

### **Distribution Agreement**

In presenting this thesis or 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 thesis or 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 thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

---

Jingzhi Wang

---

Date

**Understanding Synthetic Riboswitch Mechanism and Creating Predictive  
Biophysical Model for Designing Novel Synthetic Riboswitches**

By  
Jingzhi Wang  
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 James T. Laney School of Graduate Studies

---

Date

**Understanding Synthetic Riboswitch Mechanism and Creating Predictive  
Biophysical Model for Designing Novel Synthetic Riboswitches**

By

Jingzhi Wang  
B.S. Sichuan University

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  
2015

## **Abstract**

### **Understanding Synthetic Riboswitch Mechanism and Creating Predictive**

### **Biophysical Model for Designing Novel Synthetic Riboswitches**

By Jingzhi Wang

As an emerging class of genetic regulatory elements, synthetic riboswitch has become a handy tool that allows conditional regulation of gene expression and facilitates the downstream application with great versatility. Previous efforts have successfully expanded the repertoire of synthetic riboswitches; however, the knowledge of synthetic riboswitch mechanism remains unclear. In this thesis, we present a detailed molecular study to reveal mechanistic insights of our model theophylline riboswitch. This study helps to construct the first biophysical model of synthetic riboswitch that has predictive power. Meanwhile, the kinetic aspect of riboswitch mechanism is analyzed to support the hypothesis that riboswitch function contains both thermodynamic and kinetic components. Furthermore, we implement our biophysical model to design a novel class of synthetic tetramethylrosamine riboswitches. Detailed characterization of these riboswitches is provided to illustrate their mechanism and guide further applications. Each chapter reveals one aspect of riboswitch function and advances our ability to create more efficient synthetic riboswitches.

**Understanding Synthetic Riboswitch Mechanism and Creating Predictive  
Biophysical Model for Designing Novel Synthetic Riboswitches**

By

Jingzhi Wang  
B.S. Sichuan University

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  
2015

## **Acknowledgement**

It is such a bitter-sweet journey to study and graduate from my Ph.D. program. These seven years witnessed me growing up from a young man to a full human being with the help of all the people below.

Firstly, I would like to thank my advisor Dr. Justin P. Gallivan for his help, support, and guidance. I still remember how exciting to receive his email during my application stage for my graduate study. In the tough period when I was grieving my father's death, he generously offered his kindness and support, which makes my life much easier. His humor and wisdom for life will enlighten me beyond my path in the graduate school.

Secondly, I sincerely appreciate the help and guidance from my academic committee members Dr. David Lynn and Dr. Vincent Conticello. In the last year while I worked alone in the lab, they always kept an eye on me and had my back in a series of academic obstacles. Without their encouragement and mentoring, I will never survive through the hard times in my Ph.D. career.

I also owe many thanks to other professors as well as great staffs in our department Ann Dasher, who "babysit" me through the Ph.D. program, Dr. Cora MacBeth, who went to China, recruited me to Emory, gave me a chance to study abroad, and shared her encouragement when I'm down, Dr. Stefan Lutz and Dr. James Kindt who took me in and provided wonderful rotation projects to me, Dr. Emily Weinert and Dr. Khalid

Salaita who shared their wisdom when I defend my yearly report, Patti, Steve, and Jan, who made my research life so smooth.

My graduate life would also not be easy if I missed any of my great senior graduate students or coworkers in Gallivan Lab. Dr. Shana Topp, Dr. Shawn Desai, Dr. Sean Lynch, Dr. Colleen Reynoso, Dr. William Chiuman, Dr. Dennis Mishler, Dr. Adrienne Edwards, Dr. Joy Sinha have provided guidance and critics to me which eventually gave me the faith to go through the graduate program. Meanwhile, Dr. I-lin Wu in Conticello Lab, Dr. Zhuangqun Huang in Lian Lab, Dr. Xianyu Chen in Ryan Lab are not only my great friends, but also mentored me in both academically and mentally.

Apart from the academic life, it would be harder if there were not so many friends around me at Emory. Dr. Chunfu Xu, Dr. Wenting Wu, Dr. Ye Yang,, Chenrui Chen, Dr. Qi Shi, Xuesong Yang, Rui Liu, Dr. Nianhui Song, Dr. Chongchao Zhao, Dr. Patrrick Baldwin, Greg Goesche who shared the same memories of joining the graduate school and study together.

My amazing friends Jingxia Li, Dange Liu, Shiwei Gao, Sijia Li, Weishi Hao, Ruiqi Wu, Yun Wang, Yi Going, Pu Shi, Minghui Cao, Fei Wang, Dr. Haowei Wang, Lan Jin, Dan Kuang, Yuwei Gu, Xinxian Shao, Yao Jing, Yujie Zhou, Naifeisha Nihemaiti, Xin Wang along with other friends, help me to achieve this far. They made me realize the true meaning of life.

Last, my deepest gratitude and regret belong to my dear family especially my mother and my father. In 2010, my father passed away due to late stage liver cancer. It was a sudden and unbearable blow to me and my family. The great grief and heavy responsibility dropped onto my shoulder unexpectedly and made me suffer intensely. It was the strong hearts of my mom and my grandpa which rebuild my faith in life and make me the person who is now graduating and becomes a man. My dear dad, may you have peace in heaven.

I may not have all the talent or the money in the world, but I have you all who cherish me dearly. Thank you all!

## Contents

|  |           |
|--|-----------|
| <b>Chapter 1 - Introduction .....</b>  | <b>1</b>  |
| 1.1 Genetic Regulators and Riboswitches.....   | 1         |
| 1.2 The Discovery of Naturally Occurring Riboswitches and Their General<br>Characteristics ..... | 3         |
| 1.3 Prevalence of Riboswitches Supports the RNA World Hypothesis.....                            | 4         |
| 1.4 Synthetic Riboswitches in Gene Regulation .....  | 9         |
| 1.4.1 The Rationale behind Engineering Riboswitches.....   | 9         |
| 1.4.2 Methods for Developing Synthetic Riboswitches .....  | 11        |
| 1.4.3 Different Approaches to Design Synthetic Riboswitches .....                                | 14        |
| 1.4.4 The Performance of Riboswitch in Model Species .....                                       | 19        |
| 1.5 Different Types of Riboswitch Mechanisms.....  | 21        |
| 1.6 Conclusion.....  | 25        |
| 1.7 Reference.....   | 26        |
| <b>Chapter 2 Investigation into the Mechanism of Theophylline Riboswitches .....</b>             | <b>33</b> |
| 2.1 Introduction .....   | 33        |
| 2.2 Results .....  | 38        |
| 2.2.1 Mutation Studies Reveal Riboswitch Mechanistic Insight.....                                | 38        |

|   |           |
|---|-----------|
| 2.2.2 Characterization of Direct Interaction between RNA Molecule and<br>Theophylline Using In-line Probing Assay ..... | 41        |
| 2.2.3 Construction of Detailed Biophysical Model for Theophylline Riboswitches.   | 45        |
| 2.3 Discussion .....  | 47        |
| 2.4 Experimental .....  | 54        |
| 2.5 References .....  | 59        |
| <b>Chapter 3 Studying Kinetic Aspects of the mechanism of a Theophylline Riboswitch</b><br>.....                        | <b>61</b> |
| 3.1 Introduction .....  | 61        |
| 3.2 Results .....   | 66        |
| 3.2.1 Time Dependence of Theophylline Riboswitch Function .....   | 66        |
| 3.2.2 Investigation to the influence of ligand binding in RNA stability .....   | 69        |
| 3.2.3 <i>In vivo</i> Transcription Translation Decoupling Experiment .....  | 71        |
| 3.3 Discussion .....  | 74        |
| 3.4 Experimental .....  | 79        |
| 3.5 References .....  | 83        |
| <b>Chapter 4 Computer-assisted Screening for Tetramethylrosamine Riboswitches...</b>                                    | <b>85</b> |
| 4.1 Introduction .....  | 85        |

|  |            |
|--|------------|
| 4.2 Results .....  | 91         |
| 4.2.1 <i>In vitro</i> Characterization of Novel Tetramethylrosamine Riboswitches<br>Predicted by Computational Methods .....                           | 91         |
| 4.2.2 <i>In vivo</i> Characterization of Novel Tetramethylrosamine Riboswitches Predicted<br>by Computational Methods .....                            | 92         |
| 4.2.3 <i>In vitro</i> Transcription-Translation Decoupling Experiment of Novel<br>Tetramethylrosamine Riboswitches .....                               | 97         |
| 4.2.4 <i>In vitro</i> and <i>In vivo</i> Characterization of Novel Tetramethylrosamine<br>Riboswitches with Different DNA Template Concentration ..... | 99         |
| 4.3 Discussion .....   | 101        |
| 4.4 Experimental .....   | 107        |
| 4.5 References .....   | 113        |
| <b>Chapter 5 - Conclusions .....</b>   | <b>116</b> |
| 5.1 Summaries and Conclusions .....  | 116        |
| 5.2 References .....   | 119        |

## List of Figures

|  |    |
|--|----|
| Figure 1.1 Naturally Occurring Riboswitch Ligands. ....  | 5  |
| Figure 1.2 Predicted Riboswitch Distribution. ....   | 8  |
| Figure 1.3 SELEX Aptamer Screening Procedure. ....   | 12 |
| Figure 1.4 Synthetic Riboswitch Ligands.....   | 16 |
| Figure 2.1 Proposed Theophylline Riboswitch Mechanism.....   | 35 |
| Figure 2.2 Illustration of In-line Probing Assay.....  | 36 |
| Figure 2.3 Mutation Studies on Predicted Conformations of Theophylline<br>Riboswitch D. ....   | 40 |
| Figure 2.4 Characterization of Riboswitch-ligand Interaction with In-line Probing<br>Assay. ....                                       | 42 |
| Figure 2.5 Modified Biophysical Model of Riboswitch Function.....  | 44 |
| Figure 2.6 Gene Expression Assays of Computationally Predicted Theophylline<br>Riboswitches.....                                       | 46 |
| Figure 2.7 Correlations between Gene Expression and Free Energies in Predictive<br>Biophysical Model of Translational Riboswitch. .... | 53 |

|   |     |
|---|-----|
| Figure 3.1 Transcriptional Control Mechanism of Natural Flavin Mononucleotide<br>Riboswitches.....      | 64  |
| Figure 3.2 Ligand-dependent Response of Theophylline Riboswitch F. ....                                 | 68  |
| Figure 3.3 mRNA Half-life of Riboswitch-Containing mRNA under Different<br>Ligand Concentration. ....   | 70  |
| Figure 3.4 <i>In vivo</i> Transcription Translation Decoupling Experiments. ....                        | 73  |
| Figure 4.1 Tetramethylrosamine and TMR Aptamer. ....  | 88  |
| Figure 4.2 <i>In vivo</i> Characterization of Predicted TMR Riboswitches.....                           | 94  |
| Figure 4.3 Ligand Dose-response Gene Activation of TMR riboswitches. ....                               | 95  |
| Figure 4.4 TMR Riboswitch Performance Comparison of <i>In vivo</i> and <i>In vitro</i> Assays.<br>..... | 96  |
| Figure 4.5 <i>In vitro</i> Transcription Translation Decoupling Experiment of TMR<br>Riboswitches.....  | 98  |
| Figure 4.6 <i>In vitro</i> DNA Dose-response of TMR Riboswitches. ....                                  | 99  |
| Figure 4.7 <i>In vivo</i> Performance of TMR Riboswitches in Various Vector System.                     | 100 |
| Figure 4.8 Predicted TMR Riboswitch Mechanism. ....   | 102 |

## List of Table

|  |    |
|--|----|
| Table 4.1 Synthetic TMR Riboswitch Characteristics. .... | 93 |
|--|----|

## Chapter 1. Introduction

### 1.1 Genetic Regulators and Riboswitches

The ability to detect and respond to chemical signals is essential to the survival of living systems. Cells rely on this ability to perform various tasks, such as seeking for nutrients, avoiding harmful substances, maintaining homeostasis, as well as communicating in a chemical language.<sup>1</sup> Given its utility, nature has evolved various efficient systems of sensing chemical signals and subsequently strategizing corresponding responses.<sup>2</sup> From a cellular perspective, these molecular detection systems enable cellular machineries to be functional at an optimal level and maintain efficient energy consumption.<sup>3</sup> Taking metabolite production as an example, many cells are able to either transport small organic molecules from environment or synthesize them from simple precursors when there is a demand to do so. Each process requires a distinct set of proteins and cells often use feedback regulations by the final metabolite of the biosynthetic pathway to repress or enhance the production of relevant proteins. What, then, are the components responsible for sensing chemicals and triggering the biochemical response?

Protein sensors for small molecules were first discovered in studies of metabolic regulation.<sup>4</sup> The canonical example, the *lac* operon, enables a cell to switch between different energy consumption modes as the environment changes. When the glucose supply is limited and lactose is abundant, the cell needs to transport and break down the

disaccharide lactose into glucose and galactose. The *lac* operon encodes proteins required for importing (LacY) and hydrolyzing (LacZ) lactose, as well as an upstream repressor called LacI which senses the concentration of lactose and regulates the expression of the entire operon region in a feedback loop.<sup>4, 5</sup> Meanwhile, when glucose is available, instead of sparing extra resources to produce the  $\beta$ -galactosidase, LacI can bind to the operator region of the operon and attenuate the transcription of its downstream proteins.<sup>5</sup>

The *trp* operon is another example of a chemical-sensing system in *Escherichia coli* cells.<sup>6</sup> Similar to the *lac* operon, the *trp* operon regulates the expression of tryptophan biosynthetic proteins based on the concentration of their end-product. In contrast to *lac* operon, the attenuator, formed by tryptophan binding to the repressor protein TrpR, can bind to the upstream operator region of the operon and lead to the reduced transcription of five genes in tryptophan biosynthesis enzyme family.<sup>6</sup> At the same time, a second negative feedback mechanism is present in the *trp* operon.<sup>7</sup> A 14-amino-acid leader peptide sequence translated from the mRNA has two tryptophan codons. When there is a high level of the tryptophanyl tRNA, the read-through of tryptophan codons results in the formation of a transcription termination stem-loop and leads to the dissociation between the RNA polymerase and the DNA. However, when there is a low-level of tryptophanyl tRNA, the stalling of ribosomes on tryptophan codons allows the mRNA to fold into an alternative stem-loop which facilitates the transcription of the downstream genes.<sup>7</sup>

The ribosome-mediated mRNA attenuation mechanism of *trp* operon suggests the

possibility of regulating the genetic expression based on mRNA structural changes. Over the past two decades, the discovery of small-molecule binding RNA elements across all domains of life proved the above speculation.<sup>8</sup> Predominantly found in bacteria, these metabolite-sensing RNA regulators, called riboswitches, can regulate the metabolic genes from their mRNA transcripts. The direct binding between the riboswitch and its specific ligand is expected to exclude the involvement of protein cofactors and is capable of inducing conformational changes of mRNA molecules which results in the genetic expression regulation.<sup>8</sup>

## **1.2 The Discovery of Naturally Occurring Riboswitches and Their General Characteristics**

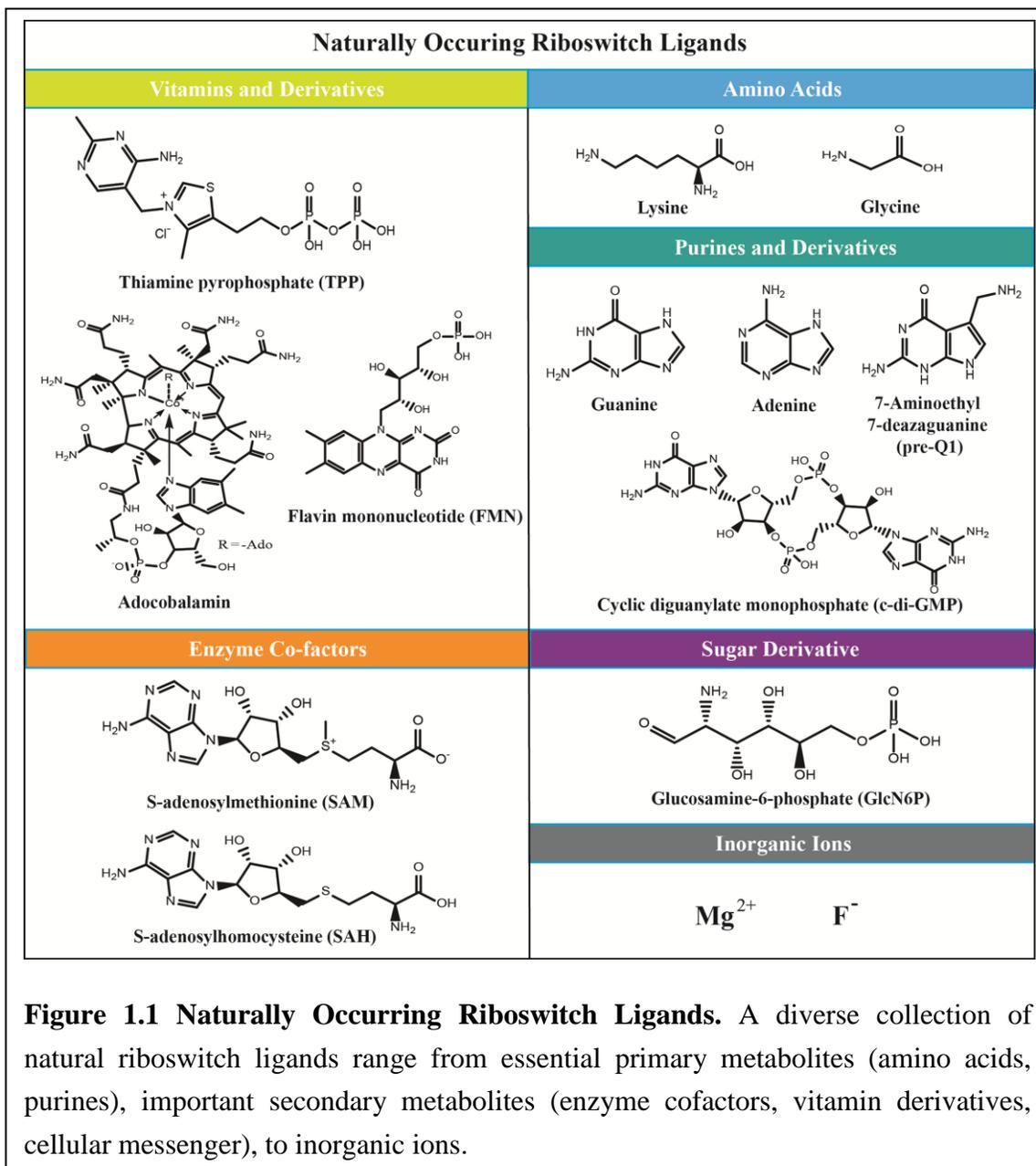
There was an enigma in the field of microbiology and molecular genetics for many years as scientists were not able to identify protein cofactors that are responsible for the regulation of some biosynthetic genes. While the conserved DNA regions (*thi* box,<sup>9</sup> B12 box,<sup>10</sup> SAM box,<sup>11</sup> and *RFN* element<sup>12</sup>) had been identified upstream of vitamin biosynthetic genes across species and were thought to be the binding site of protein repressors, the vigorous search for those hypothetical protein repressors was fruitless. The negative results, somehow, led to the speculation that the mRNA itself may directly interact with the metabolite and provide the regulatory role.<sup>12, 13, 14</sup>

Followed by the observation that the expression of the cobalamin (vitamin B12) biosynthetic *cob* operon in *Salmonella typhimurium* is repressed by the end-product,

Ravnum and Anderson found a *cis*-acting translational enhancer (TE) region upstream of the ribosomal binding site (RBS).<sup>15</sup> While they failed to provide the direct evidence of binding between cobalamin and mRNA molecule, they illustrated that the TE region can destabilize the RNA hairpin and expose the RBS in absence of cobalamin.<sup>15</sup> Similar observations from the *E. coli btuB* mRNA demonstrated that adenosylcobalamin (AdoCbl) primarily controls the translation initiation process of *btuB* by altering the level and stability of *btuB* RNA.<sup>16,17</sup>

Eventually, three vitamin derivatives, thiamine pyrophosphate (TPP),<sup>18,19</sup> flavin mononucleotide (FMN),<sup>18</sup> and adenosylcobalamin (AdoCbl)<sup>20</sup> were demonstrated to interact with their corresponding mRNA conserved regions and regulate the downstream biosynthetic operons. In Mironov's report, authors showed the formation of the FMN-mRNA complex by using real-time fluorometry.<sup>18</sup> Whereas Nahvi *et al* and Winkler *et al* monitored the *in vitro* interaction between the ligands and mRNAs with in-line probing techniques (discussed in detail in Chapter Two), and revealed important structural information for ligand binding in the conserved mRNA region.<sup>19,20</sup> These reports illustrated that the binding of metabolites to mRNA could stabilize the secondary structure of evolutionarily conserved RNA sensors (termed as aptamers) and induce an alternative folding of the downstream RNA sequence (termed as expression platform) which affects transcription termination or translation initiation.

### **1.3 Prevalence of Riboswitches Supports the RNA World Hypothesis**



**Figure 1.1 Naturally Occurring Riboswitch Ligands.** A diverse collection of natural riboswitch ligands range from essential primary metabolites (amino acids, purines), important secondary metabolites (enzyme cofactors, vitamin derivatives, cellular messenger), to inorganic ions.

Inspired by the initial discoveries of riboswitches in 2002,<sup>18,19,20</sup> much attention has been drawn to this emerging topic. Based on previous knowledge about metabolite regulation, an exploration into non-encoding DNA regions has yielded at least a dozen of naturally occurring riboswitches.<sup>8</sup> (See Figure 1.1) They respond to a wide variety of

ligands, including vitamin derivatives-AdoCbl,<sup>20</sup> TPP,<sup>19</sup> FMN,<sup>18</sup> amino acids-lysine,<sup>21</sup> glycine;<sup>22</sup> purines and their derivatives-adenine,<sup>23</sup> guanine,<sup>24</sup> 7-aminoethyl 7-deazaguanine (preQ1),<sup>25</sup> cyclic di-GMP;<sup>26</sup> phosphorylated sugar-glucosamine-6-phosphate (GlcN6P),<sup>27</sup> and enzyme cofactors-S-adenosylmethionine (SAM),<sup>28,29</sup> and S-adenosylhomocysteine (SAH).<sup>30</sup> Some riboswitches are responding to inorganic ions such as Mg<sup>2+</sup> cation;<sup>31,32</sup> and F<sup>-</sup> anion.<sup>33</sup> With each new riboswitch reported, it becomes increasingly evident that riboswitches can serve many of the same roles once predicted to be the responsibility of proteins. Through several distinct types of mechanisms (discussed later), natural riboswitches are capable of controlling gene expression in response to molecular signals ranging from primary and secondary metabolites to molecular messengers involved in intracellular communication.<sup>34</sup>

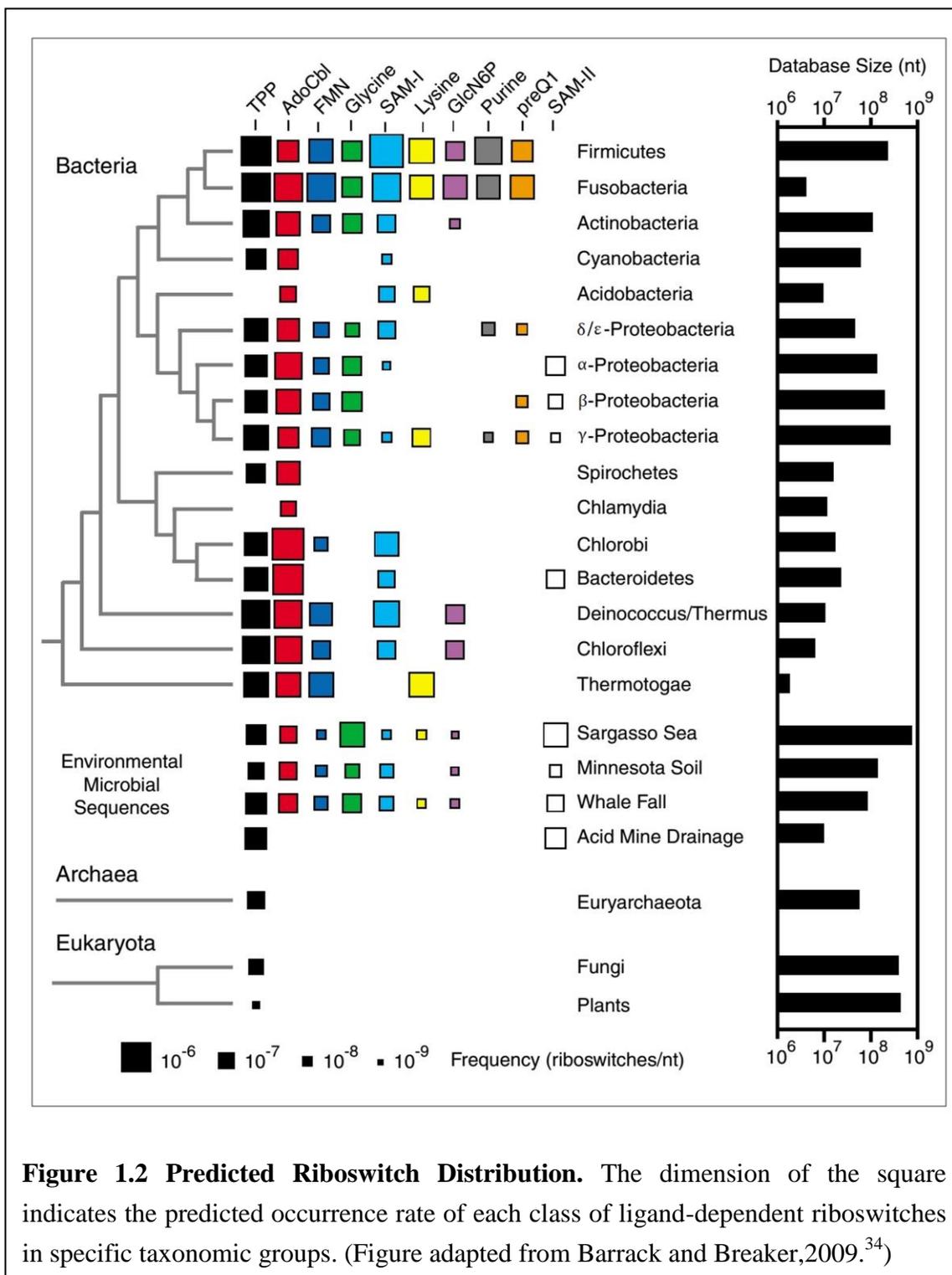
Despite the plethora of natural riboswitches, this small-molecule sensing mRNA element can be generally defined as a specific mRNA segment that can adapt to an alternative secondary structure upon binding to a small-molecule ligand and subsequently change transcriptional or translational efficiency. Most natural riboswitches are found in the conserved 5'-untranslated regions (5'-UTR) of biosynthetic genes and generally consist of two functional domains: an aptamer domain that binds to the specific ligand, and an expression platform domain that induces the mRNA conformational shift.

