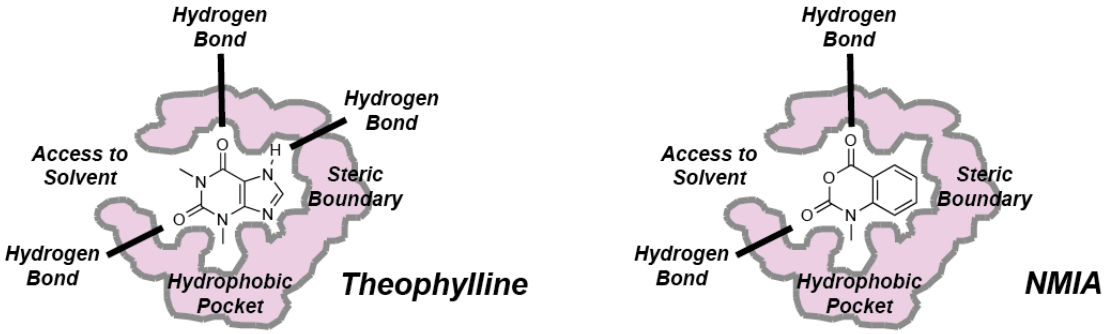

Figure 4.12: Aptamer Binding Pocket Comparison. Theophylline binds it aptamer with high affinity and specificity as a consequence of hydrogen bonding interactions, stacking interactions, and lack of steric repulsion within its binding pocket (left).[15] The NMIA reagent is similar in structure to theophylline, and may bind the aptamer at the high concentrations that are used for in vitro SHAPE studies (right). This suggestion is not based upon specific data, but is merely an observation to be considered.

However, because the RNA is folded in the absence or presence of theophylline before NMIA is added, the reagent may not shift the equilibrium significantly. Since the half-life of NMIA is short (~8 min),[21] and we have no concrete evidence to support the notion that NMIA may act as a ligand for the theophylline aptamer during SHAPE studies, we will remain aware of this possibility while proceeding under the assumption that NMIA does not stabilize the aptamer conformation.

Another potential complication arises because the ability to obtain satisfactory separation by denaturing PAGE generally requires that the transcript tested in a structure-probing experiment is a shortened version of the full-length transcript that is produced in vivo. This problem could be overcome by performing several reverse transcription reactions with primer binding sites at various locations along the transcript of interest. However, our particular experiment is further complicated by the presence of a T7 promoter within what would be the 5′-UTR under in vivo conditions. This T7 promoter sequence was an artifact of a previous experiment. Nonetheless, the consequence is that transcription with T7 RNA polymerase for in vitro structure probing studies results in the removal of a potentially important portion of the 5′ UTR.

Specifically, the position of the T7 promoter causes transcription to start within the RBS sequence, producing an in vitro transcript that is 45 bp shorter than that produced from the *tac* promoter in vivo (Figure 4.13). To test the whether SHAPE experiments using this shortened transcript provide substantially different results than those that would be observed using a more complete transcript, we introduced inactivating mutations to the inadvertent T7 promoter and performed SHAPE studies on a transcript that includes a longer portion of the 5′ UTR sequence. Although this version

also presents a potentially significant deviation from the in vivo transcript, the data are similar to those presented here, enabling us to have greater confidence in our conclusions.

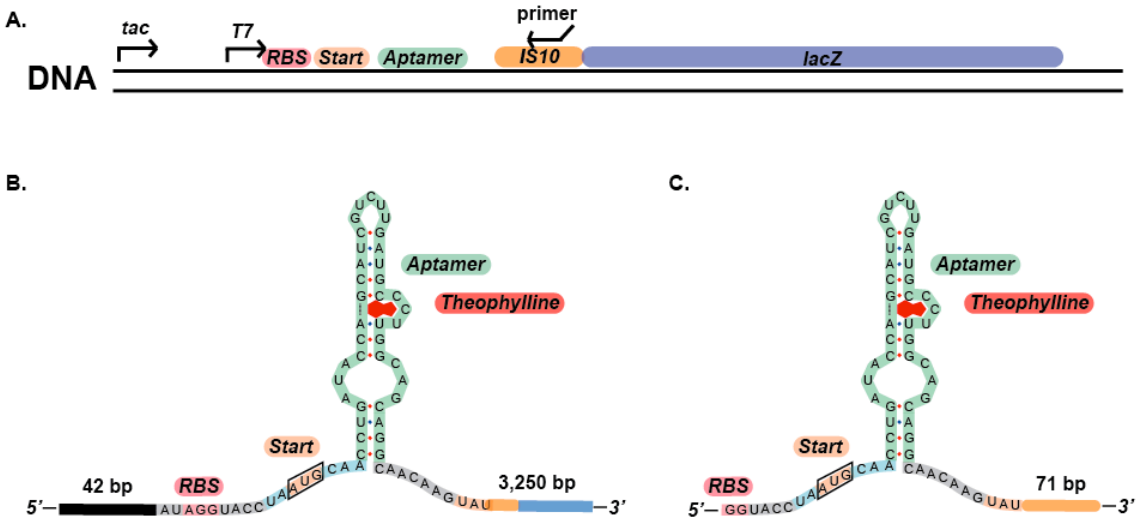

Figure 4.13: Comparison of Transcripts for Structure Probing. (A) The *tac* promoter is used by *E. coli* to initiate transcription, while a T7 promoter initiates in vitro transcription at the RBS sequence. A primer that binds within the *IS10* linker region is used to introduce a primer-binding site for reverse transcription. (B) The transcript produced in vivo extends from 43 nucleotides 5′ to the RBS though the region encoding the *IS10-lacZ* fusion (~3,250 bp). (C) The transcript produced in vitro begins within the RBS and extends through the aptamer to include 71 bases of the *IS10* sequence.

With the aforementioned caveats in mind, the data suggest that the aptamer structure forms even in the absence of ligand under these in vitro experimental conditions. Specifically, the lack of structural changes in region 4 provides evidence that the loop at the tip of the aptamer forms regardless of whether theophylline is present (Figure 4.11). This observation is consistent with the previously reported solution structure of the apo-aptamer, which showed that this loop, as well as the upper and lower stems, forms in the absence of theophylline.[23]

Other regions of the transcript exhibit distinct, theophylline-dependent changes in nucleotide accessibility. These changes generally involve decreased accessibility of

specific internal residues, which correlate with structural changes that have been observed when the aptamer binds theophylline.[23] Of substantial relevance, the presence of theophylline leads to significant decreases in the accessibility of nucleotides within the internal loop region proximal to the 5′ stem of the aptamer. The previously reported solution structure of the theophylline-bound aptamer shows that these same nucleotides form specific interactions with theophylline.[23] The data also reveal that, at least for this shortened transcript, the Shine-Dalgarno and start codon regions are highly unstructured, both in the presence and absence of theophylline.

Considered within the context of previous work, our structure-probing data brings us to the conclusion that repression is unlikely to occur by a "switching" mechanism in which the Shine-Dalgarno sequence is occluded in the absence of ligand by specific pairing interactions with other regions of the transcript. Our previously identified "on" riboswitches function by a mechanism in which the most stable conformation in the absence of theophylline provides pairing between the Shine-Dalgarno sequence and the 3′ portion of the aptamer.[8,9,24] In the presence of theophylline, the energy gained from ligand-binding favors the aptamer conformation as the most stable structure. The Shine-Dalgarno sequence is then accessible for translation initiation. Because the ribosomal-binding site (henceforth defined as the entire sequence protected by the ribosome upon translation initiation) does not typically extend much farther 5′ than the Shine-Dalgarno sequence,[25] the ligand-bound aptamer structure does not present a significant obstacle to translation initiation for these "on" switches.

In contrast, our repressor does not appear to undergo a ligand-induced conformational shift that alters the accessibility of the Shine-Dalgarno sequence. Rather,

the data support the hypothesis that repression occurs by a mechanism in which the ribosomal-binding site, which encompasses approximately 15 residues on each side of the start codon,[25-27] becomes significantly more structured in the presence of theophylline. In particular, binding of theophylline causes the 5′ internal loop to assist in the formation of a stable and compact ligand-bound structure, while the pairing interactions between the 5′ and 3′ portions of the stem also become stronger. Because the start codon of our repressor is positioned several bases from the aptamer stem, translation initiation is likely to be restricted in the presence of theophylline because the residues in the 5′ portion of the aptamer become highly structured upon ligand-binding (Figure 4.14).

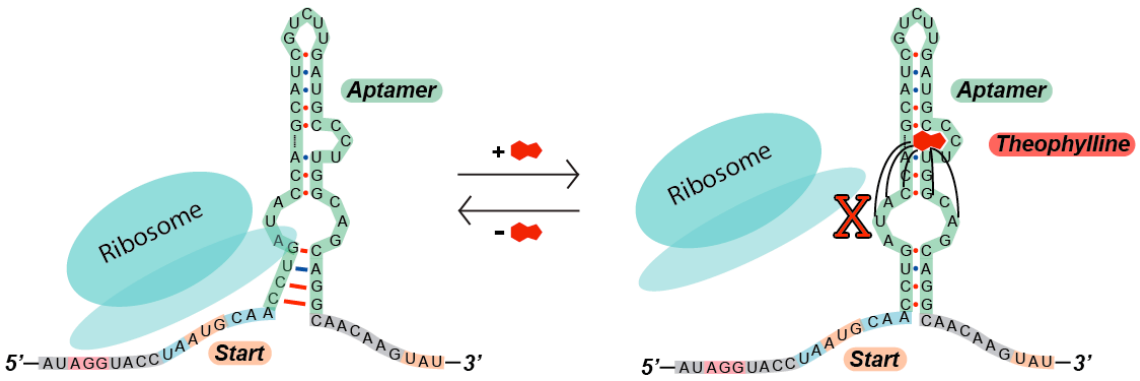

Figure 4.14: Model for Theophylline-Dependent Translational Repression. The data support a model in which the ribosome is able to access the Shine-Dalgarno sequence, as well as the entire ribosomal-binding site (~30 residues), in the absence of ligand (left). In the presence of theophylline, the Shine-Dalgarno sequence (pink) remains accessible to bind the 16S rRNA, but the ribosome is unable to initiate translation because several of the nucleotides that need to be accessible for ribosomal-binding participate in forming the highly compacted, theophylline-bound structure (right).[15,28]

This hypothesis is supported by the data we presented in section 4.2.4, which showed that moving the start codon to the base of the aptamer reduces gene expression to background level, regardless of whether theophylline is present. Since the ribosomal-binding site is centered at the start codon, moving the start codon three residues closer to the aptamer also shifts the ribosomal-binding site so that it includes an additional three

residues of the aptamer sequence. This effect potentially causes the ribosome to be sterically hindered by the apo-aptamer conformation, which features significant structure near the internal stem (Figure 4.11; region 3).

We also showed that the repressor is less able to turn off gene expression when the start codon is moved one position farther from the aptamer stem. By the same logic, this change shifts the ribosomal-binding site so that it does not stretch as far into the aptamer sequence. Because the nucleotides within the internal loop that exhibit significant structural changes upon ligand-binding are less likely to be included within the ribosomal-binding site upon this shift, the addition of ligand is less effective at repressing gene expression for these constructs. We would predict that moving the start codon an additional position farther from the aptamer would further reduce the repression observed upon ligand-binding. Since the ribosomal-binding site would extend even less far into the aptamer sequence, the entirety of the ribosomal-binding site would be relatively accessible both in the absence and presence of theophylline, permitting translation initiation in either case.

4.2.7 Repressor Switches Complement Previously Identified Riboswitches

With several ligand-dependent repressors on hand, we were interested in determining whether these genetic control elements could be used to complement the phenotypes previously conferred upon *E. coli* cells using "on" riboswitches. Specifically, we wanted to test whether we could apply the lessons learned in Chapters 2 and 3 to identify a repressor that might function to reprogram cell motility. As we considered the many repressors listed in Table 4.1, we recognized that the β-galactosidase activity of the

repressors, measured in the presence and absence of theophylline, should enable us to predict how far *E. coli* cells harboring these switches would migrate on semi-solid agar.

Based upon the correlation between β-galactosidase activity and migration distance presented in Figure 3.9, we determined that Clone 4454 might permit us to reprogram *E. coli* to migrate only when the concentration of theophylline is low. Cells harboring Clone 4454 exhibit decreases in β-galactosidase activity from 600 Miller Units to ~ 50 Miller Units when grown with 1 mM theophylline. We cloned this riboswitch in the 5′ UTR of *cheZ*, and the results confirmed the prediction that the cells should migrate ~16 mm in the absence of theophylline but remain non-motile with 1 mM theophylline.

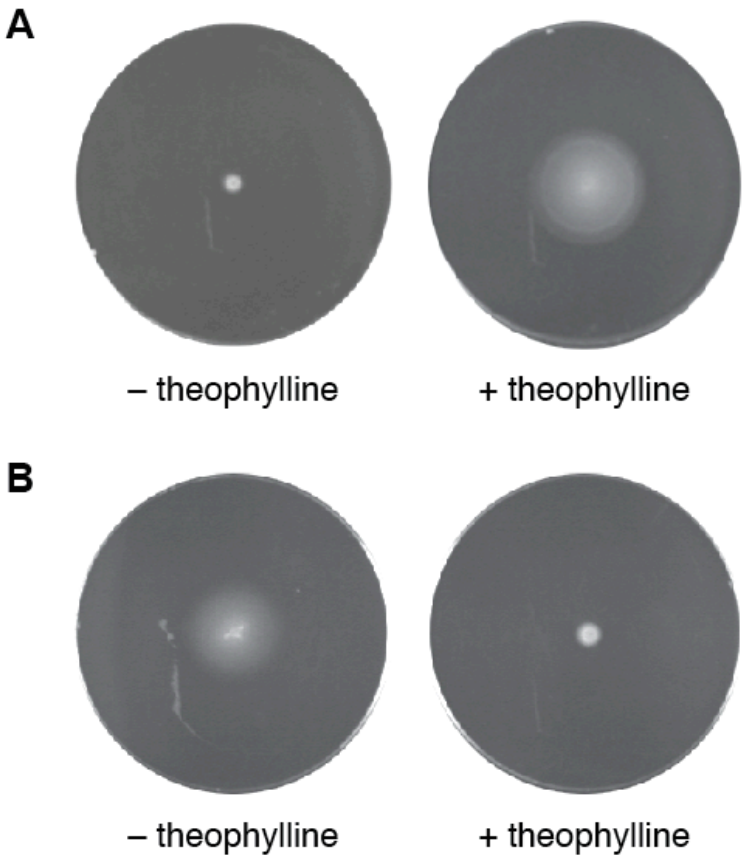

Figure 4.15: Repressor Switches Can Be Used to Reprogram *E. coli* Cells. (A) An activating riboswitch, described in Chapter 2, was used to induce cell motility in the presence of theophylline. (B) A repressor (Clone 4454, Table 4.2) can be used to reprogram cells to migrate only when the theophylline concentration is low. This result shows that these repressors can be used to complement activator switches.

4.3 Conclusion

We began this study to investigate how combinatorial libraries that were biased toward activating gene expression could be redesigned in order to identify riboswitches that repress gene expression in a ligand-dependent fashion. Through this effort, genetic selection revealed an unexpected, yet simple mechanism by which small-molecule ligands can repress gene expression in the context of synthetic riboswitches. This result serves as a reminder that biological systems often resort to unanticipated mechanisms when selective pressures are applied. Moreover, efforts to identify new RNA motifs that are modulated by ligands in natural systems may fail to detect entire classes of sequences if we remain limited to the conventional perspective that riboswitches are found within non-coding regions of RNA.

As it becomes increasingly straightforward to develop synthetic aptamers that bind new ligands using in vitro selection methods, it would be highly desirable to convert these aptamers into riboswitches with a simple procedure such as cloning the aptamer at several positions downstream of a start codon. If this mechanism of gene repression could be extended to a variety of synthetic aptamers, opportunities to engineer metabolism, create new genetic circuits, and control a variety of cell behaviors would be enhanced significantly.

4.4 Experimental

General Considerations

Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Culture media was obtained from EMD Bioscience. Theophylline,

ampicillin, chloramphenicol and *o*-nitrophenyl-β-D-galactopyranoside (ONPG), were purchased from Sigma. X-gal was purchased from US Biological. DNA polymerase, restriction enzymes, and the pUC18 cloning vector were purchased from New England BioLabs. Plasmid manipulations were performed using *E. coli* TOP10F′ cells (Invitrogen) that were transformed by electroporation. Purifications of plasmid DNA, PCR products, and enzymatic digestions were performed by using kits from Qiagen. All new plasmids were verified by DNA sequencing performed by MWG Biotech.

Construction and Screening of Genetic Libraries to Identify Riboswitches

To construct the library shown in Figure 4.1B, cassette-based PCR mutagenesis was performed using inner primers that anneal to the aptamer and the 5′ portion of the *IS10* gene, and outer primers that anneal to the regions just beyond the *Kpn*I and *Hind*III sites. The forward outer primer was used to introduce a 12-nucleotide randomized region immediately 5′ to the aptamer, as shown in Figure 4.1B. The full-length PCR product was digested with *Kpn*I and *Hind*III, and this insert was cloned between the same restriction sites within a pUC18-based vector to construct a translational fusion to the *lacZ* reporter gene. Electrocompetent TOP10F′ cells were transformed with the library, and 5,000 transformants were plated on selective media containing X-gal (25 mg per 300 mL media). Because these cells were grown in the absence of theophylline and we were interested in identifying clones that would exhibit high gene expression in the theophylline-free state, we selected colonies that were dark blue under these conditions.

To screen for riboswitches that could repress gene expression in the presence of theophylline, these clones were assayed in 96-well format, as previously described.[9] To

our surprise, a number of clones appeared to activate gene expression in the presence of theophylline. To verify this observation, we subcultured these clones in a larger volume of media, and assayed these constructs by hand. To determine why these constructs were activating gene expression, we sequenced several clones and used mfold software[29] to obtain secondary structure predictions for these RNA sequences, both in the presence and absence of theophylline. Figure 4.2 shows the predicted secondary structures for one of these sequences.

To construct the library shown in Figure 4.3, a forward primer that includes the 5′ *Kpn*I site, introduces mutations to swap two pairs of bases within the aptamer stem, and extends to within the IS10 coding region was used with a reverse primer that anneals to the IS10 fusion region and contains a *Hind*III restriction site. The PCR product was digested with *Kpn*I and *Hind*III, and the insert was cloned between the same restriction sites within a pUC18-based vector containing the *cat* gene, which confers resistance to the antibiotic chloramphenicol. Electrocompetent TOP10F′ cells were transformed with the library and plated at high density on media containing ampicillin to select for the pUC-based plasmid. To determine the approximate library size, a small aliquot of the original transformation was diluted and plated separately to reveal that the high-density plate contained ~220,000 clones. Cells from the high-density plate were scraped into liquid culture and were replated on media without theophylline, but containing ampicillin (50 μg/mL) and chloramphenicol (100 or 150 μg/mL).

Cells from each plate were scraped into separate tubes and grown in liquid culture with ampicillin (50 μg/mL). Plasmid DNA was extracted from each culture and digested with *Kpn*I and *Hind*III. Each library of inserts was cloned in front of a *lacZ* reporter gene

in a previously reported pUC18-based vector[3,6,7] that was digested with the same enzymes. TOP10F′ cells were transformed with these libraries, and cells were plated on media containing ampicillin, X-gal, and 1 mM theophylline. To identify clones that were repressed in the presence of ligand, 47 light blue clones were picked from each plate (after preselection with 100 or 150 μg/mL chloramphenicol) and were assayed by hand in 96-well plate format, as previously described. The three clones with 5-fold repression were subcultured, and plasmid was extracted for sequencing analysis.

Mutation of the ATG Codon 3′ to the Aptamer

To construct the library shown in Figure 4.5A, cassette-based PCR mutagenesis was performed with each of the three plasmids using inner primers that anneal to the region of interest to introduce the NNY library, and outer primers that anneal to the regions just beyond the *Kpn*I and *Hind*III sites. The three fully assembled PCR products were pooled and digested with *Kpn*I and *Hind*III, and the library of inserts was cloned in front of the *lacZ* reporter gene in a pUC18-based vector that was digested with the same enzymes. TOP10F′ cells were transformed with this pooled library, and cells were plated on selective media. 23 clones were picked at random and assayed by hand in plate-based format, as previously described.[9] Clones with greater than 5-fold repression were subcultured and verified by assaying a larger culture volume, as previously described. Plasmid was extracted for sequencing analysis.

To construct the library shown in Figure 4.5B, cassette-based PCR mutagenesis was performed with each of the three plasmids using inner primers that anneal to the region of interest to introduce the NNY library and remove the “AACAAG” sequence,

and outer primers that anneal to the regions just beyond the *Kpn*I and *Hind*III sites. The three fully assembled PCR products were digested separately with *Kpn*I and *Hind*III, and each library of inserts was cloned in front of the *lacZ* reporter gene in a pUC18-based vector that was digested with the same enzymes. TOP10F′ cells were transformed with each of these libraries, and cells were plated on selective media. For each of the three libraries, 32 clones were picked at random and assayed by hand in plate-based format, as previously described.[9] Clones with greater than 5-fold repression were subcultured and verified by hand-based assay. Plasmid DNA was extracted for sequencing analysis.

Mutation of the ATG Codon 5′ to the Aptamer

To construct the 'NTG' clones, a forward primer that includes the 5′ *Kpn*I site, introduces the NTG mutation, and extends into the aptamer sequence was used with a reverse primer that anneals to the IS10 fusion region and contains a *Hind*III restriction site. The PCR product was digested with *Kpn*I and *Hind*III, and the insert was cloned in front of the *lacZ* reporter gene in a pUC18-based vector that was digested with the same enzymes. TOP10F′ cells were transformed with the 4-member library, and cells were plated on selective media. 10 clones were picked and grown in liquid media. DNA was extracted from these cultures, revealing that each of the four clones was isolated at least once. The clones were subcultured and assayed by hand.

To convert the triplet immediately following the 5′ ATG to a stop codon, a forward primer that includes the 5′ *Kpn*I site, introduces the point mutation, and extends into the aptamer sequence was used with a reverse primer that anneals to the IS10 fusion region and contains a *Hind*III restriction site. The PCR product was digested with *Kpn*I

and *Hind*III, and cloned in front of the *lacZ* reporter gene as above. Single clones were picked and grown in liquid media. After confirmation by sequencing analysis, the construct was assayed by hand, as previously described.[9]

Repositioning of the Start Codon

To reposition the start codon at various positions before the aptamer, as shown in Figure 4.8A and Figure 4.8B, forward primers that include the 5′ *Kpn*I site, introduce the indicated mutations, and extend into the aptamer sequence were used with a reverse primer that anneals to the IS10 fusion region and contains a *Hind*III restriction site. The PCR products were digested with *Kpn*I and *Hind*III, and cloned in front of the *lacZ* reporter gene in a parent vector, which was digested with the same enzymes and dephosphorylated. TOP10F′ cells were transformed with these ligations, and cells were plated on selective media. Single clones were grown in liquid media, and the constructs were assayed by hand, as previously described.[9] The plasmids were sequenced to verify the sequence of the single clone shown in Figure 4.8A and to determine the sequence of clones derived from the library shown in Figure 4.8B.

Preparations for Mass Spectrometry Analysis

To determine whether the riboswitch is located within the coding region using mass spectrometry, a translational fusion was first constructed between the ~70 amino acid portion of the IS10 gene and GFPuv. To clone GFPuv into the vector containing the 27-fold repressor, pGFPuv was used as a template for a forward primer containing a *Hind*III site and a reverse primer that makes a silent mutation to remove a *Sac*I site within the GFPuv coding region, adds a C-terminal hexahistidine tag and a stop codon at the end

of the gene, and adds a 3′ SacI site for cloning purposes. The SDS-PAGE results and mass spectrometry data suggested (Figure 4.16), and the literature confirmed,[18] that the IS10 fusion protein contains a protease-sensitive post-translational cleavage site. Because the purified protein lacked its N-terminus due to post-translational cleavage, we used this vector to make a modified version of the construct with a shortened IS10 fusion.

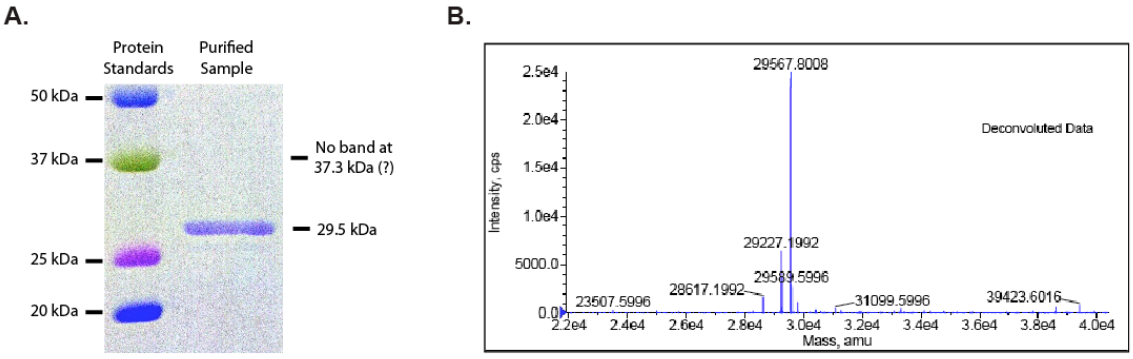

Figure 4.16: Purified Protein Is Smaller than Anticipated. (A) 12% SDS-PAGE analysis suggests that the purified protein is significantly smaller than the expected size (37.3 kDa). (B) ESI mass spectrometry confirmed that the purified protein is smaller than expected. Protease cleavage at a previously reported site within the IS10 linker region[18] is expected to yield a protein of 29.5 kDa, which is consistent with our data.

To remove the protease-sensitive site, the 27-fold repressor was amplified with a forward primer that includes the 5′ *Kpn*I site and a reverse primer that contains a *Hind*III restriction site and anneals to a short sequence at the beginning of the IS10 gene. The PCR product and the GFP construct described above were digested with *Kpn*I and *Hind*III, and the insert containing the repressor and shortened IS10 fusion was cloned into the vector containing the histidine-tagged GFPuv gene. TOP10F′ cells were transformed with this ligation, and cells were plated on selective media. Single clones were picked and grown in liquid media. DNA was extracted and confirmed by sequencing analysis.

To obtain protein for characterization by mass spectrometry, cells were grown in 100 mL cultures with 0 mM or 1 mM theophylline for 7 hours. Cultures were chilled on

ice and pelleted at 2 °C for 10 min at 12,000 rcf. When observed under a hand-held UV light, only the cells grown in the absence of theophylline were fluorescent. The cells grown without theophylline were suspended in lysis buffer and sonicated on ice. The histidine-tagged protein was purified using Ni^{2+}-affinity chromatography according to the manufacturer's instructions (G-Biosciences). Elution fractions, which were fluorescent, were combined and concentrated with a Millipore Ultrafree 0.5 mL centrifugal filter device. The concentrated protein was run on a 10% SDS-PAGE gel, and the protein-containing band was cut from the gel and fully trypsinized as described.[19] The tryptic peptides were extracted and analyzed as previously reported[20] using an LTQ-Orbitrap ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) at the Emory Center for Neurodegenerative Disease Proteomics Core. The results revealed that the N-terminal peptide was consistent with the expected sequence for the repressor, indicating that the riboswitch is within the translated region.

RNA Structure Probing Studies of the 27-Fold Repressor

To obtain an RNA transcript to probe the structure of our 27-fold repressor, we used PCR to generate a 182 base pair product. This PCR product was gel purified and used in transcription reactions with the Ampliscribe T7 transcription kit, according to the manufacturer's instructions (Epicentre Biotechnologies). After digestion of the DNA template using DNaseI, the transcription reactions were combined and subjected to 10% denaturing PAGE. The gel bands containing RNA were located by UV shadowing, cut into small pieces, added to water (4 mL), and frozen at -80 °C for 15 min. The RNA was permitted to passively elute at 4 °C overnight. The eluted RNA was filtered, precipitated,

and dissolved in water to obtain 7.7 pmol/μL. Figure 4.16 shows the sequence of the transcript, which includes the theophylline aptamer, part of the RBS (Shine-Dalgarno sequence), and a 3′ primer-binding site.

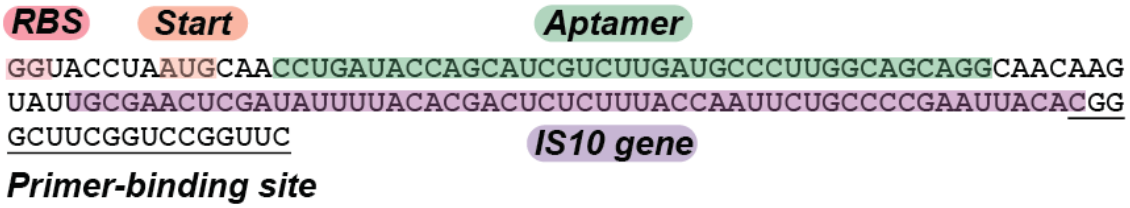

Figure 4.17: Sequence of RNA Transcript Probed by SHAPE. The 5′ portion of the transcript begins within the RBS (Shine-Dalgarno sequence) and continues through aptamer (green) and a portion of the *IS10* fusion gene (purple). A primer-binding site (underlined) was included for reverse transcription after modification with the NMIA reagent.

To probe the structure of this transcript, 15 pmol of RNA was used for SHAPE reactions with NMIA, as previously reported.[21] We also prepared a negative control sample, to which neither theophylline nor NMIA was added. After reaction with NMIA at 37 °C for 50 min, each sample was precipitated and dissolved in 10 μL of nuclease-free water. Simultaneously, a DNA primer complementary to the 3′-primer binding site (underlined in Figure 4.17) was end-labeled with [γ-^{32}P]ATP (7,000 Ci/mmol) using T4 polynucleotide kinase (New England BioLabs). The radiolabeled primer was purified using a Qiagen Nucleotide Extraction Kit and eluted with 100 μL H_2O.

Primer extension reactions with the radiolabeled primer were performed as previously described.[21] The radiolabeled primer was also used to perform sequencing reactions with dideoxynucleotides on unmodified RNA transcript. The resulting cDNA fragments were subjected to 8% denaturing PAGE. The gel was transferred to filter paper (Whatman), dried, and exposed overnight to a phosphor screen. The screen was scanned using a phosphorimager, and band intensities were quantified using ImageQuant

analysis software (Molecular Dynamics).

Reprogramming Cell Behavior with a Repressor

To test whether a repressor can be used to complement activator switches in the context of reprogramming cell behavior, we cloned the repressor identified as '4454' in Table 4.1 into the 5′ UTR the *cheZ* gene. Specifically, we performed a restriction digest with *Kpn*I and *Hind*III using the '4454' plasmid to obtain an insert containing the theophylline-dependent repressor. This insert was ligated between the *IS10* promoter and the *cheZ* gene in a pUC-based plasmid that confers resistance to ampicillin. JW1870 *E. coli* cells were transformed directly with this ligation and plated on selective media. A single clone was picked and grown in LB to isolate plasmid for sequencing analysis. We then subcultured these cells in selective tryptone broth until the culture reached mid-log phase. 600,000 cells were spotted of motility plates containing 25 mL selective media (tryptone broth with 0.25% agar, 50 μg/mL ampicillin, and 0 mM or 1 mM theophylline), which were prepared in Petri dishes (85 mm dia.). These plates were incubated at 30 °C for 14 hours (Figure 4.15).

4.5 References

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

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

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

(4) Bauer, G.; Suess, B., *J. Biotechnol.* **2006,** *124*, 4-11.

(5) Shine, J.; Dalgarno, L. *Nature* **1975,** *254*, 34-38.

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

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

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

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

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

(11) Topp, S.; Gallivan, J.P. *J. Am. Chem. Soc.* **2007,** *129*, 6807-6811.

(12) 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.

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

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

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

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

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

(18) Kwon, D.; Chalmers, R.M.; Kleckner, N. *Proc. Natl. Acad. Sci.* **1995,** *92*, 8234-8238.

(19) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996,** *68*, 850-858.

(20) Peng, J.M.; Gygi, S.P. *J. Mass Spectrom.* **2001,** *36*, 1083-1091.

(21) Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. *Nat. Protoc.* **2006,** *1*, 1610-1616.

(22) Pan, T.; Sosnick, T. *Annu. Rev. Biophys. Biomol. Struct.* **2006,** *35*, 161-175.

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

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

(25) Sergiev, P.V.; Lavrik, I.N.; Wlasoff, V.A.; Dokudovskaya, S.S.; Dontsova, O.A.; Bogdanov, A.A.; Brimacombe, R. *RNA* **1997,** *3*, 464-475.

(26) Steitz, J.A. *Nature* **1969,** *224*, 957-964.

(27) Steitz, J.A.; Jakes, K. *Proc. Natl. Acad. Sci.* **1975,** *72*, 4734-4738.

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

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

CHAPTER 5 – Exploring the Portability of Synthetic Riboswitches for use as Genetic Tools in Diverse Bacterial Species

5.1 Introduction

The development of a variety of ligand-inducible expression systems for use in *E. coli* has enabled tremendous advancements in genetics, biochemistry, molecular biology, and biotechnology. Such tools would be of significant value for facilitating more efficient studies to probe the genetics and behavior of lesser-understood bacteria. For example, there are few options for the induction or repression of genes in *Mycobacterium tuberculosis* or other, non-pathogenic mycobacteria.[1] Studies of pathogenesis would be greatly enhanced by the ability to selectively induce expression of a specific gene (or set of genes) at a desired time. However, efforts to develop such tools in bacteria that are less tractable suffer from the inherent problem that less is known about these species. Furthermore, with the exception of the tetracycline-inducible expression system,[2] most protein-based ligand-inducible expression systems that function well in *E. coli* have proven difficult to transport into a broad range of bacteria. We hypothesized that synthetic riboswitches might present an alternative to protein-based genetic regulatory systems for enabling inducible gene expression in such bacteria.

Over the past decade, synthetic riboswitch technologies have been developed in common model organisms such as *E. coli*,[3-6] *B. subtilis*,[7] *S. cerevisiae*,[8,9] and Chinese hamster ovary (CHO) cells.[10] As discussed in Chapter 2, such synthetic riboswitches provide a mechanism by which cellular behavior can be reprogrammed.[11] Although more advanced applications for these riboswitches are only in the incipient stages of

development, we believe that it is now possible to harness the potential of these powerful genetic control elements to expand the genetic toolset for a variety of less tractable species. Because synthetic riboswitches are ligand-inducible expression systems that do not require accessory proteins, riboswitch components may be tailored more easily to the needs of various species. For example, riboswitches that function by a translational mechanism regulate gene expression by providing ligand-dependent access to the ribosome-binding site. Therefore, these riboswitches may be modified for use in other species based upon knowledge of the 16S rRNA sequence and other requirements for ribosomal binding or translation initiation.

There is good reason to believe that synthetic riboswitches can be used to regulate gene expression in a variety of species: Natural riboswitches have been discovered in many species of bacteria, fungi, and eukaryotes, such as: *E. coli*,[12] *B. subtilis*,[13] *Streptococcus pneumoniae*,[14] *Agrobacterium tumefaciens*,[15] *Clostridium perfringens*,[16] *Clostridium acetobutylicum*,[17] *Vibrio vulnificus*,[16] *Vibrio cholerae*,[18] *Neurospora crassa*[19] and *Arabidopsis thaliana*.[20] Moreover, riboswitches are predicted to exist and to function in the important role of metabolite sensing across an extremely diverse range of bacteria, as well as in Archaea and Eukaryotes (Figure 5.1).[21] Therefore, it is apparent that the basic requirements for RNA-based genetic switching are met across all domains of life.

We hypothesized that synthetic riboswitches, in some form, should also be functional in a variety of species. However, a synthetic riboswitch that has been optimized for small-molecule dependent gene expression in *E. coli* may not enable optimal function in a different species. For example, it is well known that despite strong conservation of the 3′ end of the 16S rRNA, many Gram-positive bacteria, such as

Bacillus, *Mycobacterium*, and *Streptococcus* require substantially stronger interactions between the 16S rRNA and the Shine-Dalgarno region upstream of the start codon.[22,23] Additionally, optimal spacing between the Shine-Dalgarno sequence and the start codon may vary from species to species.[24,25] Therefore, we need to identify ways to make rational changes to riboswitches identified in *E. coli* to enable them to work in other species. Alternatively, we can improve library design to reduce these pools to a more manageable size for the *de novo* identification of synthetic riboswitches in other bacteria.

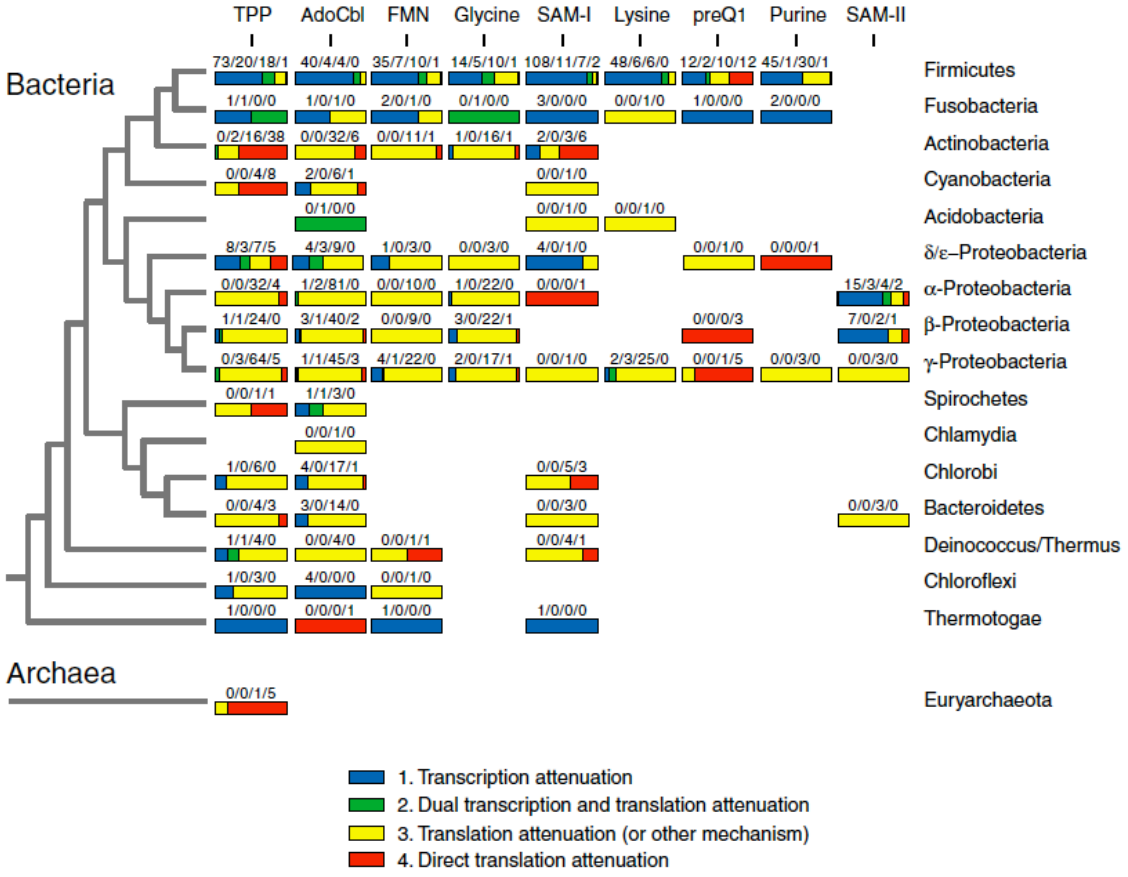

Figure 5.1: Distribution of Natural Riboswitches Across Taxonomic Groups. Natural riboswitches from various taxonomic groups are predicted to function by the indicated mechanisms.
(Reproduced from: Barrick and Breaker, 2007; Ref # 21)

Here, we present a series of experiments designed to identify a set of principles that can be applied to introduce, redesign, or screen for riboswitches in new species of bacteria. If successful, this study may enable the introduction of a new set of tools into less tractable species for genetic studies. We pursue a path beginning with *E. coli* and extend our studies to *Acinetobacter baylyi*, *Agrobacterium tumefaciens*, *Bacillus subtilis*, and *Mycobacterium smegmatis*, as we explore whether synthetic riboswitches can be transferred from one species to another. We hope that these studies can be generalized for a variety of other species, and particularly those for which high-throughput screens for the *de novo* identification of synthetic riboswitches would be unrealistic due to low transformation efficiencies, slow growth rates, difficult culture systems, or other factors.

5.2 Results and Discussion

We initially approached the question of riboswitch portability by cloning synthetic riboswitches that were previously discovered in *E. coli* into shuttle vectors that could be used to transform other species of bacteria. Preliminary results proved inconclusive. However, with access to more optimal shuttle vectors and promoters, we achieved good success in transporting synthetic riboswitches between distantly related species of bacteria. Notably, this study enabled us to identify a synthetic riboswitch that could provide ~ 50-fold activation of gene expression in *M. smegmatis* when cells were grown in the presence of 2 mM theophylline.