Riboswitches have been discovered in bacteria, archaea, fungi, and plantae,

indicating that riboswitches provide prevalent gene regulation mechanism. With the help of bioinformatics, each of the natural riboswitches has been searched among existing genomic databases. The results showed that riboswitch regions may be present in various species in all three superkingdoms of life (See Figure 1.2).<sup>34</sup> While the bulk of existing riboswitches are found in bacteria, TPP-dependent riboswitches are also found to regulate TPP synthesis in fungi and green algae via different mechanisms.<sup>35,36</sup> In fungi, the TPP riboswitch is found inside of intron regions and affects the gene expression by inducing an alternative intron splicing of mRNA molecule.<sup>35</sup> In green algae, a similar alternative intron splicing strategy is adopted; however, the TPP riboswitch is identified inside of coding region.<sup>36</sup> Finally, a similar TPP riboswitch was discovered in the 3'-UTR of the thiamine biosynthesis genes in the genome of Arabidopsis.<sup>37</sup>

A recent report showed that adenosine-binding aptamers created by *in vitro* selection techniques can be mapped to the intron regions of *RAB3C* and *FDG3* genes of human genomes.<sup>38</sup> And the hypothesized *RAB3C* riboswitch was shown to have binding affinity ten-fold lower than the physiological concentration of ATP, whereas the hypothesized *FDG3* riboswitch could function in a co-transcriptional fashion.<sup>38</sup> Encouraging as these results are, however, the biologic function of such aptamers remained unclear.

The ubiquitous presence of mRNA sensors along with their non-reliance on protein cofactors lead to the idea that riboswitches may be the remnants of ancient metabolic ribozymes in a RNA world.<sup>39,40</sup> The RNA world hypothesis states that RNA may have



been chosen as the genetic and enzymatic biomolecule prior to the protein era.<sup>41</sup> There are four essential processes to constitute a satisfying account for the evolution in the

RNA world-replication, cellularization, translation, and metabolism.<sup>42</sup> Widely accepted evidence, such as RNA-catalyzed RNA synthesis,<sup>43,44</sup> along with well-studied RNA-based translational reactions (amino acid activation, aminoacyl-RNA synthesis, and peptide bond formation) have answered most of the vexing problems and set forth the concept of RNA world. However, the remaining question is how could RNA perform the function as a small molecule sensor and regulator of gene expression to maintain a proper metabolism for the living system?

The results from riboswitch researches fill in this missing gap. As riboswitches possess crucial small-molecule binding functions similar to protein regulators, they are postulated as an important component in the RNA world hypothesis. Remarkably, RNA only utilizes the 4 possible canonical bases (versus 20 canonical choices of amino acid in the proteins) to bind tightly and selectively to a wide variety of small molecules. Over 15 classes of natural riboswitches exhibit a variety of target ligand-vitamin derivatives, purines, amino acids, to even inorganic cation and anion, a versatile collection of regulatory mechanism, and a large response range.<sup>45</sup>

## **1.4 Synthetic Riboswitches in Gene Regulation**

### **1.4.1 The Rationale behind Engineering Riboswitches**

The value of fundamental cellular biochemical mechanism research should never be underestimated. Knowledge gained in molecular biology and biochemistry has empowered fruitful applications of many basic gene regulatory elements such as the *lac*

operon and the pBAD promotor system.<sup>46</sup> The implementation of those gene regulatory systems not only creates more insightful research tools but also expands our scope to a broader range of bacteria hosts. However, despite the wonderful utility, these systems are limited by the small molecules that they can respond to. The challenge is, then, how can we engineer those gene regulatory systems to respond to any small molecule of one's choosing.

The directed evolution of protein sensors has been successful in creating novel proteins that can recognize desired small molecules with regulatory functions. As an example of protein-switch engineering, a maltose binding protein (MBP) was evolved to change its binding specificity from the natural ligand-maltose to an alternative ligand-sucrose.<sup>47</sup> Briefly, the creation of the sucrose version MBP involved a number of iterative screening processes on MBP's randomized substrate binding pocket. The library was screened in the presence of sucrose, but the absence of maltose, and surviving clones were cultivated to the next round of screening process.<sup>47</sup> While the *in vivo* method was successful in changing the specificity of MBP, it remains very challenging to engineer protein-based systems to recognize novel molecules that share little chemical similarity to natural metabolites.

Another consideration for designing novel small-molecule-inducible gene regulatory system is that one would need to engineer a cascade of proteins that has both components for sensing small molecule and ones that can translate the chemical signal into a change

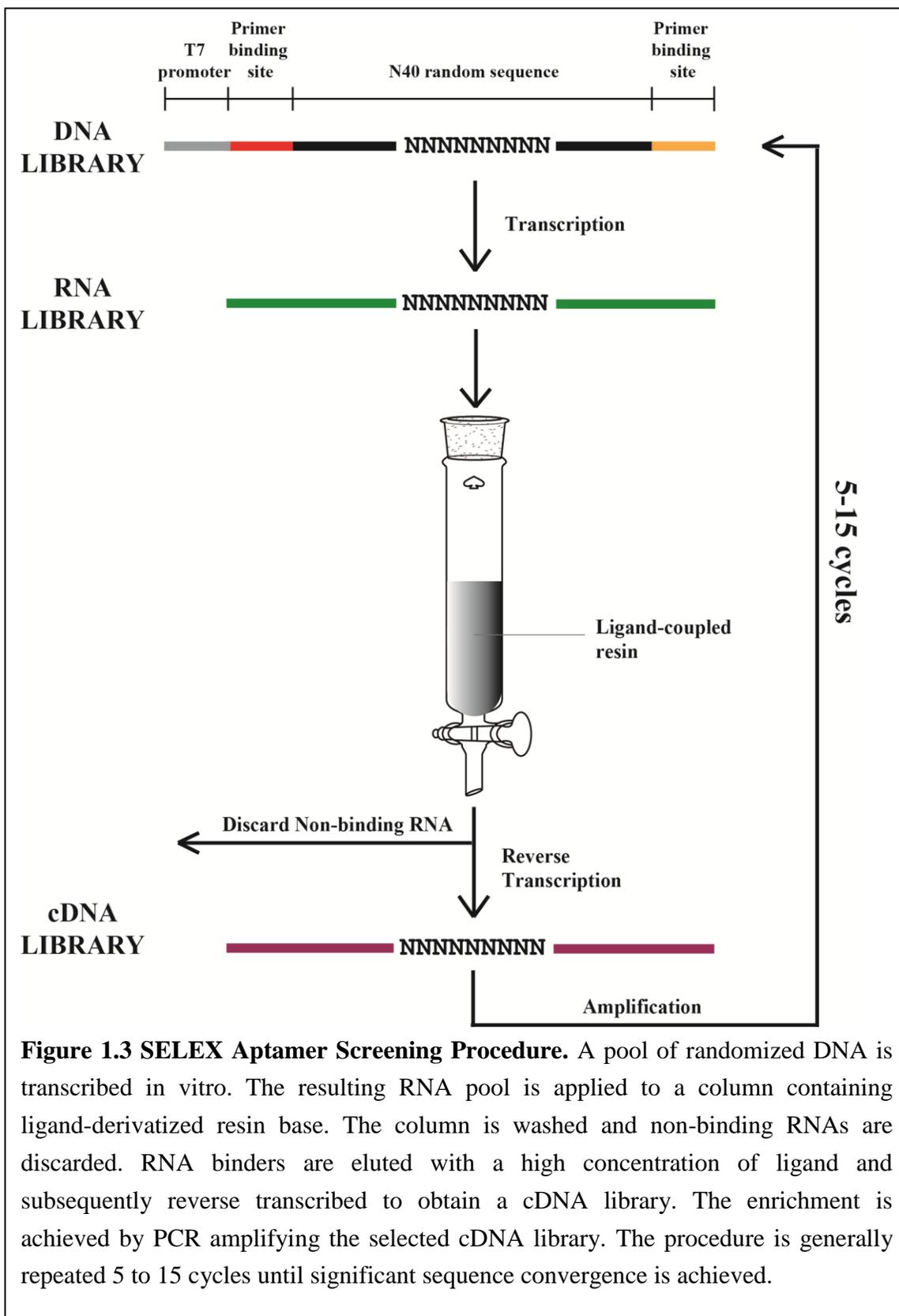
in gene expression.

With above considerations, it is quite attempting to utilize the concept of a RNA-based riboswitch to accomplish the small-molecule-induced gene regulation task. As a stand-alone system, the functioning of riboswitch does not depend on the presence of any extra protein cofactor. Meanwhile, the gene expression regulation only requires the secondary and tertiary structural changes of mRNA molecules. The only question is, then, how can we create RNA sequences that can sense small molecules and induce gene expression changes?

#### **1.4.2 Methods for Developing Synthetic Riboswitches**

Fortunately, prior to the discovery of riboswitches, methods developed in the early 1990s for *in vitro* selection and evolution of RNA sequences were well-established.<sup>48</sup> These methods have demonstrated that RNAs could be generated to bind a wide range of protein and small-molecule ligands with no need for a pre-existing template.<sup>49,50</sup>

Commonly referred as Systematic Evolution of Ligands through EXponential enrichment (SELEX) (See Figure 1.3), the iterative screening process begins with a chemically synthesized pool of randomized DNA. The randomized region is normally 20 to 60 nucleotides long, and is flanked by a pre-defined 5'- and 3'- primer sequence for PCR amplification.<sup>48</sup> Theoretically,  $\sim 10^{15}$  distinct DNA sequences can be efficiently transcribed to a RNA pool *in vitro* with the aid of a T7 promoter placed in the upstream of the random sequences. To remove non-specific RNA binders (e.g. those that bind to



the column base material), a negative selection step is applied where the random pool is washed over a column without the ligand. Any non-binding RNA sequences are then applied to a binding column where the ligand is chemically immobilized on the column. After washing away the non-binding RNA molecules, the candidate RNA sequences that are bound to the ligand are eluted from the column with a ligand-containing buffer, and are reverse transcribed to obtain the corresponding cDNA pool. The enrichment is achieved by re-amplifying the cDNA pool with PCR and iteratively repeating these processes until the RNA pool reaches a convergence, typically 5-15 rounds. Because all the steps are performed *in vitro*, the selection is not bounded by the theoretical size limit *in vivo* and is considered to be a high-throughput screening method.

To ensure the binding specificity, a counter selection strategy is optionally adopted in some screening processes. One or more structurally similar molecules (other than the ligand) are applied to the RNA-bound column in order to remove the RNA sensors that can non-specifically bind to a pool of small molecules. After discarding the non-discriminative RNA aptamers, the column once again is eluted with a high concentration of the desired ligand. Thus, specific RNA aptamers can be enriched.

Since the first report of this screening method, a series of aptamers against a plethora of different targets have been generated including ions, organic compounds, proteins, viruses and even whole cells.<sup>51</sup> The growing repertoire of aptamers gives scientists more available building blocks for constructing functional riboswitches. Meanwhile, even

though different aptamers may respond to different ligands and have different scaffolds, they often adopt to final three-dimensional structures only upon binding to their ligands. Therefore, an aptamer can function as both sensing and regulation-inducing domains at the same time.

### **1.4.3 Different Approaches to Design Synthetic Riboswitches**

The first example of converting an artificial small-molecule aptamer into a riboswitch made an inspiring leap forward for the RNA gene regulatory studies. Wetstuck and Green incorporated an aptamer that recognize Hoechst dye 33342 into the 5'-UTR of a reporter gene and observed repression in the reporter gene expression when the ligand molecule is present.<sup>52</sup> This interesting finding demonstrated for the first time that the binding-induced RNA structural changes can interfere with the downstream gene expression by forming an inhibitory complex.

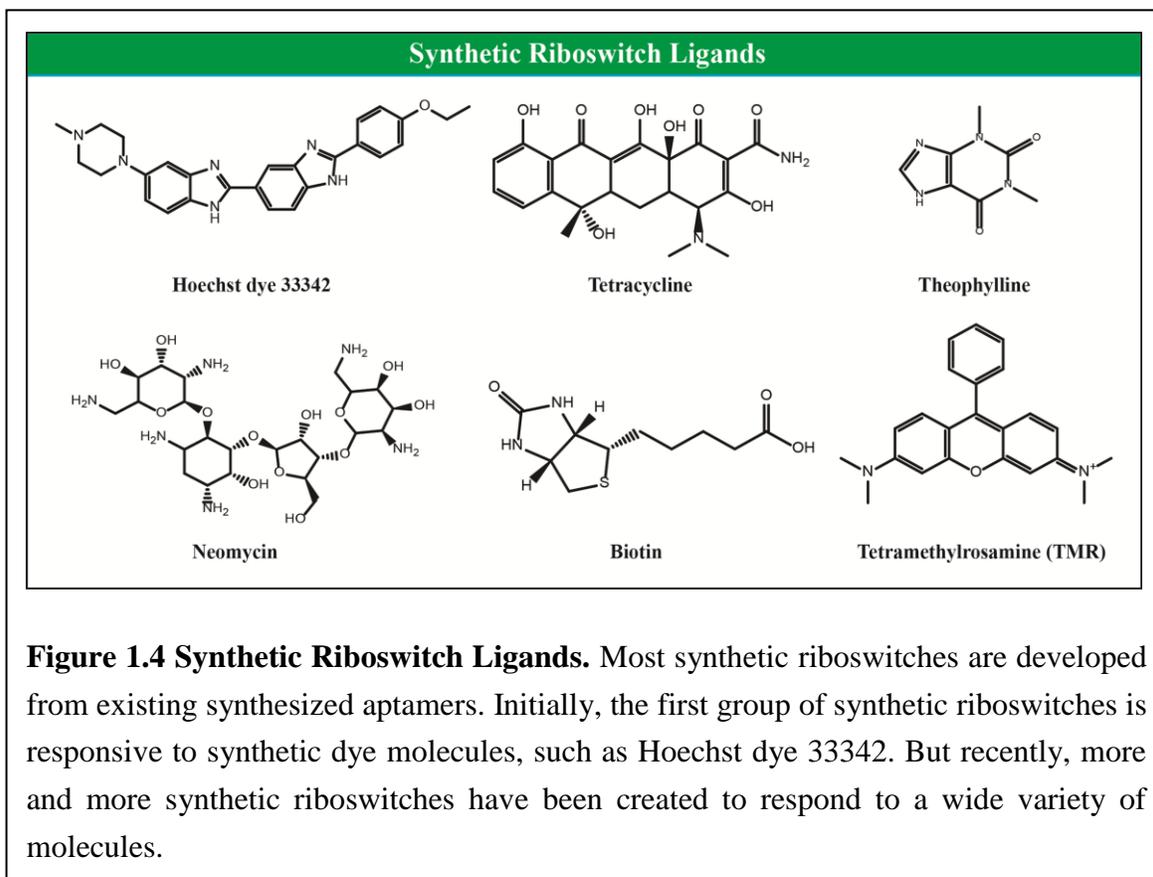
Inspired by this roadblock mechanism, Suess and co-workers developed a class of tetracycline riboswitches to regulate translation initiation in yeast.<sup>53</sup> The aptamer-tetracycline complex in the 5'-UTR can prevent the binding of the small ribosome unit, thus attenuating translation. It has also been shown that the closer distance to the starting codon, the more efficient regulation an aptamer can accomplish.<sup>53</sup> Furthermore, the same group engineered this riboswitch in tandem and greatly improved the repression power and the dynamic range of the regulation. Other efforts in converting neomycin, malachite green, biotin, and theophylline-binding aptamers into roadblock

riboswitches in wheat germ and reticulocyte extracts, yeast or *Xenopus* oocytes were also reported.<sup>54,55</sup>

Another strategy involving modulation of pre-mRNA splicing has also been adopted in several successful designs. Placing a riboswitch near sequences essential for splicing such as 5'-splicing site (SS), branching point, or 3'-SS, the presence of riboswitch could impair the effectiveness of pre-mRNA self-splicing and render a low turnover rate for pre-mRNA molecule. Gaur and co-workers inserted a theophylline-binding aptamer into a 3'-SS and observed a 4-fold theophylline-dependent reduction for gene expression in HeLa nuclear extract.<sup>56</sup> Suesse and co-workers also employed a tetracycline aptamer to interfere with splicing at a 5'-SS. The slightly larger tetracycline aptamer is able to achieve a 16-fold reduction in the gene regulation in the presence of tetracycline.<sup>57</sup> In a recent report, the Gaur lab positioned the theophylline aptamer in the branching point of splicing sequence.<sup>58</sup> The engineered intron led to two-fold higher alternative splicing rate for the downstream adjacent exon in HeLa nuclear extract which could potentially give insight into the modulation of pre-mRNA alternative splicing.<sup>58</sup>

At the meantime, our lab used a different approach to engineer riboswitches. Desai and co-workers adopted a proof-of-principle screening method where the signal of sensing small-molecule ligand is crucial for the survival of bacteria cells.<sup>59</sup> A theophylline aptamer and its C27A mutant aptamer (altered specificity to structurally related 3-methylxanthine) were separately placed in the 5'-UTR of a chloramphenicol

acetyl transferase gene (conferring chloramphenicol resistance). Even in the most dilute case where the parent aptamer is only representing 0.0001% of the total constructs, they were able to isolate the theophylline sensing plasmid when the transformed cultures were selected on a chloramphenicol-supplemented agar plate.



Later on, our lab developed a  $\beta$ -galactosidase-based screening procedure to select for functional riboswitches from a randomized library.<sup>60</sup> The detailed information about this procedure will be provided in later chapters. We also developed a riboswitch selection method using a cell motility-based approach.<sup>61</sup> Briefly, the randomized library is placed upstream of the *CheZ* gene which senses a smooth-swimming signal to control flagellar

rotation. If a candidate in the randomized library can transform a theophylline-binding signal into an activation of translation initiation of *CheZ* gene, the cell containing a functional riboswitch can move away from the original location and achieve separation in the presence of theophylline. However, the drawback of this motility-based screening method is that when the *CheZ* gene is overexpressed, cells can become non-motile due to embedding in the agar media.<sup>62</sup> Therefore the possibility of discovering highly activating riboswitches is limited.

A solution for the above drawback is to extend the detection range of the reporter gene. Consequently, our lab developed a high-throughput screening procedure to screen dynamic, functional riboswitches using fluorescence activated cell sorting (FACS) technique, which can detect robust activation while maintaining a large theoretical library size of  $\sim 10^8$  members.<sup>63</sup> The screening library is designed to harbor the randomized sequence between the theophylline aptamer and the red fluorescent protein (*DsRed*) gene.

These efforts led to the discovery of a series of versatile theophylline riboswitches named A through G.<sup>64</sup> These riboswitches exhibit different gene activation efficiencies, dynamic ranges and can regulate gene expression in different organisms (discussed later). One of our riboswitches can induce a 100-fold increase in gene expression in *E. coli*, when theophylline is present which remain the most robust of the natural or synthetic riboswitches published.<sup>63</sup>

More efforts have also been gathered to develop new screens and selections for

synthetic riboswitches by other groups. Weigand and Süss developed a screening method involving a different reporter gene-green fluorescent protein (*gfp*)-in the yeast *Saccharomyces cerevisiae*.<sup>65</sup> Fowler and Li used FACS to screen for transcriptional inhibition riboswitches in *E. coli*, similar to the method described by Lynch and Gallivan above.<sup>66</sup>

Recently, a number of allosterically regulated ribozymes have been reported to expand the riboswitch repertoire. These allosteric self-cleaving ribozymes can attenuate post-transcriptional stability of mRNA by regulating the degradation rate of mRNA molecules. As no species-specific element of transcription or translation is involved, this mechanism can be considered as species-independent and may be portable to different organisms. The first example of such a self-cleaving ribozyme is based on the minimal hammerhead ribozyme.<sup>67</sup> The Breaker lab inserted an ATP aptamer in a rationally designed hammerhead cleavage RNA device where ATP can both allosterically activate and inhibit the self-cleavage. In another construct where they fused a theophylline aptamer in the same site, however, the resulting ribozyme can only attenuate the cleavage when the ligand is present. Despite the promising results from these riboswitches, the ribozyme approach is not ideal for cellular applications due to the lack of flexibility in natural hammerhead ribozyme structure that mediates tertiary loop-loop interactions.

Alongside riboswitches controlling on translational level, there also are a few examples that function at the transcription of level. Based on their previous discovery of a

transcriptional activator sequence, Liu and co-workers fused a tetramethylrosamine aptamer to the RNA activator sequence and generated a transcriptional activating riboswitch.<sup>68,69</sup> Similarly, Suess and co-workers engineered a theophylline aptamer into a TetR repressor protein-dependent transcriptional activator RNA sequences and achieved a theophylline-dependent transcription repressor.<sup>70</sup>

#### **1.4.4 The Performance of Riboswitch in Model Species**

As a conceptually simple and increasingly versatile system, engineered riboswitches have attracted attentions of scientists from different fields such as synthetic biology, microbiology, as well as biosynthetic engineering. However, since most of the existing natural riboswitches are found in a number of bacterial genomes, the biggest challenge for the underlying riboswitch applications is, then, whether engineered riboswitches can be universally functional in different species, especially the ones with physiological or developmental significance. Their characterization demands fine-tuned inducible expression systems to ensure that the products to be analyzed or synthesized are only produced when desired.

To overcome this challenge, a number of successful attempts have been taken to test riboswitches in well-characterized model organisms such as *E. coli*, *B. subtilis*, *S. cerevisiae* and in mammalian cell lines.<sup>42</sup> Among these endeavors, our lab optimized our theophylline-inducible riboswitches to control gene expression in diverse species of Gram-negative and Gram-positive bacteria, including the model organisms *E.coli*,

*B.subtilis*, *Mycobacterium smegmatis*, and the plant and human pathogens *Agrobacterium tumefaciens*, *Acinetobacter baumannii*, and *Streptococcus pyogenes*.<sup>64</sup> Meanwhile, dose-dependent riboswitches for controlling protein expression have also been verified in model systems such as Cyanobacterium *Synechococcus elongatus*—a model organism for photosynthesis and biomaterials generation,<sup>71</sup> *Streptomyces coelicolor*—an industrial producer of biomedical products,<sup>72</sup> as well as human pathogens *Francisella tularensis*<sup>73</sup> and *Mycobacterium tuberculosis*<sup>74</sup>.

Even though different theophylline riboswitches in our repository behave differently in various organisms, the optimal regulatory factors of at least 25-fold in all tested species were observed.<sup>64</sup> A greater-than-100-fold gene expression increase in *M. smegmatis* and *S. elongatus* were in the same range as observed in the engineered TetR transcriptional regulation system.<sup>74,75</sup> In the medically relevant pathogens *F. tularensis* and *M. tuberculosis*, the riboswitch regulation device can match or surpass the performance of other induction systems in a macrophage infection system with much lower cost for labor and material.<sup>73,74</sup>

In the meantime, tetracycline-regulated riboswitches have been implemented as inducible control elements in *Methanosarcina acetivorans*, an anaerobic model organism from the *Archaea* by Sues and coworkers.<sup>76</sup> So far, natural riboswitches have not been validated experimentally in *Archaea*, although identical gene elements for fluoride and TPP riboswitches have been discovered. Nevertheless, the exciting proof-of-principle

results from the functional tetracycline riboswitch in *M. acetivorans* have demonstrated the effectiveness of RNA regulation in such superkindom of life.<sup>76</sup>

An interesting experiment conducted by Tozawa and co-workers in *S. elgongatus* is to use engineered riboswitches to control the circadian rhythm of the cell.<sup>71</sup> Authors grafted a theophylline riboswitch to the 5'-upstream of the circadian clock *kaiC* gene. The fine-tuning capability of theophylline riboswitches yielded an adequately adjusted expression level of KaiC, thus restored complete circadian rhythm in the *kaiC*-deficient arrhythmic mutant.

The engineering of artificial riboswitches has come a long way since its initial reports with the roadblock Hoechst dye aptamer system approximately 15 years ago.<sup>52</sup> A wide variety of riboswitch systems published by a number of groups have demonstrated the great diversity and potential utility in establishing simple but potent regulatory modules. Currently, remaining gaps in developing this riboswitch toolbox are made closer with the discoveries of new aptamer sensors and the advancement of novel methodologies. The increasing demand of “off the shelf” gene regulatory elements and “plug-and-play” approaches enabled by riboswitches drives our attention to focus on the mechanistic understanding of riboswitch systems.

### **1.5 Different Types of Riboswitch Mechanisms**

A large number of riboswitches discovered and engineered in the last few years suggested the existence of not only a huge diversity of regulatory ligands but also

versatile genetic mechanisms of regulation. The mechanistic information gathered from individual riboswitch reports indicated that there are four primary mechanisms by which riboswitches regulate gene expression. All four types of mechanisms involve binding of a specific ligand and a subsequent rearrangement of RNA conformation.

Some riboswitches regulate gene expression on the transcription level. The adenine riboswitch found in *B. subtilis* represents the transcriptional anti-terminator mechanism. It adopts a terminator structure in the absence of ligand, but upon binding to adenine the riboswitch region shifts to alternative secondary conformation thereby allowing transcription of a purine efflux pump.<sup>77</sup> Quite interestingly, the guanine riboswitch found to regulate the purine biosynthesis genes in *B. subtilis* represents the opposite mechanism.<sup>78</sup> In the absence of guanine, the transcription of purine biosynthesis gene is allowed to generate necessary guanine. However, when guanine is abundant inside the bacterial cell, production of guanine is turned down by the binding between guanine and its riboswitch which stabilizes an alternative transcriptional terminator conformation.<sup>78</sup>

Riboswitches can also target the translation process to regulate gene expression. Also referred as helix-slipping mechanism, the FMN riboswitch found in *E. coli* is capable of sequestering the ribosome binding site (RBS) upon binding to FMN and preventing the initiation of translation process.<sup>8</sup> On the other hand, the theophylline riboswitches developed by our lab adopt a similar concept but work in an opposite direction where the binding of ligand stabilizes the unsequestered RBS conformation allowing the initiation

of translation.