5.2.1 Addressing the Conundrum of Riboswitches in *Acinetobacter baylyi*

Previous work in our lab focused on transporting synthetic riboswitches that were identified in *E. coli* into the closely related species, *Acinetobacter baylyi* (strain ADP1).[26]

Like *E. coli*, *A. baylyi* is a γ-proteobacterium, making it a good choice for preliminary studies of riboswitch portability.[27] Although *Acinetobacter* species are generally non-pathogenic, *A. baumannii* is an infectious agent that can be acquired readily by immune-compromised patients during hospitalization in intensive care units.[28] The ability to introduce riboswitches as genetic control elements in *A. baylyi* may serve as a first step toward developing these tools for studies of *A. baumannii* pathogenesis. Additionally, *A. baylyi* strain ADP1 is naturally competent, which has piqued the interest of scientists who believe that this feature may be useful for a variety of biotechnology applications.[27]

Since *A. baylyi* is a close relative of *E. coli*, we hypothesized that our previously identified synthetic riboswitches would function in ADP1. Surprisingly, initial studies using a shuttle vector derived from pWH1266[29] (IMBB-kanR-pWH1266 ori-del *Hind*III; Ichiro Matsumura) suggested that transporting synthetic riboswitches between bacterial species might not be a straightforward task. In fact, it appeared that some of the best riboswitches (50- to 100-fold induction) our lab had identified in *E. coli* using advanced, high-throughput screening methods failed to function optimally in *A. baylyi*.[26] This observation motivated efforts to screen directly for riboswitches in ADP1, which led to the discovery of a synthetic riboswitch that appeared to function well.[26] This riboswitch featured a purine-rich Shine-Dalgarno sequence (RBS). These results suggested that it may be necessary to identify synthetic riboswitches *de novo* in new species of bacteria.

However, further studies revealed that several complicating factors may have led us to the premature conclusion that riboswitches identified in *E. coli* do not function well in *A. baylyi*. Specifically, the *tac* promoter, which was used in some of the studies described above, has not been previously characterized in *Acinetobacter*. Additionally, it

is known that the origin of replication in the pWH1266-derived vector promotes transcription of a promoterless reporter gene inserted within the multiple cloning site (Anton Bryksin, unpublished). Therefore, we sought to use a shuttle vector that would replicate in *A. baylyi*, but that would feature transcriptional terminator sequences on both sides of the multiple cloning site. Fortunately, we became aware of such a vector, which also appeared to replicate in a variety of bacteria, including several Gram-positive species (pBAV1K; Anton Bryksin). We anticipated that it might be possible to test a set of synthetic riboswitches in a variety of Gram-negative and Gram-positive bacteria, without the need to clone the switches into different shuttle vectors. We hoped that such a study might enable us to develop a set of principles for the transfer of synthetic riboswitches across closely- or distantly-related species of bacteria.

5.2.2 Selection of the Study Species

To determine whether synthetic riboswitches can be transferred easily across bacterial species, we chose to evaluate their function in several species of Gram-negative and Gram-positive bacteria. Joining *E. coli* and *A. baylyi* in the γ-proteobacteria class, the pathogenic bacteria *Francisella tularensis* would be a good test case to determine if synthetic riboswitches discovered in *E. coli* or *A. baylyi* could be used in a less-studied, but closely related, species of bacteria.[30] We also chose to test riboswitches in *Agrobacterium tumefaciens*, which is an α-proteobacterium.[31] Although *Agrobacterium* is somewhat less closely related to *E. coli*, it is an agriculturally relevant species that has been the focus of many studies.[31] Therefore, *A. tumefaciens* may provide an opportunity to evaluate the requirements for riboswitch portability more readily than might be possible using a species that has been less well studied.

To test whether synthetic riboswitches selected in *E. coli* function in Gram-positive bacteria, we included *Bacillus subtilis* and *Mycobacterium smegmatis* in our portability study. *B. subtilis*, classified as a firmicute (low G + C), has been widely embraced as the archetypical Gram-positive species.[32] Therefore, we should be able to draw from the wealth of knowledge provided by decades of research to: 1) Make predictions regarding which riboswitches isolated in *E. coli* may remain functional upon transfer; 2) Make rational modifications to optimize riboswitches for Gram-positive bacteria; and 3) Select for riboswitches directly in this easily manipulated model species. Furthermore, at least 29 natural riboswitches have been identified in the *B. subtilis* genome,[21] and a synthetic riboswitch based upon a helix-slipping mechanism has been engineered for this Gram-positive species.[7] Thus, the potential of small-molecule sensing riboswitches to function in *B. subtilis* is not in question; rather, our focus is centered on whether it is possible to transfer a riboswitch that has been identified in *E. coli* directly to *Bacillus*, or whether riboswitch function in these Gram-positive bacteria may be subject to a different set of requirements.

Our final species for riboswitch portability studies was *Mycobacterium smegmatis*, which is a laboratory-friendly model species that is closely related to the pathogenic *Mycobacterium tuberculosis* species.[33] Although the thick waxy coat of *Mycobacteria* spp. provides resistance to Gram-staining, these species are definitively Gram-positive because they lack outer membranes. Thus, mycobacteria are classified as actinobacteria (high G + C, Gram-positive).[33] There is strong interest in the development of ligand-inducible expression systems for basic and applied studies of mycobacteria, as few regulated expression systems have been developed for these species.[1]

The first inducible expression system for mycobacteria was developed after the discovery that expression of an acetamidase enzyme could be induced by aliphatic amides such as acetamide and butyramide.[34] This system was a tremendous advancement for basic studies of mycobacterial genetics, but it was not suitable for inducing gene expression during an infection. To address this shortcoming, Schnappinger et al. demonstrated that the ubiquitous TetR-controlled expression system functions in *M. smegmatis* as well as in *M. tuberculosis*.[2,34] This system can be used during infections of mammalian cells in culture or in whole animals.

This relatively recent advancement highlights the importance of developing additional tools to facilitate more complex studies of mycobacterial genetics and pathogenesis. If synthetic riboswitches prove to be readily transferable from *E. coli* or *B. subtilis* to *Mycobacterium* spp., the ability to construct ligand-inducible expression systems that respond to new ligands would become straightforward. Therefore, we have chosen to include *M. smegmatis* in our studies of riboswitch portability because, in addition to serving as proof-of-principle, these ligand-inducible genetic control elements would enrich the relatively meager toolset of scientists in the mycobacterial field.

5.2.3 Testing Previously Identified Riboswitches in Several Species

To begin our studies of riboswitch portability, we focused on testing a synthetic riboswitch that was previously identified in *E. coli* (22-fold induction),[6] as well as the purine-rich riboswitch identified through a *de novo* screen in *Acinetobacter baylyi* (25-fold induction).[26] We decided to determine how well each of these synthetic riboswitches functions in the easily manipulated study species: *E. coli*, *A. baylyi*, *A. tumefaciens*, and

B. subtilis. Because the broad-host-range vector is able to transform each of these species, we cloned both synthetic riboswitches into a derivative of pBAV1K that harbored the *lacZ* reporter gene. Transcription was initiated from the T5 promoter, which could be repressed in some strains of *E. coli* and *Bacillus* by the presence of the *lac* operator sequence.[35] The data reveal that the *E. coli* riboswitch functions quite poorly in *A. baylyi*, *B. subtilis*, and *A. tumefaciens*. In contrast, the riboswitch with a purine-rich Shine-Dalgarno sequence, which was identified in *A. baylyi*, provides at least 10-fold induction in all four species (Table 5.1).

Name	Linker	*E. coli*	*A. baylyi*	*Bacillus*	*Agro*
8.1 (*E. coli*)	CCGCT**G**CA**AG**A	22	3	2	1
Purine-rich (*A. baylyi*)	CTGA**GAAGGGG**	25	17	28	10

Table 5.1: Activation Ratios of Two Synthetic Riboswitches in Several Species. We tested the portability of a riboswitch that was previously identified in *E. coli* (8.1), as well a riboswitch that was identified in *A. baylyi*. The sequence linking the aptamer to the constant region preceding the start codon is shown, with the putative Shine-Dalgarno sequences (RBS) emboldened. The activation ratio is listed for each switch in various bacterial species.

Since the purine-rich riboswitch identified *de novo* in *Acinetobacter* performs reasonably well in all species tested, yet previously identified riboswitches from *E. coli* performed relatively poorly in other species, we hypothesized that *E. coli* may actually be a poor choice for identifying riboswitches that are to be transferred directly to another species. Numerous studies have shown that *E. coli* is quite promiscuous in its ability to initiate translation, even when there is little complementarity between the RBS (Shine-Dalgarno sequence) and the 16S rRNA.[23] Therefore, riboswitches that may function perfectly well in *E. coli* may not feature RBS sequences that are recognized by other species. Thus, even if conformational switching occurs in other species, the RBS

sequence that is revealed upon ligand-binding may not be sufficiently complementary to the 16S rRNA to permit translation initiation.

This problem is likely to arise during efforts to transport riboswitches from *E. coli* to Gram-positive bacteria, which generally require much stronger interactions between the 16S rRNA and the mRNA.[22] Therefore, we decided to identify riboswitches that would function in *B. subtilis* by combining semi-rational design with a low-throughput screen. We hoped that riboswitches identified in this manner would function in a broader range of Gram-negative and Gram-positive bacteria than those identified in *E. coli*.

5.2.4 Semi-Rational Riboswitches for Gram-Positive Bacteria

Guided by the data from the many high-throughput screens and selections our lab has performed to identify theophylline-dependent riboswitches in *E. coli*,[4-6] we designed a 256-member library to identify synthetic riboswitches in *B. subtilis*. As shown in Figure 5.2, we randomized the two bases preceding a strong Shine-Dalgarno sequence (RBS), as well as the two bases following this RBS sequence. These randomized regions are flanked by constant regions, which incorporate sequences that pair with the theophylline aptamer to block the ribosomal-binding site and prevent gene expression in the absence of ligand.[5] The predicted difference in free energy between the aptamer conformation and the RBS-paired conformation is ~9 kcal/mol, while the free energy of theophylline binding to the aptamer is -8.9 kcal/mol.[36] Therefore, we anticipated that screening this library would reveal specific riboswitch sequences with optimal differences in free energy between the ligand-free and ligand-bound states to permit theophylline-dependent conformational switching.

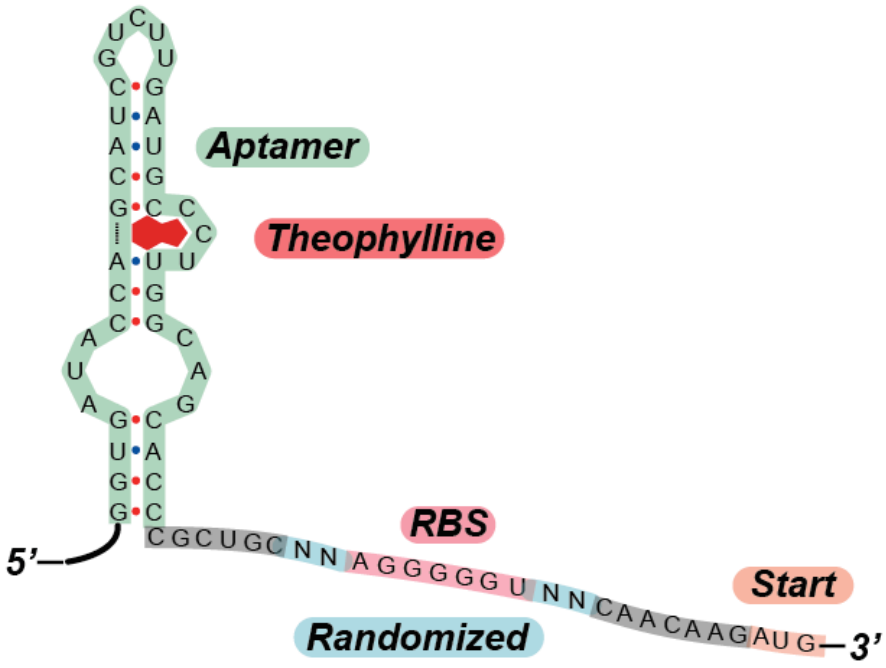

Figure 5.2: Semi-Rational Library to Identify Riboswitches for *B. subtilis*. A 256-member library was designed for identifying synthetic riboswitches that function in *B. subtilis*. A Shine-Dalgarno sequence (RBS) with high complementarity to the 16S rRNA was preceded and followed by two randomized bases. These randomized bases were flanked by constant regions, which are complementary to portions of the aptamer sequence.

We cloned this library into the broad-host-range vector (pBAV1K), which was described in Section 5.2.1. MDS42 *E. coli* cells[37] were transformed with this ligation, and we picked 47 colonies to assay for riboswitch activity in 96-well format. Encouragingly, more than half of these clones exhibited greater than 10-fold increases in gene expression when *E. coli* cells were grown with 1 mM theophylline. Next, we wanted to test whether these riboswitches, which had been designed to have adequate Shine-Dalgarno sequences for expression in Gram-positive bacteria, are portable from *E. coli* to *B. subtilis*.[23] Therefore, we isolated plasmid from three clones that had better than 20-fold activation in *E. coli* and one that had only 6-fold activation, and we transformed *B. subtilis* strain JH642 with these riboswitch constructs.

The subsequent assays of *B. subtilis* cells harboring these plasmids revealed that the three riboswitches that functioned well in *E. coli* also functioned well in *Bacillus* (65, 25, and 18-fold activation), while the 6-fold riboswitch from *E. coli* functioned as a 4-fold riboswitch in *Bacillus*. Secondary structure predictions[38] for each of the three riboswitches that performed well in *Bacillus* suggest that the anticipated conformational changes enable theophylline-dependent gene expression. Figure 5.3 depicts the predicted secondary structures for the 65-fold riboswitch in the ligand-free and ligand-bound states.

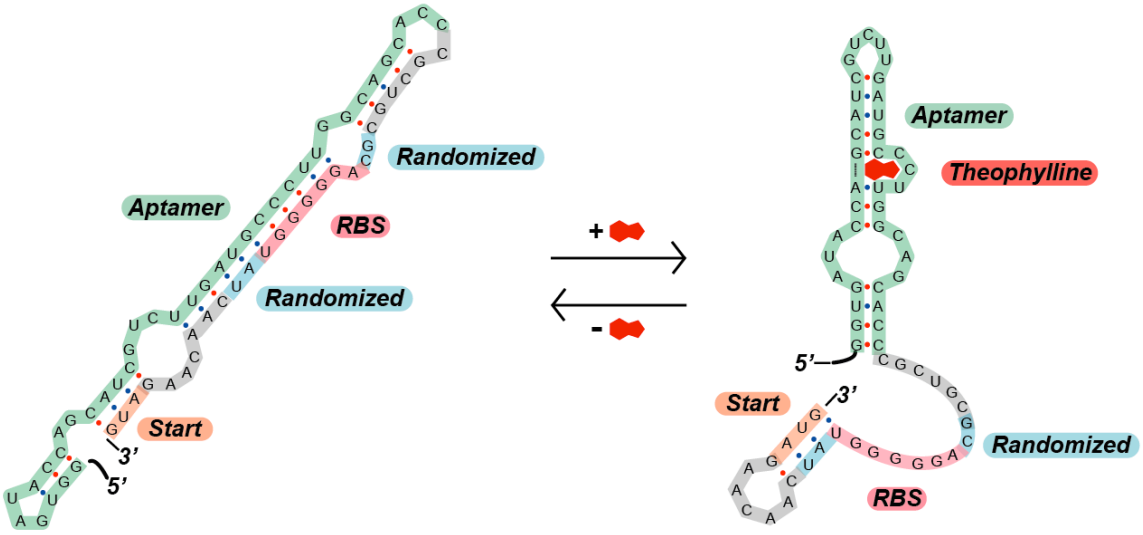

Figure 5.3: Predicted Secondary Structures for a *Bacillus*-Optimized Riboswitch. Predicted folds for a *Bacillus*-optimized synthetic riboswitch in the absence of ligand (left) and presence of theophylline (right). The aptamer is shaded in green, the Shine-Dalgarno sequence (RBS) is shown in pink, and the randomized region is highlighted in blue.

This interesting result suggested that it might be possible to screen for synthetic riboswitches in *E. coli* if appropriate care is taken to ensure that the Shine-Dalgarno sequences of the resulting riboswitches are likely to be recognized by the 16S rRNA of other bacterial species. To test whether the 65-fold *Bacillus*-optimized riboswitch might also function well in other species, we transformed *A. tumefaciens* and *A. baylyi* with the

plasmid and assayed these Gram-negative bacteria for β-galactosidase activity in the presence and absence of theophylline. The riboswitch activated gene expression 13-fold in *A. tumefaciens* and 32-fold in *A. baylyi*, suggesting that some riboswitches identified in *E. coli* may have the potential to function well in a variety of bacterial species. To further test this hypothesis, we have made preparations to test this *Bacillus*-optimized riboswitch in *Francisella tularensis*, and we are hopeful that it will function well in this Gram-negative pathogen. Additionally, we were interested in determining whether this riboswitch functions in mycobacteria, which will be discussed in the next Section (5.2.5).

5.2.5 Testing a 'Riboswitch Package' in *Mycobacterium smegmatis*

It has been especially challenging to develop ligand-inducible genetic control systems for mycobacteria because many molecules are unable to pass through their waxy coat.[33] Although theophylline is similar in size and structure to other purine heterocycles (such as uric acid) that are known to pass into mycobacterial cells,[39,40] we wanted to assess whether theophylline can enter these cells. This would become a particularly important experiment if we were to obtain negative results when testing our riboswitches. In this situation, it would be otherwise impossible to determine whether the riboswitches do not function in mycobacteria, or whether theophylline was simply not getting into the cells at adequate concentrations to activate the riboswitches. Therefore, before testing the riboswitches in *M. smegmatis*, we grew cells in increasing concentrations of theophylline. We hypothesized that the observation of very slow growth or toxicity at high theophylline concentrations would likely indicate that the ligand passes into the cell at concentrations that would be sufficient to activate the riboswitches.

To determine whether the ligand is toxic to cells grown over a range of theophylline concentrations (0 to 45 mM), we grew cells in 24-well format and monitored their growth over a 30 hour time period (Jessica C. Seeliger). As shown in Figure 5.4, *M. smegmatis* cultures showed no significant differences in growth for theophylline concentrations up through, and including, 2 mM. The rate of growth then decreased steadily as the concentration of theophylline was increased from 3.5 mM to 20 mM, at which point further increases in theophylline concentration delayed cell growth so substantially that no growth was observed after 30 hours. While one could imagine that toxicity might occur without the ligand crossing into the cell, such mechanisms are less likely, especially in light of the waxy mycobacterial coat. Therefore, Occam's razor guides us to the conclusion that toxicity suggests the passage of ligand into the cells.

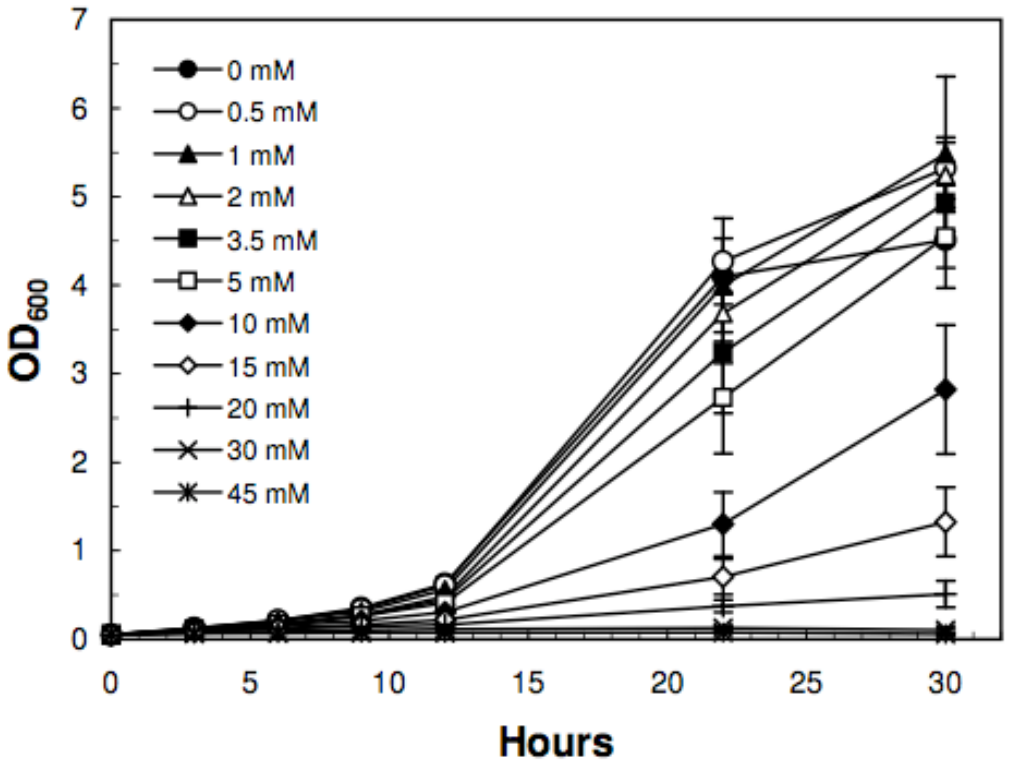

Figure 5.4: Growth of Cells in Various Theophylline Concentrations. *M. smegmatis* cells were grown in 24-well format at 37 °C in 7H9 liquid media containing various theophylline concentrations. Substantial deceases in growth rate are observed for cells grown in media containing more than 5 mM theophylline.

Having established that theophylline likely enters *M. smegmatis* cells, we began to develop a package of riboswitch constructs to test in this 'Gram-positive' species (Table 5.2). The package was comprised of five riboswitches, two of which had been identified previously in *E. coli* using fluorescence activated cell sorting (FACS)[5] or motility-based selection methods.[6] Although these riboswitches performed well in *E. coli* cells (95-fold and 24-fold induction), we believed that their weak Shine-Dalgarno sequences would not provide sufficient levels of gene expression in *M. smegmatis*. However, we noticed that it might be possible to redesign the switch identified by FACS to introduce a stronger Shine-Dalgarno sequence. Specifically, we observed that mutation of the 'TAA' sequence to 'AGG' would introduce an ideal Shine-Dalgarno sequence for Gram-positive bacteria. It should be noted that this mutation reduces the spacing between the Shine-Dalgarno sequence and the start codon by three nucleotides. Importantly, secondary structure predictions suggested that this mutation would maintain pairing of the RBS in the ligand-free state. Thus, we anticipated that the Gram-positive FACS riboswitch might perform well in *Mycobacteria* spp.

Name	Aptamer	Linker	Constant	Start
Motility switch	...AGCACC	TTC**GAGG**GCA	CACT	ATG
FACS switch	...AGCACC	CTGCTA**AGGT**AA	CAACAAG	ATG
Gram+ FACS	...AGCACC	CTGCT**AAGGAGG**	CAACAAG	ATG
A. baylyi switch	...AGCACC	CTGA**GAAGGGG**	CAACAAG	ATG
B. subtilis switch	...AGCACC	CGCTGCGC**AGGGGGT**AT	CAACAAG	ATG

Table 5.2: Riboswitches Tested in *Mycobacterium smegmatis*. The sequences of the five riboswitches that we tested in *M. smegmatis* are listed, with their putative Shine-Dalgarno sequences (RBS) shown in bold. The linker sequences are aligned to show complementarity with the 16S rRNA.

Joining these riboswitches, the fourth and fifth riboswitches in our package had been identified by screening libraries directly in *A. baylyi* or by semi-rational design for

B. subtilis, as described in Sections 5.2.3 and 5.2.4, respectively. The riboswitch identified in *A. baylyi* had a purine-rich Shine-Dalgarno sequence, while the *Bacillus*-optimized riboswitch featured an RBS sequence with good complementarity to the 16S rRNA consensus sequence for Gram-positive bacteria. Although we anticipated that both of these riboswitches could potentially provide theophylline-dependent gene expression in *M. smegmatis*, we hypothesized that the *Bacillus*-optimized riboswitch might activate gene expression more effectively than the switch that had been identified in *A. baylyi*. While the riboswitch sequences used to develop constructs for *M. smegmatis* were identical to those tested in *E. coli*, *B. subtilis*, *A. baylyi*, and *A. tumefaciens*, we needed to make modifications in the choice of vector, reporter gene, and promoter sequence:

Although we were hopeful that the broad-host-range vector might propagate in mycobacteria, a preliminary test to transform *M. smegmatis* with this plasmid suggested that it would not be a suitable choice. Therefore, we decided to clone our riboswitches into the known mycobacteria shuttle vector pMV261.[41] Secondly, to enable more efficient measurements of induction, we chose to use eGFP as our reporter protein in place of β-galactosidase. In addition to eliminating the need to lyse these waxy cells to perform an enzymatic assay, the use of eGFP would enable us to more easily measure changes in gene expression over time.

Finally, since mycobacteria typically use promoters that are very distinct from those used by other species of bacteria, we anticipated that the T5 phage promoter might not function well in *M. smegmatis*. Therefore, we cloned the riboswitches listed in Table 5.2 downstream of the constitutively active mycobacterial groEL promoter (*hsp60*), which produces high levels of gene expression in mycobacteria.[42] However, the 5′ UTR

transcribed from the native *hsp60* promoter is extremely structured, which may be non-ideal for riboswitch function (Figure 5.5). Therefore, we decided to eliminate this structure by shortening the 5′ UTR so that it would include only the first 15 bases transcribed by the native *hsp60* promoter. As we will see, our studies demonstrated that this modification does not lead to devastating decreases in maximum induction when cells are grown under standard experimental conditions.

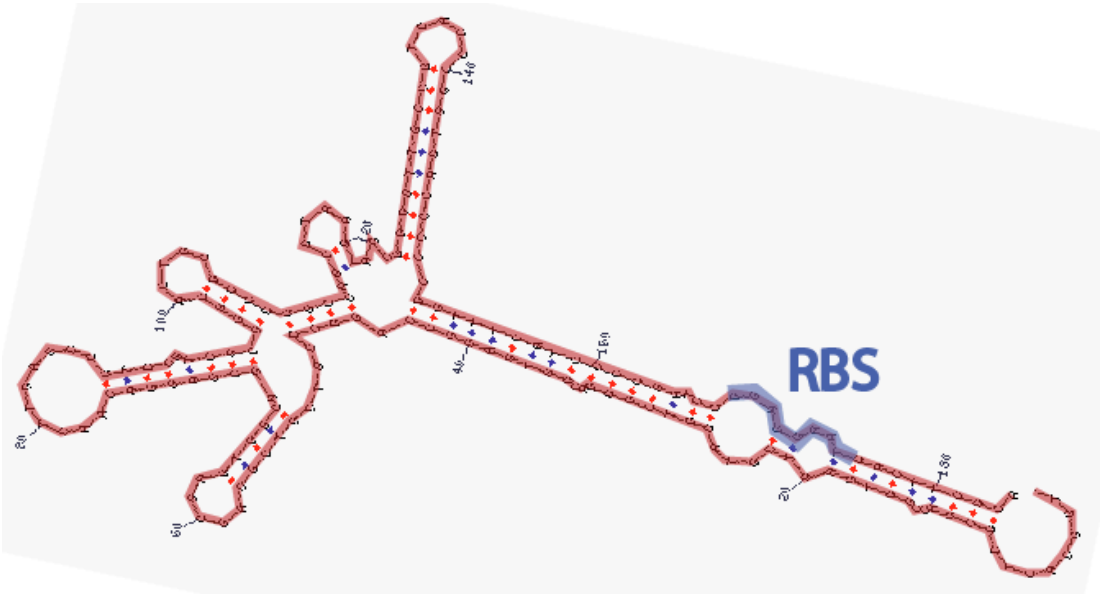

Figure 5.5: Predicted Secondary Structure of the *hsp60* Promoter 5′ UTR. Predicted secondary structure of the 5′ UTR (184 bases) transcribed from the native *hsp60* promoter. The Shine-Dalgarno sequence (RBS) is highlighted in purple.

In addition to constructing plasmids to test the five riboswitches listed in Table 5.2, we also made a control construct in which eGFP was expressed from the *hsp60* promoter with the shortened 5′ UTR sequence. All of these vectors were constructed using standard cloning methods in MDS42 *E. coli* cells. After confirming the sequences, *M. smegmatis* cells were transformed with each construct by electroporation. Cells were also transformed with an empty vector, to obtain a negative control for eGFP expression.

Additionally, cells were transformed with a positive control construct that enables constitutive expression of eGFP from the unmodified, *hsp60* promoter. We measured the relative fluorescent units (RFU) of *M. smegmatis* cells harboring the synthetic riboswitches or control constructs after induction for 6 hours (two doubling times) in liquid media containing 0, 0.5, 1, or 2 mM theophylline (Figure 5.6, top). The cells were assayed in 24-well format, and displayed identical growth at all theophylline concentrations tested. We also determined the "fold-increase in RFU" for each construct by dividing the RFU measured at 0.5, 1, or 2 mM theophylline by the RFU measured for that same construct at 0 mM theophylline (Figure 5.6, bottom)

As anticipated, cells harboring the empty vector showed background levels of gene expression at all concentrations tested. Additionally, cells harboring the positive control vector with the native *hsp60* promoter expressed high levels of eGFP regardless of whether theophylline was present. We were quite gratified to see that cells harboring the Gram-positive FACS riboswitch exhibited very low levels of gene expression when grown in the absence of ligand, but showed ~50-fold induction of gene expression when grown in the presence of 2 mM theophylline. Additionally, we found the dynamic range of this riboswitch to be quite satisfactory, as its RFU at 2 mM theophylline was greater than 70% of the RFU exhibited by the positive control.

We observed ~10-fold induction of gene expression for cells harboring both the riboswitch that was identified in *A. baylyi*, as well our *Bacillus*-optimized riboswitch. Interestingly, the two riboswitches identified in *E. coli* showed less than 2-fold induction when *M. smegmatis* cells were grown in the presence of 2 mM theophylline. Although these riboswitches provided very good activation ratios in *E. coli* cells (96-fold and 24-

fold, respectively), and the mRNA is likely to switch conformations in *M. smegmatis*, the Shine-Dalgarno sequences (RBS) for these switches do not provide sufficient interaction with 16S rRNA to enable translation in the presence of theophylline.

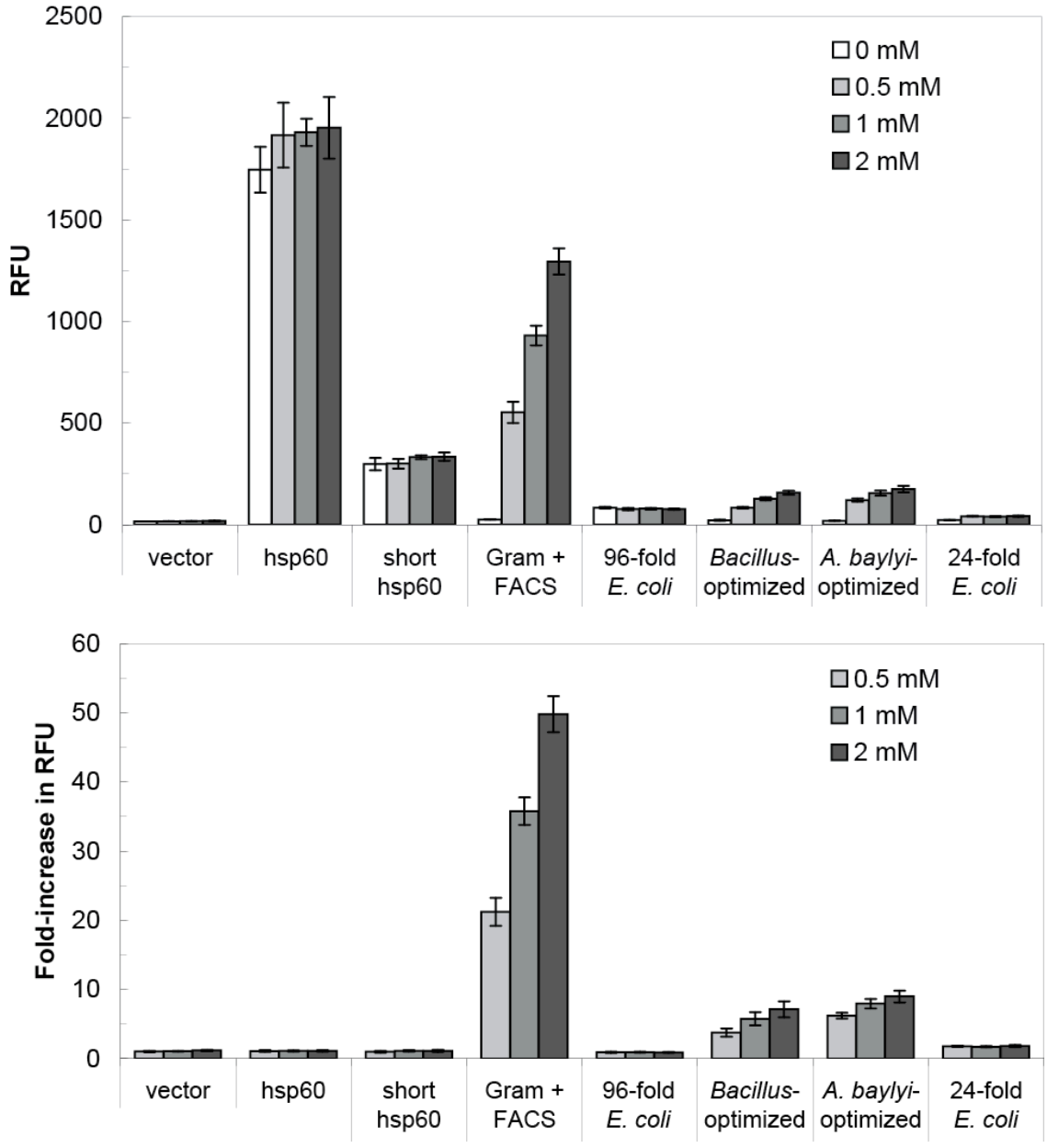

Figure 5.6: Performance of Riboswitches in *Mycobacterium smegmatis*. Top: Relative fluorescent units (RFU) at 0, 0.5, 1, and 2 mM theophylline for each riboswitch and the control vector. Bottom: Fold-increase from the RFU without ligand to the RFU at 0.5, 1, and 2 mM theophylline for each riboswitch. The Gram-positive riboswitch shows ~50-fold induction at 2 mM theophylline, while the *A. baylyi* and *Bacillus*-optimized switches exhibit ~10-fold induction at 2 mM theophylline. (Data obtained by J.C.S).

While the data presented thus far were consistent with our expectations, we observed lower than anticipated RFU for cells harboring the shortened *hsp60* control vector, which was constructed to verify that gene expression would remain high if the native 5′ UTR were shortened. However, upon further consideration, we realized that the context of the RBS sequence may change dramatically from control to riboswitch construct, such that a positive control may not necessarily reflect the maximum level of gene expression from a given promoter. Moreover, each riboswitch has a different RBS. Secondary structure predictions for the shortened-promoter control support this hypothesis, as the 5' UTR is relatively structured and the RBS and start codon appear to be somewhat inaccessible for translation initiation (Figure 5.7).

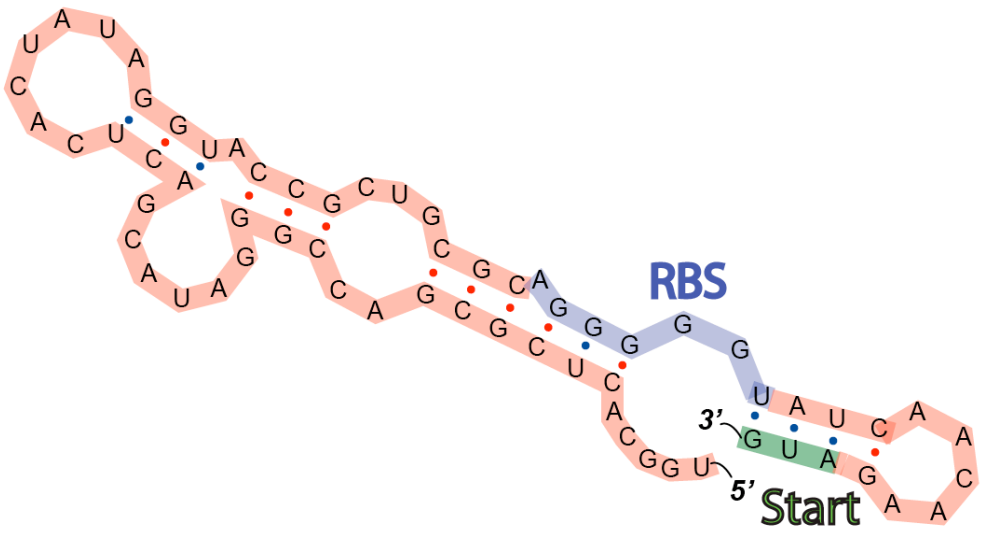

Figure 5.7: Predicted Secondary Structure of a Not-So-Positive Control. Predicted secondary structure of the 5′ UTR transcribed from the shortened *hsp60* promoter for a 'positive control' construct. The Shine-Dalgarno sequence (RBS) is highlighted in purple, while the start codon is highlighted in green.

To further characterize the Gram-positive riboswitch in *M. smegmatis*, we determined its induction response at increasing theophylline concentrations. To obtain this dose-response profile, we grew cells harboring the riboswitch, a positive control

(pMWS114), or a negative control (pMV261) in media containing various concentrations of theophylline. Figure 5.8 shows the relative fluorescence of cells harboring these vectors at 0-10 mM theophylline. The relative fluorescence was calculated by dividing the normalized fluorescence intensity of each data point by the normalized fluorescence intensity of cells harboring the same vector at 0 mM theophylline. Since expression of eGFP in cells harboring either control vector is not theophylline dependent, both provide a relative fluorescence of ~1 at all ligand concentrations. In contrast, the semi-rationally designed Gram-positive FACS riboswitch produces dose-dependent increases in relative fluorescence, reaching a maximum of ~70-fold induction at 10 mM theophylline.

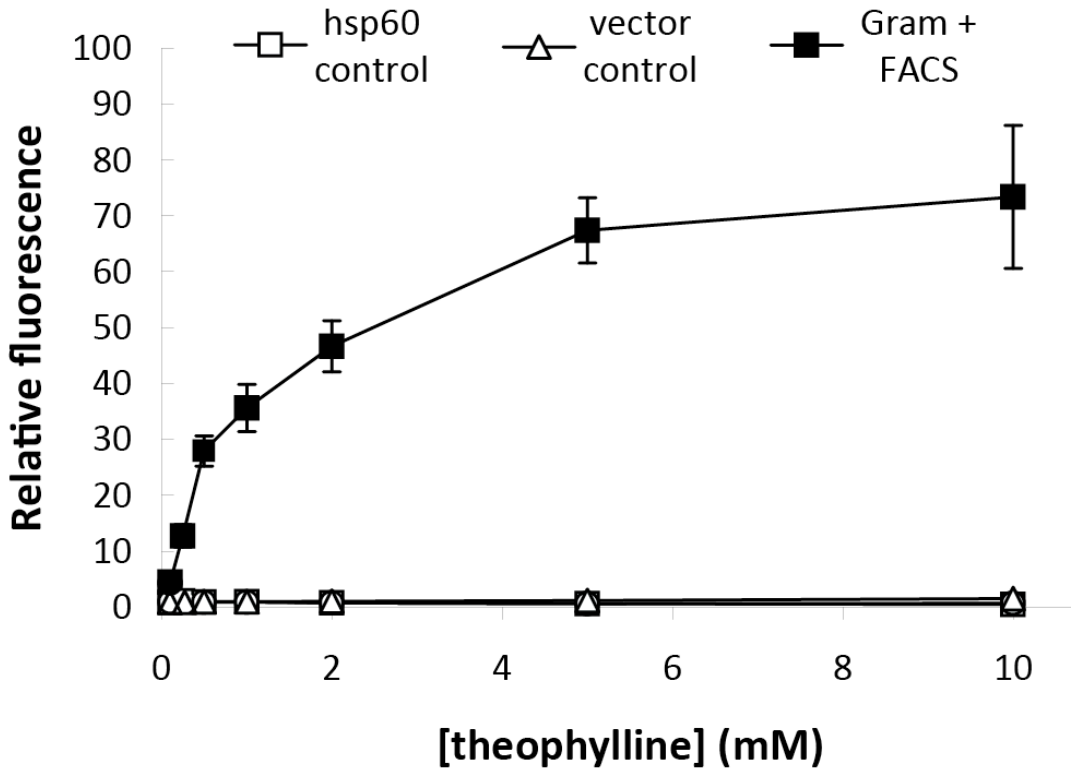

Figure 5.8: Dose-Response Profile of a Semi-Rationally Designed Gram-Positive Riboswitch. The relative fluorescence of cells harboring the riboswitch (filled squares), positive control (open squares), or negative control (triangles) are plotted as a function of theophylline concentration. Relative fluorescence was calculated by dividing the normalized fluorescence intensity of each data point by the normalized fluorescence intensity of cells harboring the same vector at 0 mM theophylline. (Data obtained by J.C.S.)

5.3 Conclusion

We initiated this study with the hypothesis that synthetic riboswitches identified in *E. coli* could be transferred with relative ease to other species of bacteria. The data revealed that the conformational changes required for riboswitch function can be obtained in a wide variety of bacterial species. However, many *E. coli* riboswitches feature Shine-Dalgarno sequences (RBS) that may not enable sufficiently stable pairing with the 3′ end of the 16S rRNA to enable translation initiation in other bacterial species; this feature is especially critical for riboswitch function in Gram-positive bacteria.[22] In contrast, riboswitches that provide strong interactions between the RBS and 16S rRNA may function well across a broad range of Gram-negative and Gram-positive bacteria.