The first engineered riboswitches by Werstuck and Green represent a roadblock mechanism.<sup>52</sup> The Hoechst dye aptamer was incorporated into the 5'-upstream of a reporter gene. Translation of the reporter gene is repressed by the formation of aptamer-ligand complex which serves as an inhibitor to the assembly of the ribosome-mRNA complex. Inspired by this idea, a number of similar riboswitches were created using tetracycline, malachite green, theophylline and biotin aptamers to control gene expression in wheat germ and reticulocyte extracts, yeast or *Xenopus* oocytes.<sup>54,55</sup>

Moreover, some riboswitches are involved in regulating pre-mRNA intron splicing. Found and engineered in more sophisticated organisms, the Gaur lab successfully incorporated an artificial theophylline aptamer to the 3'-splicing site of a model pre-mRNA (AdML-Theo29AG) and enabled repression of self-splicing by the addition of theophylline.<sup>56</sup> Later, several exciting experiments were carried out by choosing different sites for the aptamer insertion as well as different ligand-responsive aptamers, and the regulation of gene expression was quite successful.<sup>57,58</sup>

A good example to demonstrate the diversity in riboswitch mechanisms are the thiamine pyrophosphate (TPP) riboswitches. In Gram-negative *E. coli*, negative feedback regulation of the biogenesis of TPP and thiamine can be fulfilled by TPP riboswitch sequestering RBS upon binding to ligand, thus inhibiting the translation of TPP-related biosynthetic genes.<sup>19</sup> However, in Gram-positive *B. subtilis*, the underlying mechanism

differs significantly for the feedback control of this biosynthesis process. Mironov and coworkers observed that the full length of mRNA can be synthesized in vitro when TPP is present while the truncated mRNA is the major product with the absence of TPP.<sup>18</sup> This observation indicates the TPP riboswitch can be shifted to transcriptional terminator and attenuate the transcription efficiency.

Some variations based on the above four primary mechanism also exist among natural and synthetic riboswitches. In some cases, the RBS can be integrated in the aptamer region, making conformational shifts in the mRNA unnecessary for changes in expression.<sup>79,80</sup> Alternatively, a dual functional riboswitch system can be constructed by burying the RBS in a sequestered segment in the secondary structure of a transcriptional terminator. This particular construct can inhibit the translation of downstream genes even if transcription is not hindered.

Less represented mechanisms, such as a trans-acting small RNA mechanism,<sup>81</sup> are also very scientifically interesting. The first case of the trans-acting small RNA mechanism was reported by Putzer and coworkers.<sup>81</sup> The small RNA fragment generated by pre-mature transcription termination of an S-adenosylmethionine(SAM)-dependent riboswitch in *B. subtilis* is capable of utilizing the alternative endoribonuclease (RNase Y) system to increase the turnover rate for mRNA molecules. However, these mechanisms lack the simplicity for universal applications and will not be discussed in further detail here.

## **1.6 Conclusion**

It has been over 15 years since the initial proof of the existence of natural riboswitches. Since their discovery, researchers have discovered a variety of riboswitches in different species. And while discoveries of new natural riboswitches are underway, many researchers have started to utilize these elegant elements and have crossed the natural boundary to generate synthetic riboswitches. Research on riboswitches has led us to gain more insight into this once mysterious genetic control element and to create a versatile and “off-shelf” genetic regulation toolbox. As new research results emerge, many more discoveries in the riboswitch field will be published shortly.

## 1.7 Reference

1. Boor, K.J., *PLoS Biol*, **2006**. 4: e23.
2. Stoebel, D.M., Dean, A.M. and Dykhuizen, D.E., *Genetics*, **2008**. 178: 1653-60.
3. DeBusk, R., *Nutr Clin Pract*, **2010**. 25: 627-33.
4. Jacob, F. and Monod, J., *J Mol Biol*, **1961**. 3: 318-56.
5. Jobe, A. and Bourgeois, S., *J Mol Biol*, **1972**. 69: 397-408.
6. Monod, J. and Cohen-Bazire, G., *C R Hebd Seances Acad Sci*, **1953**. 236: 530-2.
7. Jackson, E.N. and Yanofsky, C., *J Mol Biol*, **1973**. 76: 89-101.
8. Nudler, E. and Mironov, A.S., *Trends Biochem Sci*, **2004**. 29: 11-7.
9. Miranda-Rios, J., Morera, C., Taboada, H., Davalos, A., Encarnacion, S., Mora, J. and Soberon, M., *J Bacteriol*, **1997**. 179: 6887-93.
10. Lundrigan, M.D., Koster, W. and Kadner, R.J., *Proc Natl Acad Sci U S A*, **1991**. 88: 1479-83.
11. Grundy, F.J. and Henkin, T.M., *Mol Microbiol*, **1998**. 30: 737-49.
12. Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I. and Perumov, D.A., *Trends Genet*, **1999**. 15: 439-42.
13. Miranda-Rios, J., Navarro, M. and Soberon, M., *Proc Natl Acad Sci U S A*, **2001**. 98: 9736-41.
14. Stormo, G.D. and Ji, Y., *Proc Natl Acad Sci U S A*, **2001**. 98: 9465-7.
15. Ravnum, S. and Andersson, D.I., *Mol Microbiol*, **1997**. 23: 35-42.

16. Nou, X. and Kadner, R.J., *J Bacteriol*, **1998**. *180*: 6719-28.
17. Nou, X. and Kadner, R.J., *Proc Natl Acad Sci U S A*, **2000**. *97*: 7190-5.
18. Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. and Nudler, E., *Cell*, **2002**. *111*: 747-56.
19. Winkler, W., Nahvi, A. and Breaker, R.R., *Nature*, **2002**. *419*: 952-6.
20. Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. and Breaker, R.R., *Chem Biol*, **2002**. *9*: 1043.
21. Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S. and Breaker, R.R., *Genes Dev*, **2003**. *17*: 2688-97.
22. Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. and Breaker, R.R., *Science*, **2004**. *306*: 275-9.
23. Mandal, M. and Breaker, R.R., *Nat Struct Mol Biol*, **2004**. *11*: 29-35.
24. Batey, R.T., Gilbert, S.D. and Montange, R.K., *Nature*, **2004**. *432*: 411-5.
25. Roth, A., Winkler, W.C., Regulski, E.E., Lee, B.W., Lim, J., Jona, I., Barrick, J.E., Ritwik, A., Kim, J.N., Welz, R., Iwata-Reuyl, D. and Breaker, R.R., *Nat Struct Mol Biol*, **2007**. *14*: 308-17.
26. Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H. and Breaker, R.R., *Science*, **2008**. *321*: 411-3.
27. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J.K. and Breaker, R.R., *Proc Natl*

- Acad Sci U S A*, **2004**. *101*: 6421-6.
28. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. and Breaker, R.R., *Nat Struct Biol*, **2003**. *10*: 701-7.
29. Gilbert, S.D., Rambo, R.P., Van Tyne, D. and Batey, R.T., *Nat Struct Mol Biol*, **2008**. *15*: 177-82.
30. Wang, J.X., Lee, E.R., Morales, D.R., Lim, J. and Breaker, R.R., *Mol Cell*, **2008**. *29*: 691-702.
31. Cromie, M.J., Shi, Y., Latifi, T. and Groisman, E.A., *Cell*, **2006**. *125*: 71-84.
32. Dann, C.E., 3rd, Wakeman, C.A., Sieling, C.L., Baker, S.C., Irnov, I. and Winkler, W.C., *Cell*, **2007**. *130*: 878-92.
33. Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B. and Breaker, R.R., *Science*, **2012**. *335*: 233-5.
34. Barrick, J.E. and Breaker, R.R., *Genome Biol*, **2007**. *8*: R239.
35. Cheah, M.T., Wachter, A., Sudarsan, N. and Breaker, R.R., *Nature*, **2007**. *447*: 497-500.
36. Croft, M.T., Moulin, M., Webb, M.E. and Smith, A.G., *Proc Natl Acad Sci U S A*, **2007**. *104*: 20770-5.
37. Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K. and Breaker, R.R., *Plant Cell*, **2007**. *19*: 3437-50.
38. Vu, M.M., Jameson, N.E., Masuda, S.J., Lin, D., Larralde-Ridaura, R. and Luptak,

- A., *Chem Biol*, **2012**. *19*: 1247-54.
39. Sudarsan, N., Barrick, J.E. and Breaker, R.R., *RNA*, **2003**. *9*: 644-7.
40. Breaker, R.R., *Cold Spring Harb Perspect Biol*, **2012**. *4*.
41. Vlassov, A.V., Kazakov, S.A., Johnston, B.H. and Landweber, L.F., *J Mol Evol*, **2005**. *61*: 264-73.
42. Yarus, M., *Curr Opin Chem Biol*, **1999**. *3*: 260-7.
43. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. and Bartel, D.P., *Science*, **2001**. *292*: 1319-25.
44. McGinness, K.E. and Joyce, G.F., *Chem Biol*, **2003**. *10*: 5-14.
45. Serganov, A. and Nudler, E., *Cell*, **2013**. *152*: 17-24.
46. Helling, R.B. and Weinberg, R., *Genetics*, **1963**. *48*: 1397-410.
47. Guntas, G., Mansell, T.J., Kim, J.R. and Ostermeier, M., *Proc Natl Acad Sci U S A*, **2005**. *102*: 11224-9.
48. Ellington, A.D. and Szostak, J.W., *Nature*, **1990**. *346*: 818-22.
49. Robertson, D.L. and Joyce, G.F., *Nature*, **1990**. *344*: 467-8.
50. Tuerk, C. and Gold, L., *Science*, **1990**. *249*: 505-10.
51. Lee, J.F., Hesselberth, J.R., Meyers, L.A. and Ellington, A.D., *Nucleic Acids Res*, **2004**. *32*: D95-100.
52. Werstuck, G. and Green, M.R., *Science*, **1998**. *282*: 296-8.
53. Suess, B., Hanson, S., Berens, C., Fink, B., Schroeder, R. and Hillen, W., *Nucleic*

*Acids Res*, **2003**. *31*: 1853-8.

54. Grate, D. and Wilson, C., *Bioorg Med Chem*, **2001**. *9*: 2565-70.
55. Harvey, I., Garneau, P. and Pelletier, J., *RNA*, **2002**. *8*: 452-63.
56. Kim, D.S., Gusti, V., Pillai, S.G. and Gaur, R.K., *RNA*, **2005**. *11*: 1667-77.
57. Weigand, J.E. and Sues, B., *Nucleic Acids Res*, **2007**. *35*: 4179-85.
58. Kim, D.S., Gusti, V., Dery, K.J. and Gaur, R.K., *BMC Mol Biol*, **2008**. *9*: 23.
59. Desai, S.K. and Gallivan, J.P., *J Am Chem Soc*, **2004**. *126*: 13247-54.
60. Lynch, S.A., Desai, S.K., Sajja, H.K. and Gallivan, J.P., *Chem Biol*, **2007**. *14*: 173-84.
61. Topp, S. and Gallivan, J.P., *Chembiochem*, **2008**. *9*: 210-3.
62. Wolfe, A.J. and Berg, H.C., *Proc Natl Acad Sci U S A*, **1989**. *86*: 6973-7.
63. Lynch, S.A. and Gallivan, J.P., *Nucleic Acids Res*, **2009**. *37*: 184-92.
64. Topp, S., Reynoso, C.M., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R., Scott, J.R. and Gallivan, J.P., *Appl Environ Microbiol*, **2010**. *76*: 7881-4.
65. Weigand, J.E., Sanchez, M., Gunnesch, E.B., Zeiher, S., Schroeder, R. and Sues, B., *RNA*, **2008**. *14*: 89-97.
66. Fowler, C.C., Brown, E.D. and Li, Y., *Chembiochem*, **2008**. *9*: 1906-11.
67. Soukup, G.A., Emilsson, G.A. and Breaker, R.R., *J Mol Biol*, **2000**. *298*: 623-32.
68. Buskirk, A.R., Kehayova, P.D., Landrigan, A. and Liu, D.R., *Chem Biol*, **2003**. *10*:

533-40.

69. Kehayova, P.D. and Liu, D.R., *Chem Biol*, **2007**. *14*: 65-74.
70. Auslander, D., Wieland, M., Auslander, S., Tigges, M. and Fussenegger, M., *Nucleic Acids Res*, **2011**. *39*: e155.
71. Nakahira, Y., Ogawa, A., Asano, H., Oyama, T. and Tozawa, Y., *Plant Cell Physiol*, **2013**. *54*: 1724-35.
72. Rudolph, M.M., Vockenhuber, M.P. and Suess, B., *Microbiology*, **2013**. *159*: 1416-22.
73. Reynoso, C.M., Miller, M.A., Bina, J.E., Gallivan, J.P. and Weiss, D.S., *MBio*, **2012**. *3*.
74. Seeliger, J.C., Topp, S., Sogi, K.M., Previti, M.L., Gallivan, J.P. and Bertozzi, C.R., *PLoS One*, **2012**. *7*: e29266.
75. Huang, H.H. and Lindblad, P., *J Biol Eng*, **2013**. *7*: 10.
76. Demolli, S., Geist, M.M., Weigand, J.E., Matschiavelli, N., Suess, B. and Rother, M., *Archaea*, **2014**. *2014*: 725610.
77. Lemay, J.F., Penedo, J.C., Tremblay, R., Lilley, D.M. and Lafontaine, D.A., *Chem Biol*, **2006**. *13*: 857-68.
78. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. and Breaker, R.R., *Cell*, **2003**. *113*: 577-86.
79. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S., *J Biol Chem*,

**2002.** 277: 48949-59.

80. Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S., *Trends Genet*,

**2004.** 20: 44-50.

81. Shahbadian, K., Jamalli, A., Zig, L. and Putzer, H., *EMBO J*, **2009.** 28: 3523-33.

## **Chapter 2. Investigation into the Mechanism of Theophylline**

### **Riboswitches**

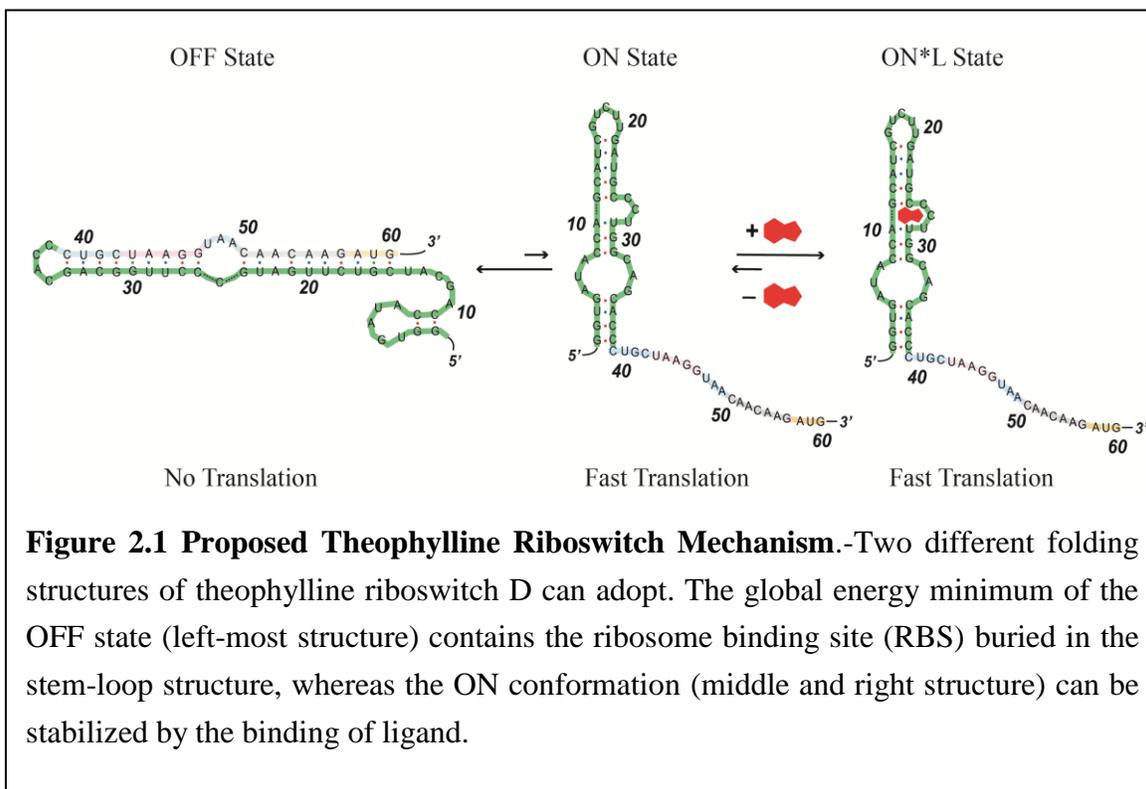
#### **2.1 Introduction**

Riboswitch is an efficient and elegant device to regulate genetic expression. Since its functionalities only rely on the structural shift of RNA molecules and do not require additional cofactors, a riboswitch can be regarded as a ‘stand-alone’ element to be implemented under most desired genetic control schemes. While the concept of a riboswitch is intuitively simple, three essential aspects remain to be addressed to fully utilize this powerful tool. First, the successful creation of an engineered riboswitches for a desired ligand remains challenging. Normally, a two-step approach is necessary for constructing a functional riboswitch: finding a RNA aptamer to bind small molecule ligand, and turning the aptamer into a riboswitch. Even though more and more riboswitch screening methods have been invented to accomplish a higher screening efficiency,<sup>1,2,3</sup> such screening processes may not be successful. Second, the adaptation of riboswitches for downstream applications in different microbial species can demand arduous efforts from both molecular engineering and microbiology. In the case of naturally occurring riboswitches, nature evolves the same riboswitch sequence into variants to meet the different demands of genetic expression regulation in various species. Likewise, a synthetic riboswitch needs engineering efforts to adapt to versatile roles that vary in ligand dose-response range, genetic expression dynamic range, and up/down-regulation

efficiency (activation/repression ratio). Finally, while mechanisms for most synthetic riboswitches have been proposed, a quantitative mechanism with predictive powers remains opaque.

Desai and Gallivan used the minimal TCT8-4 theophylline aptamer discovered by Jenison et al.<sup>4</sup> and created a theophylline-dependent up-regulatory riboswitch.<sup>5</sup> Their initial result suggested that this riboswitch can activate the downstream  $\beta$ -galactosidase gene expression by 12-fold.<sup>5</sup> Using a number of advanced selection methods developed in our lab, a plethora of theophylline-responsive riboswitches have been created that exhibit distinct characteristics.<sup>1,3,8</sup> Termed A through G, these theophylline riboswitches vary in their activation ratio and reporter gene expression regulation ranges in *E. coli*. Uniformly based on the same theophylline aptamer, these riboswitches vary in their linker region between aptamer sequence and starting codon also known as expression platform. A more recent report demonstrated that these riboswitches performed well in various model organisms with minimal modification.<sup>8,9</sup>

As ‘proof-of-principle’ reports have suggested the feasibility of creating and adapting novel riboswitches into various tasks,<sup>10</sup> then, the next hurdle for riboswitch research is how can we account for the mechanism of these riboswitches and use this knowledge to further improve our design for synthetic riboswitches.



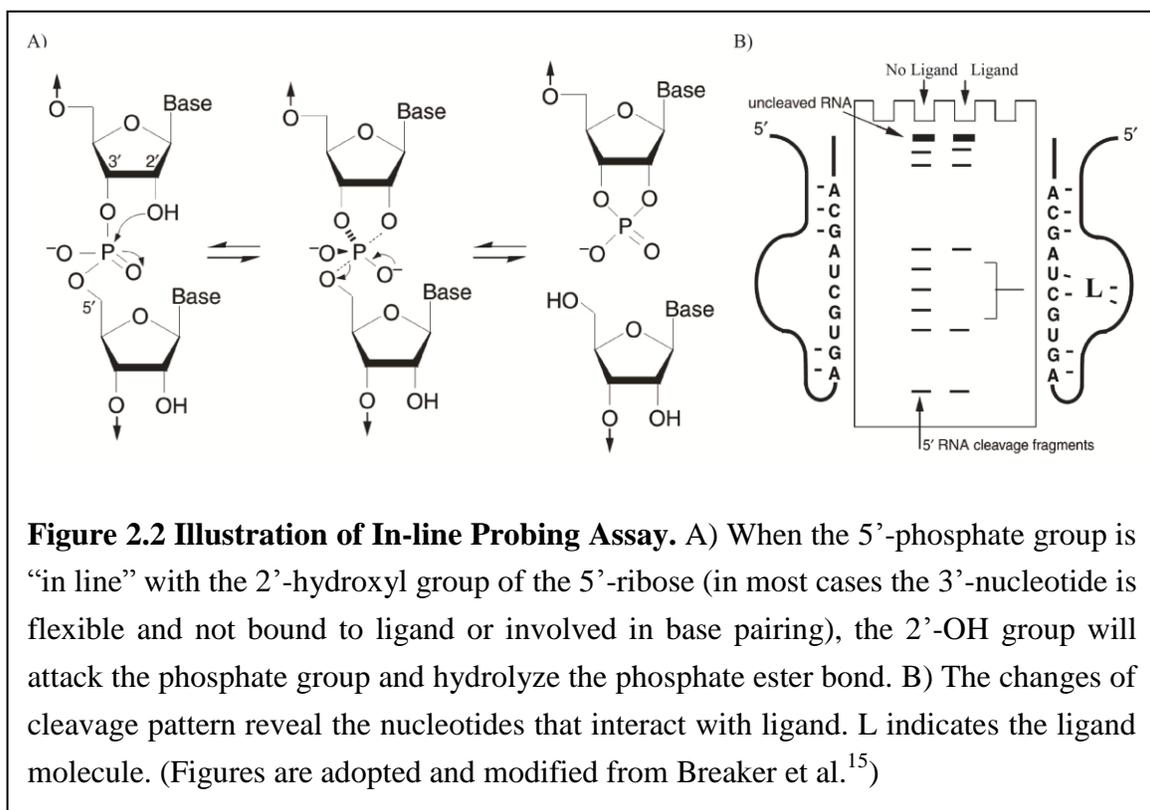
**Figure 2.1 Proposed Theophylline Riboswitch Mechanism.**-Two different folding structures of theophylline riboswitch D can adopt. The global energy minimum of the OFF state (left-most structure) contains the ribosome binding site (RBS) buried in the stem-loop structure, whereas the ON conformation (middle and right structure) can be stabilized by the binding of ligand.

In their theophylline riboswitch mechanism research, Desai and Gallivan constructed a simple model for the synthetic theophylline riboswitch.<sup>11</sup> (See Figure 2.1) They proposed two conformations of the riboswitch sequence based on the structural information of the aptamer and the predicted folding structure of riboswitch with mFold.<sup>5</sup>

<sup>12</sup> For example, in our theophylline riboswitch D, the riboswitch sequence in the 5'-untranslated region (UTR) adopts the most stable folding scheme called 'OFF' conformation in ligand-free conditions, which has 17 Watson-Crick-like pairs (A-U, C-G) or sub-optimal pairs (U-G). However, if the ligand is present, the riboswitch region can fold into an 'ON' conformation with a stem-loop which contains 12 base pairs and a binding pocket for theophylline. The alternative 'ON' conformation is thought to be

stabilized by the negative free energy from ligand binding, and the exposed ribosome binding site (RBS) in the new conformation can lead to the activated expression of downstream gene. (See Figure 2.1)

To support this hypothesized mechanism, Lynch and Gallivan later designed a covariant mutation experiment studying riboswitch F that yielded valuable insight into the theophylline riboswitch mechanism.<sup>3</sup> They focused on the base pairs in the predicted stem-loop structure of unbound RNA molecules. When there is a single mutant that can interrupt the pairing in the stem-loop, the downstream reporter showed a higher basal expression level. When a covariant mutant restores the Watson-Crick-like nucleotide pairs, the basal expression can be reduced to the parent level. The authors believed that



the stability of the ‘OFF’ conformation affected the equilibrium concentration of RNA molecules thus leading to a difference in basal expression.

The results from Lynch and Gallivan clearly gains favors for our proposed mechanism, however, it only provided a qualitative explanation of the complex mechanism of a single theophylline riboswitch.<sup>3</sup> Specially, the direct involvement of the ligand in the ligand-responsive RNA conformation changes has not been proven.

A widely accepted method used to investigate riboswitch mechanisms is the in-line probing assay. Breaker and co-workers utilized this method to validate the interaction between the first class of TPP and AdoCbl riboswitches and their ligands.<sup>13,14</sup> This method reveals different hydrolysis rates of nucleotides under ligand-free and ligand-bound conditions. (See Figure 2.2) When the RNA molecule adopts a secondary structure that allows the flexibility of a ribonucleotide (not in base-pairing or bound to ligand), the 3'-phosphate ester bond of unbound nucleotide can rotate into an “in-line” position where 2'-OH group can serve as a nucleophile under mildly basic conditions. The difference in hydrolysis rates can reveal conformational information about RNA and can explain the conformation shift upon adding ligand. A detailed procedure for in-line probing assay was reported by Breaker et al.<sup>15</sup>, and is described in the Experimental section.

Meanwhile, recent developments in computational RNA structure modeling have been able to predict the folding structure and free energy for riboswitches.<sup>16,17,18</sup> Progress

has been made to ensure higher accuracy on predicting the partition function over all possible secondary structures. Based on those computational models, we can separately calculate the folding free energy for different conformations in our proposed model for the theophylline riboswitch. Combining our experimental data with advanced computational RNA modeling, a more detailed riboswitch function model can be constructed that has quantitative prediction power.

In this chapter, we extend mutagenesis studies to reveal mechanistic insights of our model theophylline riboswitch D, provide evidence for direct interaction between the riboswitch and its ligand by *in vitro* in-line probing experiments, and proposed a more comprehensive biophysical model for theophylline riboswitch based on computational models. These results can provide more comprehensive details about the ligand-dependent conformational shift in theophylline riboswitches at the molecular level and provide quantitative models to predict riboswitch function. We anticipate that our efforts described below will extend the understanding of riboswitch mechanisms and improve the rational design of synthetic riboswitches.

## **2.2 Results**

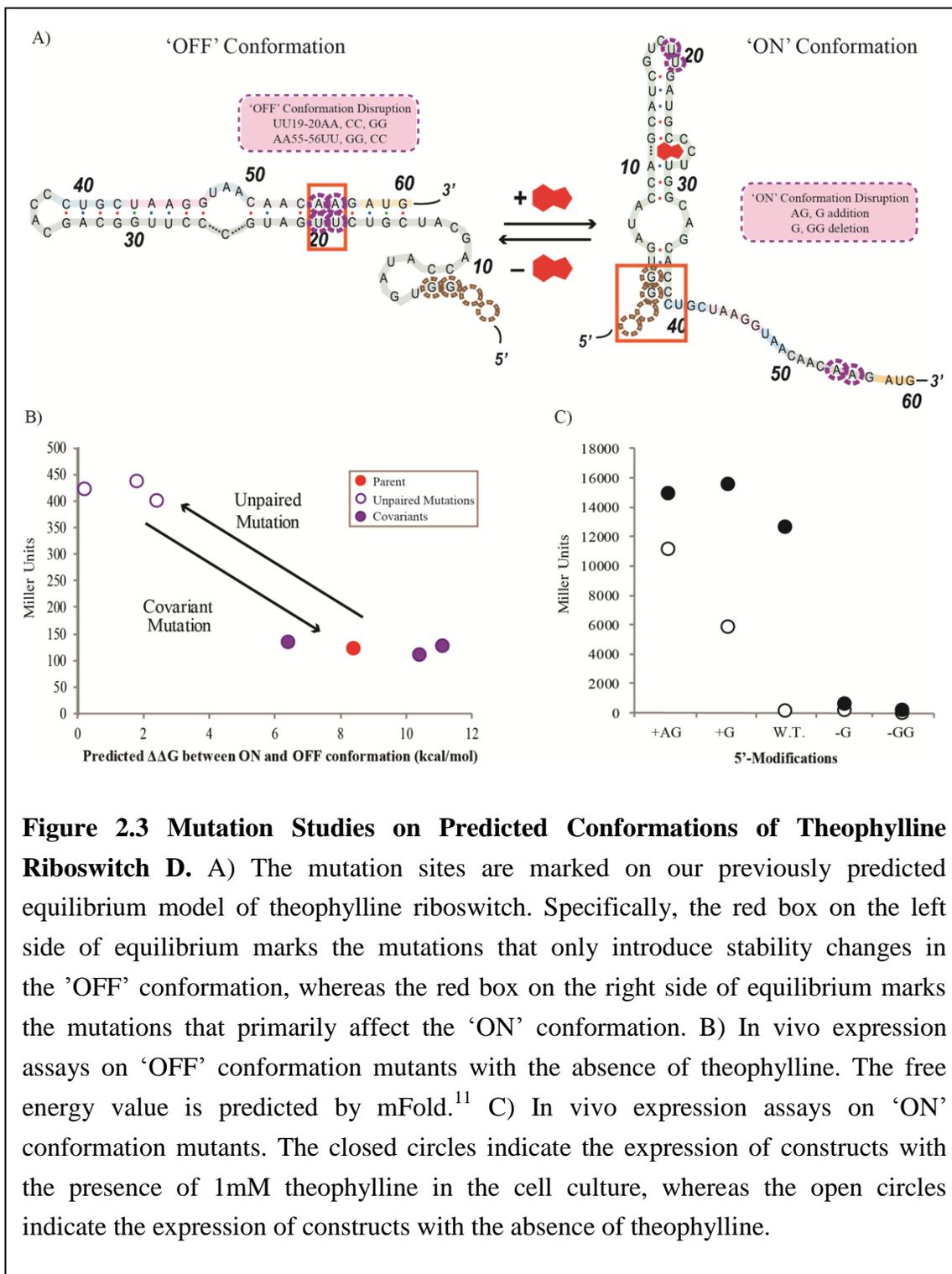
### **2.2.1 Mutation Studies Reveal Riboswitch Mechanistic Insight.**

To extend Lynch's mutagenesis study, we designed a set of mutations to explore the effect of both 'OFF' and 'ON' conformation stability on riboswitch function. Based on our previous equilibrium model outlined in Figure 2.2A, two sets of mutations are located

at predicted stem-loop regions in ‘OFF’ conformation and ‘ON’ conformation respectively, and intended to disrupt one stem-loop structure without affecting the other.

Firstly, we focused on UU19-20 and AA55-56 which form two AU pairs in the ‘OFF’ conformation (shown in the red box of the left ‘OFF’ conformation in Figure 2.2A) but do not participate in any base-pairing in the ‘ON’ conformation (purple-circled). The mFold algorithm predicts a 8.4 kcal/mol free energy difference between ‘ON’ and ‘OFF’ conformations ( $\Delta\Delta G$ ) in the parent D switch construct (also known as pSAL12.1 construct) (closed red circle in Figure 2.2B) and 0.2 to 2.4 kcal/mol  $\Delta\Delta G$  in the single-sided UU19-20 mutants (open purple circles in Figure 2.2B). A decrease in  $\Delta\Delta G$  results in less stable ‘OFF’ conformations for the mutants comparing the parent construct. Whereas the free energy difference  $\Delta\Delta G$  can be restored to 6.4 and 11.1 kcal/mol by implementing a covariant AA55-56 mutation (closed purple circles in Figure 2.2B). Thus, the ‘OFF’ conformation stability of the covariant mutants should be close to the parent level. In vivo assays revealed that all unpaired mutations (UU19-20AA, UU19-20CC, UU19-20GG) strains showed a 4-fold higher basal expression level under ligand-free conditions. This increased basal expression can be restored to the original level by mutating the AA55-56 region to the complimentary nucleotides to form the more stable ‘OFF’ conformation stem-loop (Figure 2.2B).

Next, we modified the 5’-end of riboswitch to fine-tune the stability of ‘ON’ conformation without affecting the ‘OFF’ conformation. We either inserted additional



**Figure 2.3 Mutation Studies on Predicted Conformations of Theophylline**

**Riboswitch D.** A) The mutation sites are marked on our previously predicted equilibrium model of theophylline riboswitch. Specifically, the red box on the left side of equilibrium marks the mutations that only introduce stability changes in the 'OFF' conformation, whereas the red box on the right side of equilibrium marks the mutations that primarily affect the 'ON' conformation. B) In vivo expression assays on 'OFF' conformation mutants with the absence of theophylline. The free energy value is predicted by mFold.<sup>11</sup> C) In vivo expression assays on 'ON' conformation mutants. The closed circles indicate the expression of constructs with the presence of 1mM theophylline in the cell culture, whereas the open circles indicate the expression of constructs with the absence of theophylline.

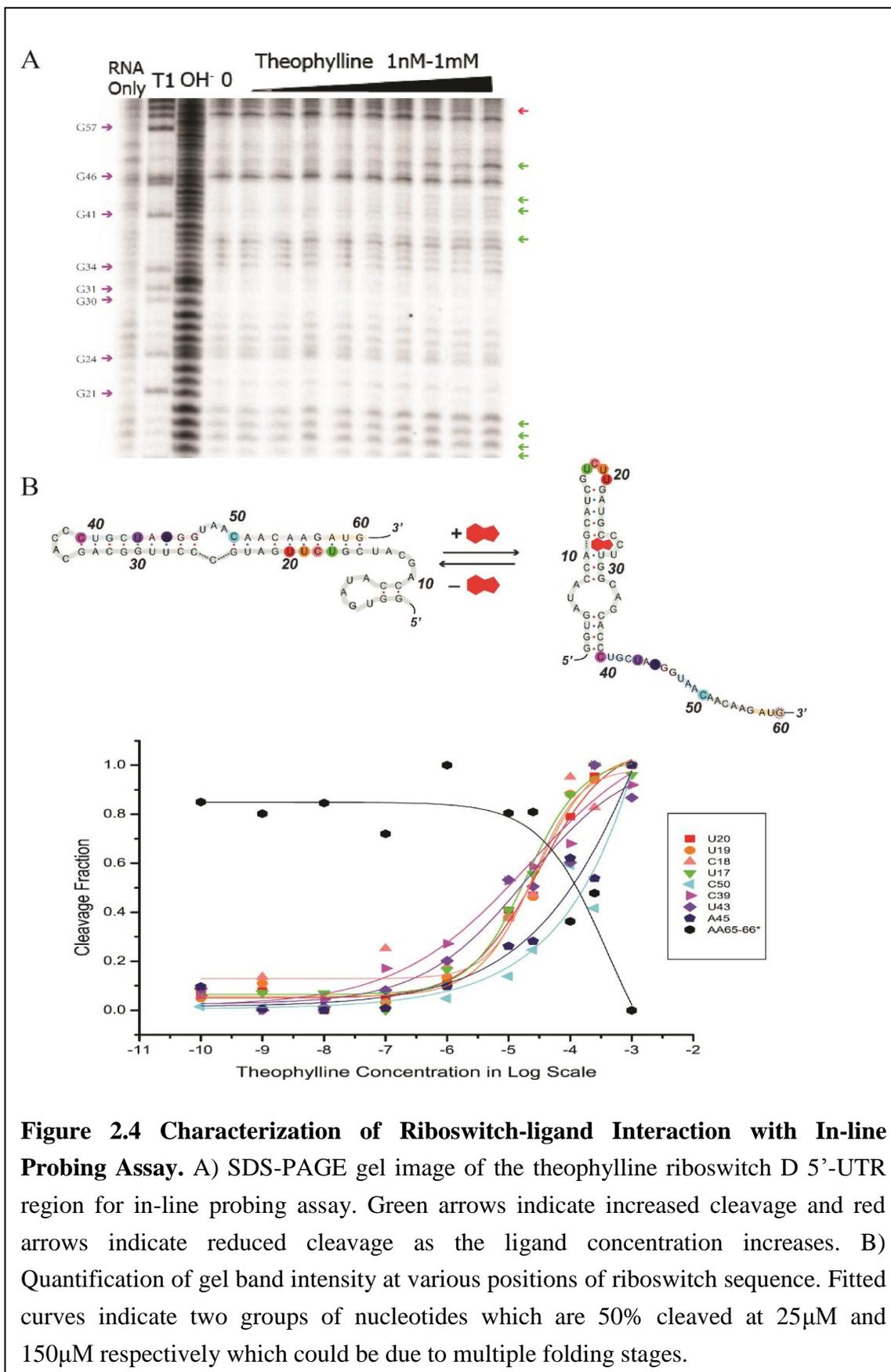
nucleotides (AG and G) to increase the length of stem-loop structure or deleted nucleotides (G and GG) in the 5'-end to change the number of base pairing in the 'ON'

conformation. According to our previous model, additional base pairs at the 5'-end will stabilize the 'OFF' conformation stem-loop structure, thus shifting the equilibrium towards the species that have fast translation. The mFold server predicts free energy differences ( $\Delta\Delta G$ ) between 'ON' and 'OFF' to be 4.5, 5.4, 8.4, 9.4, and 11.3 kcal/mol as we remove the 5'-ribonucleotides. *In vivo* assays of these mutations suggest that expression levels in both the absence (open circles in Figure 2.3C) and the presence (closed circles in Figure 2.3C) of theophylline are increased as we introduce more base pairs in the 5'-end of riboswitch.

In addition, we created the CC8-9 and GG30-31 covariant mutations, which are adjacent to the proposed binding pocket of aptamer sequence. These mutants completely disrupted the riboswitch function and implied the conserved regions of our theophylline riboswitches. Meanwhile, we removed the aptamer region and modified the RBS region to build no-aptamer control sequence for our covariant mutants. These variants yielded different expression levels that coincide with our discoveries in various microbes that have different preferences in RBS sequence and altered functionality of riboswitches.

### **2.2.2 Characterization of Direct Interaction between RNA Molecule and Theophylline Using In-line Probing Assay**

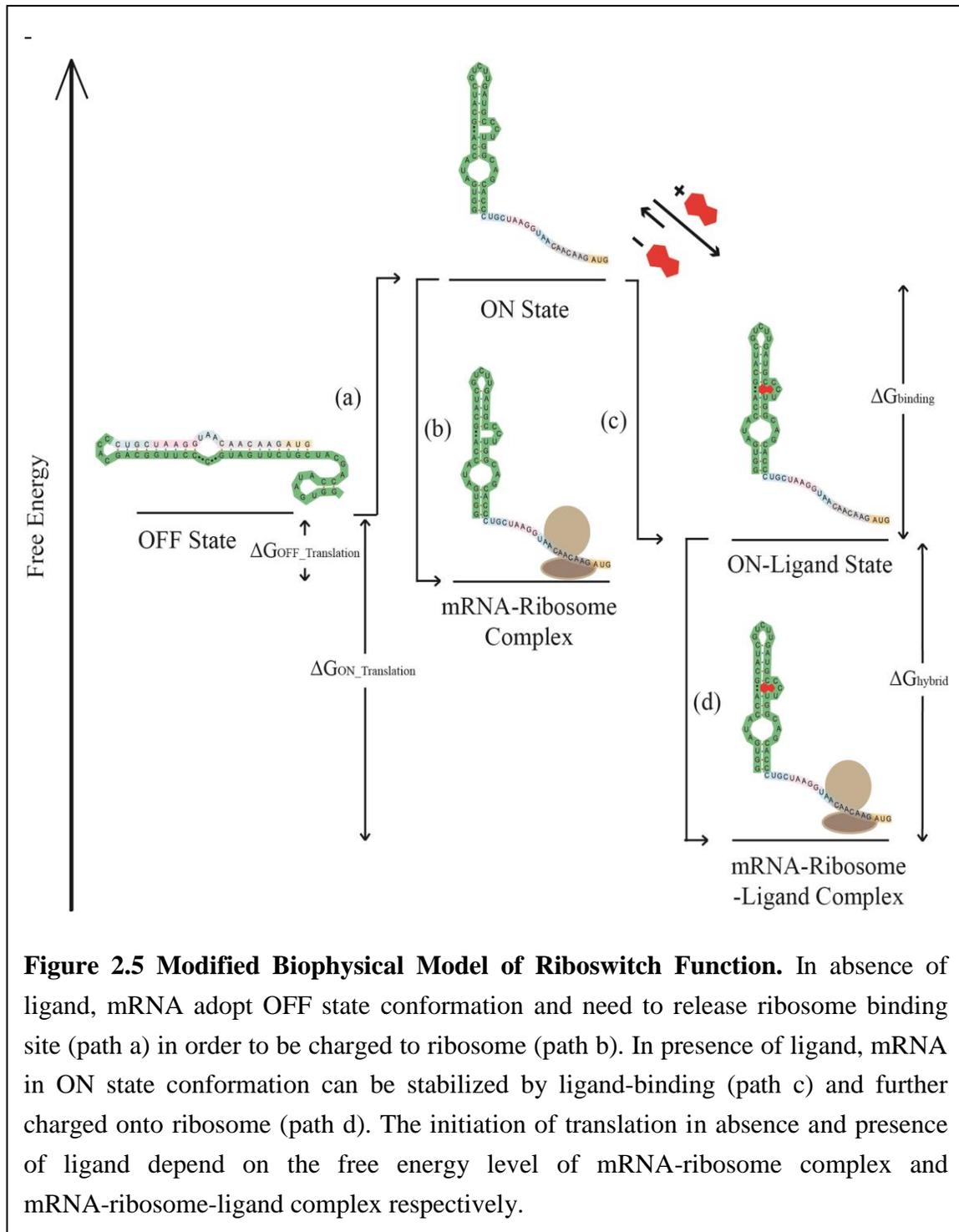
While the nuclease digestion experiments by Lynch and Gallivan validate their proposed simple mechanism for theophylline riboswitch, they alone do not fully explain the ligand-binding process. To confirm the ligand-dependent equilibrium model, we performed an in-line probing assay to directly characterize the binding between



theophylline riboswitch and its ligand. The intensity of bands on the gel represents the cleavage efficiency of flexible nucleotides in mRNA molecules under different ligand concentration. Normally, if a nucleotide is in a base pair or bound to ligand, the cleavage will be limited, thus yielding a light band. And if the ribonucleotide is not incorporated in a fixed position, the flexible 2'-OH group will catalyze the self-cleavage reaction, thus rendering a dark band. With the increasing concentration of ligand, we expect mRNA molecules will shift from 'OFF' conformation to 'ON' conformation in the equilibrium. The gel image of riboswitch regions showed increasing (green arrows in Figure 2.3A) or decreasing (red arrow in Figure 2.3A) cleavage at different nucleotide position as the concentration of theophylline is increasing. The shifts in cleavage patterns of highlighted nucleotides complied with our hypothesized model and explained the predicted 'ON' and 'OFF' secondary structures at a molecular level.

Additionally, the figure summarizing dose-response cleavage has shown consistent cleavage patterns at nine different nucleotide positions of theophylline riboswitch D. (Shown in Figure 2.3B) The fitted curves converged into a narrow ligand concentration range and suggest that there might be two affinity coefficients for the binding of theophylline riboswitch and its ligand. The lower affinity coefficient of theophylline riboswitch which is exhibited by the 5'-nucleotides is at 25 $\mu$ M, while the higher one is measured at 150 $\mu$ M. These apparent  $K_d$  values are two-to-three orders of magnitude higher than the prototype theophylline aptamer and may indicate a two-step folding in the

conformational change during mRNA activation which will be discussed in the discussion section.

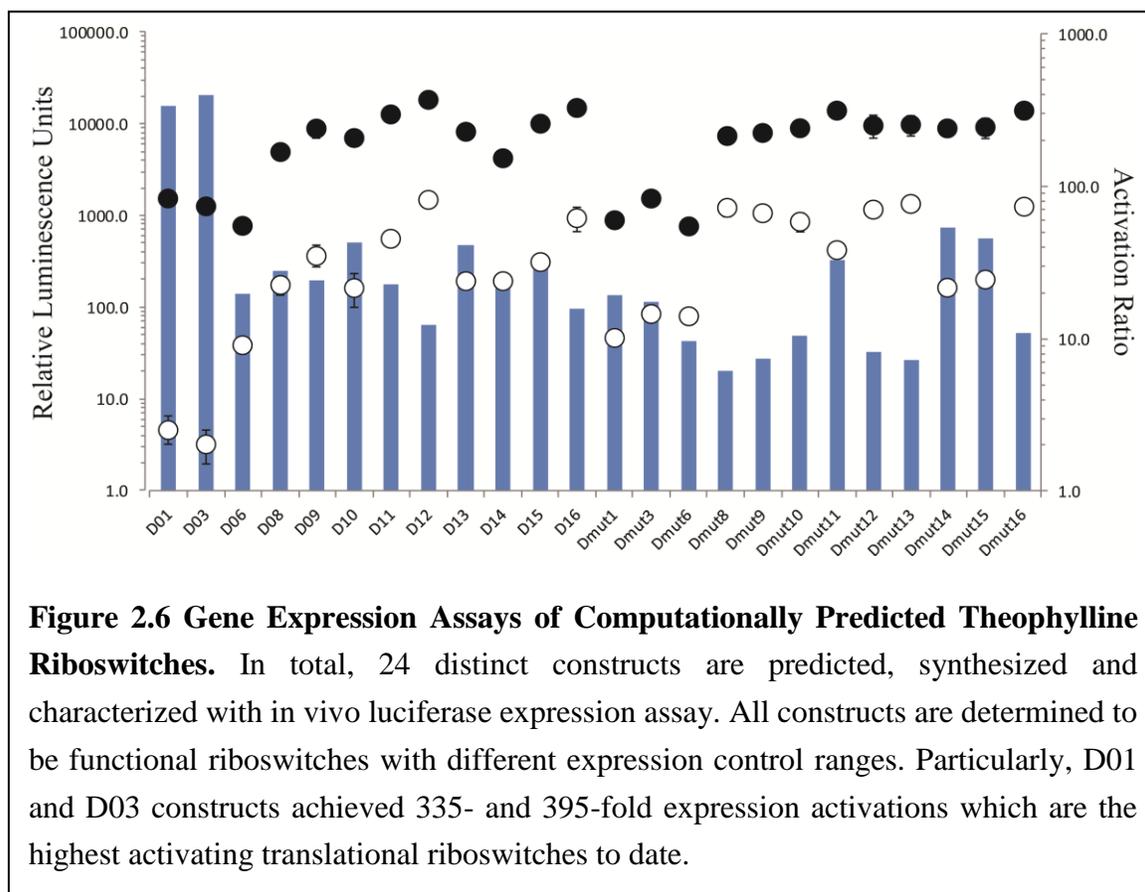


### 2.2.3 Construction of Detailed Biophysical Model for Theophylline Riboswitches

Due to the qualitative nature of our previous model, its predictive power is limited. A quantitative model would help to advance our understanding about riboswitches and give us the tool to rationally design novel riboswitches. In collaboration with the Salis lab, we proposed a more comprehensive biophysical model for riboswitch functionalities. (See Figure 2.5) In this new model, we fully account for another important player in the riboswitch regulatory process—the ribosome. The intuition of including ribosome interaction in our new model is that mRNA needs to be charged onto ribosome and form a complex structure to initiate translation, thus the free energy associated with the RNA-ribosome interaction is important for riboswitch function.

Briefly, when ligand is absent, the majority of mRNA molecules stay in the minimal energy state-OFF state. Translation in the no ligand condition will only happen when mRNA adopts a new conformation that has an exposed ribosome binding site (RBS) (Figure 2.5 path **a**) and reach to mRNA-ribosome complex state (Figure 2.5 path **b**) with a free energy change ( $\Delta G_{\text{OFF\_Translation}}$ ) from  $\Delta G_{\text{OFF}}$  to  $\Delta G_{\text{mRNA-Ribosome Complex}}$ . Whereas, when the ligand is present, the suboptimal ON state can be stabilized by binding to ligand with a negative  $\Delta G_{\text{binding}}$  (Figure 2.5 path **c**) and subsequently charged onto a ribosome (Figure 2.5 path **d**). Thus, the translation initiation under abundant ligand condition happens with a greater decrease of free energy ( $\Delta G_{\text{ON\_Translation}}$ ) from  $\Delta G_{\text{OFF}}$  to  $\Delta G_{\text{mRNA-Ribosome-Ligand Complex}}$ .

Previous computational studies of RNA folding have provided handy tools to predict the free energies on different conformations of RNA molecules as well as the free energies associated with RNA-ribosome interactions<sup>17,18,19</sup>. Insights from computational research in the Salis lab allow us to predict the free energies for mRNA folding energy ( $\Delta G_{\text{mRNA}}$ ), ribosome-charging processes ( $\Delta G_{\text{hybrid}}$ , the binding free energy of Shine-Dalgarno (SD) sequence of mRNA and anti-SD sequence in 16s rRNA, and  $\Delta G_{\text{standby}}$ , the binding free energy of 30S ribosome subunit), the free energy penalty of the spacing between RBS and starting codon ( $\Delta G_{\text{spacing}}$ ), as well as the free energy of tRNA binding ( $\Delta G_{\text{start}}$ ). Using these predicted values to approximate the free energy changes for



translation initiation processes under our biophysical riboswitch model, we are able to predict functional riboswitch constructs *in silico* and yield a set of 24 distinct putative theophylline riboswitches (Unpublished results). These distinct constructs are then synthesized *in vitro* and cloned into plasmid to control the downstream luciferase reporter gene in *E.coli*. All 24 constructs are functional riboswitches. (Summarized in Figure 2.6) Two particular constructs exhibit activation ratios of 395 and 335-fold which are the highest among theophylline riboswitches found.

### 2.3 Discussion

In our previous equilibrium model for the theophylline riboswitch illustrated by Figure 2.1, the performance of riboswitches are explained in Equation 1 developed by Lynch and co-workers, where  $k$  indicates the translation initiation rate of different constructs, [ON] and [OFF] indicate the concentration of ‘ON’ and ‘OFF’ conformations in the absence of ligand, and [ON]’, [OFF]’, [ON•L]’ represent the concentration of ‘ON’, ‘OFF’, and ligand-bound structures when the ligand is present.<sup>3</sup> In that model, we made several assumptions. First, we used the mRNA population and its relative translation rate as a proxy to estimate the expression efficiency for each condition. Second, to estimate the relative translation rate of each conformation, we considered the accessibility of the RBS and expected a much lower translation initiation rate for the ‘OFF’ conformation than the ‘ON’ conformation ( $k_{OFF} \ll k_{ON}$ ) and a similar translation rate between the ‘ON’ conformation and the ligand-bound ‘ON’ structure ( $k_{ON} \cong k_{ON-L}$ ).

With the above assumptions, we can qualitatively estimate the population of each conformation based on their folding free energy and the binding energy ( $\Delta G_{\text{binding}}$ ) which shifts the equilibrium in Figure 2.1. Therefore, the riboswitch regulation happens when the addition of ligand leads to a much higher ligand-bound population than the ‘ON’ population in absence of ligand ( $[ON \cdot L]' > [ON]$ ) and, consequently, the translation occurs much more readily upon adding the riboswitch ligand.

$$\text{Activation Ratio} = \frac{\text{Expression}_{ON}}{\text{Expression}_{OFF}} = \frac{[OFF]' * k_{OFF} + [ON]' * k_{ON} + [ON \cdot L]' * k_{ON \cdot L}}{[OFF] * k_{OFF} + [ON] * k_{ON}} \quad (1)$$

Our results on measuring various mutants that have altered conformational stabilities provided validity for our 3-state equilibrium model for theophylline riboswitches. The mutations introduced to alter the ‘OFF’ conformation stability essentially lowers the energy barrier for mRNA to enter the ‘ON’ state, thus, higher basal expression is observed for those mutants in absence of ligand. When the covariant mutations were made to restore the stable ‘OFF’ conformation, the energy barrier is once again preventing the riboswitch sequence from adopting the ‘ON’ conformation and yielding low basal expression. Meanwhile, the step-wise reduction of base pairs by removing ribonucleotide at the 5’-end in the ‘ON’ conformation showed that decreasing stability of ‘ON’ structure will increase the energy barrier for riboswitch to escape from the ‘OFF’ conformation, thus leading to a lower basal expression in absence of ligand while

diminishing the activating function of the riboswitch. Taking both results together, we conclude that both folding free energies of 'OFF' and 'ON' conformations ( $\Delta G_{\text{OFF}}$  and  $\Delta G_{\text{ON}}$ ) contribute to the overall performance of a riboswitch. The free energy difference between two states ( $\Delta\Delta G$ ) with regard to the ligand-binding free energy ( $\Delta G_{\text{binding}}$ ) plays a delicate role in determining the performance of riboswitches.