The results presented here demonstrate that it is indeed possible that riboswitches identified in one species (*A. baylyi* or *B. subtilis*, in these studies) can be transferred to other species of bacteria without the need to perform *de novo* screens for riboswitches in each species. Whereas riboswitches identified in *E. coli* failed to activate gene expression in our other study species, the synthetic riboswitch identified in *A. baylyi*, as well as the semi-rational *Bacillus*-optimized switch, provided at least 10-fold induction for all species tested in this study. Furthermore, we demonstrated that semi-rational design was useful for improving a synthetic riboswitch that was discovered by high-throughput screening in *E. coli*. Most notably, a riboswitch identified by FACS provided 95-fold induction in *E. coli* cells but was nonfunctional when tested in *M. smegmatis*. However, rational mutation of this switch to enable stronger interactions between the Shine-Dalgarno sequence and 16S rRNA yielded a synthetic riboswitch that provided ~50-fold induction in *M. smegmatis* cells grown in the presence of 2 mM theophylline.

We are currently pursuing studies to identify factors that may require optimization to achieve at least 20-fold induction in *Agrobacterium tumefaciens*. Toward this end, we will capitalize on semi-rational design strategies to modify a synthetic riboswitch presented in Chapter 4 of this thesis (Figure 4.2). The preliminary data are encouraging, and serve as further evidence that semi-rational design of existing synthetic riboswitches is a promising approach toward obtaining synthetic riboswitches for use in other bacterial species (Ian Goldlust). In addition, we are currently testing our 'riboswitch package' in *Francisella tularensis*; the ability to transport synthetic riboswitches to *Francisella* would significantly enhance the genetic toolset available for study of this pathogen.

Finally, we are pursuing further studies to compare our semi-rationally designed, Gram-positive FACS riboswitch to the currently available TetR expression system in *M. smegmatis*. We are also testing its ability to function as a ligand-inducible genetic control element in *M. tuberculosis.* We anticipate that it will be possible to develop synthetic riboswitches to alternative ligands (starting with single aptamers to these ligands) for use in other Gram-positive bacteria through the combination of semi-rational design and relatively low-throughput screening in *B. subtilis*.

5.4 Experimental

General Considerations

Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Culture media was obtained from EMD Bioscience. Theophylline and *o*-nitrophenyl-β-D-galactopyranoside (ONPG) were purchased from Sigma. Kanamycin was purchased from Fisher Scientific. X-gal was purchased from US Biological. DNA

polymerase, restriction enzymes, and the pUC18 cloning vector were purchased from New England BioLabs. Plasmid manipulations were performed using *E. coli* MDS42 cells (Scarab Genomics) that were transformed by electroporation. Purifications of plasmid DNA, PCR products, and enzymatic digestions were performed by using kits from Qiagen. All new plasmids were verified by DNA sequencing performed by MWG Biotech.

Plasmid Construction

To construct the riboswitch package for *M. smegmatis*, each riboswitch sequence was cloned within the 5′ UTR of the eGFP gene within the pMWS114 shuttle vector. All other constructs were derived from pBAV1K (Anton Bryksin) and featured a T5 promoter with the lac operator. Each pBAV1K-derived plasmid expresses the *lacZ* reporter gene. The sequence of each riboswitch is shown in Table 5.2; the mTCT8-4 aptamer[36] was employed in these studies.

Transformation Protocols

Competent *B. subtilis* cells (strain JH642) were prepared and transformed using the Spizizen method.[43] Cells were grown overnight in 5 mL LB at 37 °C with shaking (250 rpm). 500 μL of saturated culture was added to 4.5 mL SPI and was allowed to grow at 37 °C with shaking (250 rpm) until the $OD_{540} \geq 1.2$. This culture was transferred to a centrifuge tube containing 20 mL prewarmed SPII and was allowed to grow at 37 °C for 1.5 h with gentle aeration (100 rpm). Cells were pelleted by centrifugation at 5,000 *g* (rt) and were resuspended in 2.5 mL of spent media. 5-10 μL of plasmid DNA was

added to 500 μL of cells, which were then grown at 37 °C with gentle shaking (100 rpm) for 1 h. 300 μL was plated on selective media.

A. baylyi strain ADP1 was grown overnight in 5 mL LB at 30 °C, with shaking at 250 rpm. In the morning, 3 mL fresh LB was inoculated with 200 μL from the overnight culture. These cells were grown at 30 °C for 90 min, at which time the culture was divided into 300 μL aliquots. 3 μL plasmid DNA was added to each tube of cells. These cultures were grown with shaking at 30 °C. After 3 h, 200 μL of these cultures was plated on selective media.

Competent *A. tumefaciens* cells (strains C58 and A136) were prepared and transformed by the method of Cangelosi et al.[44] Cells were grown in 5 mL LB at 28 °C with shaking (250 rpm). After 18 h, 10-100 μL of saturated culture was added to fresh LB (50-150 mL) and allowed to grow until the OD_{600} = 0.4. The cells were then pelleted three times by centrifugation at 5000 *g,* washing twice with water and once with 10% glycerol. 1 μL of plasmid DNA was mixed with 50 μL of cells, which were transformed by electroporation at 1800 V (Eppendorf Electroporater 2510; Westbury, NY). Electroporated cells were permitted to recover in 500 μL SOB for 4 h. Finally, 10-100 μL of culture was plated on selective media. All experiments in *A. tumefaciens* were performed by Ian S. Goldlust.

Riboswitch Activity Assays

Assays for β-galactosidase activity in *B. subtilis* were performed after permeabilization with toluene. Briefly, the OD_{595} was measured for each culture. 1 mL of culture was centrifuged at top speed in a microcentrifuge. The supernatant was

discarded, and the pellet was frozen in a dry ice/ethanol bath. The pellet was suspended in 1 mL Z buffer, to which 10 μL toluene was added. The suspension was vortexed for 15 s and was then warmed to 30 °C for 5 min. 200 μL ONPG solution (4.0 mg/mL) was added, and the reaction was permitted to proceed at 30 °C until a yellow color appeared. Reactions were stopped with 500 μL 1 M Na_2CO_3, and cellular debris was removed by centrifugation at top speed in a microcentrifuge for 5 min. The absorbance at 420 nm (A_{420}) was then measured. Miller Units were determined by the following formula:

Miller Units = (1000 x A_{420}) / (min reaction time x OD_{595})

Assays for β-galactosidase activity in *E. coli*, *A. baylyi*, and *A. tumefaciens* were performed according to standard protocol, as previously described.[4]

Mycobacterium Studies

Media, strains and plasmids. M. smegmatis mc^2155[45] was grown at 37 °C in 7H9 liquid media or on 7H11 agar (Difco) containing 0.5% glycerol, 0.5% glucose, 0.05% Tween 80 and 20 μg/mL kanamycin unless otherwise noted. pMWS114 contains the eGFP gene (S65T/F64L) cloned as a EcoRI-HindIII fragment into pMV261.[41] pGS-eGFP is a derivative of pMWS114 in which the *hsp60* promoter was removed and replaced by the glutamine synthase promoter cloned as a XbaI-PstI fragment from the plasmid pPGSY (gift from Lee Riley).[46]

Whole-cell GFP fluorescence assays. Plasmids were electroporated into *M. smegmatis* and plated on selective media for 3 days. Transformants were grown overnight in 7H9 to an optical cell density in 1cm at 600 nm (OD_{600}) of 1-2. Cells were exchanged into fresh

7H9 containing 0-2 mM theophylline to OD_{600} 0.3 in 1 mL per media condition in 24-well plates and incubated for 6 hrs (2 doubling times) at 215 rpm. The OD_{600} was measured prior to resuspending the pelleted cells in 200 μL PBS containing 0.05% Tween 80. Fluorescence was measured in a plate reader at 510 nm with 450 nm excitation and 495 nm cutoff (SpectraMax Gemini XPS, Molecular Devices) and normalized to OD_{600}. Cells transformed with the plasmid pMV261 were used as control for background fluorescence. Each growth condition was performed in triplicate. The relative fluorescence was calculated by dividing by the normalized fluorescence intensity observed in control cells at 0 mM theophylline. All experiments in *M. smegmatis* were performed by Jessica C. Seeliger.

5.5 References

(1) Ehrt, S.; Guo, X.Z.V.; Hickey, C.M.; Ryou, M.; Monteleone, M.; Riley, L.W.; Schnappinger, D. *Nucl. Acids Res.* **2005,** *33*, e21.

(2) Carroll, P.; Muttucumaru, D.G.; Parish, T. *Appl. Environ. Microbiol.* **2005,** *71*, 3077-3084.

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

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

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

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

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

(8) Bayer, T.S.; Smolke, C.D. *Nat. Biotech.* **2005,** *23*, 337-343.

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

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

(11) Topp, S.; Gallivan, J.P. *J. Am. Chem. Soc.* **2007,** *129*, 6807-11.

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

(13) Mandal, M.; Boese, B.; Barrick, J.E.; Winkler, W.C.; Breaker, R.R. *Cell* **2003,** *113*, 577-586.

(14) Meyer, M.M.; Roth, A.; Chervin, S.M.; Garcia, G.A.; Breaker, R.R. *RNA* **2008,** *14*, 685-695.

(15) Corbino, K.A.; Barrick, J.E.; Lim, J.; Welz, R.; Tucker, B.J.; Puskarz, I.; Mandal, M.; Rudnick, N.D.; Breaker, R.R. *Genome Biol.* **2005,** *6*, R70.

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

(17) Rodionov, D.A.; Vitreschak, A.G.; Mironov, A.A.; Gelfand, M.S. *Nucl. Acids Res.* **2004,** *32*, 3340-3353.

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

(19) Cheah, M.T.; Wachter, A.; Sudarsan, N.; Breaker, R.R. *Nature* **2007,** *447*, 497-U7.

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

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

(22) Band, L.; Henner, D.J. *DNA* **1984,** *3*, 17-21.

(23) Vellanoweth, R.L.; Rabinowitz, J.C. *Mol. Microbiol.* **1992,** *6*, 1105-1114.

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

(25) Ma, J.; Campbell, A.; Karlin, S. *J. Bacteriol.* **2002,** *184*, 5733-5745.

(26) Desai, S.K. **(2008)** Engineering Control of Gene Expression in Bacteria Using RNA-Small Molecule Interactions. Doctoral Dissertation, Emory University.

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

(28) Gasink, L.B.; Brennan, P.J. *Curr. Opin. Infect. Dis.* **2009,** *22*, 339-344.

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

(30) Barker, J.R.; Klose, K.E. *Ann. N. Y. Acad. Sci.* **2007,** *1105*, 138-159.

(31) Wood, D.W. et al. *Science* **2001,** *294*, 2317-2323.

(32) Kunst, F. et al. *Nature* **1997,** *390*, 249-256.

(33) Connell, N.D. *Curr. Opin. Biotechnol.* **2001,** *12*, 446-449.

(34) Parish, T.; Mahenthiralingam, E.; Draper, P.; Davis, E.O.; Colston, E.O. *Microbiology* **1997,** *143*, 2267-2276.

(35) Monod, J.; Changeux, J.-P.; Jacob, F. *J. Mol. Biol.* **1963,** *6*, 306-329.

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

(37) Posfai, G. et al. *Science* **2006,** *312*, 1044-1046.

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

(39) Rouf, M.A.; Lomprey, R.F., Jr. *J. Bacteriol.* **1968,** *96*, 617-622.

(40) Vogels, G.D.; Van der Drift, C. *Bacteriol. Rev.* **1976,** *40*, 403-468.

(41) Stover, C.K. et al. *Nature* **1991,** *351*, 456-460.

(42) Dellagostin, O.A.; Esposito, G.; Eales, L.J.; Dale, J.W.; Mcfadden, J. *Microbiology* **1995,** *141*, 1785-1792.

(43) Cutting, S. M., and C. Harwood. **1990.** Molecular biological methods for *Bacillus*. John Wiley & Sons Ltd., Chichester, England.

(44) Cangelosi, G.A., Best, E.A., Martinetti, G., Nester, E.W. **1991.** Genetic Analysis of *Agrobacterium*. In *Methods Enzymol.*, *204*, 384-397.

(45) Snapper, S.B.; Melton, R.E.; Mustafa, S.; Kieser, T.; Jacobs, W.R., Jr. *Mol. Microbiol.* **1990,** *4*, 1911-1919.

(46) Harth, G.; Horwitz, M.A. *J. Biol. Chem.* **1997,** *272*, 22728-22735.

Part II

A Genetic Toolbox for Creating Reversible Ca^{2+}-Sensitive Biomaterials

CHAPTER 6 – A Genetic Toolbox for Creating Reversible Ca^{2+}-Sensitive Biomaterials

6.1 Introduction

A major goal in materials science is to produce stimuli-sensitive polymers that respond predictably to various environmental signals such as changes in pH, temperature, light, the concentrations of salts or other molecules, and electrical or magnetic fields.[1,2] These materials can be used for diverse applications such as phase separations, affinity precipitations, bioactive surfaces, permeability switches, bioreactors, medical diagnostics, and drug delivery systems.[3] Engineering a stimuli-sensitive polymer to have specific functionality for such applications requires the ability to control structure and composition. Thus, the synthesis of materials that respond to stimuli in predictable ways is critically dependent on methods capable of producing polymers with defined mass, composition, stereochemistry, and topology.[4]

Traditional approaches to constructing stimuli-responsive polymeric materials often rely on the standard solution phase techniques of polymer synthesis, such as the ring opening polymerization of α-amino acid-*N*-carboxyanhydrides, or the radical polymerization of *N*,*N*-diisopropyl acrylamide.[3] However, these polymerizations may produce unpredictable structures as a result of the lack of control over polymer chain end reactivity. In response to this problem, controlled copolymerization has been developed to synthesize products of defined length and structure. It has been shown that the use of transition metal complexes as end groups to the previously uncontrolled chain ends

---

Adapted with permission from: S. Topp, V. Prasad, G.C. Cianci, E.R. Weeks, and J.P. Gallivan. *J. Am. Chem. Soc.*, **2006**, *128*, 13994-13995.

permits regulation of the chain propagation reaction.[5] Notwithstanding the utility of this improvement to traditional methods of polymer synthesis, recombinant DNA technology quickly has become a significant tool for designing and engineering biomaterials with precise composition, stereochemistry, and molecular weight.[6]

The first step in producing a stimuli-sensitive biomaterial by recombinant methods involves synthesis of a gene encoding for the protein. While short sequences of oligonucleotides can be synthesized chemically, it is generally more cost-effective to construct longer genes through stepwise ligation or by using PCR to isolate a sequence from a biological source. The constructed gene is then ligated into a plasmid vector that includes an inducible promoter, which allows conditional expression of the target gene. The plasmid is then used to transfrom a bacterial host such as *E. coli*.

When expression is induced, the host cells typically produce the target protein in modest to significant quantities. The recombinant protein may then be purified by a method such as immobilized metal affinity chromatography (IMAC).[3] This technique commonly uses a column with Ni^{2+} ions attached to agarose beads to bind a target protein with 5 or 6 histidine residues incorporated at its terminus. After washing to remove contaminating bacterial proteins, the recombinant protein is eluted from the column by changing the pH or by washing with a competing ligand. Bioengineered polymers that have been produced using this approach range from simple chain polymers to complex, multidomain proteins that exhibit highly sensitive responses to external stimuli.[7]

Complex linear polymers have been designed to control the biological activity of proteins in response to stimuli such as light or pH, and hydrogels offer equally useful applications. A critical factor in the formation of a functional hydrogel depends upon

interactions between chains that are sufficiently strong to form junction points, while preventing the exclusion of solvent to avoid precipitation out of solution. With the goal of developing a hydrogel material exhibiting strong interactions between proteins at junction points while maintaining chain solubility, Petka et al. employed recombinant DNA methods to engineer triblock polymers with multidomain characteristics.[8] Two binding domains were flanked by a water soluble, random coil peptide with amino acid sequence $(AGAGAGPEG)_{10}$. The resulting synthetic polymer maintained a phase equilibrium in which either a gel or a viscous liquid is favored as a consequence of changes in pH and temperature (Figure 6.1). This polymer is just one example of biomaterials that have been engineered by molecular genetics to respond predictably to these specific stimuli.

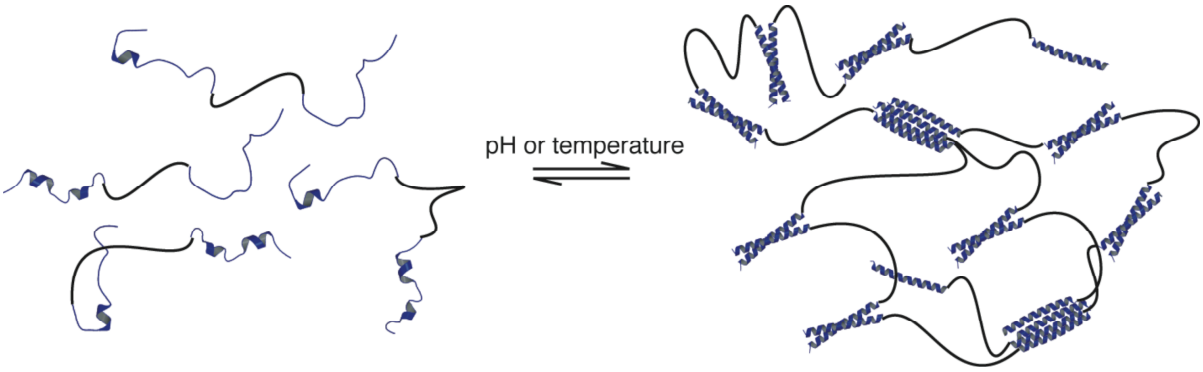

Figure 6.1: Temperature and Acid-Sensitive Hydrogel. Triblock proteins are comprised of a soluble, random coil peptide that is flanked by leucine zipper domains. Solutions of these triblock proteins exhibit visco-elastic properties at high pH and temperature (left). Decreases in pH or temperature favor the formation of leucine zipper interactions, shifting the equilibrium toward the hydrogel phase (right)
(Adapted from Ref #8)

While many reversible self-assembling polymers mimic or directly incorporate elements of natural biomaterials to provide mechanical strength, few of these materials capture the exquisite chemical-sensing properties of biological systems.[3] Yet, the ability

to sense and respond to environmental changes is a hallmark of living systems, which often employ proteins to recognize and relay chemical signals. For example, changes in Ca^{2+} concentration mediate several biological functions, including neuronal communication and muscle contraction. Many Ca^{2+}-sensitive pathways are regulated by the protein calmodulin (CaM), which upon binding four Ca^{2+} ions, undergoes a conformational change that allows it to bind to one of over 100 different calmodulin binding domains (CBDs) within other proteins.[9-12] This process reverses upon removal of calcium, and it is well characterized both structurally and biochemically.

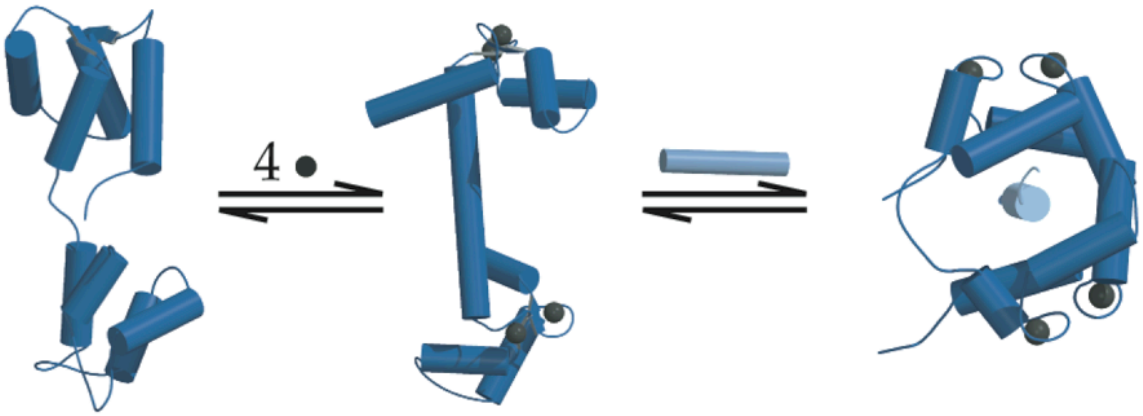

Figure 6.2: Calmodulin Structures. Calmodulin (CaM) undergoes a conformational change in the presence of calcium ions to reveal a methionine-rich hydrophobic surface. Introduction of a CaM-binding domain (CBD) induces another conformational change as Ca^{2+}-loaded CaM complexes with the CBD.