After validating our riboswitch function model with mutagenesis studies, we attempted to confirm our model at a molecular level. Prior to our investigation, the only available structural information we can identify was provided by imino proton NMR characterization of the theophylline RNA aptamer TCT8-4.<sup>4</sup> Therefore, an in-line probing assay can help to provide evidence that theophylline TCT8-4 aptamer-derived riboswitch D can induce a structural shift upon binding theophylline according to our proposed model. In our experiment, a number of ribonucleotides exhibit cleavage pattern shifts under different ligand concentration (marked in Figure 2.4). These shifts in the riboswitch region agreed well with the two predicted secondary structures of RNA sequence and support our proposed mechanism.

Furthermore, we observed a much lower ligand affinity of our riboswitch than its prototype aptamer (2 to 3 orders of magnitudes). We expect that the length of RNA molecule can affect the  $K_d$  value (the measurement of ligand affinity). When RNA gets longer, it can usually adopt multiple alternative folding structures with lower free energy and achieve higher stability in physiological conditions. Therefore, the free energy

provided by ligand binding would become less effective in changing the secondary structure of the riboswitch sequence, which renders a higher apparent  $K_d$ . In the report of prototype theophylline aptamer, the  $K_d$  value (320 nM) is determined by using a RNA sequence of minimum aptamer length (42 nt).<sup>4</sup> However, in our experiment, a ~200nt fragment of RNA molecule (riboswitch D linked with IS10 region) is used to perform the in-line probing assay and achieves a  $K_d$  of 25  $\mu$ M. We estimate that the mRNA that is transcribed in full-length will have even lower ligand affinity than the apparent  $K_d$  value we measured in the in-line probing assay. Therefore, in practice, the working concentration of riboswitch ligand (1 mM) may often exceed several orders of magnitude than the inherent aptamer binding affinity.

Additionally, the fitted curves of cleavages at different nucleotide positions converge into two groups indicating two different affinity constants ( $K_d$ ) at 25  $\mu$ M and 150  $\mu$ M respectively. We found that the nucleotides that converge at a low  $K_d$  (25  $\mu$ M) are located toward 5'-end and near the proposed binding pocket of the aptamer, whereas the nucleotides that showed a higher  $K_d$  (150  $\mu$ M) are located towards 3'- downstream near the reporter gene. This observation suggests that the nucleotides in the riboswitch sequence may have uneven affinities towards ligand. Within the association/dissociation equilibrium between RNA aptamer and its ligand, 5'-end nucleotides may form tighter bonds with the ligand (resulting in a lower  $K_d$ ) and 3'-end nucleotides may play a minor role in ligand-binding due to the structural characteristics of the binding pocket. Even

though this hypothesis would not significantly change the mechanism of our theophylline riboswitch, a more detailed structural shift scheme in the riboswitch-ligand interaction can be expected.

Finally, based on the qualitative results above and recent research on mRNA-ribosome interactions,<sup>17</sup> we proposed a quantitative biophysical model for riboswitch function (described in Figure 2.5). The core concepts of our new model were developed upon our previous equilibrium model, where the three key states are kept in our new model and their folding free energy ( $\Delta G_{\text{OFF}}$  and  $\Delta G_{\text{ON}}$ ) can be estimated by computational models.<sup>12,19</sup> In addition, we introduced two more states—translation complexes with and without ligand—to account for the role of ribosome in our new model. These translation complexes are the species that can be readily translated, thus, the free energy change from the original ‘OFF’ state to these complexes dictates the translation initiation rates.

From a quantitative perspective, we can estimate the free energy level of ligand-free mRNA-ribosome complex ( $\Delta G_{\text{mRNA-ribosome complex}}$ ) by starting from computationally calculate the folding energy of ‘ON’ conformation ( $\Delta G_{\text{ON}}$ ) and adding two more energy components to account for the ribosome-charging process ( $\Delta G_{\text{hybrid}}$  for base pairing with anti-SD sequence in 16S subunit and  $\Delta G_{\text{standby}}$  for incorporation of 30S subunit), one parameter for penalizing spaces between RBS and starting codon ( $\Delta G_{\text{spacing}}$ ), and one parameter for tRNA incorporation ( $\Delta G_{\text{start}}$ , fixed value).<sup>15,16</sup> (Equation 2) In a similar

fashion, we can estimate the free energy level of ligand-bound mRNA-ribosome-ligand complex ( $\Delta G_{\text{mRNA-ribosome-ligand complex}}$ ) by simply adding the ligand-binding free energy ( $\Delta G_{\text{binding}}$ ).

$$\Delta G_{\text{mRNA-ribosome complex}} = \Delta G_{\text{ON}} + \Delta G_{\text{hybrid}} + \Delta G_{\text{standby}} + \Delta G_{\text{spacing}} + \Delta G_{\text{start}} \quad (2)$$

Therefore, under ligand-free conditions, the translation initiation requires a free energy from path **a** to **b** in Figure 2.5 ( $\Delta G_{\text{OFF\_Translation}}$  expressed in Equation 3). The rate for translation initiation process should be, in theory, proportional to  $e^{-\Delta G_{\text{OFF\_Translation}}}$ . Meanwhile, under abundant ligand condition, the translation initiation requires a free energy from **path a, then to c and d** ( $\Delta G_{\text{ON\_Translation}}$  expressed in Equation 4). Thus, the translation initiation rate will be proportional to  $e^{-\Delta G_{\text{ON\_Translation}}}$

$$\Delta G_{\text{OFF\_Translation}} = \Delta G_{\text{mRNA-ribosome complex}} - \Delta G_{\text{OFF}} \quad (3)$$

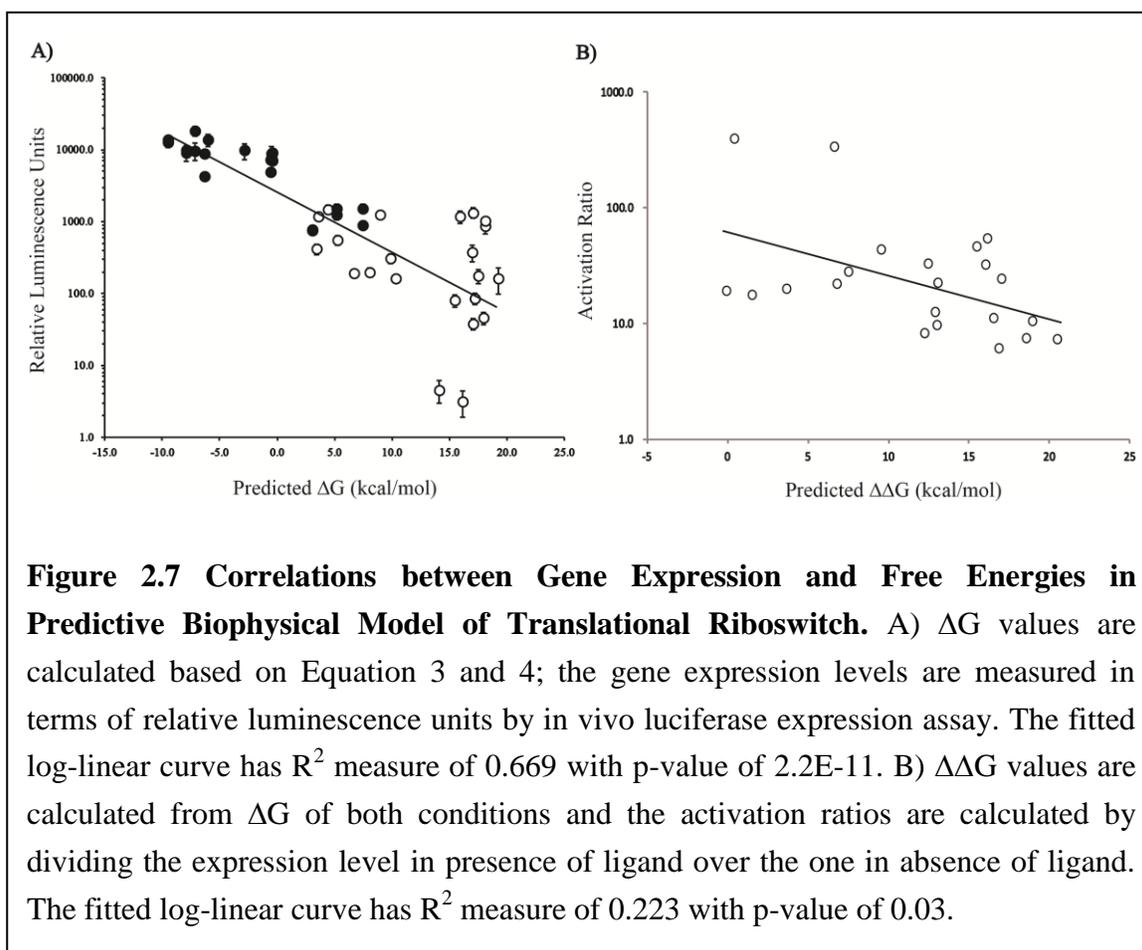
$$\Delta G_{\text{ON\_Translation}} = \Delta G_{\text{mRNA-ribosome-ligand complex}} - \Delta G_{\text{OFF}} \quad (4)$$

Based on the experimental results of our predicted theophylline riboswitches, we plotted the free energy levels of each conformation ( $\Delta G_{\text{ON\_Translation}}$  and  $\Delta G_{\text{OFF\_Translation}}$ ) against their reporter gene expression levels. (See Figure 2.7A, ON

conditions in closed circles, OFF conditions in open circles) The result matches nicely with our proposed quantitative model as the gene expression level decreases log-linearly with the increase of the predicted  $\Delta G$ . The coefficient of correlation  $R^2$  is resolved to be 0.669 with p-value of 2.2E-11.

$$\text{Activation Ratio} = \frac{\text{Expression}_{ON}}{\text{Expression}_{OFF}} = \frac{\beta e^{-\Delta G_{OFF\_Translation}}}{\beta e^{-\Delta G_{ON\_Translation}}} \quad (5)$$

Furthermore, the measurement on riboswitch function-activation ratio-should be



log-linearly proportional to the difference between the translation free energy under ON/OFF conditions ( $\Delta\Delta G$ ). (Equation 5) Thus, we plot the logarithmic activation ratio of each riboswitch against the difference ( $\Delta\Delta G$ ) between  $\Delta G_{ON\_Translation}$  and  $\Delta G_{OFF\_Translation}$ . The plot showed an approximate log-linear relationship with a  $R^2$  value of 0.223 and p-value of 0.03.

In summary, our biophysical model can explain how theophylline riboswitches function from a free energy perspective in a quantitative fashion. We expect the experience gained from this study provide insights to design other synthetic riboswitches in a similar mechanism.

## 2.4 Experimental

### *General Considerations*

All plasmid manipulations were conducted with standard cloning techniques and all constructs were verified by DNA sequencing (MWG Operon, Huntsville, AL). Purifications of plasmid DNA, PCR products, and enzyme digestions were facilitated by using kits from Qiagen (Germantown, MD), New England Biolabs (Ipswich, MA). 5-bromo-4-chloro-indolyl- $\beta$ -D-glucuronide X-gal was purchased from US Biological (Swampscott, MA). Radioactive [ $\gamma$ - $^{32}P$ ]-ATP was purchased from MP Biomedicals (Santa Ana, CA). Theophylline, *o*-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG), ampicillin, and all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Bacto<sup>TM</sup> Agar was purchased from BD (Sparks, MD). Synthetic oligonucleotides primers were

purchased from IDT (Coralville, IA). All PAGE supplies were purchased from Bio-Rad Laboratories (Hercules, CA). All experiments were performed in *E. coli* TOP10 F' cells (Invitrogen, Carlsbad, CA) cultured in media obtained from EMD Bioscience (Merck KGaA, Darmstadt, Germany).

### ***Mutant Plasmid Constructions***

A cassette mutagenesis strategy was used to generate UU19-20, AA55-56, and 5'-modification mutants of theophylline riboswitch D. A PCR product was synthesized by using riboswitch D as template with forward primer which contains mutations. The forward primer also contains a XbaI site at the same position of the wild-type sequence. The reverse primer anneals to the IS10 region. The product is overlap with the reporter gene sequence to build the full length insertion fragment. The full-length cassette product was digested with XbaI and HindIII and ligated to the backbones bearing the same restriction sites to yield plasmids with desired mutations.

### ***In vivo $\beta$ -galactosidase expression assay***

TOP10F' cells were transformed with plasmid that was synthesized by cassette mutagenesis. A single colony is isolated on ampicillin-containing agar plate and verified by sequencing. Cells harboring desired plasmids were grown at 37 °C to reach mid-log phase ( $OD_{600} \approx 0.5$ ) in the absence and presence of theophylline (1 mM). Then, the cells are cooled on ice bath for 15 min, lysed in buffer solution containing chloroform and SDS. The  $\beta$ -galactosidase activity was measured by monitoring the biocatalytic

hydrolysis of ONPG using spectrometry. Assays were conducted in triplicate and repeated on different days. The Miller Units were calculated by the following formula:

$$\text{Miller Units} = \text{OD}_{420} / (\text{OD}_{600} \times \text{hydrolysis time} \times [\text{volume of cell lysate} / \text{total volume}])$$

### ***In-line Probing Assay***

To synthesize the transcription template, the forward primer for PCR amplification is designed to contain a fused T7 RNA polymerase promoter sequence and anneal to theophylline aptamer region. Using 1  $\mu\text{g}$  of the linear DNA template of theophylline riboswitch D, a 20  $\mu\text{L}$  in vitro transcription reaction was prepared using the AmpliScribe™ T7-Flash™ Transcription Kit from Epicentre Biotechnologies and incubated at 37 °C for 1 h. After transcription process, 1  $\mu\text{L}$  of DNaseI was added to the reaction mixture to remove the DNA template. Transcribed RNA was purified by phenol/chloroform extraction procedure and precipitated in ethanol.

The RNA was resuspended in nuclease-free water and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs). The resulting dephosphorylated RNA molecules were once again purified by phenol/chloroform extraction followed by ethanol precipitation. Resuspended RNA was labeled in the 5'-end using T4 polynucleotide kinase (New England Biolabs) and 0.5  $\mu\text{L}$  [ $\gamma$ -<sup>32</sup>P]ATP (5000 Ci/mmol, 150 mCi/ml). Radiolabeled RNA was purified using denaturing 8% polyacrylamide gel electrophoresis and eluted by crush/soak method.

Radiolabeled RNAs were then aliquoted into 30 kcpm per each sample. 10 samples were diluted in mild base buffer solutions (pH = 8.3) with different concentration of ligand and incubated at room temperature for 40 hrs. Right before loading samples on the 12% PAGE, the same amounts of sample are hydrolyzed by sodium carbonate solution (pH = 9) at 90°C for 5min and digested by RNase T1 at 55°C for 10min desperately. All the reactions are mixed with 2x loading dye and loaded on the denaturing polyacrylamide gel and run for approximate 3.5-4 hrs until the dye bromophenol blue reaches the bottom of the gel. The gel is then dried in a gel drier and analyzed with a GE phosphor imaging film and scanner. Data were measured by GelAnalyzer (GelAnalyzer.com) and analyzed by Origin (OriginLab.com).

### ***In vivo luciferase expression assay***

TOP10F' cells were transformed with plasmid that has luciferase gene under riboswitch control. A single colony is isolated on ampicillin-containing agar plate and verified by sequencing. Cells harboring desired plasmids were grown at 37 °C to reach mid-log phase ( $OD_{600} \approx 0.3$ ) and split to grow in the absence and presence of theophylline (2 mM) for 30 min. Then, the cell culture is assayed for its luciferase activity (protocol from Promega). Briefly, 90  $\mu$ L of cell culture are diluted with 10  $\mu$ L luciferase reagent buffer and put on dry-ice for 10 min. The mixture is then thawed in a room temperature in water bath for 10 min, and lysed with 300  $\mu$ L Cell Culture Lysis Buffer for 10 min. An aliquot of 20  $\mu$ L lysate and 20  $\mu$ L of luciferase substrate are mixed

right before the measurement using spectrometry. The chemo-luminescent signal in relative luminescent units (RLU) were calculated by dividing the spectrometer readings by the cell density (OD600) and normalizing to a standard no-aptamer control sample. Assays were conducted in triplicate and repeated on three different days.

## 2.5 References

1. Lynch, S.A., Topp, S. and Gallivan, J.P., *Methods Mol Biol*, **2009**. 540: 321-33.
2. Topp, S. and Gallivan, J.P., *Chembiochem*, **2008**. 9: 210-3.
3. Lynch, S.A., Desai, S.K., Sajja, H.K. and Gallivan, J.P., *Chem Biol*, **2007**. 14: 173-84.
4. Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B., *Science*, **1994**. 263: 1425-9.
5. Desai, S.K. and Gallivan, J.P., *J Am Chem Soc*, **2004**. 126: 13247-54.
6. Topp, S. and Gallivan, J.P., *J Am Chem Soc*, **2007**. 129: 6807-11.
7. Lynch, S.A. and Gallivan, J.P., *Nucleic Acids Res*, **2009**. 37: 184-92.
8. Topp, S., Reynoso, C.M., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R., Scott, J.R. and Gallivan, J.P., *Appl Environ Microbiol*, **2010**. 76: 7881-4.
9. Reynoso, C.M., Miller, M.A., Bina, J.E., Gallivan, J.P. and Weiss, D.S., *MBio*, **2012**. 3.
10. Serganov, A. and Nudler, E., *Cell*, **2013**. 152: 17-24.
11. Nudler, E. and Mironov, A.S., *Trends Biochem Sci*, **2004**. 29: 11-7.
12. Zuker, M., *Nucleic Acids Res*, **2003**. 31: 3406-15.
13. Winkler, W., Nahvi, A. and Breaker, R.R., *Nature*, **2002**. 419: 952-6.
14. Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. and Breaker, R.R., *Chem Biol*, **2002**. 9: 1043.

15. Regulski, E.E. and Breaker, R.R., *Methods Mol Biol*, **2008**. 419: 53-67.
16. Hofacker, I.L., *Methods Mol Biol*, **2014**. 1097: 71-84.
17. Espah Borujeni, A., Channarasappa, A.S. and Salis, H.M., *Nucleic Acids Res*, **2014**. 42: 2646-59.
18. Salis, H.M., Mirsky, E.A. and Voigt, C.A., *Nat Biotechnol*, **2009**. 27: 946-50.
19. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L., *Nucleic Acids Res*, **2008**. 36: W70-4.

## **Chapter 3. Studying Kinetic Aspects of the mechanism of a Theophylline Riboswitch**

### **3.1 Introduction**

RNA research can be broadly divided into three major categories: folding, structure, and function. These three aspects of RNA influence each other in a reciprocal way.<sup>1</sup> As an emerging type of RNA element, synthetic riboswitches possess distinct properties comparing to natural riboswitches and require intensive investigation on each of the topics above. Recent publications have been focused on improving the function of synthetic riboswitches and expanding the choice of ligands.<sup>2</sup> In contrast, research on intrinsic properties of synthetic riboswitches demands more attention.

In the previous chapter, we investigated structure-related properties and depicted a sketch of a biophysical model to explain how a theophylline riboswitch functions. The model states that the transition between two different genetic regulation modes is based on the thermodynamic equilibrium between different conformations of the same RNA sequence under different ligand concentrations. In that case, the underlying assumption is that the ligand binding event happens after the full length of mRNA is synthesized. However, in details, the process of transcription and translation process are coupled in bacteria.<sup>3</sup> Translation initiates once the 5'-end of mRNA molecule is available for ribosome binding, which happens long before the full-length mRNA is synthesized. This coupling phenomenon in bacteria imposes great challenges for our thermodynamic

equilibrium model. Firstly, and most importantly, when does the ligand-binding event happen relative to the transcription process? Secondly, if ligand binds to nascent RNA molecule during the transcription process, how does it affect the folding, structure, and function of riboswitch sequences? With these questions, we need to take a deeper look at the transcription process where riboswitches might be able to execute their function.

There are three major steps in the transcription process: initiation, elongation, and termination. In the initiation step, RNA polymerase binds to the “promoter” sequence (e.g. *tac* promoter) upstream of the unwound DNA template coding for the RNA transcripts. The incorporation of promoter and RNA polymerase sensor sequence allows the initiation of transcription from a specific start position. After synthesizing ~10 polynucleotides, RNA polymerase detaches from the promoter and enters the elongation stage. In the elongation stage, RNA polymerase synthesizes the RNA at a much faster speed (20–80 nt/s for a bacterial RNA polymerase) until recognizing a termination sequence (intrinsic termination sequence or rho-dependent termination site). In the termination step, RNA polymerase is paused by the termination structure formed by termination sequences and dissociates from the DNA template, and subsequently the RNA transcript is released from the polymerase.

Even though the transcription speed is often considered as a passive parameter for RNA folding, varying speeds of elongation can provide crucial time windows to influence the folding process. Considering the formation of a secondary structure that is

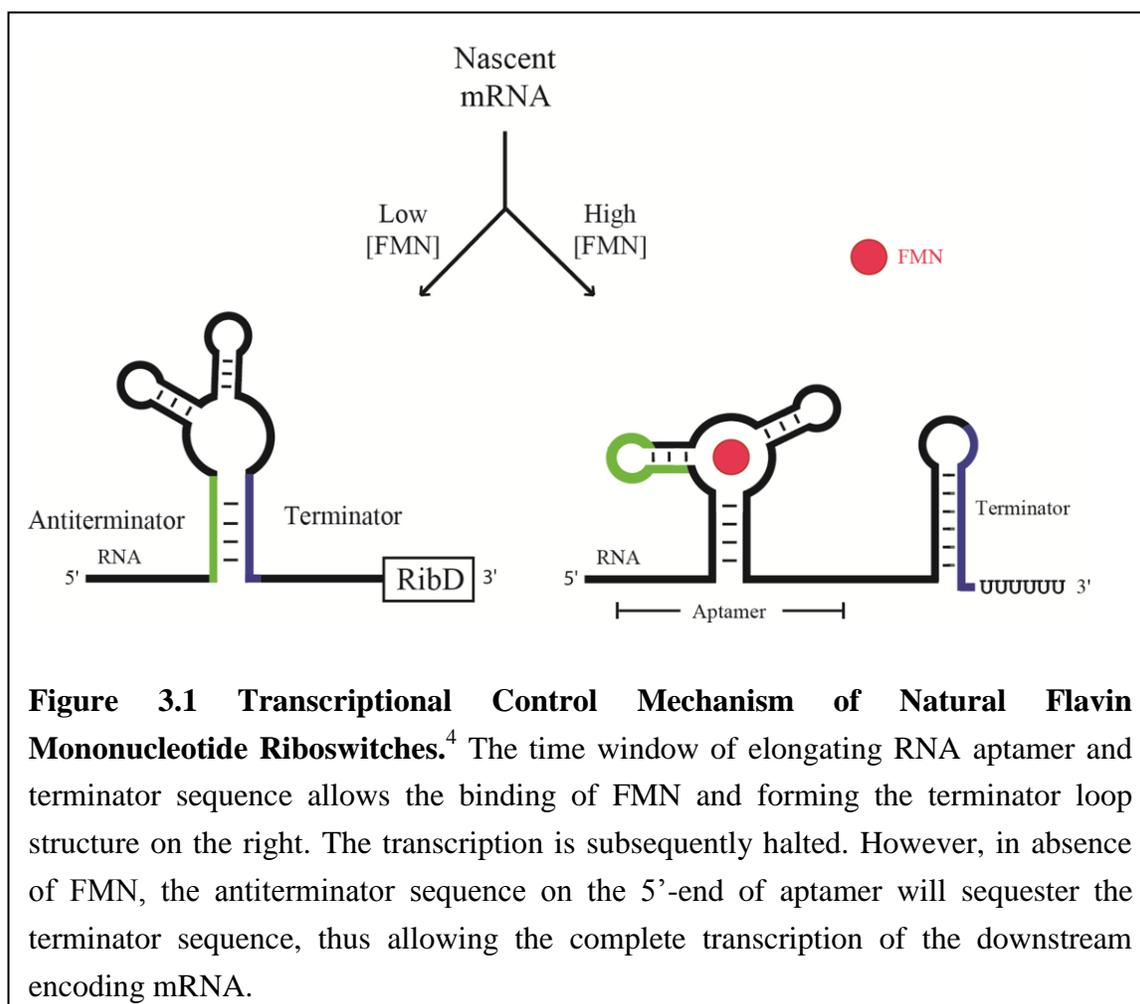
100 nt apart, the 5'-segment is synthesized 1 to 5 seconds prior to the complete synthesis of the 3'-segment depending on the cellular conditions and the choice of RNA polymerase. During this period, the 5'-segment may be involved in a non-native scaffold of existed polynucleotide, which prevents the normal function of native secondary structure.

In fact, transcriptional control riboswitches found in nature provide excellent examples to demonstrate the influence of the transcription process over riboswitch activities.<sup>4,5,6</sup> These transcriptional control riboswitches often require the binding of ligand to the nascent RNA polynucleotides to trigger the termination signal. For example, the flavin mononucleotide (FMN) RNA riboswitch is capable of forming an antiterminator structure when the RNA folds with limited ligand during transcription.<sup>4</sup> (See Figure 3.1) This antiterminator structure allows the successful transcription of the full length mRNA, which is then translated. Alternatively, in the presence of ligand, the nascent RNA molecule can fold into a ligand-bound secondary structure that sequesters antiterminator sequence. In the meantime, the crucial time window allows the formation of a terminator hairpin which halts the transcription process and releases the incomplete RNA transcript along with the RNA polymerase.

A report from Crothers *et al.* elegantly analyzed the folding process during transcription of a FMN riboswitch *in vitro*.<sup>7</sup> The authors discovered two transcriptional pausing sites, each with duration of up to 60 s, in 5'-end of the riboswitch sequence. The

effectiveness of these pausing sites is found to be affected by an elongation factor named NusA. Given the ample time window, the folding of FMN riboswitch sequence is kinetically driven with the strong dependence on the transcription speed and the ligand binding kinetics. Since the halting signal requires that the ligand binding occur during the synthesis of the riboswitch sequence, it rules out the possibility of reaching equilibrium. Later on, similar approaches have been used to study adenine-responsive riboswitches and revealed a similar mechanism.<sup>8</sup>

Even though our theophylline riboswitches appeared to exhibit a different



mechanism, the folding process during transcription may affect the translational control of exposing ribosome binding site upon binding to ligand. The kinetically driven process may change the overall folding landscape for the full-length RNA molecule, creating suboptimal conformations with kinetic traps, and ultimately influencing the riboswitch activity. These possibilities cannot be captured by our thermodynamic equilibrium model and require us to extend our investigation of riboswitch folding during the transcription process.

In a recent report, our co-workers published results indicating the signal of gene activation by our theophylline riboswitches can be triggered during transcription.<sup>9</sup> They decoupled the transcription and translation process using *in vitro* expression experiments and used RNA transcripts as the input to observe riboswitches activity. Interestingly, they found most of our riboswitches with high activation ratios tend to form a kinetic trap upon ligand binding, and to fold into a conformation that promotes translation of the downstream genes.

Results from the above report clearly suggest that there is a strong kinetic component among our theophylline riboswitches, and the equilibrium dissociation constant ( $K_d$ ) of the riboswitch may not accurately explain the riboswitch function. It also implied that traditional screening methods, which depend on the expression of the reporter gene over an extended time period, could favor translational control riboswitches over riboswitches that adopt kinetically driven mechanism.

Thus, in this chapter, we focus our attention on investigating the kinetic aspect of riboswitch function and extending our knowledge about the mechanisms of theophylline riboswitches. First, we investigate our theophylline riboswitch function with different time resolution. The results could reveal the initiation time required for riboswitch activation *in vivo* and the time-dependent regulatory control with a larger time span. Combined with empirical transcription and translation rate data, we can infer more precisely when the ligand binding event happens. Second, we try to study the difference in stability of ligand-free and ligand-bound RNA molecules by quantitative PCR experiment. The decay rates under different ligand concentrations would answer whether a thermodynamic equilibrium can be achieved. Last, we push beyond the *in vitro* method in Mishler et al.'s report<sup>9</sup> and create a procedure to decouple the transcription and translation process *in vivo* in *E. coli* to study the kinetic component of our theophylline riboswitches function.

## **3.2 Results**

### **3.2.1 Time Dependence of Theophylline Riboswitch Function**

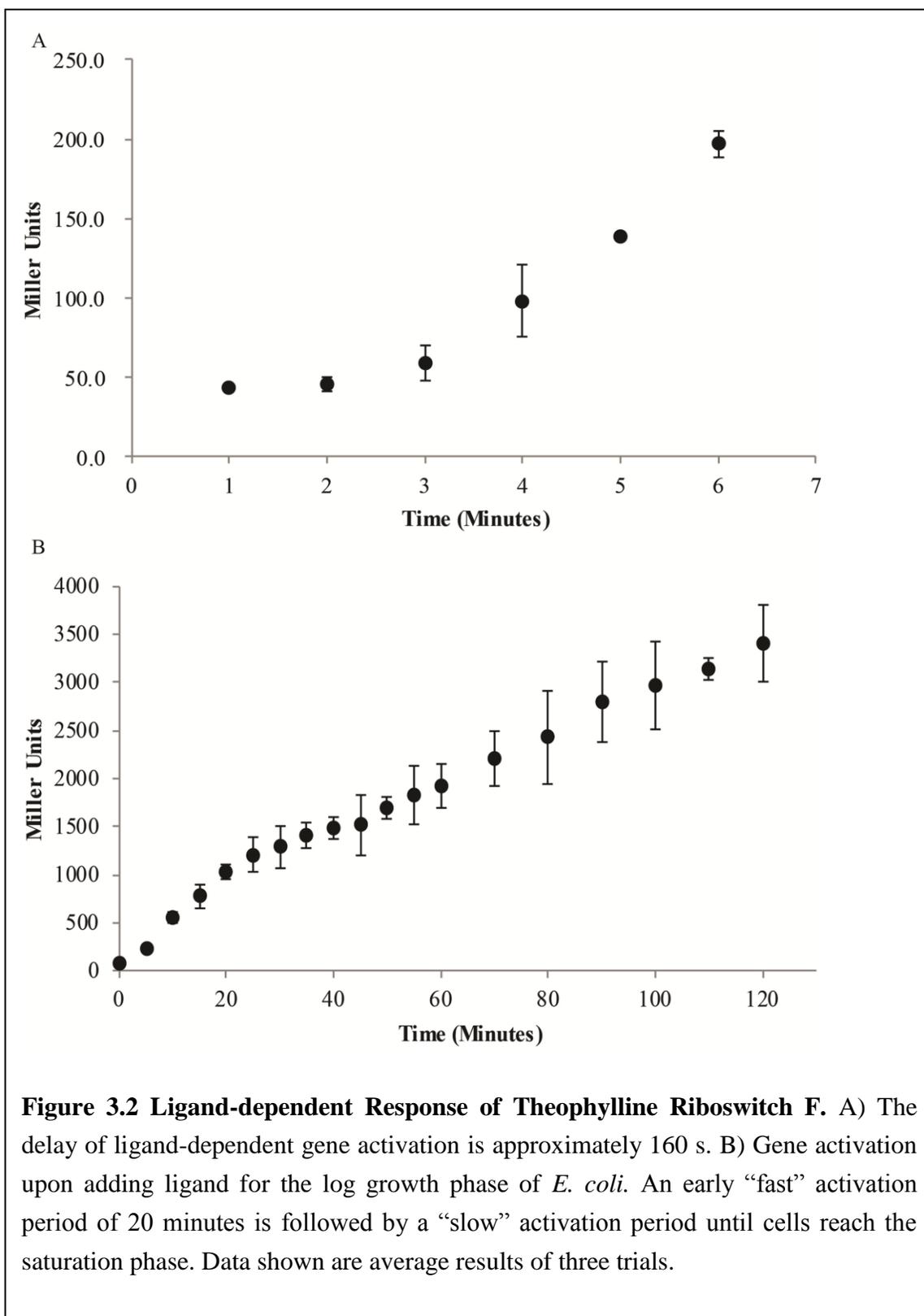
The question of how long a riboswitch takes to activate is interesting not only to engineers who want to create various devices with riboswitches, but also to researchers who try to reveal riboswitch mechanism. Here, we design a simple experiment to measure the delay between adding ligand and the signal activation of gene expression.

We use our existing riboswitch-*lacZ* fusion construct to perform this assay. *E. coli*

cells harboring the riboswitch-*lacZ* plasmid are grown into mid-log phase (OD<sub>600</sub> ~0.5). Then, 20 mM theophylline solution is added to culture 1:19 to make the final theophylline concentration of 1 mM. At 0, 1, 2, 3, 4, 5, 6 minutes after adding theophylline, an aliquot of culture is taken from the cell culture and assayed with the standard *lacZ* assay described in Chapter 2 (See Figure 3.2A). The result indicates that the delay time for riboswitch function is approximately 160 seconds, including transcription, translation, and protein folding, which will be discussed in Section 3.3. For the next 4 minutes, the gene expression increases in a linear fashion.

Next, we investigate the ligand-dependent reporter gene activation in a larger time span. Even though a large number of factors including transcription/translation rate, mRNA/protein hydrolysis rate influence the activation of translation and accumulation of reporter protein, nevertheless, it would be useful to examine the time-dependent gene activation for the reference of later applications.

Similarly, *E. coli* cells harboring the riboswitch-*lacZ* plasmid are grown to early to mid-log phase (OD<sub>600</sub> ~0.3). Then, 20 mM theophylline solution is added to culture 1:10 to make the final theophylline concentration of 2 mM. Within 120 minutes after adding theophylline, an aliquot of cell culture is taken from the cell culture and assayed with standard *lacZ* assay described in Chapter 2 (See Figure 3.2B). The result indicates that the activation of reporter gene under theophylline riboswitch control happens in a two-stage fashion. The first “fast” activation stage of about 20 minutes is followed by a “slow”



activation stage for a time up to 100 minutes where the cell enters into an early stationary phase.

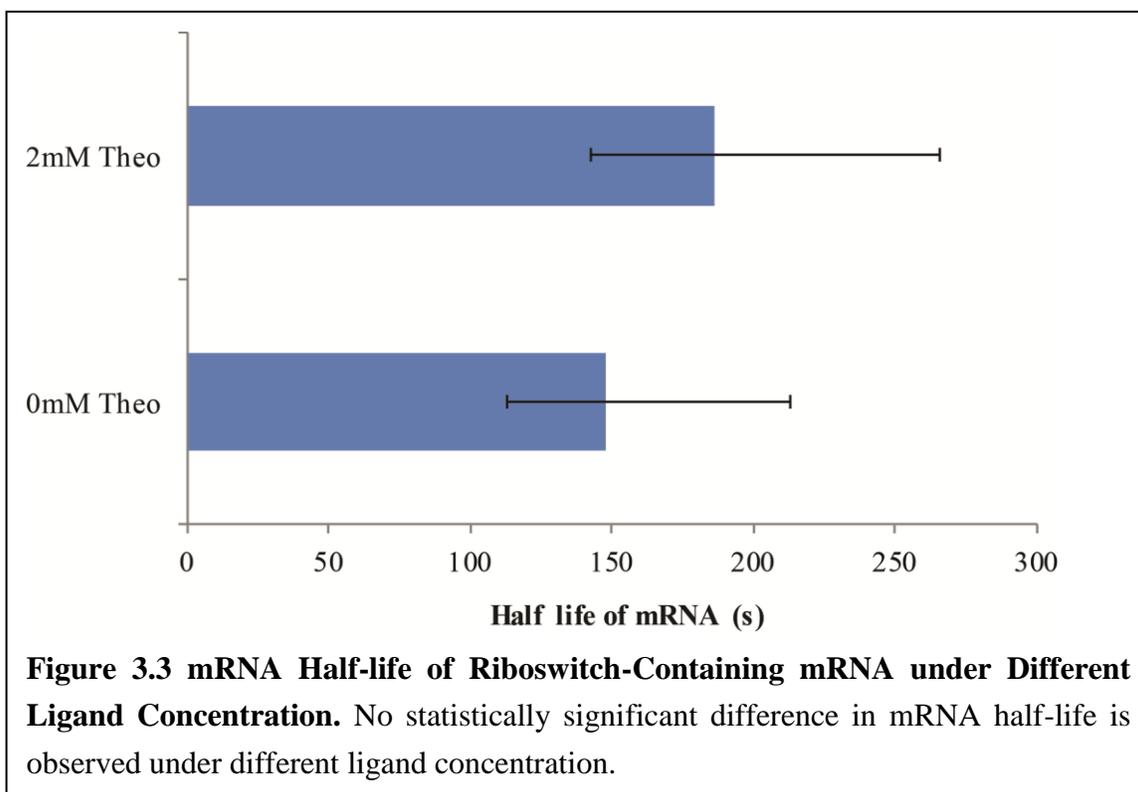
### **3.2.2 Investigation to the influence of ligand binding in RNA stability**

The stability of an mRNA molecule has a great impact on its function as genetic messenger. As one transcript of mRNA can be utilized multiple times to translate the encoded gene, the life span of mRNA can be an important factor to dictate the genetic expression level in cell.

Even though the riboswitch sequence only represents a small portion of the entire mRNA molecule, the ligand-bound complex may prevent mRNA hydrolysis by a ribonuclease, and, consequently, influence the gene expression by changing the decay rate of an mRNA. Also, determining the decay rate of an mRNA under ligand-available conditions would yield the insight that whether mRNA molecules have enough time to reach equilibrium.

One way to study the stability of RNA molecules against ribonuclease digestion is using reverse transcription PCR (RT-PCR). The RT-PCR technique is commonly used in molecular biology to detect RNA expression. In conjunction with quantitative PCR, *in vivo* RNA expression can be quantitatively determined by fluorescent probes. Comparing to other RNA quantification methods, such as a northern blot, the qRT-PCR method is considered the most powerful and sensitive assay for measuring RNA level.

In our research, we designed two probe sites at our target riboswitch-reporter gene



sequence, one at the junction of the riboswitch and reporter gene, and another located in the middle of the luciferase reporter gene. As published in a previous report, the bacterial 16S rRNA expression is taken as our reference due to its extremely stable nature.<sup>10</sup> Each of the probing sites is designed to be 139-140 nt long and flanked with primer sites. Under PCR thermal cycles, the amplification by DNA polymerase will release probes on the primers and register a fluorescent signal. When the fluorescent signal crosses certain threshold, the thermal cycle number (crossing point, Cp) is recorded and can be converted to the absolute concentration of its target RNA in the sample. After normalizing to the reference 16S rRNA concentration, a relative expression level can be achieved for the desired reporter gene.

In our experiment, we used rifampicin to stop the normal transcription in *E. coli* cells and achieve mRNA hydrolysis. At different time points, we measure the relative reporter gene mRNA concentration comparing to our stable 16S rRNA concentration and further fit to an exponential decay model to determine the half-life. Without ligand, the half-life of the mRNA is measured at 148 seconds (95% confidence interval is 113-213 s). Whereas, in the presence of 2 mM ligand, the half-life time of mRNA is measured at 186 seconds (95% confidence interval is 143-265 s). Even though the mRNA half-life difference under ligand-free and ligand-available conditions are not statistical significant, the result gives us the answer about the time frame for riboswitch to function as the genetic regulator.

### **3.2.3 *In vivo* Transcription Translation Decoupling Experiment**

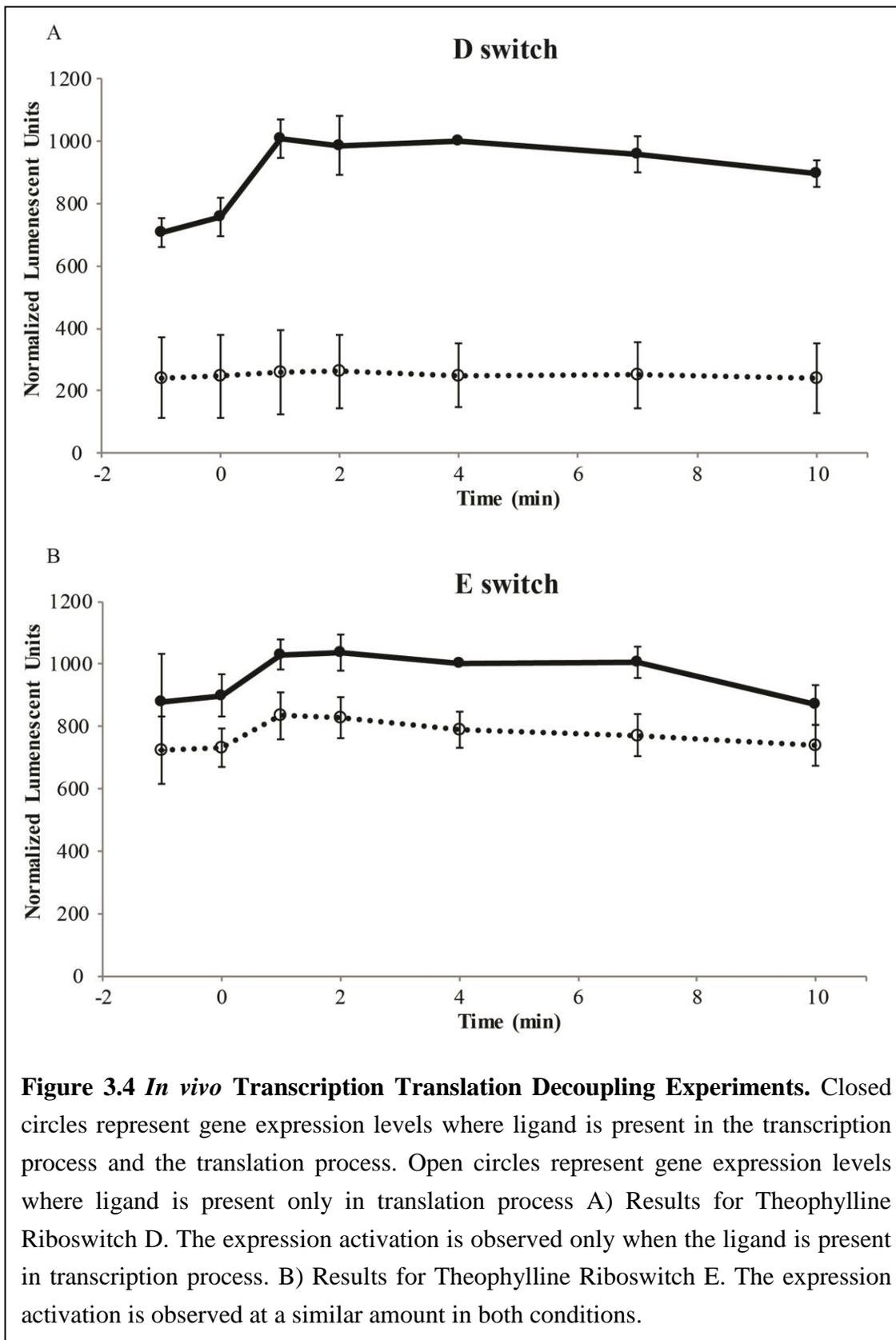
The *in vitro* transcription translation decoupling experiment is elegantly carried out by our co-workers.<sup>9</sup> The report demonstrated the existence of kinetic component in the riboswitch function; however, it would be more convincing if we can carry a similar assay *in vivo*.

The RNA polymerase inhibitor rifampicin provided a way to halt the transcription in cells without disrupting other life process. Thus, cellular activities, such as translation, are still able to happen within a short period of no transcription. Decoupling translation and transcription is relatively easy to achieve under *in vitro* condition by withholding of amino acids from the bacteria extract. In *in vivo* experiment, we attempted to use two

approaches to achieve the desired separation of transcription and translation.

Initially, we adopted an unnatural amino acid as the trigger for translation.<sup>11</sup> We inserted one amber codon in the 5'-end of reporter gene to suppress normal translation. Thus, mRNA transcripts can be synthesized by RNA polymerase but are not capable of synthesizing the reporter protein in the absence of the unnatural amino acid. After halting transcription with rifampicin, a pyrrolysine derivative (*Nε*-(*tert*-Butoxycarbonyl)-L-lysine, also known as Boc-lysine) is added to allow the translation process to produce the desired protein with the help of the tRNA<sup>Lys</sup> - LysRS pair (tRNA and synthetase pair) from *Archaea*.<sup>12</sup> In principle, this method can efficiently separate the transcription and translation processes; however, it suffered from extremely low expression of the target reporter protein after halting transcription. Clearly, this strategy cannot provide enough sensitivity to analyze the proposed kinetic component in riboswitch function.

Next, we tried to create an artificial nutrient-deficient environment to achieve decoupling by suppressing the translation process. After growing cells to mid-log phase, the nutrient-rich cell culture media is removed from the cell by centrifugation. Cells are resuspended in minimal media which lacks amino acids. During the transcription-only stage, cellular machinery is still able to produce riboswitch-containing mRNA transcripts with limited nutrients. After halting transcription with rifampicin, the translation is permitted by adding amino acid supplement. The expression levels of reporter gene are



measured using standard luciferase assay to monitor the gene activation signal.

This strategy, though not optimal, provided enough sensitivity to investigate the kinetic component of the riboswitch function. We applied this method to survey our theophylline riboswitch A-F. The results indicate two distinct types of responses from six different riboswitches. (See Figure 3.4) Represented by our D switch, one type of response suggested the riboswitch-induced expression activation is completely destroyed if we add the ligand in the translation-only stage (Figure 3.4A). Represented by our E switch, another type of response suggested the riboswitch-induced expression activation is still present when adding ligand post-transcriptionally. These results reckoned well with the in vitro data in our earlier publication (Figure 3.4B).

### **3.3 Discussion**

The initial delay time of gene expression activation is often ignored in riboswitch research because the riboswitch function is often measured in a much larger time frame. However, knowing the dynamics of riboswitch function would reveal useful information about riboswitch mechanism and help to optimize riboswitches for downstream applications. In principle, cells need to transport ligand, transcribe RNA, translate protein, and allow protein to be folded to express reporter gene signal. Therefore, for example, if riboswitches function is determined during transcription process, it would take much longer initial delay than those with translationally controlled riboswitches.

Given the transcription speed of 20-80 nt/sec in bacteria<sup>3</sup> and the translation rate of

~20 amino acid/sec in fast growing phase of *E. coli*<sup>13</sup>, a riboswitch-*lacZ* fusion DNA fragment of about 3200-bp would take about 40 to 160 seconds to transcribe, and ~50 seconds to translate. Additional cellular processing time includes ligand diffusion, initiation delays of transcription and translation (a couple of seconds<sup>14</sup>), protein folding (which happens co-translationally), and formation of a bioactive tetramer (two steps of second-order reactions with large rate constants ( $>10^5 \text{ M}^{-1}\text{s}^{-1}$ )<sup>15</sup>). Thus, if the expression activation is based on a thermodynamic equilibrium which includes the ligand diffusion time (short), the translation step (~50 s) and the post-translational processing steps (<10 s), we should expect the reporter gene expression signal appears within 60 seconds. In contrast, if the riboswitch exerts its genetic control power during transcription, the initial delay time would significantly increase by the amount used for transcribing RNA. In our experiment, a delay time about 160 second is detected for our theophylline-dependent riboswitches which is twice as long as the expected translation-only delay (60 s). In another word, the experiment result strongly suggests the gene activation happen during transcription process.

With a larger time span to monitor the riboswitch function, we observed the gene activation effect extended up to 2 hours. Since ligand is added in the early log phase of *E. coli* culture, the assay captured the entire fast growth phase. These results indicate that it may take longer than we previously anticipated for riboswitches reaching maximal gene activation. However, since the reporter protein concentration is determined by a large

number of factors, including the protein recycling rate, it is hard to provide a quantitative model to predict the gene activation level given the time elapsed.

Meanwhile, during our *in vivo* gene activation assay for the extended time span, we noticed a 2-stage activation pattern in our theophylline riboswitch F (also called 8.1). A fast activation period of 20 minutes is followed by a slightly slower activation period extending to 120 minutes after adding ligand molecules. Even though more detailed research is required to explain this phenomenon, we anticipate that the early fast activation period may include a thermodynamic component whereas the later slow activation may only contain the kinetic component of riboswitch activation.

In our qRT-PCR experiment, we determined the half-life of riboswitch-*lacZ* fusion to be 148-180 seconds which is comparable to historic data (141 seconds).<sup>16</sup> Even though we cannot make conclusive argument in our experiment, the additional sequence and secondary structure of riboswitch may provide extra protection from RNase and increase the half-life of the mRNA molecule.

Moreover, we observed a 38-second but not statistically significant increase in half-life of mRNA molecule in the presence of theophylline. We expect that the formation of riboswitch-ligand complex may slow down the degradation process by rendering a more stable structure. The inherent noise of the qRT-PCR experiment did not yield ideal resolution for mRNA half-life, however, it at least implies that the difference in half-lives under different ligand concentrations is not a major contributor to the

riboswitch function. We also speculate that the tighter interactions between ligand and mRNA will lead to more significant changes in mRNA stability thus leading to a greater distribution of mRNA half-lives.

Previous transcription-translation decoupling experiments with *E. coli* S30 extract system have shown that although our theophylline riboswitches regulate the initiation of protein translation, the gene expression activation signal can be determined during transcription. The possibility of forming kinetic traps to promote efficient translation indicates an alternative explanation of riboswitch functions and a change of screening strategy for “kinetic” riboswitches.

The survey of our theophylline riboswitch pack (A-F) for *in vivo* transcription and translation decoupling experiment leads to two distinct groups of riboswitches similar to the results achieved from *in vitro* examination. The first group can be considered as equilibrium driven riboswitches which is represented by our riboswitch E. Similar to the *in vitro* experiment, riboswitch E exhibits the same magnitudes of expression activation under ligand absence or presence during the transcription process. However, another group can be considered as kinetically driven riboswitches, which are represented by our riboswitch C and D. They exhibit no gene expression activation when the ligand is absent during transcription, which is consistent with the *in vitro* results. Meanwhile, riboswitches A, B, and F fall into the middle area where they exhibit modest gene activation when the ligand is absent during the transcription, and much stronger gene

activation when the ligand is present during the transcription.

Comparing those two groups of riboswitches, we found that the kinetically driven group contains riboswitches that have higher gene activation ratios and are considered as better performing riboswitches in our previous screening practices. On the other hand, the equilibrium driven riboswitch E exhibits much wider dynamic range for genetic control. This correlation implies that an efficient synthetic riboswitch may adopt kinetic traps to enhance its efficiency. Thus, in designing new screening methods, especially *in silico* methods, we should not only consider the thermodynamic parameters of different riboswitch conformations but also need to take kinetic traps into our considerations.

It is also noteworthy that different groups of riboswitches correlate their performance in different species of bacteria. In our previous report on riboswitch performance in different bacteria,<sup>17</sup> we observed that our typical “kinetic” riboswitch C, D cannot function in *A. baylyi*, *S. pyogenes*, *M. smegmatis*, and *B. subtilis*, but riboswitch E achieved close gene activation performances comparing to *E. coli*. Even though more detailed characterization is required to investigate this interesting correlation, we believe that different types of riboswitches could yield different performances when the cellular machinery changes.

One of the challenges for performing *in vivo* transcription translation decoupling experiments is that mRNA degradation and protein recycling process are active inside the living cell, where *E. coli* extract can achieve ideal RNase-free conditions *in vitro*. Indeed,

in our experiment, the observed gene expression activation signal is limited due to the restrained living condition of *E. coli* cells. However, we believe that the significant difference between two riboswitch groups can reflect their different mechanisms.

### **3.4 Experimental**

#### ***General Consideration***

All plasmid manipulations were conducted with standard cloning techniques and all constructs were verified by DNA sequencing (MWG Operon, Huntsville, AL). Purifications of plasmid DNA, PCR products, and enzyme digestions were facilitated using kits from Qiagen (Germantown, MD), New England Biolabs (Ipswich, MA). 5-bromo-4-chloro-indolyl- $\beta$ -D-glucuronide X-gal was purchased from US Biological (Swampscott, MA). Theophylline, *o*-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG), ampicillin, and all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Bacto<sup>TM</sup> Agar was purchased from BD (Sparks, MD). Synthetic oligonucleotides primers were purchased from IDT (Coralville, IA). All electrophoresis supplies were purchased from Bio-Rad Laboratories (Hercules, CA). All experiments were performed in *E. coli* TOP10 F<sup>'</sup> strain cells (Invitrogen, Carlsbad, CA) or MG1655 strain (ATCC, Manassas, VA) cultured in media obtained from EMD Bioscience (Merck KGaA, Darmstadt, Germany).