Ehrick et al.[13] recently created a Ca^{2+}-sensitive hybrid hydrogel by a free-radical copolymerization of a chemically-modified CaM with acrylamide, bis-acrylamide, and an acryloyl-substituted phenothiazine. Removal of Ca^{2+} disrupted physical cross-links between the pendent CaM and phenothiazine groups and swelled the gel. While some bulk properties of these materials could be tuned by adjusting the monomer stoichiometries, the placement of functional groups in a given polymer chain is

essentially random. Moreover, because the binding of phenothiazine to CaM mediates the Ca^{2+} response, opportunities for systematic tuning may be limited to subtle modifications of the phenothiazine scaffold.

We anticipated that a fully genetic approach would provide several important advantages to creating and systematically studying Ca^{2+}-sensitive self-assembling materials. First, genetic engineering provides precise control over the composition of a material. Second, gene sequences are highly modular and amenable to rapid combinatorial materials synthesis. Finally, genetic engineering provides access to the great diversity of natural and engineered CBDs, which can bind to CaM as monomers or dimers, with affinities spanning 5-orders of magnitude (K_d 10^{-7}–10^{-12} M), and with varying Ca^{2+}-dependencies.[9,10,14-17] This diversity enables systematic variations in network topology, junction strength, and Ca^{2+} sensitivity to create materials with new properties.

Here, we describe a genetic toolbox of natural and engineered protein modules that can be rationally combined in manifold ways to create reversible self-assembling materials that vary in their composition, architecture, and mechanical properties. Using this toolbox, we created several proteinaceous materials that reversibly self-assemble in the presence of Ca^{2+} and characterized these materials using microrheology.

6.2 Results and Discussion

6.2.1 A Genetic Toolbox for Materials Design

The reversible sensing behavior of CaM and the wide availability of CBDs suggested that it may be possible to create Ca^{2+}-sensing biomaterials based on CaM. To

build reversible Ca^{2+}-sensitive networks, we created a genetic toolbox comprised of several modules that were chosen to perform specific functions in a material (Figure 6.3). CaM serves as our calcium sensor, adopting a conformation capable of binding CBDs in the presence of micromolar concentrations of Ca^{2+}. The wide diversity of CBDs enables us to vary the stoichiometry and strength of the calcium-sensitive junction points. Leucine zipper domains, which may be acid-sensitive, can be combined with other modules to increase the number of junctions in a given network.[8] Finally, linkers of varying lengths and flexibility can be used to attach one module to another to construct triblock proteins. All of these modules can be encoded genetically and expressed in *E. coli* cells, either individually or as components of multi-block proteins.

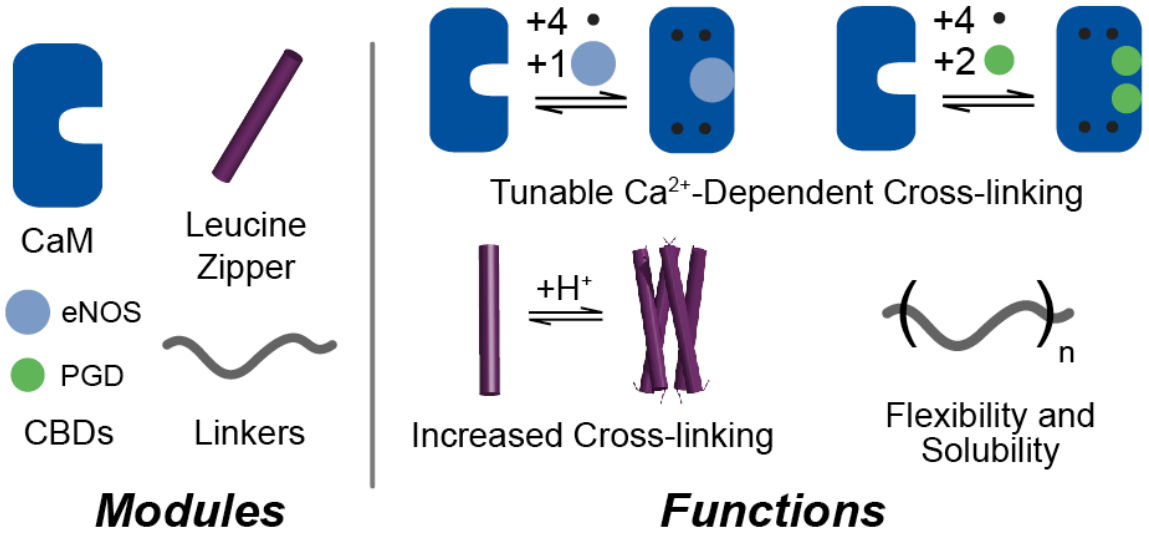

Figure 6.3: A Genetic Toolbox for Stimuli-Sensitive Biomaterials. Each module in the genetic toolbox encodes a specific function. Modules can be combined in many ways to produce functional materials.

The toolbox enables the construction of many architectures. For example, one might imagine covalently linking CaM to a CBD through a flexible spacer to create a triblock polymer. Addition of Ca^{2+} would permit the CaM module of one triblock to bind

the CBD module of another triblock, resulting in a linear polymerization of the material (Figure 6.4). The length and structural characteristics of the linker can be varied to favor intermolecular binding over nonproductive, intramolecular binding. Removal of Ca^{2+} through chelation would cause the polymer to separate into monomer units, leading to a solution of diminished viscosity.

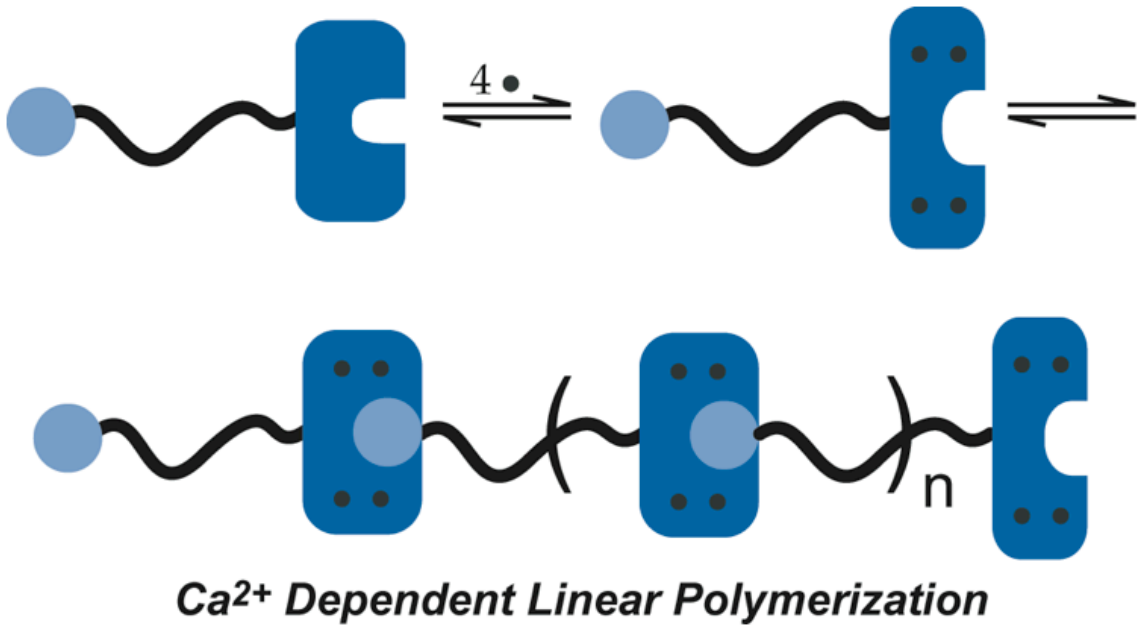

Figure 6.4: Ca^{2+}-Dependent Linear Polymerization. A triblock protein comprised of a CBD linked to CaM polymerizes in the presence of Ca^{2+} to form a linear polymer. Removal of the calcium ions through chelation induces a return to the unpolymerized triblock units.

Returning to our toolbox, the anticipated architecture can be varied to produce more complex polymers. While most CBDs bind to CaM in a 1:1 stoichiometry, several bind CaM in other stoichiometries. A peptide sequence from petunia glutamate decarboxylase (PGD) is known to bind to CaM in a 2:1 stoichiometry.[16] Therefore, the incorporation of this peptide into the above framework would result in a hyperbranched

topology upon Ca^{2+} binding (Figure 6.5). This hyperbranched polymer could be further crosslinked with a triblock comprised of a CBD linked to another CBD. These are just a few examples of the many architectures that one might choose to investigate using our genetic toolbox. Importantly, triblock proteins containing CaM should not bind CBDs during expression in *E. coli*, since calcium binding and activation of CaM would not be favored at intracellular Ca^{2+} concentrations (< 300 nM).

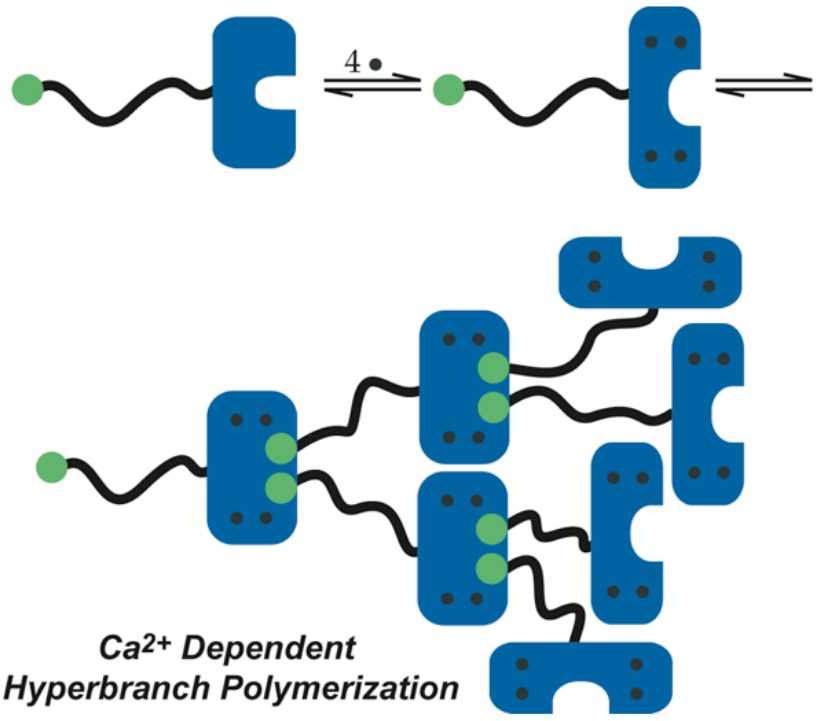

Figure 6.5: Ca^{2+}-Dependent Hyperbranched Polymerization. A triblock protein comprised of a CBD linked to CaM polymerizes in the presence of Ca^{2+} to form a linear polymer. Removal of the calcium ions through chelation induces a return to the unpolymerized triblock units.

While this toolbox enables the construction of many modular proteins, our initial studies employed several classes of triblock proteins to explore the effects of 1) varying the stoichiometry of junction points, and 2) varying the interaction strength of junction

points. To pursue these studies, genes encoding the modules were constructed with compatible restriction sites to allow rapid assembly.[18,19] Full-length genes were expressed in *E. coli*, and the resulting proteins were purified in high yields (typically >100 mg/L).

6.2.2 Varying the Stoichiometry of Junction Points

To form Ca^{2+}-sensitive junctions, CaM was used in tandem with either the CBD from human endothelial NO synthase,[14,20] which binds as a monomer ($K_d \approx 3\times10^{-9}$ M, eNOS), or with the CBD from petunia glutamate decarboxylase,[15,16] which binds as a dimer ($K_d \approx 2\times10^{-8}$ M, PGD). A previously described tetrameric leucine zipper domain[8] (Zip) and hydrophilic protein sequence[8] $((AG)_3PEG)_8$ were used to provide additional branching and to link components together, respectively (Figure 6.6).

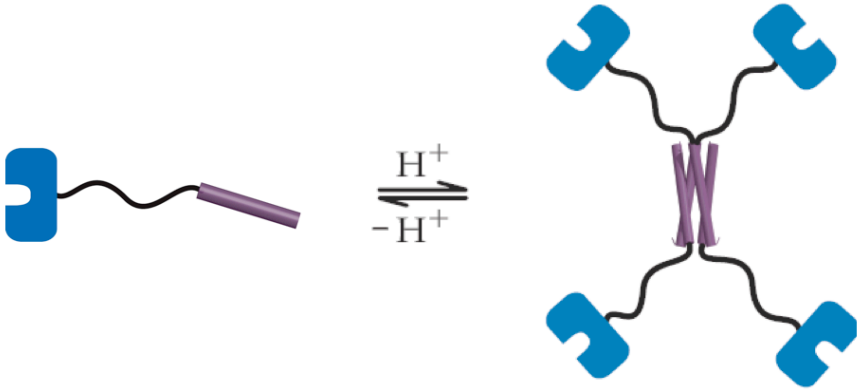

Figure 6.6: Leucine Zipper Domains Provide Additional Branching. A triblock protein comprised of a leucine zipper peptide linked to CaM forms an acid-sensitive tetrameric complex. This complex provides additional branching for improved network formation.

To form Ca^{2+}-sensitive junction points for network formation, we constructed triblock crosslinkers by using the hydrophilic linker to connect a CBD that binds CaM in

a 1:1 stoichiometry to a different CBD that binds CaM in a 2:1 ratio (Figure 6.7). Neither of these CBDs binds CaM in the absence of calcium ions. We anticipated that we could create a more viscous network by increasing the crosslinking potential of the system. Toward this end, a second crosslinker was designed to incorporate two identical 2:1 CBDs at each end of the hydrophilic linker (Figure 6.7). We constructed two lengths for each of these crosslinkers, such that the hydrophilic linker between the two CBDs consisted of either 8 or 40 repeats of a 9 amino acid sequence.

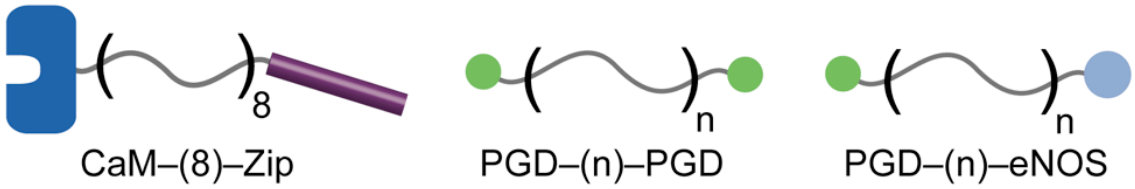

Figure 6.7: Triblock Proteins to Vary Junction Point Stoichiometry. A triblock protein comprised of a leucine zipper peptide linked to CaM forms an acid-sensitive tetrameric complex (left). These tetramers can form extended networks with triblock crosslinkers composed of a 2:1 CBD linked either to an identical 2:1 CBD (middle) or to a 1:1 CBD (right).

After expressing and purifying these proteins in high yield, particle-tracking microrheology was used to determine the viscosity of these samples in the presence and absence of Ca^{2+} or Mg^{2+} (Figure 6.8).[21,22] Although the 2:1-1:1 crosslinker is unable to cap a single CaM protein, the samples did not become viscous for the short linker length. The binding properties of these CBDs support a model where 2 different crosslinkers bind first at their 1:1 CBD end. Then, the 2:1 CBD ends of each crosslinker are at high effective concentrations and bind the third CaM, since these 2:1 CBDs are known to bind strongly only in pairs. Thus, if an additional 1:1 CBD binds, the tetramer is entirely nonproductive for extending network formation. If two 2:1 CBPs bind the fourth CaM

instead, the result is a linear architecture at this tetramer junction. When we extended the length of the 2:1-1:1 crosslinker so that it could span the tetramer distance at the experimental concentration, network formation does occur. This result indicates that intra-tetramer crosslinking is prevalent, and that perhaps increasing concentration significantly would improve inter-tetramer crosslinking for this bifunctional triblock. We tested this hypothesis and determined that as the total concentration and the crosslinker ratio are increased, the viscosity also increases for this system.

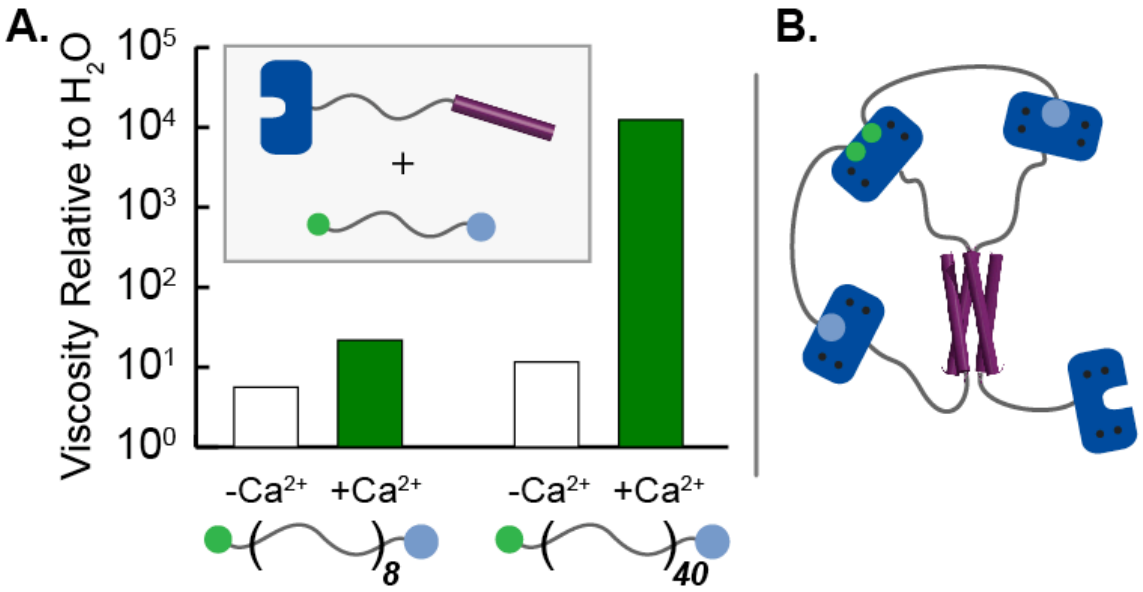

Figure 6.8: Crosslinker Length Can Affect Network Formation. (A) Viscosities of equimolar mixtures of PGD–(n)–eNOS and CaM–(8)–Zip in the absence and presence of Ca^{2+}. (B) Model for capping of CaM–(8)–Zip with the short PGD–(8)–eNOS cross-linker.

While the length of the hydrophilic linker spanning each CBD pair proved to be extremely significant for 1:1-2:1 crosslinker, the length of the hydrophilic linker did not greatly affect the ability of the 2:1-2:1 crosslinker to form extended networks. Despite the potential that the 2:1-2:1 crosslinker could exhibit significant capping of single CaMs

at short linker lengths, these 2:1-2:1 crosslinkers formed highly crosslinked networks for both lengths (Figure 6.9). These results indicated that intra-tetramer crosslinking occurs infrequently for the 2:1-2:1 architecture, or is less likely to lead to entirely nonproductive or linear architectures at the concentrations used in this experiment. We also found that superstoichiometric concentrations of the 2:1-2:1 crosslinker relative to CaM were less ideal than when we tested network formation of crosslinker to CaM at stoichiometric concentrations.

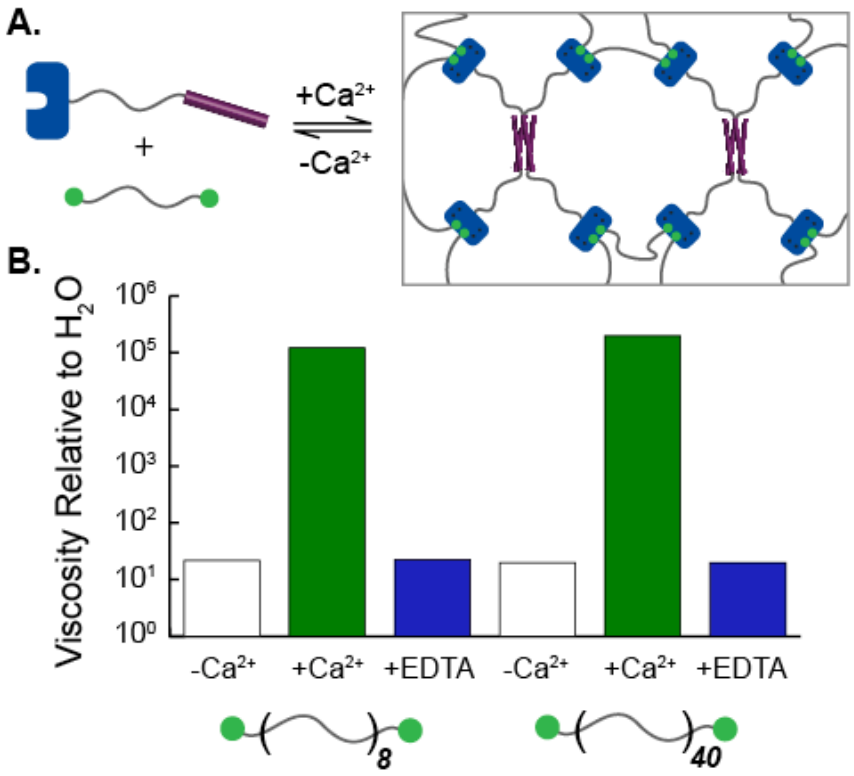

Figure 6.9: Increasing Junction Point Stoichiometry Enhances Network Formation. (A) Idealized representation of network formation between an equimolar mixture of PGD–(n)–PGD and CaM–(n)–Zip. (B) Viscosities of equimolar mixtures of PGD–(n)–PGD and CaM–(8)–Zip in the absence of Ca^{2+}, the presence of Ca^{2+} (30 mM), and after Ca^{2+} chelation by EDTA.