#### ***Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)***

RNA Extraction. TOP10F<sup>'</sup> cells harboring desired plasmids were grown at 37 °C to

reach mid-log phase ( $OD_{600} \approx 0.5$ ) under different ligand concentration. 1 mL of cell culture is harvested and spun down at 14000 rpm using centrifuge for 1 min. The culture media is removed and the cells are washed with 1 mL RNase-free water to avoid mRNA degradation. Then, the cell is resuspended in 330  $\mu$ L RNase-free water and 1 mL TRIZOL reagent (Life Technologies). The mixture is vortexed for 20 s to achieve cellular lysis. The lysate is incubated for 5 min at room temperature to allow complete dissociation of the nucleoprotein complex. Consequently, the RNA molecule is extracted by adding 0.2 mL chloroform and co-precipitated with 2  $\mu$ L glycogen in the aqueous phase. The total cellular RNA in the pellet is washed by 75% ethanol and dried in vacuum in room temperature.

**DNase Digestion.** RNA pellets are dissolved in 50  $\mu$ L water and the solution containing RNA is digested by DNase I in DNase buffer under 37  $^{\circ}$ C for 10 min. The DNase I is then deactivated under 75  $^{\circ}$ C for 10min and dilute to 200  $\mu$ L. Then, the RNA in the digestion reaction dilution is precipitated with Ethanol precipitation and dried under vacuum at room temperature. The resuspended RNA concentration is then measured by UV-Vis spectrometers and diluted for further experiments. (Optional: repeat DNase digestion to remove DNA contaminants and achieve a better qRT-PCR result)

**Reverse Transcription.** An aliquot of resuspended RNA solution is added to 1  $\mu$ L 10 mM dNTP and 1  $\mu$ L random hexamer primer, and heat up to 65  $^{\circ}$ C for 2 min and cool down in ice bath for 1 min. 4  $\mu$ L 5X FS buffer, 2  $\mu$ L 0.1M DTT and 1  $\mu$ L Superscript III

reverse-transcriptase are sequentially added in to the RNA/dNTP/primer mixture and kept at 25 °C for 10 min, then followed by 50 min at 50 °C to allow efficient reverse transcription. The reaction is terminated at 85 °C for 5 min and chill at 0 °C. 1 µL of RNase H is added to each tube and incubate for 20 min at 37 °C for removing bacteria RNA. Subsequently, synthesized cDNA is extracted and purified using Qiagen MiniElute columns and measured on Nanodrop spectrometer.

qRT-PCR reaction. A mixture of 10 µL SYBR green PCR master mix, 0.5 µL primers containing fluorescent labels, 8.5 µL water and 1 µL of diluted cDNA sample is added to a 96-well qPCR reaction plate and sealed with transparent plastic film. The reaction plate is inserted into Roche Lightcycler 480 qPCR machine and programmed for 35 thermal cycles to obtain a fluorescent amplification signal. The data is then analyzed using the manufacturer's software.

### ***In vivo Transcription Translation Decoupling Experiment***

*E. coli* MG1655 cells harboring desired plasmids were grown at 37 °C to reach mid-log phase ( $OD_{600} \approx 0.5$ ). 20 mL cell culture is harvested and spun down at 5000 rpm in centrifuge for 5 min at 4 °C. The culture media is removed and the cells are washed twice with 20 mL 0.9% NaCl solution to remove nutrients. Then, the cell is resuspended in 20 mL minimal media (7.5 mM  $(NH_4)_2SO_4$ , 8.5 mM NaCl, 55 mM  $KH_2PO_4$ , 100 mM  $K_2HPO_4$ , 1 mM  $MgSO_4$ , 1 mg/L  $CaCl_2$ , 20 mM glucose, 10 mg/L thiamine hydrochloride, 10 mg/L biotin in PBS) without amino acids. The minimal media culture is further

incubated for 20 minutes to deplete all amino acids.

Then, the cell culture is split into 4 culture tubes for 4 different conditions. The co-transcriptional trial is conducted with ligand added in the transcription-only phase (10 minutes prior to the halting of transcription by 250  $\mu$ L rifampicin). The post-transcriptional trial is performed with ligand added during the translation-only phase (2 minutes after the halting of transcription). Non-decoupled controls with the presence or absence of ligand are performed with another tubes of minimal media culture (not shown in Figure 3.4). At 0, 1, 2, 4, 6, 10 minutes in the translation-only phase, an aliquot of cell culture is taken and frozen in dry ice. A luciferase assay is subsequently performed as described in Chapter 2.

### 3.5 References

1. Pan, T. and Sosnick, T., *Annu Rev Biophys Biomol Struct*, **2006**. 35: 161-75.
2. Berens, C. and Suess, B., *Curr Opin Biotechnol*, **2014**. 31C: 10-15.
3. Gowrishankar, J. and Harinarayanan, R., *Mol Microbiol*, **2004**. 54: 598-603.
4. Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. and Nudler, E., *Cell*, **2002**. 111: 747-56.
5. McDaniel, B.A., Grundy, F.J., Artsimovitch, I. and Henkin, T.M., *Proc Natl Acad Sci U S A*, **2003**. 100: 3083-8.
6. Henkin, T.M. and Grundy, F.J., *Cold Spring Harb Symp Quant Biol*, **2006**. 71: 231-7.
7. Wickiser, J.K., Winkler, W.C., Breaker, R.R. and Crothers, D.M., *Mol Cell*, **2005**. 18: 49-60.
8. Wickiser, J.K., Cheah, M.T., Breaker, R.R. and Crothers, D.M., *Biochemistry*, **2005**. 44: 13404-14.
9. Mishler, D.M. and Gallivan, J.P., *Nucleic Acids Res*, **2014**. 42: 6753-61.
10. Cole, S.E. and LaRiviere, F.J., *Methods Enzymol*, **2008**. 449: 239-59.
11. Srinivasan, G., James, C.M. and Krzycki, J.A., *Science*, **2002**. 296: 1459-62.
12. Ibba, M., Losey, H.C., Kawarabayasi, Y., Kikuchi, H., Bunjun, S. and Soll, D., *Proc Natl Acad Sci U S A*, **1999**. 96: 418-23.
13. Young, R. and Bremer, H., *Biochem J*, **1976**. 160: 185-94.

14. Rajala, T., Hakkinen, A., Healy, S., Yli-Harja, O. and Ribeiro, A.S., *PLoS Comput Biol*, **2010**. *6*: e1000704.
15. Matsuura, T., Hosoda, K., Ichihashi, N., Kazuta, Y. and Yomo, T., *J Biol Chem*, **2011**. *286*: 22028-34.
16. Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M. and Rosenow, C., *Genome Res*, **2003**. *13*: 216-23.
17. Topp, S., Reynoso, C.M., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R., Scott, J.R. and Gallivan, J.P., *Appl Environ Microbiol*, **2010**. *76*: 7881-4.

## Chapter 4. Computer-assisted Screening for Tetramethylrosamine Riboswitches

### 4.1 Introduction

The creation of synthetic riboswitches made programming cells to perform desired functions more feasible. These novel genetic regulators are featured with their plug-and-play characteristics under versatile control schemes. One of the successful examples is our synthetic theophylline riboswitch family.<sup>1</sup> However, as engineers from the downstream area, such as synthetic biology, begin to create more realistic and complex systems, expanding the diversity of riboswitches becomes a large challenge.<sup>2,3,4</sup>

Within the last years, several families of engineered riboswitches have been reported including tetracycline riboswitches, 2, 4-dinitrotoluene riboswitches.<sup>5,6,7,8</sup> As described in the first chapter, these novel synthetic riboswitches respond to a wide variety of ligands and cover a large spectrum of regulatory ranges. Due to commonalities in the aptamer domains, most synthetic riboswitches shared a similar two-stage screening discovery process. An *in vitro* selection process for the molecule-sensing aptamer domain is often followed by an *in vivo* screening process for the regulating domain which induces the alternative conformation under different ligand availability.

In most cases, the first *in vitro* selection process is achieved by the SELEX (systematic evolution of ligands by exponential enrichment) procedure, but the second screening procedure often has multiple choices as various efforts have been made to

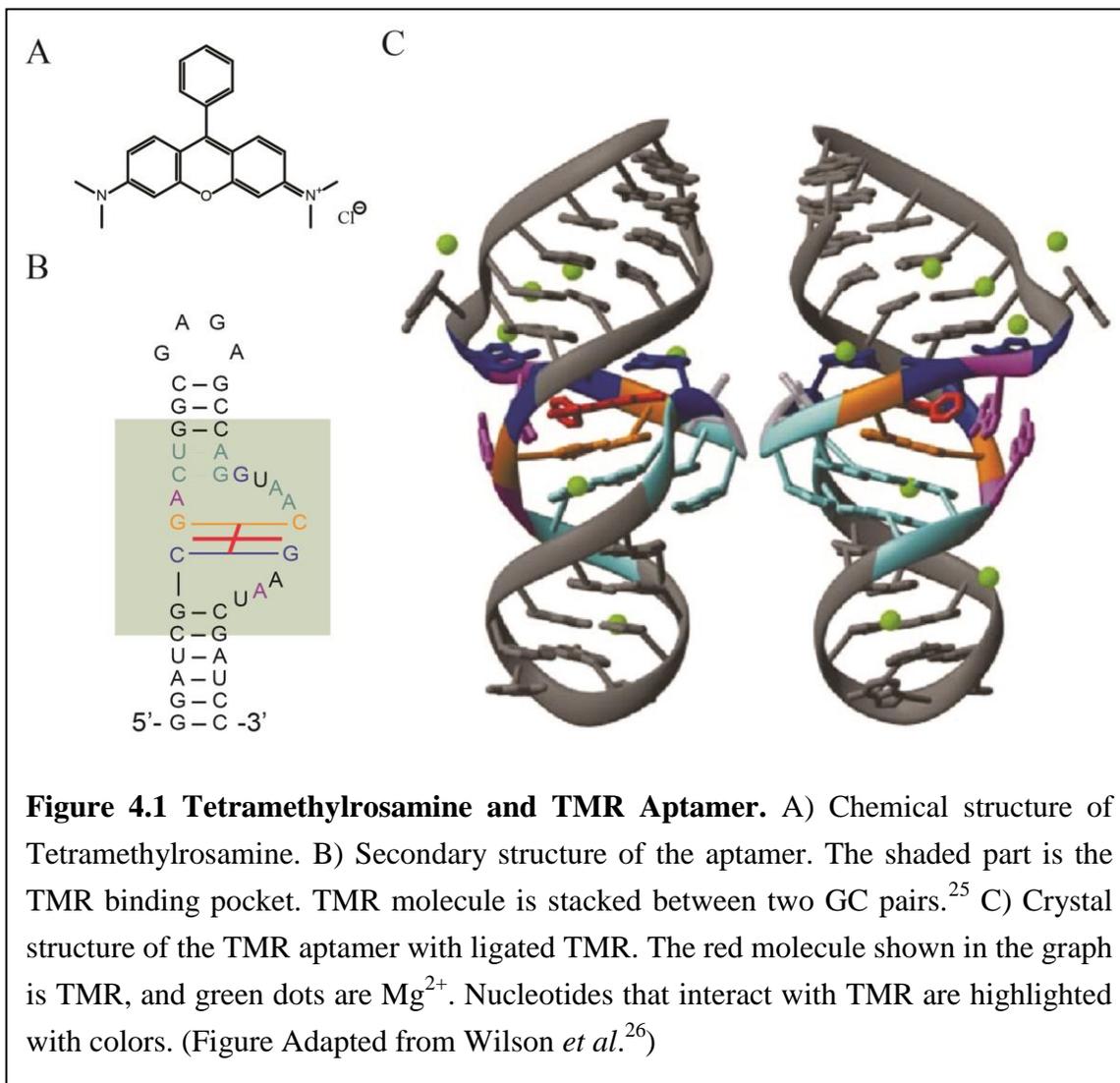
create novel and efficient screening methods. Our co-worker utilized a *lacZ* reporter screening procedure to discover our first theophylline riboswitch.<sup>1</sup> Later on, a robot-assisted high throughput screening method, flow-cytometry procedure, and cell motility screening were used to discover riboswitches.<sup>9,10,11</sup> Other groups reported the dual screening method and rational design method for the discovery of other synthetic riboswitches.<sup>12,13,14,15</sup> Even though these screening processes are quite successful in identifying new functional riboswitches, the laborious repetition during the screening process clearly lacks of efficiency.

One of the modern methods to improve screening efficiency and accuracy is to use computer-assisted process. Due to the rapidly growing speed of computers for the last several decades, a number of complicated biological screening tasks were made possible. For example, Barrick and Breaker used computational sequence searches to predict the distribution of ten widespread natural riboswitches in the genomes of organisms from all three domains of life.<sup>16</sup> With the bioinformatic prediction, the same group later conducted experiments to verify a putative preQ1 riboswitch in a previously unknown non-coding region.<sup>17</sup> Computational methods can also help to predict the correlation between nucleic acid sequence and structure. A web-based server QGPS can predict G-quadruplex structures within DNA sequences.<sup>18</sup> Another computational study revealed the correlation between RNA primary structure and its function by predicting functional introns that can efficiently disrupt bacterial genes.<sup>19</sup>

However, using computers to screen for riboswitches raises a large challenge for computational biologists because of the limited knowledge about the correlation between RNA structure and riboswitch function. As an emerging RNA regulatory element, riboswitch function is determined not only by the primary RNA sequence but also the secondary and tertiary structures of the RNA molecules. Besides, different families of riboswitches possess unique correlations among RNA sequence, RNA structure, and riboswitch function. Thus it is hard to provide a unifying biophysical model with the same set of parameters for predicting novel riboswitches.

Despite the case-by-case nature of riboswitch prediction, previous studies have provided enough empirical data to predict secondary structures of RNA molecules. Computational tools, such as mFold<sup>20</sup> and the Vienna RNA web server<sup>21</sup>, granted experimental scientists an accurate method to predict RNA structures without conducting laborious experiments.

A step further from predicting RNA secondary structures is to predict the function of putative riboswitches based on their sequences. Our riboswitch research results promoted the creation of the first biophysical model of our theophylline riboswitch family and provided the input parameters of a computational model which successfully predicts a number of functional theophylline riboswitch variants. Given the success in extending our existing theophylline riboswitch family, our next question is: can we translate our previous knowledge about theophylline riboswitch to efficiently create other types of



riboswitches using our computer-assisted method.

One good candidate ligand is tetramethylrosamine (TMR). TMR is one of a large family of commercial triphenylmethane-based dyes that is primarily used as a fluorophore in biological assays and fluorescence microscopy (See Figure 4.1A). It can serve as a fluorescent marker for mitochondria with an excitation peak at 550 nm and the emission peak at 574 nm.<sup>22</sup> One of unique properties for TMR and its analogs is that they can effectively photosensitize chemosensitive and multidrug-resistant cell lines.<sup>23,24</sup>

In 1999, Grate and Wilson reported an RNA aptamer for malachite green (MG) which is also a triphenylmethane-based dye.<sup>25</sup> (See Figure 4.1B) Absorbance wavelength shifting assay suggested a ligand binding affinity  $K_D$  less than 1  $\mu\text{M}$  for the RNA binding motif and mutation studies revealed the minimal scaffold of the binding aptamer.

Later on, the crystal structure and the NMR solution structure of this MG aptamer was reported by Wilson group.<sup>26,27</sup> (See Figure 4.1C) The authors found the MG-binding pocket is defined by an asymmetric internal loop, flanked by a pair of helices. They also tested the aptamer binding affinity for other available triphenylmethane dyes and identified that a tighter binding between the aptamer motif and the analogous TMR with a  $K_D$  of  $\sim 20$  nM. The comparison between the crystal structure and the NMR solution structure revealed that the same RNA molecule can adaptively bind to both planar (TMR) and nonplanar (MG) ligands.

Another research team conducted a thorough investigation of the MG aptamer with isothermal titration calorimetry, equilibrium dialysis, and fluorescence titration.<sup>28</sup> Their results revealed that the entropy of complex formation plays an important part in determining binding affinity and ligand specificity. In conjunction with the structural information of the MG aptamer by Wilson group, the authors showed that metal ions, such as  $\text{Mg}^{2+}$ , are required to stabilize the RNA complexes with MG-derivatives, such as TMR and crystal violet, whereas the RNA complex with the original MG substrate is stable at low salt concentration and in the absence of divalent metal ions.

Given distinct chemical properties of these triphenylmethane dyes, the MG/TMR aptamer has been adopted as a model to study RNA-related biological processes. In the original paper which reported the MG aptamer, the authors created an mRNA fusion where an encoding mRNA sequence is capped with the MG aptamer at the 5'-end. After forming the aptamer-ligand complex, a laser is used to excite MG molecule and induce the degradation of the mRNA *in vitro*.<sup>25</sup> This laser-mediated degradation of mRNA molecules yielded possibilities of several potential *in vitro* and *in vivo* applications.

Also, research on *in vitro* nanoparticle-based RNA folding<sup>29</sup> and thermodynamics of ligand-aptamer binding<sup>30</sup> have been subsequently carried out with the MG/TMR aptamer system. These mechanistic insights inspired a series of downstream applications including mercury detection using the resonance scattering spectral assay<sup>31</sup> and triphenylmethane dyes detection in fish tissues<sup>32</sup>.

Meanwhile, cellular studies have paved the way for the usage of TMR in *in vivo* applications.<sup>33</sup> It has been demonstrated that TMR and its analogues can be transported by P-glycoprotein into chemo-resistant CR1R12 cells, and the cardiovascular drug Verapamil can enhanced the uptake of TMR into the cells by roughly 3-fold.<sup>33</sup> Even though thio- and seleno- analogues of TMR exhibit high phototoxicity when cells are exposed to light, TMR itself does not effectively photosensitize and kill cells.

With above illustrated characteristics, TMR becomes an excellent choice of riboswitch ligand. In fact, the Wilson group constructed a TMR aptamer insertion into

5'-untranslated region of mRNA and reported that the synthetic mRNA construct can regulate *S. cerevisiae* cell cycle.<sup>34</sup> However, the mechanism of such a primitive TMR riboswitch is not illustrated due to the limited understanding about riboswitches at the time of publication.

Thus, in this chapter, we design and characterize novel TMR riboswitches with our computational tools. Firstly, we used a computational method to predict a set of putative TMR riboswitches and performed *in vitro* and *in vivo* reporter gene expression assays to characterize the gene activation function for our predicted TMR riboswitches. Then, we adopted *in vitro* transcription translation decoupling experiments to provide insights to the mechanism of our newly synthesized TMR riboswitches. Moreover, we designed an experiment that investigates whether the concentration of DNA template could influence the riboswitch function both *in vitro* and *in vivo* by assaying riboswitches with different DNA concentrations in S30 cell extract reactions or in plasmids with different cellular copy numbers in cell culture.

## **4.2 Results**

### **4.2.1 *In vitro* Characterization of Novel Tetramethylrosamine Riboswitches Predicted by Computational Methods**

Based on the parameters from our theophylline riboswitch biophysical model, we formulated a new TMR riboswitch prediction model. We continued to use parameters for RNA folding ( $\Delta G_{\text{mRNA}}$ ), ribosome-charging processes ( $\Delta G_{\text{hybrid}}$  and  $\Delta G_{\text{standby}}$ ), the binding

of 30S ribosome subunit and tRNA ( $\Delta G_{\text{spacing}}$  and  $\Delta G_{\text{start}}$ ), but adjusted the ligand-binding free energy based on the difference in binding affinities of two different aptamers ( $\Delta G_{\text{binding}}$ ).

We designed a pre-aptamer region for about 11-12 nt before aptamer sequence and a post-aptamer region for about 20-30 nt after the aptamer. The post-aptamer region can be further divided into functional regions of linker sequence, Shine-Dalgarno sequence, and spacing sequence. These above functional regions around the aptamer are filled with random ribonucleotides and are calculated with given parameters to obtain their biophysical properties. Ten putative TMR riboswitches are screened out by our collaborator Salis group. (See Table 4.1)

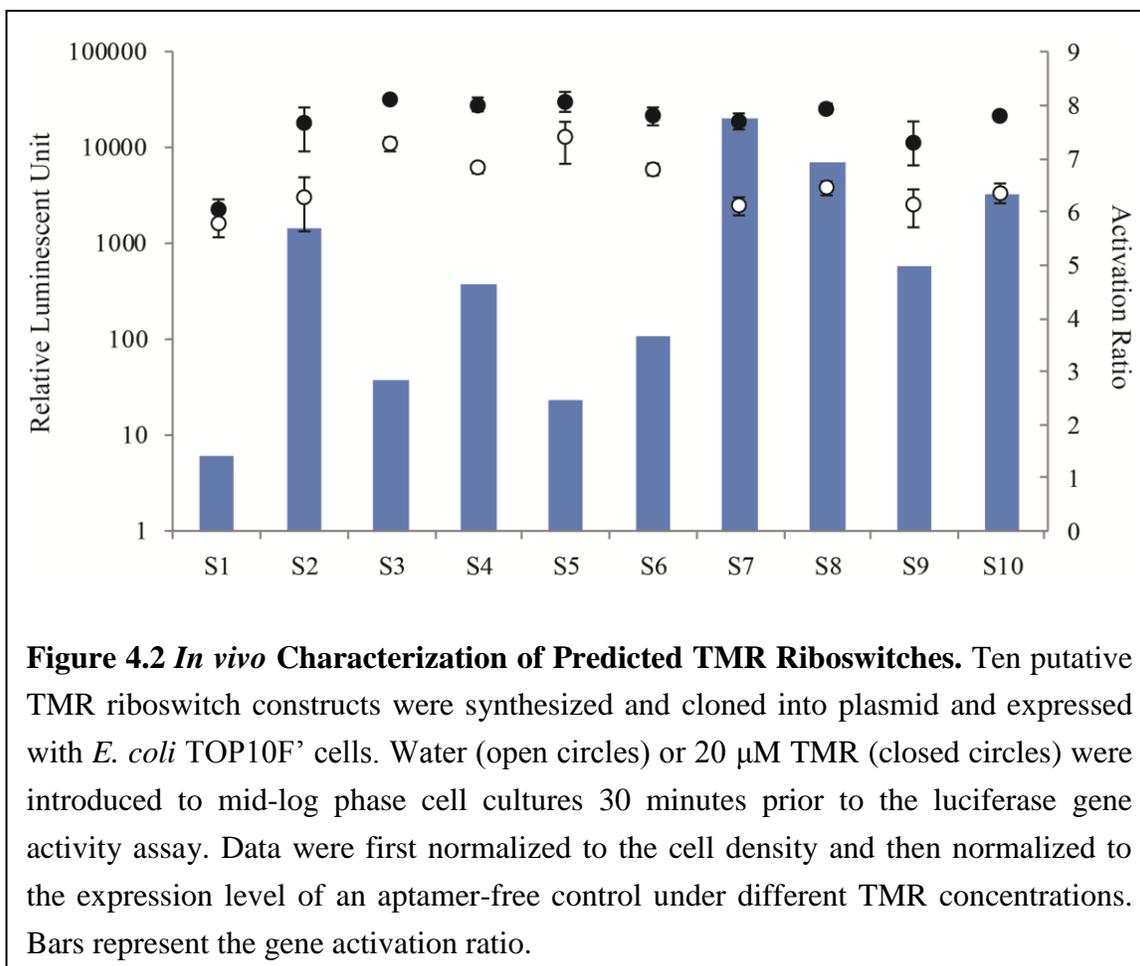
These ten distinct sequences are then synthesized *in vitro* and cloned into a plasmid to control the downstream luciferase reporter gene in *E.coli* TOP 10F'. The plasmids were amplified *in vivo*, then assayed in cell extract. *In vitro* results are normalized to an aptamer-deletion construct and summarized in Table 4.1, where 9 out of 10 exhibit significant gene activation function and 7 out of these 9 riboswitches achieved a 5-fold or higher performance. Among these TMR riboswitches, S2 is the most efficient construct, which can induce 17-fold gene activation under 30  $\mu\text{M}$  TMR.

#### **4.2.2 *In vivo* Characterization of Novel Tetramethylrosamine Riboswitches Predicted by Computational Methods**

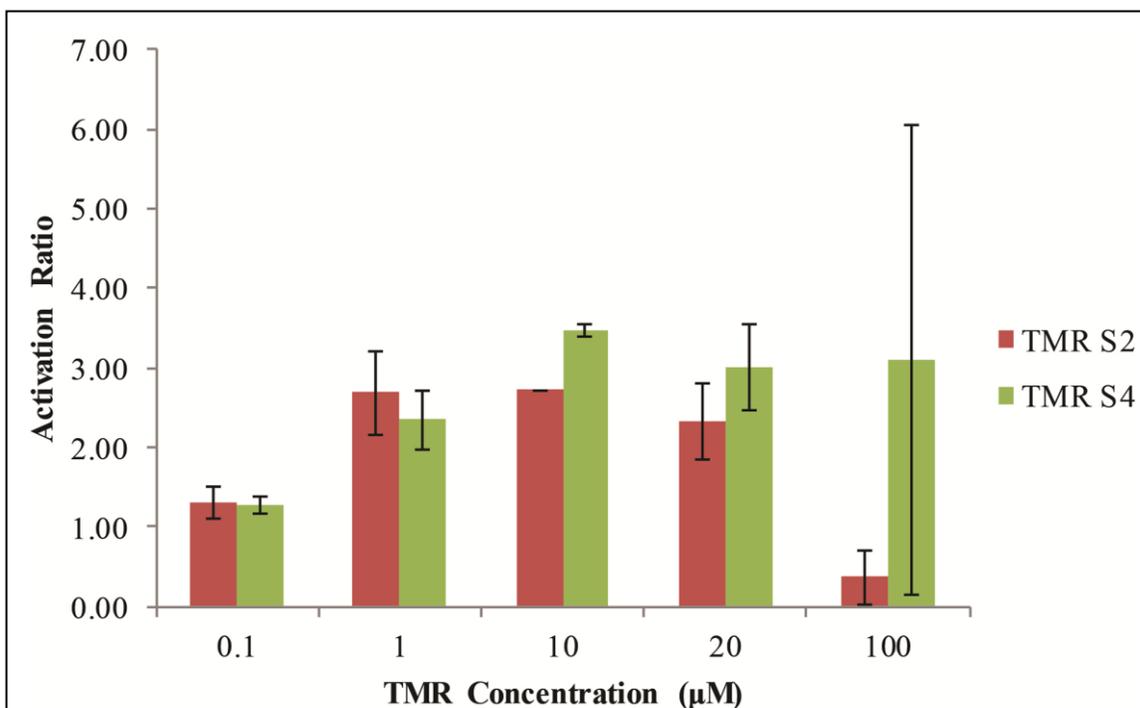
The same ten putative TMR riboswitches are transformed into *E.coli* TOP 10F'. Cells are grown into mid-log phase and are induced by 20  $\mu\text{M}$  TMR for 30 minutes. Then

| Clone | Pre-sequence  | Linker            | Spacing | RLU (0μM) | RLU (30μM) | Activation Ratio |
|-------|---------------|-------------------|---------|-----------|------------|------------------|
| S1    | GCGGATTCTAC   | GCATCGATTTCGT     | TAAGTCC | 389       | 269        | 0.7              |
| S2    | GATTCCCCGTACC | AAGATTTCGT        | CCAT    | 5571      | 96740      | 17.4             |
| S3    | GCAGGCTCCCAC  | AGTTCCACCGATTTCGT | AATCA   | 9822      | 19993      | 2.0              |
| S4    | GTCACCATACCA  | ACAGATTTCGT       | CCCACT  | 3089      | 21826      | 7.1              |
| S5    | GCAGGCTCCCAC  | AGTACCCACGATTTCGT | AATCA   | 3639      | 5484       | 1.5              |
| S6    | GAAACCATACCA  | AAAGATTTCGT       | CCGAGAT | 2105      | 15405      | 7.3              |
| S7    | GCAGGCTCCCAC  | AGTTCCACCGATTTCGT | AATCC   | 1393      | 9761       | 7.0              |
| S8    | GCAGGCTCCCAC  | AGTACCCACGATTTCGT | AATCC   | 1482      | 8079       | 5.5              |
| S9    | GCAGGCTCCCAC  | AGTTCCACCGATTTCGT | GATCA   | 1463      | 5929       | 4.1              |
| S10   | GCAGGCTCCCAC  | AGTACCCACGATTTCGT | GATCA   | 807       | 4903       | 6.1              |

**Table 4.1 Synthetic TMR Riboswitch Characteristics.** Ten TMR riboswitches were predicted by computational model. The sequence is arranged by the order of pre-sequence, aptamer, linker, SD sequence, spacing, then starting codon for luciferase reporter gene. In all constructs, the aptamer sequence and the SD sequence are constant. (Aptamer = “GGATCGGACTGGCGAGAGAGCCAGGTAACGAATCGATCC” and SD sequence = “TAAGGAGGT”). The Relative Luminescent Unit (RLU) from *in vitro* expression reactions in the presence and absence of TMR, as well as the resulting activation ratios are also shown. (All entries are measured with *E. coli* S30 cell extract)



standard luciferase assay protocol is used to determine the gene activation effects of these putative TMR riboswitches. The results are summarized in Figure 4.2 where 10 out of 10 putative riboswitch constructs exhibit different degrees of gene activation *in vivo*. Among these 10 constructs, S2, S7, and S8 were the best performers and able to achieve 5.7-, 7.7-, and 6.9-fold of gene activation respectively. These TMR riboswitch constructs worked at the same magnitude of gene expression as the control construct, but exhibited basal luciferase reporter activities at ~20% to ~120% of the aptamer-free control construct, which are much higher than our theophylline riboswitches.