We were somewhat surprised that networks formed with PGD-(8)-PGD appeared to form highly crosslinked networks equal to those formed with PGD-(40)-PGD. We

anticipated that there should be a concentration at which PGD-(8)-PGD could no longer span the distance from one tetramer to it nearest neighbor. At this concentration, the CaM-containing tetramers would be so distant that the free end of a crosslinker with one CBD already bound would be more likely to bind a CaM protein on the same tetramer than one on a neighboring tetramer. This event would severely reduce network formation. To test this idea, we prepared concentrated solutions of CaM-(8)-Zip and PGD-(8)-PGD and diluted this solution to measure the viscosities of samples over a range of concentrations (Figure 6.10). We observed an abrupt increase in the ability to form highly crosslinked networks when the concentration of CaM reaches 2.1 mM (equivalent to a CBD concentration of 4.2 mM), which is consistent with a simple model (rms distances of the random coil linkers) to predict the concentration at which this effect could be observed.

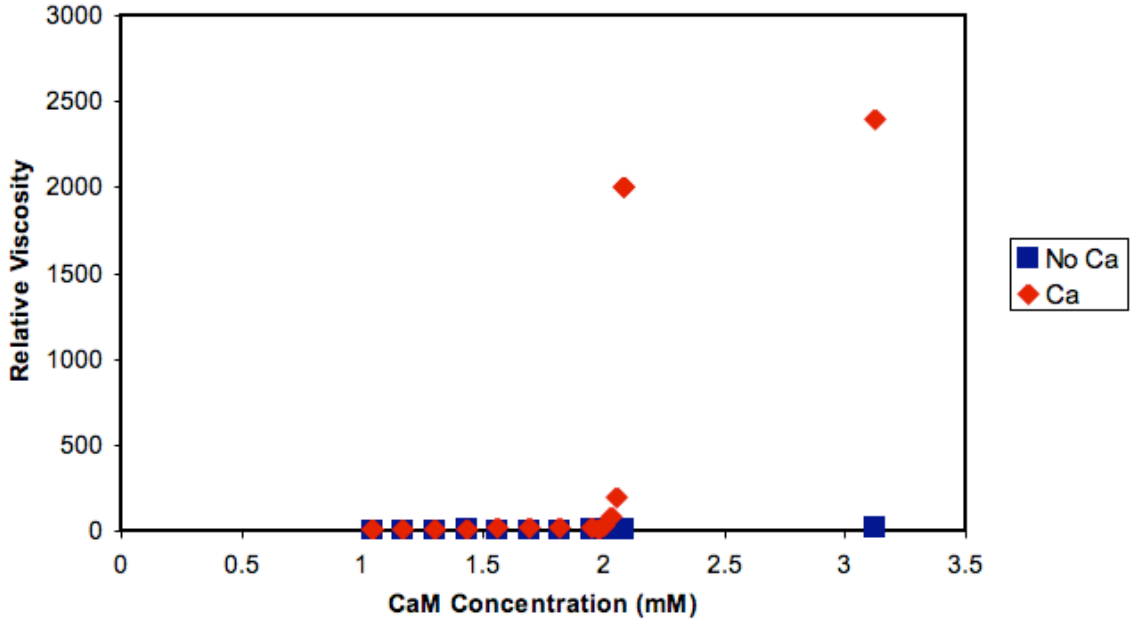

Figure 6.10: Protein Concentration Is Critical for Network Formation. Solutions of CaM-(8)-Zip and PGD-(8)-PGD do not have the potential to form extended networks until the concentration of CaM reaches 2.1 mM (red diamonds). The viscosity of the solutions before Ca^{2+} addition is shown for each concentration (blue squares).

In contrast to the 2:1-1:1 crosslinkers, the 2:1-2:1 crosslinkers of both lengths formed Ca^{2+}-sensitive, reversible networks that maintained their shape against gravity when probed in bulk with a pipet tip (Figure 6.11) Both 2:1-2:1 crosslinkers formed networks that were extremely viscous (100-200 thousand times more viscous than water) at long timescales. Additionally, microrheology showed that the tracer particles were motionless at our resolution limit during timescales shorter than 10 seconds.

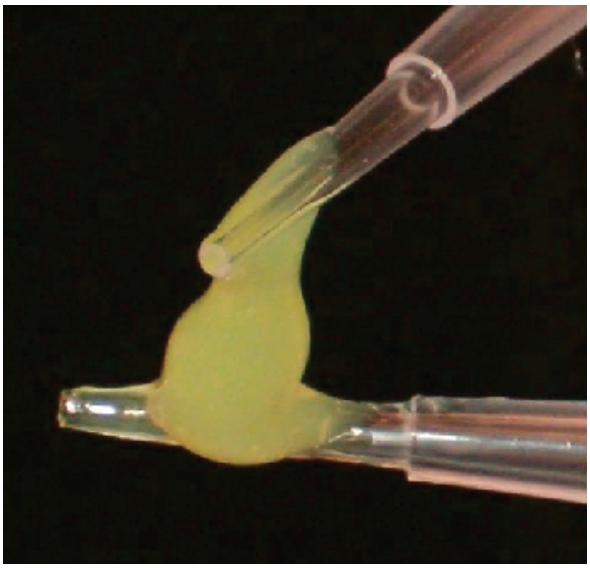

Figure 6.11: A Ca^{2+}-Sensitive Material Maintains Shape. The materials described in Figure 6.9 maintain their shape against gravity on the timescale of several minutes.

Interestingly, we were able to use a microscope to watch the effects of Ca^{2+} and EDTA on the motion of tracer particles within these samples in real time. It was evident that the particles became nearly immobile as the Ca^{2+} front passed through the sample and permitted network formation via CBD-CaM binding. We observed the opposite effect when EDTA was added to chelate the Ca^{2+}, reversing network formation and causing the tracer particles to become highly diffusive.

To further establish that the observed viscosity changes were the result of the specific assemblies described above rather than other less-specific interactions, we performed several additional control experiments. To confirm the essential nature of the leucine zipper domains, we prepared mixtures analogous to those in Figure 6.9A, but with the Zip domain deleted (Cam-(8) + PGD–(8)–PGD). Addition of Ca^{2+} to this solution produced no detectable change in viscosity (Figure 6.12A).

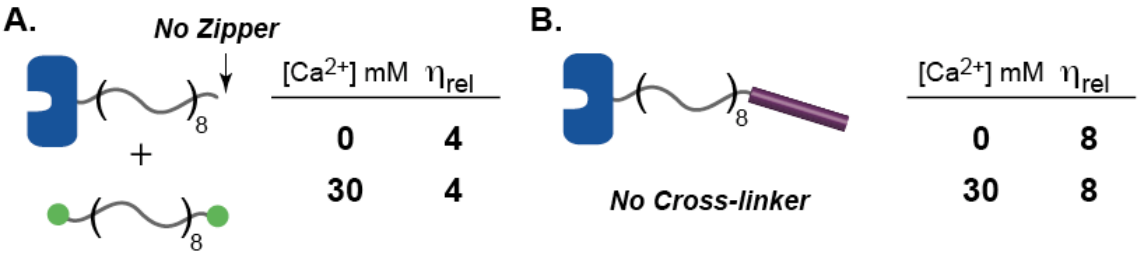

Figure 6.12: Solutions Lacking Key Modules Remain Non-Viscous. (A) An equimolar mixture of CaM-(8) + PGD-(8)-PGD (each 1.7 mM) does not show significant changes in viscosity upon addition of Ca^{2+} (30 mM) indicating the importance of the zipper domain. (B) A solution of CaM-(8)-Zip (2.6 mM) does not show significant changes in viscosity upon addition of Ca^{2+} (30 mM) indicating the necessity of the cross-linker for extended network formation. Viscosities are reported relative to H_2O.

To establish the role of the bifunctional cross-linker, we prepared a solution of CaM-(8)-Zip that lacked the cross-linker. Addition of Ca^{2+} had no significant effect on viscosity, which indicates the necessity of the cross-linker (Figure 6.12B). Addition of Ca^{2+} to solutions of the cross-linkers alone induces aggregation of the anionic colloidal particles used for particle tracking studies; this effect is not observed in solutions containing CaM, which presumably results from its high affinity for Ca^{2+}. Although this aggregation prevented us from obtaining microrheological data, visual analysis revealed that these samples acted like liquids: The solutions could be pipetted easily in the presence of Ca^{2+}, whereas the samples reported in Figure 6.9 cannot be pipetted in the

presence of Ca^{2+}. We note that cationic microspheres are available, however, our protein linkers are polyanions and we wished to avoid non-specific charge-charge interactions with the microspheres.

6.2.3 Varying the Binding Strength of Junction Points

The diversity of natural and engineered CBDs provides ample opportunity to assess whether the properties of a protein-based material can be tuned in a rational and predictable manner by varying the junction point strength. We anticipated that networks featuring weak junction points may exhibit weak visco-elastic behavior, while networks with strong junction points may enable the formation of a material with hydrogel properties. To test this hypothesis, we performed a study in which CaM-(8)-Zip was crosslinked with bifunctional crosslinkers featuring CBD domains that bind Ca^{2+}-loaded CaM with different affinities (K_d=4.5 x 10^{-7} M for CAD[17]; K_d=3.7 x 10^{-9} M for mLCK[12]; K_d=2.2 x 10^{-12} M for the N9A mutant[12] of the mLCK peptide) (Figure 6.13).

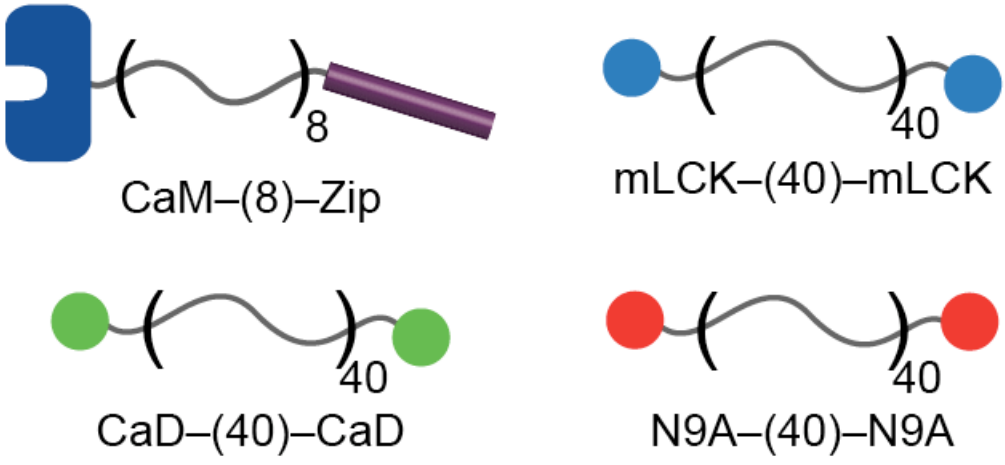

Figure 6.13: Triblock Proteins to Vary Junction Point Binding Strength. CaM-(8)-Zip was tested with three bifunctional crosslinkers, each of which binds CaM with different affinities in the presence of Ca^{2+}.

We anticipated that the crosslinker featuring the CaD CBD would form weaker networks in the presence of calcium ions than the crosslinker featuring the mLCK CBD or the mLCK mutant, N9A. To test this hypothesis, we expressed and purified these proteins, and used particle-tracking microrheology to determine the viscosity of these samples in the presence and absence of Ca^{2+}. Solutions containing CaD-(40)-CaD showed slight increases in viscosity upon addition of Ca^{2+} (~200 η_{water}), while solutions containing either mLCK-(40)-mLCK or N9A-(40)-N9A showed more complex behavior. Figure 6.14 shows an image of solutions containing these proteins 2.5 hours after the addition of Ca^{2+}, as well as 2 days later.

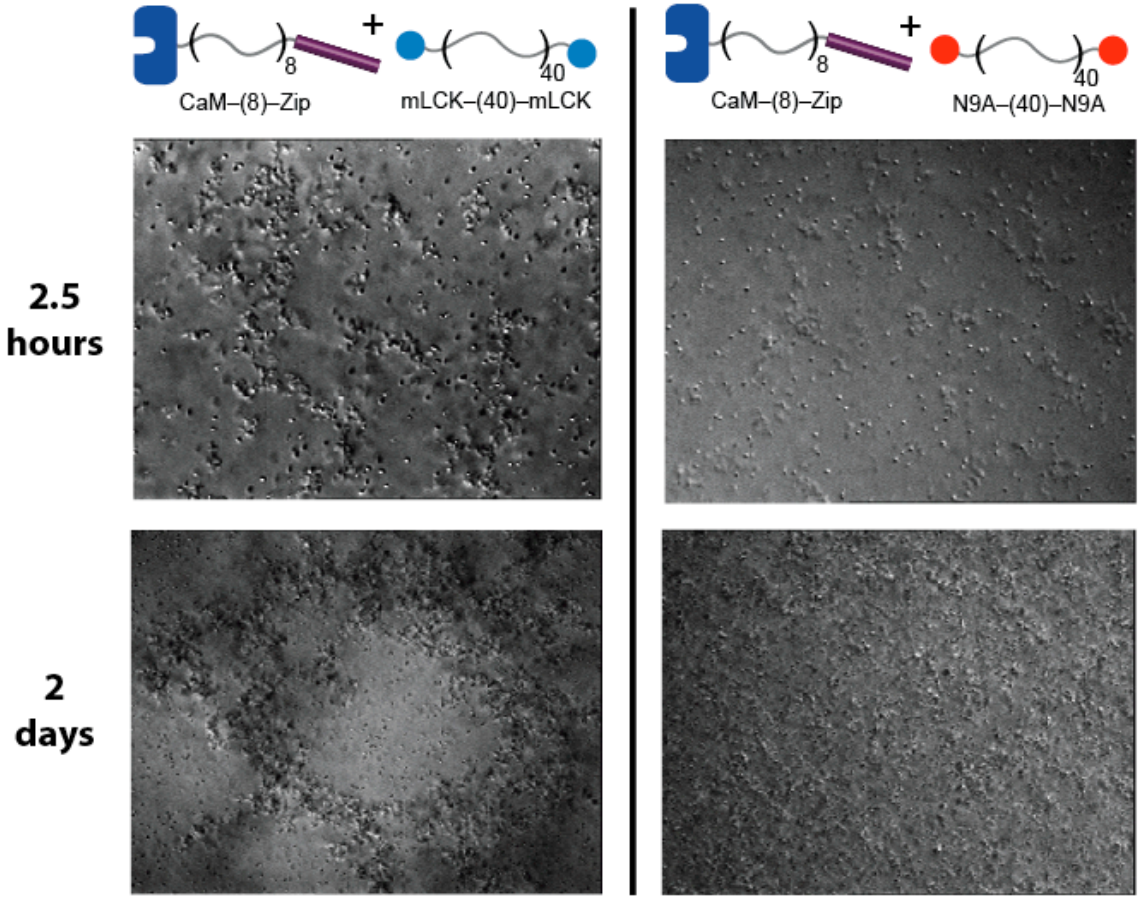

Figure 6.14: Images of Solutions Containing Nanomolar or Picomolar Crosslinkers. CaM-(8)-Zip was tested with mLCK-(40)-mLCK or N9A-(40)-N9A, which appear to form different macrostructures.

The solution on the left contains the bifunctional crosslinker with nanomolar affinity for CaM (mLCK), while solution on the right contains the crosslinker that binds CaM with picomolar affinity (N9A). Although both solutions contained fluorescent tracer particles for tracking, samples prepared without these particles formed similar structures, indicating that the structures were not caused by tracer particle aggregation. Interestingly, clumps of material begin to form for both solutions, although the solution containing the mLCK bifunctional crosslinker appears to form more distinct structures. We repeated this experiment with freshly prepared proteins on several occasions, and the results were similar each time.

The analysis of these samples proved challenging, as the complex structures and channels provided numerous paths through which the tracer particles could flow. Figure 6.15 shows an image taken 4 days after adding Ca^{2+} to a solution of CaM-(8)-Zip and mLCK-(40)-mLCK (left). The picture on the right shows the path of tracer particles over the course of the video. Some particles flow through the channels, while other particles are embedded within the structured material and remain relatively motionless.

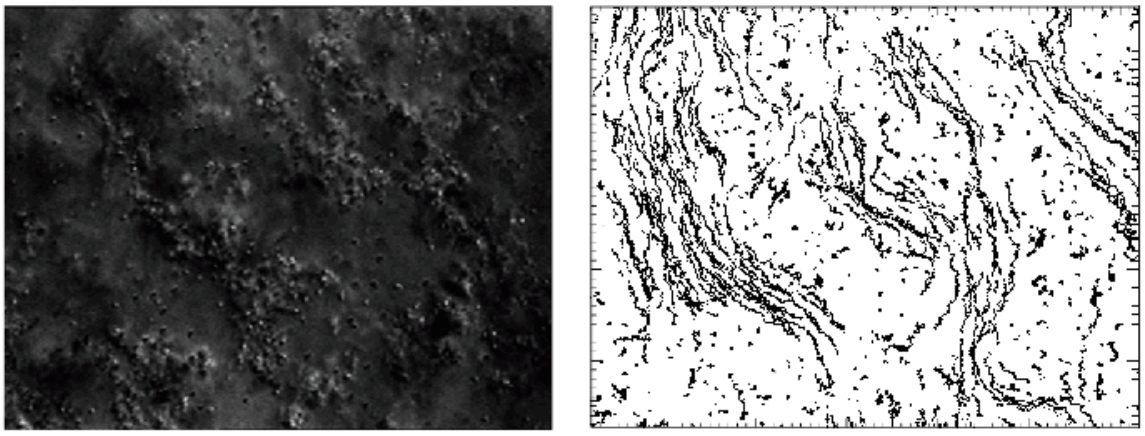

Figure 6.15: Tracer Particles Exhibit Complex Behavior in Structured Materials. CaM-(8)-Zip was tested with mLCK-(40)-mLCK (left). Some tracer particles flow, while others remain very still (right).

Typically, it is straightforward to subtract flow from the motion of tracer particles to calculate their diffusion. However, when flow is not constant across an image, as is especially true in this case, it is impossible to accurately quantity the viscosity of the solution. However, it is relatively simple to analyze the behavior of particles embedded within the material structure by removing particles that exhibit flow from our calculations. After isolating the motions of embedded particles from those of the flowing particles, we calculated the mean square displacement (Δr^2) at various lag times (Δt). This data is then used to determine the viscosity of the solution (black line) or to assess its elastic properties (blue line). Particles embedded within this material appear to reach a maximum Δr^2 at long lag times, suggesting gel behavior (Figure 6.16). This was an intriguing preliminary result that would require further investigation.

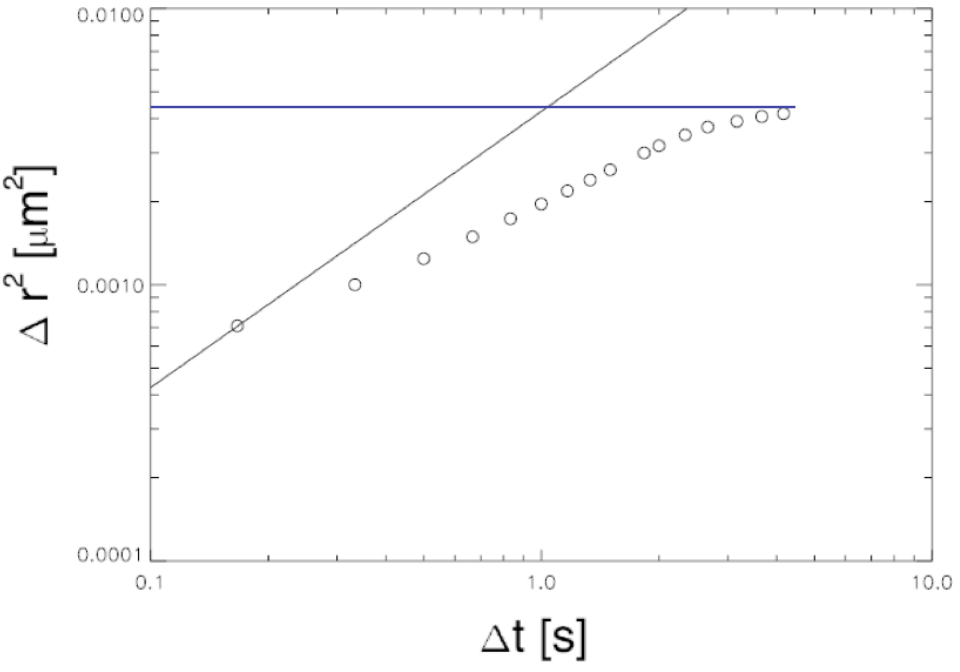

Figure 6.16: Tracer Particles Suggest Gel Behavior. MSD versus lag time for tracer particles embedded in a material formed by adding Ca^{2+} to solutions containing CaM-(8)-Zip and mLCK-(40)-mLCK. A purely viscous material would follow the black line. The data deviate from this line to approach a constant MSD at long lag times. The blue line can be used to assess the elastic component of the material.

6.3 Conclusion

We used a genetic toolbox comprised of natural and engineered protein modules to create several materials that form reversible networks in the presence of Ca^{2+}. Our explorations of binding stoichiometry enabled us to rationally control the architectures, sensing behavior, and rheological properties of protein-based biomaterials. The modular nature of this system is highly amenable to creating precisely tuned materials through variations in composition, network architecture, and junction strength, which may provide opportunities to explore the universality of the gel transition in a controlled and systematic way. Further explorations will be necessary to explore the formation of networks with bifunctional crosslinkers comprised of CBDs with various binding affinities and off-rates.

The ability to develop a firm correlation between the dilute solution behavior of the proteins and the mechanical properties of the materials would provide opportunities to engineer or evolve proteins with desirable self-assembling properties for incorporation into new biomaterials. These efforts may be facilitated by the tremendous diversity of natural and engineered CBD sequences, as well as by the development of CaM variants with altered ion sensitivities or binding properties. Additionally, we anticipate that this modular framework, in which a natural protein sensor directly mediates self-assembly, can be extended to create a variety of new sensing materials using modules from the diverse array of signaling pathways that are mediated by reversible protein-protein interactions. Because these materials incorporate elements that evolved to sense specific biological signals in complex, physiologically relevant environments, they may be useful in the fields of tissue engineering and controlled drug delivery.

6.4 Experimental

Materials and General Considerations

Synthetic oligonucleotides were purchased from Integrated DNA Technologies. Lysozyme, PMSF, and DTT were obtained from Sigma. Antibiotics, $CaCl_2$, $MgCl_2$, and EDTA were purchased from Fisher. Tris base, Tris-HCl, NaCl, and culture media were purchased from EMD Bioscience. IPTG was obtained from US Biological. Imidazole was obtained from Fluka. DNaseI was purchased from Roche. Nickel-agarose resin was obtained from GBioscience. DNA polymerase and restriction enzymes were purchased from New England Biolabs. Purifications of plasmid DNA, PCR products, and enzymatic digestions were performed using kits purchased from Qiagen. Centrifugal concentrating devices were obtained from Millipore. Yellow-green fluorescent (505/515) carboxylate-modified polystyrene spheres (0.5 μm diameter) for particle-tracking microrheology were purchased from Molecular Probes. The gene encoding chicken calmodulin was kindly provided by Prof. A. R. Means (Duke University).

Plasmid manipulations were performed using *E. coli* TOP10F′ cells (Invitrogen) that were transformed by electroporation. All new constructs were verified by DNA sequencing at the NSF-supported Center for Fundamental and Applied Molecular Evolution at Emory University.

Gene Construction

Genes encoding the $((AG)_3PEG)_2$ linker, the leucine zipper, and CBDs were constructed by the extension of overlapping synthetic oligonucleotides with DNA polymerase to create modules that contained restriction sites compatible with seamless

ligation.[18,19] Digestion with appropriate enzymes (typically *Bbs*I or *Bsm*BI) produced fragments that lacked the recognition site and could be subcloned into plasmids with compatible sticky ends to produce the desired in-frame protein sequences. The linker sequences $((AG)_3PEG)_8$ and $((AG)_3PEG)_{40}$ were created in pUC18 by a seamless recursive directional ligation approach beginning with a synthetic sequence encoding the homodimeric sequence $((AG)_3PEG)_2$.

To create vectors for protein expression, we created triblock cassettes (CaM-(n)-Zip, PGD-(n)-PGD, PGD-(n)-eNOS, mLCK-(n)-mLCK, and N9A-(n)-N9A), in which a sacrificial DNA sequence was incorporated as the central block. These cassettes were subcloned into a derivative of the pET14b expression vector (Novagen) that introduced an N-terminal hexahistidine tag. Digestion of these vectors with *Bsm*BI excised the sacrificial sequence and revealed sticky ends that were compatible with linker sequences digested with *Bbs*I and *Bsm*BI. This modular strategy allowed various linkers to be cloned directionally and seamlessly to give genes encoding the desired triblock proteins in a single step.

Protein Expression and Purification

For protein expression, *E. coli* BL21(DE3)/pLysS cells (Novagen) were transformed by electroporation with the desired vector, plated onto selective LB/agar containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL), and grown overnight at 37 °C. Single colonies were used to inoculate 5 mL starter cultures of selective LB broth containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL), which were grown overnight with shaking (300 rpm, 37 °C) and then used to inoculate 250 mL or

600 mL of selective TB containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL) which were grown with shaking (300 rpm, 37 °C). Overexpression of each protein from the T7 promoter was induced by the addition of IPTG (1 mM final concentration) during exponential growth phase (OD_{600} 0.65-0.85). Cultures were chilled on ice after approximately 3 hours, and cell pellets were harvested by centrifugation at 2 °C. Pellets were washed with lysis buffer (50 mM Tris-Cl, 300 mM NaCl, 5 mM imidazole, 0.03 mM EDTA, 1 mM PMSF, pH 7.4), repelleted by centrifugation at 2 °C, and stored at -80 °C.

Pellets were thawed on ice for 10 min, lysis buffer (15-20 mL) containing lysozyme (1 mg/mL) was added, and the suspensions were homogenized by sonication. Following sonication, DNaseI was added to a concentration of 5 μg/mL, the mixture was incubated on ice for 10 min, and the lysate was centrifuged at 2 °C. The cleared lysate was applied to a column containing nickel resin that was pre-equilibrated with lysis buffer. After all lysate had entered the resin, the column was washed with additional lysis buffer (5 column volumes), followed by wash buffer (5 column vol., 50 mM Tris-Cl, 300 mM NaCl, 10 mM imidazole, 0.03 mM EDTA, pH 7.4). Finally, the column was washed with elution buffer (10–12 mL, 50 mM Tris-Cl, 300 mM NaCl, 200 mM imidazole, 0.03 mM EDTA, pH 7.4), and proteins were collected in 1.5 mL fractions that were analyzed by SDS-PAGE with Coomassie Blue staining. Fractions containing pure proteins were combined and concentrated using 4 mL or 15 mL centrifugal concentrating devices with appropriate molecular weight cutoffs. Several cycles of dilution with sample buffer (25 mM Tris-Cl, 0.05 mM EDTA, 1 mM DTT, pH 7.0) and concentration

exchanged the buffers. The concentration of each protein solution was determined by absorbance at 280 nm.

Microrheology Studies

Proteins (final concentrations indicated in main text) were mixed for each experiment, and HCl (1 N) was added to each protein mixture to lower the final pH to 6.0 for each sample. Fluorescent polystyrene spheres (0.5 μm) were added to obtain a final concentration of ~10^7 particles/μL of sample, which resulted in 50-100 particles in the field of view at 100x magnification. Samples were mixed by stirring and were transferred into a chamber (15 mm x 4 mm x 340 μm) constructed with a glass slide and cover slips. Samples were imaged in two dimensions (93 nm/pixel resolution, 30 Hz frame rate) using an inverted microscope (Leica) equipped with a mercury arc lamp and a 100X objective (N.A. = 1.4). In all experiments, data were collected far from the glass surfaces to eliminate wall effects.

To measure cation-dependent viscosity changes, stock solutions of either $CaCl_2$ or $MgCl_2$ (1 M) were added to the samples to give a final concentration of 30 mM (~4 equivalents per Ca^{2+} binding site). For samples in which the addition of the ion did not visibly alter particle diffusion, data were collected after 15 min (6 Hz frame rate). For samples in which the diffusion of the particles slowed dramatically, we could typically observe the Ca^{2+} front in real-time (Figure 6.17), and data were collected after 60 min (2 Hz frame rate). The reversible behavior of the Ca^{2+}-dependent systems was probed by the addition of excess EDTA, which rapidly diffused through the samples; data were collected after the sample appeared homogeneous.

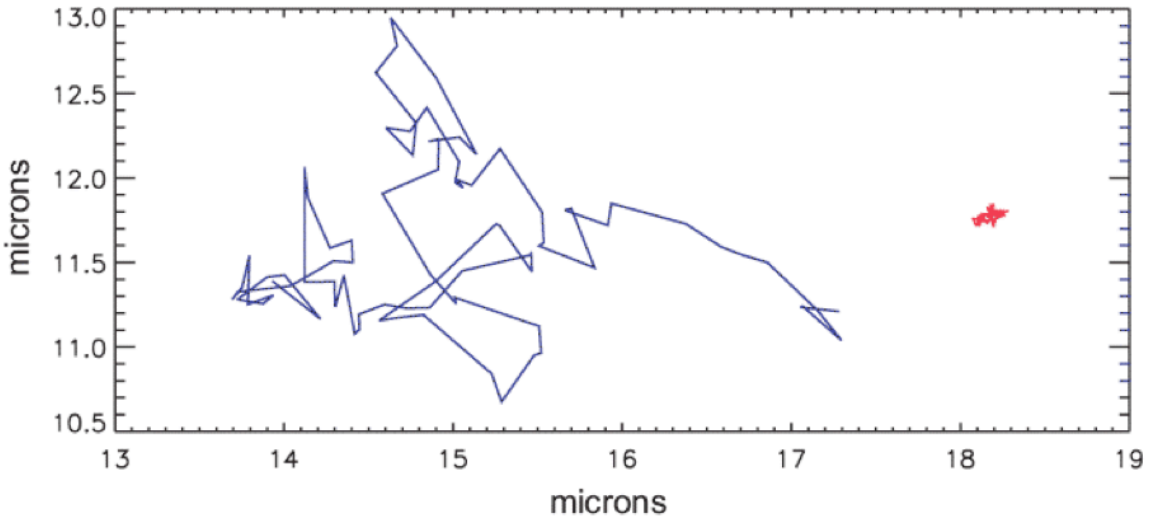

Figure 6.17. Comparison of Particles Trajectories at the Ca^{2+}-Front. The trajectories of two particles on opposite sides of the diffusing Ca^{2+} front produced by the addition of Ca^{2+} to an equimolar mixture of CaM-(8)-Zip and PGD-(40)-PGD. The duration of the trajectories is 12 s. The particle at left (blue) is in a region where the sample is still fluid, while the particle at right (red) is in a region where Ca^{2+} ions have induced polymerization and increased the local viscosity by approximately 500 times. Both trajectories were observed simultaneously approximately 30 μm apart; they are artificially displaced to be close to one another in the figure.

Images were analyzed using macros within the Interactive Data Language (IDL), whereby particles were identified and their positions tracked as a function of time. The mean-square displacements (MSDs) of the particles were extracted from their positions. The viscosities of the samples were determined by extrapolating the long lag time linear behavior of the MSDs using the Stokes-Einstein relation for solutions without Ca^{2+} and with Ca^{2+} (Figure 6.18). These materials are purely viscous at long time scales. The highly viscous samples shown in Figure 6.18 (bottom) deviate from this line at short time scales because particles diffuse less than our resolution limit can detect.

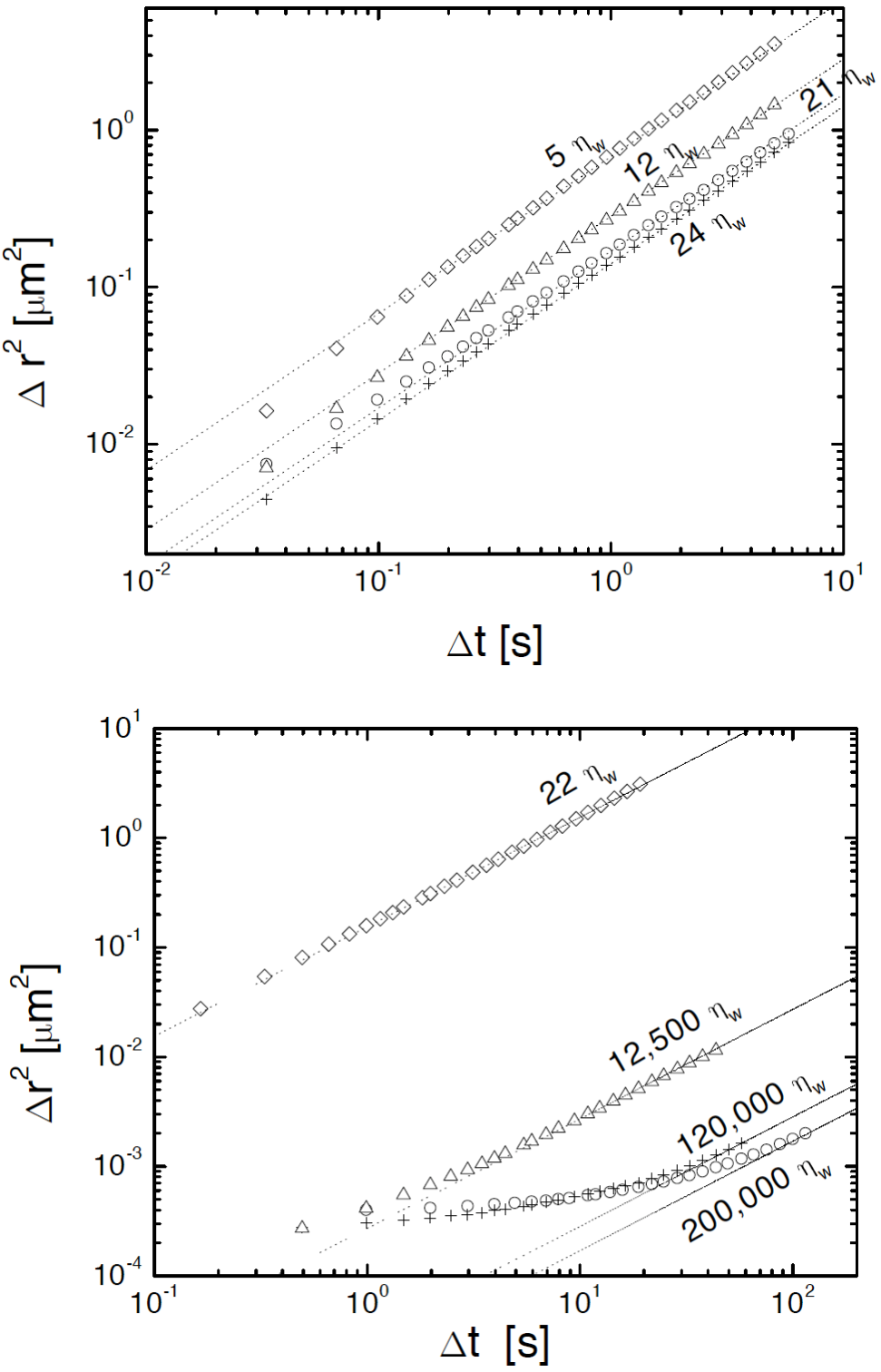

Figure 6.18. Plots of MSD Versus Lag Time. Mean-square displacements versus lag time for samples before (top) and after (bottom) addition of Ca^{2+}. Lines with a slope equal to unity on the log-log plot were extrapolated from the data at long lag times to calculate the viscosities relative to water.
◇ = CaM-(8)-Zip + PGD-(8)-eNOS; △ = CaM-(8)-Zip + PGD-(40)-eNOS;
+ = CaM-(8)-Zip + PGD-(8)-PGD; ○ = CaM-(8)-Zip + PGD-(40)-PGD.

Protein Sequences

MH$_6$-**CaM-**(8)-*Zip*

MHHHHHH**AADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTI**
DFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDG
QVNYEEFVQMMTAK(AGAGAGPEG)$_8$*AGSGDLENEVAQLEREVRSLEDEAAELEQKVSRLKNEIEDLKAE*

MH$_6$-**PGD**-(n)-**PGD**

MHHHHHH**AGHKKTDSEVQLEMITAWKKFVEEKKKK**(AGAGAGPEG)$_n$**AGHKKTDSEVQLEMITAWKKFVEEK**
KKK

MH$_6$-**PGD**-(n)-*eNOS*

MHHHHHH**AGHKKTDSEVQLEMITAWKKFVEEKKKK**(AGAGAGPEG)$_n$*AGRKKTFKEVANAVKISASLMGAER*
LI

MH$_6$-**CaD**-(n)-**CaD**

MHHHHHH**AGGVRNIKSMWEKGNVFSS**(AGAGAGPEG)$_n$**AGGVRNIKSMWEKGNVFSS**

MH$_6$-**mLCK**-(n)-**mLCK**

MHHHHHH**AAARWKKNFIAVSAANRFKKIS**(AGAGAGPEG)$_n$**AAARWKKNFIAVSAANRFKKIS**

MH$_6$-**N9A**-(n)-**N9A**

MHHHHHH**AAARWKKAFIAVSAANRFKKIS**(AGAGAGPEG)$_n$**AAARWKKAFIAVSAANRFKKIS**

6.5 References

(1) Langer, R.; Tirrell, D.A. *Nature* **2004,** *428*, 487-492.

(2) Roy, I.; Gupta, M.N. *Chem. Biol.* **2003,** *10*, 1161-1171.

(3) Kopecek, J. *Eur. J. Pharm. Sci.* **2003,** *20*, 1-16.

(4) McMillan, R.A.; Lee, T.A.T.; Conticello, V.P. *Macromolecules* **1999,** *32*, 3643-3648.

(5) Deming, T.J. *Nature* **1997,** *390*, 386-389.

(6) Nagarsekar, A.; Crissman, J.; Crissman, M.; Ferrari, F.; Cappello, J.; Ghandehari, H. *Biomacromolecules* **2003,** *4*, 602-607.

(7) Haider, M.; Megeed, Z.; Ghandehari, H. *J. Control. Release* **2004,** *95*, 1-26.

(8) Petka, W.A.; Harden, J.L.; McGrath, K.P.; Wirtz, D.; Tirrell, D.A. *Science* **1998,** *281*, 389-392.

(9) Bhattacharya, S.; Bunick, C.G.; Chazin, W.J. *Biochim. Biophys. Acta.* **2004,** *1742*, 69-79.

(10) Crivici, A.; Ikura, M. *Annu. Rev. Biophys. Biomol. Struct.* **1995,** *24*, 85-116.

(11) Ikura, M.; Ames, J.B. *Proc. Natl. Acad. Sci.* **2006,** *103*, 1159-1164.

(12) Montigiani, S.; Neri, G.; Neri, P.; Neri, D. *J. Mol. Biol.* **1996,** *258*, 6-13.

(13) Ehrick, J.D.; Deo, S.K.; Browning, T.W.; Bachas, L.G.; Madou, M.J.; Daunert, S. *Nat. Mater.* **2005,** *4*, 298-302.

(14) Venema, R.C.; Sayegh, H.S.; Kent, J.D.; Harrison, D.G. *J. Biol. Chem.* **1996,** *271*, 6435-6440.

(15) Yap, K.L.; Yuan, T.; Mal, T.K. *J. Mol. Biol.* **2003,** *328*, 193-204.

(16) Yuan, T.; Vogel, H.J. *J. Biol. Chem.* **1998,** *273*, 30328-30335.

(17) Zhuang, S.; Wang, E.; Wang, C.-L.A. *J. Biol. Chem.* **1995,** *270*, 19964-19968.

(18) Goeden-Wood, N.L.; Conticello, V.P.; Muller, S.J.; Keasling, J.D. *Biomacromolecules* **2002,** *3*, 874-879.

(19) Padgett, K.A.; Sorge, J.A. *Gene* **1996,** *168*, 31-35.

(20) Matsubara, M.; Hayashi, N.; Titani, K.; Taniguchi, H. *J. Biol. Chem.* **1997,** *272*, 23050-23056.

(21) Crocker, J.C.; Grier, D.G. *Phys. Rev. Lett.* **1996,** *77*, 1897-1900.

(22) Mason, T.G.; Weitz, D.A. *Phys. Rev. Lett.* **1995,** *74*, 1250-1253.