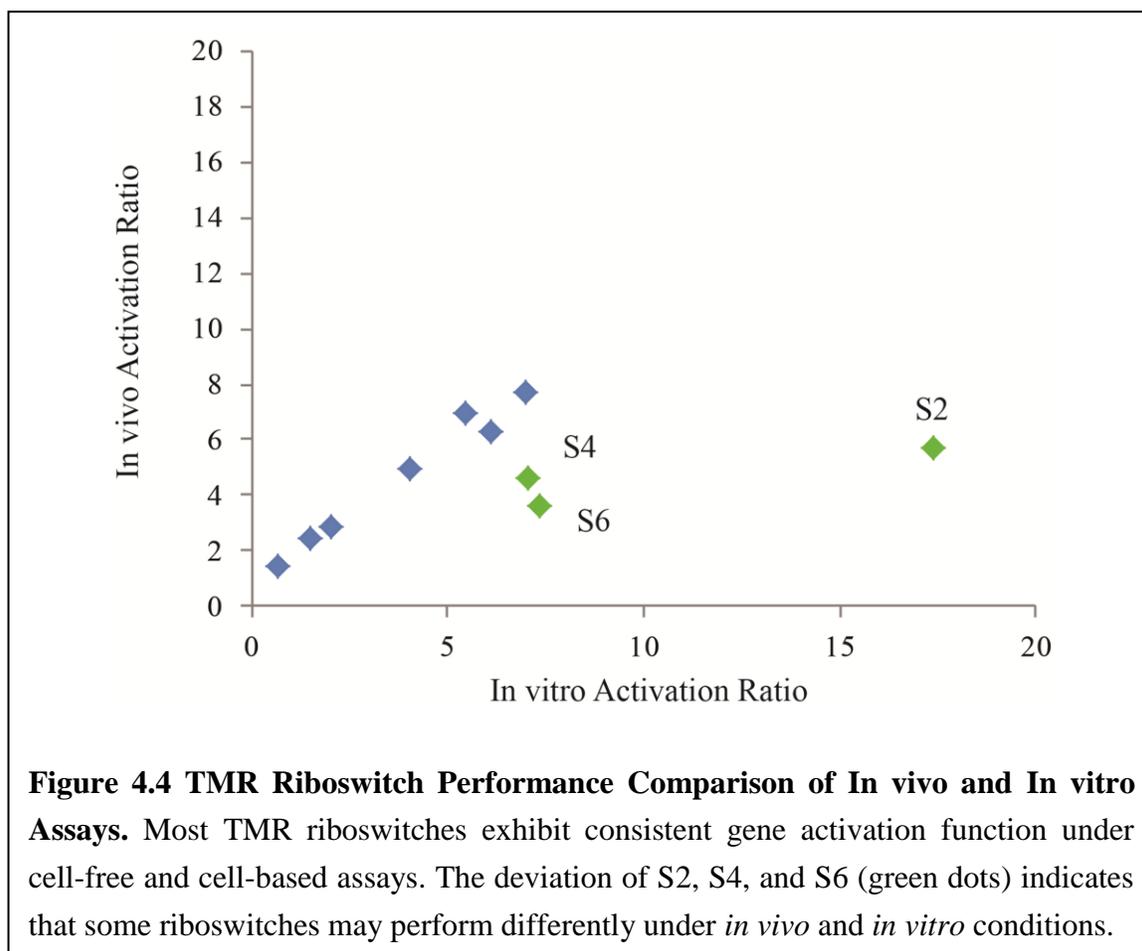


**Figure 4.3 Ligand Dose-response Gene Activation of TMR riboswitches.** The activation ratio is calculated by dividing the adjusted expression with the presence of TMR by the adjusted expression with the absence of TMR. Results are consistent when the TMR concentration is below 20  $\mu\text{M}$ . The maximal TMR gene activation concentration lies between 1 and 20  $\mu\text{M}$ . Red bars represent results of S2 construct and green bars represent results of S4 construct. Experiments are repeated three times and the large error bars at 100  $\mu\text{M}$  are due to significant cytotoxicity of TMR.

Meanwhile, we conducted an *in vivo* ligand dose-response assay on two of our TMR riboswitches (S2 and S4) (See Figure 4.3). In our preliminary research, a ~50% reduction in reporter expression was observed with a 20  $\mu\text{M}$  TMR for 30 minutes activation and reduction aggravated when the TMR concentration or the activation time increased. Thus, in our ligand dose-response assay, TMR concentrations were set to be 100 nM, 1  $\mu\text{M}$ , 10  $\mu\text{M}$ , 20  $\mu\text{M}$ , 100  $\mu\text{M}$  and activation time was set to be 20 minutes. We observed weak but noticeable (1.2-fold) gene activation at the very low TMR concentration of 100 nM in a

relatively short period (20 minutes). The reporter activity reaches maximum activation at 1-20  $\mu\text{M}$  TMR after normalizing to the control expression. When the TMR concentration reaches 100  $\mu\text{M}$ , significant cytotoxicity is noticed and the *E. coli* population declines, and, consequently, leads to the failure of activating gene expression. Low concentrations of TMR thus provide a more efficient riboswitch system for downstream engineering applications.

Interestingly, even though most TMR riboswitches exhibited similar gene activation functions *in vitro* and *in vivo*, several of them (S2, S4, and S6) failed to perform consistently. (See Figure 4.4) The discrepancies put a challenge for the downstream



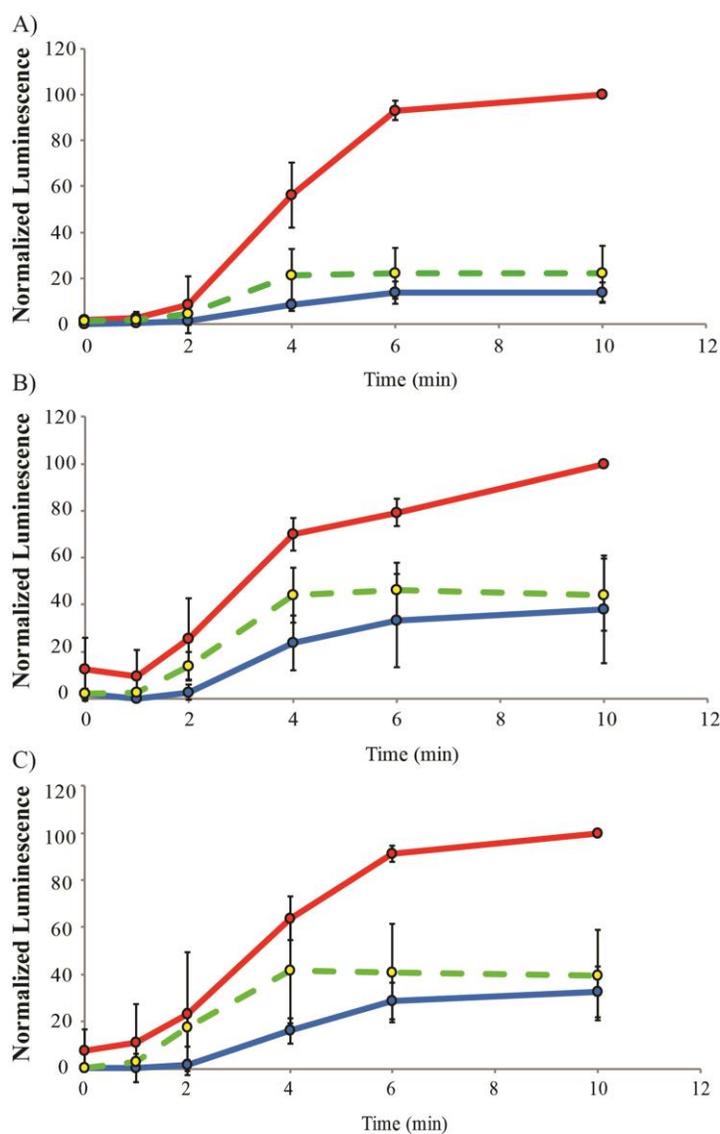
application that some of our *in vitro* selected and characterized riboswitches may not function similarly *in vivo*. Detailed analysis for this phenomenon will be presented in the discussion section.

#### **4.2.3 *In vitro* Transcription-Translation Decoupling Experiment of Novel Tetramethylrosamine Riboswitches**

Based on our previous knowledge of riboswitch mechanism, riboswitches can be classified as equilibrium-driven or kinetically driven riboswitches. One way to investigate different types of riboswitches is *in vitro* transcription translation decoupling experiment design by our co-worker.<sup>35</sup>

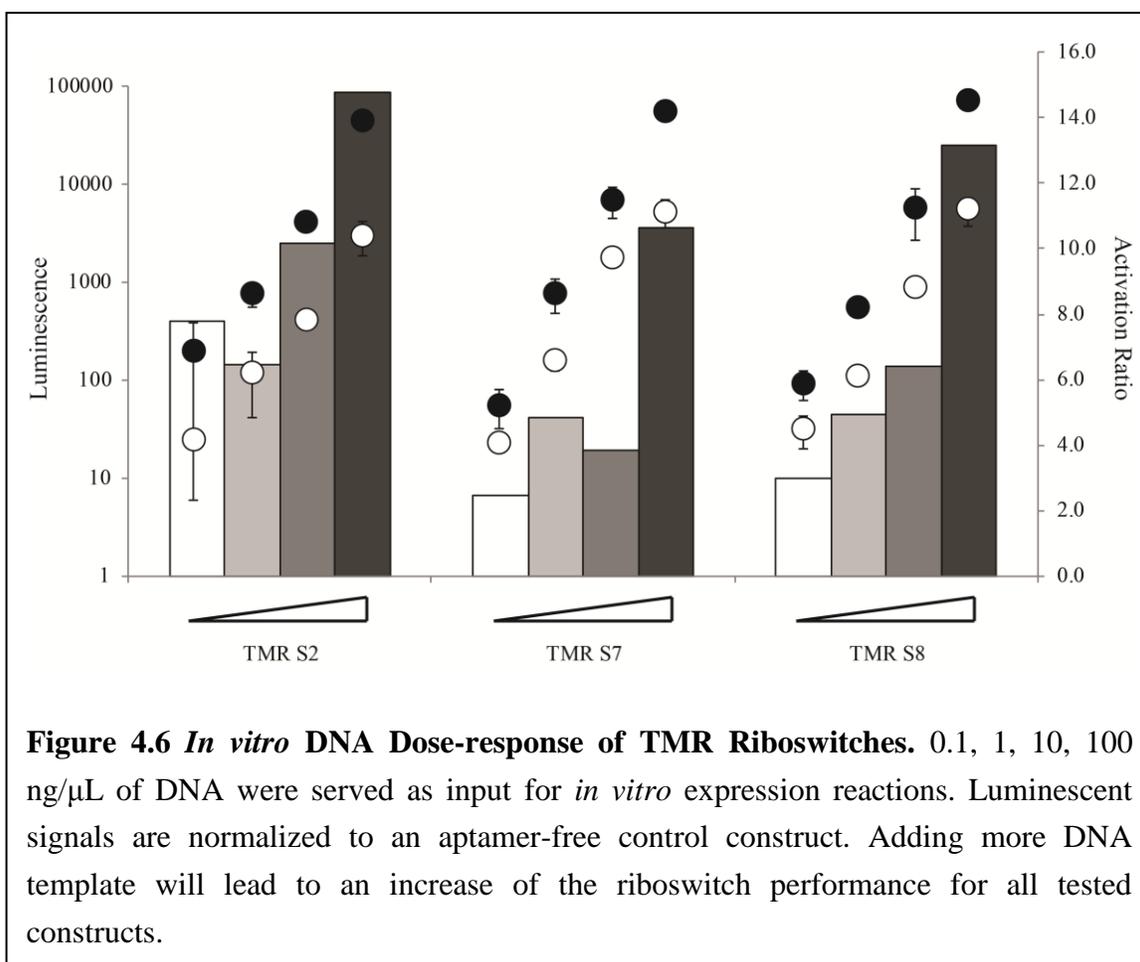
In a modified version of the original decoupling experiment, we used inherent cellular machinery to produce RNA template instead of supplying PAGE-purified RNA template. The extracted riboswitch-containing plasmid is used as DNA template to initiate the transcription in the S30 cell extract where translation is inhibited by the lack of amino acids. Then, the transcription is halted by addition of 250  $\mu\text{g/ml}$  rifampicin and, subsequently, the translation is initiated by adding an amino acid solution. The luciferase activity is assayed for different conditions where ligand is present or absent in the transcriptional stage.

We selected the best performing TMR riboswitches S2, S7, S8 to be assayed in our *in vitro* transcription translation decoupling experiment. We found that our best *in vitro* TMR riboswitch S2 is kinetically driven, as its gene activation function is only observed when ligand is present in transcriptional stage. The other two of the mechanism tend to



**Figure 4.5** *In vitro* Transcription Translation Decoupling Experiment of TMR Riboswitches. A) The result of S2 construct. B) The result of S7 construct. C) The result of S8 construct. Red dots correspond to the co-transcriptional trials where TMR is present during both of the transcription and translation stage. Blue dots correspond to the post-transcriptional trials where TMR is only present during the translation stage. Yellow dots correspond to the negative control where no TMR is added.

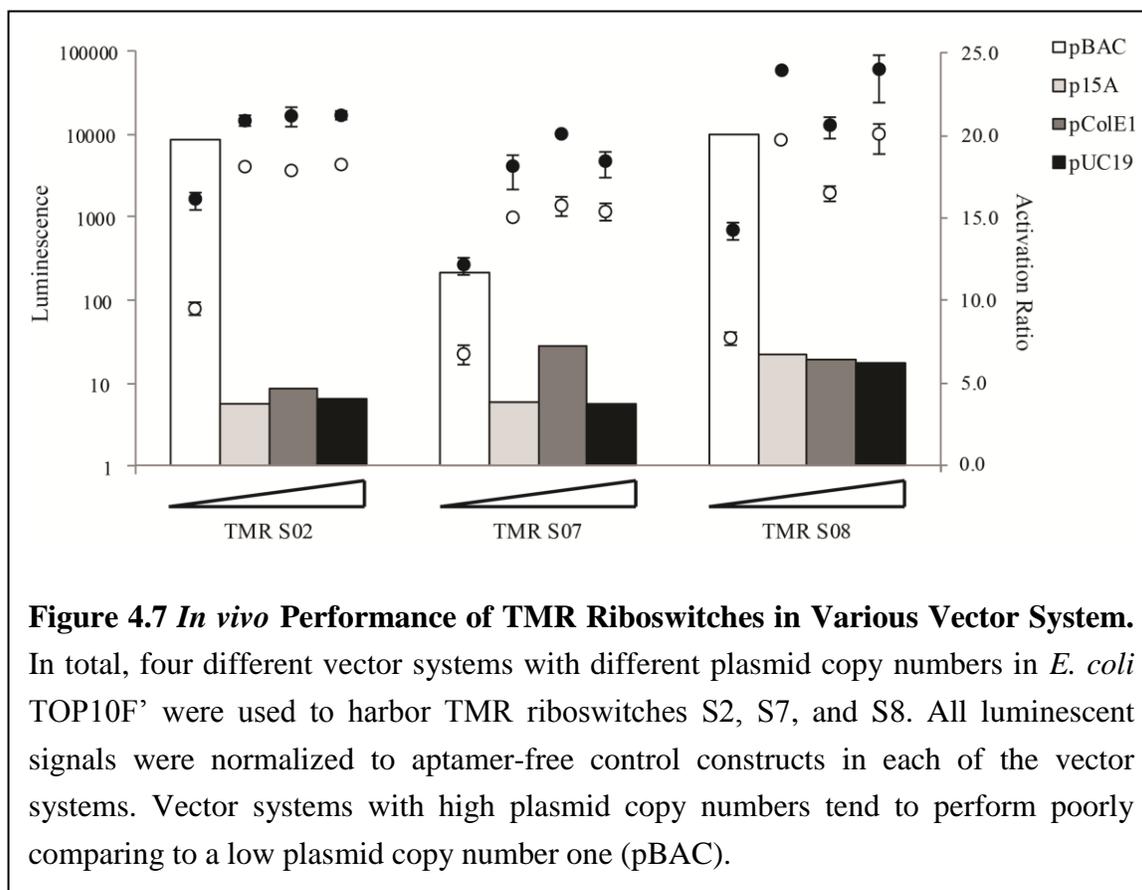
exhibit both equilibrium and kinetic components, but the equilibrium component takes a minor role in the gene activation. (See Figure 4.5, blue curves account for ~30% of gene



activation in S7 and S8)

#### 4.2.4 *In vitro* and *In vivo* Characterization of Novel Tetramethylrosamine Riboswitches with Different DNA Template Concentration

We speculated that the difference in DNA template could drastically change the concentration of mRNA thus leading to a different proportion of equilibrium component and kinetic component of gene activation. A smaller mRNA-to-ligand ratio would increase the possibility of ligand-binding during the transcription stage thus rendering a higher kinetic component of gene activation. Meanwhile, the knowledge of riboswitch function under different DNA concentrations would help engineers to design their gene



switching devices accordingly.

In our *in vitro* characterization, we simply varied the input amount of DNA template. The *in vitro* luciferase assay results are summarized in Figure 4.6. The results indicated that the riboswitch gene activation functions can be enhanced by the increasing amount of DNA template in the *in vitro* expression reactions.

In our *in vivo* characterization, we inserted our TMR riboswitches into four vector systems with different replication origins- pBAC, p15A, pColE1, and pUC19, and applied *in vivo* luciferase assay to measure the gene activation function. The *in vivo* luciferase assay results are summarized in Figure 4.7. Among all the vector systems,

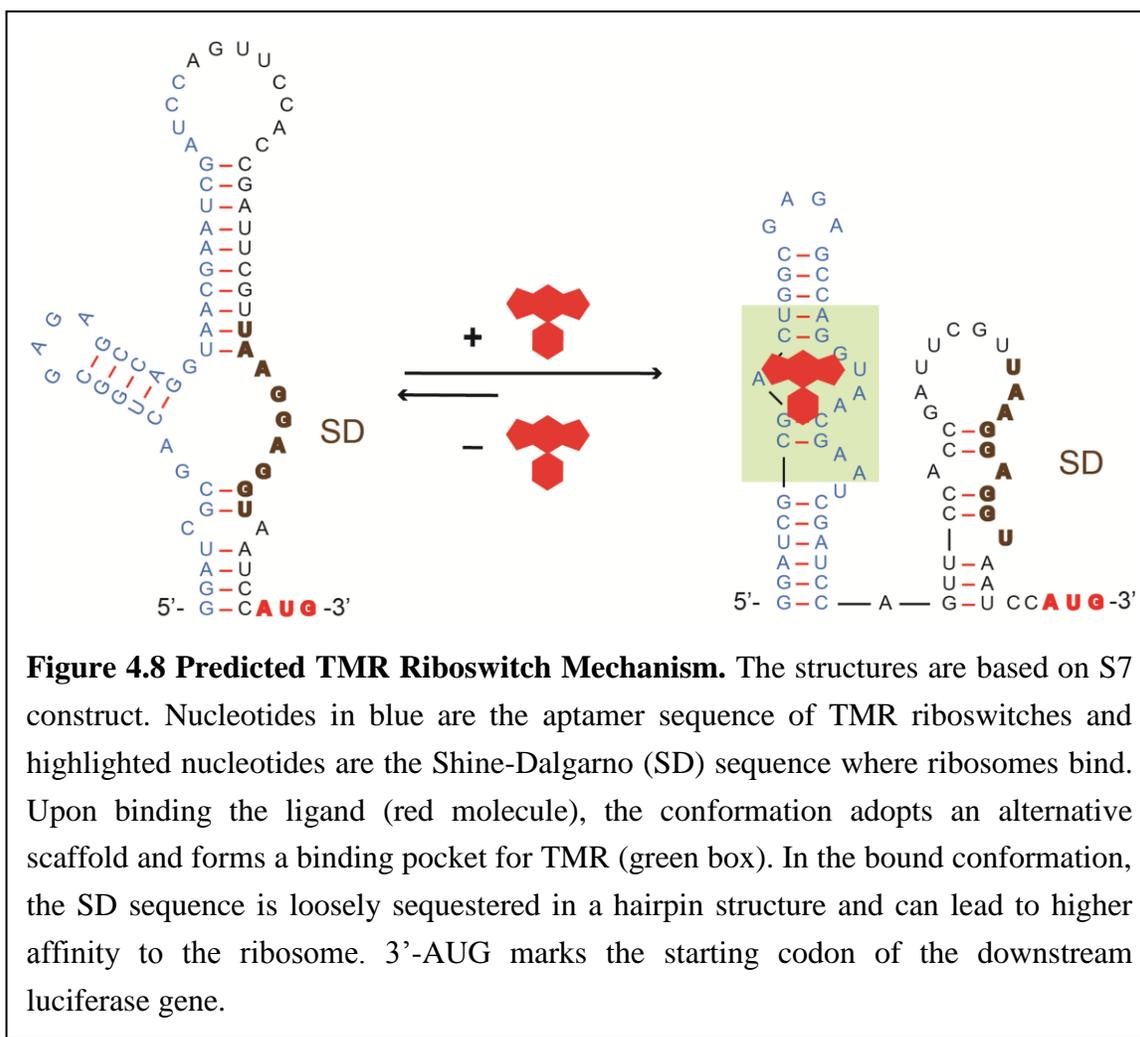
pBAC vector system achieved the highest gene activation effect. As the cellular copy numbers of plasmid increase (pBAC < p15A < pColE1 < pUC19), surprisingly, we observed a decline in gene activation effect *in vivo* rather than an increase as we encountered in our *in vitro* assays.

### 4.3 Discussion

The success in identifying functional TMR riboswitches with our computational model provides another piece of evidence to support the validity of the biophysical model we constructed for our theophylline riboswitches. The efficient method precludes the necessity of laborious bench experiments for the screening of expression platforms. Besides, the success rate for our computation screening process is impressively high (9 out of 10 predicted TMR riboswitches can induce genetic activation both *in vivo* and *in vitro*). Nearly all putative constructs are characterized with gene activation functionalities and more than half of them can achieve higher than five-fold gene activation regulation.

On a broader view, the freedom gained from the riboswitch biophysical model facilitated efficient rational design for riboswitches. Most *in vitro* selected RNA aptamers have approximated binding affinities and lack experimentally characterized structures. Our computational method would help to bridge the gaps between the aptamer creation and the downstream riboswitch applications. As the experimental data accumulate, we believe our biophysical model will enjoy the abundant data and make more accurate predictions in the future.

Using our computational prediction tools, we predicted the secondary folding structure for both ON and OFF states of our TMR riboswitches. The structures shown in Figure 4.8 illustrate the conformational changes upon adding the ligand TMR. Compared to our theophylline riboswitches on which our biophysical predictive model is built, TMR riboswitches have much more complicated secondary structures and form a branched hairpin structures. In the OFF conformation, 17 base pairs are formed to stabilize the branched hairpin structure, whereas the aptamer forms two individual hairpins upon



binding the ligand. In the ON conformation, the binding pocket hairpin is stabilized by 13 base pairs, but the SD sequence is only stabilized by 7 base pairs, which makes it more accessible for ribosomes to bind.

Comparing to the sharp difference in ribosome accessibility between two conformations of theophylline riboswitches, the SD sequence of TMR riboswitches does not pose a sharp contrast. In both conformations shown in Figure 4.8, the SD sequence is sequestered in hairpin structures. Even though the binding affinity of TMR and its aptamer is quite high, the ligand binding free energy cannot be used to open up the adjacent hairpin structure. Therefore, these *in silico* selected constructs may not exhibit potent gene activation effect as observed in our theophylline riboswitches. This postulation of secondary structure of TMR riboswitch hints the possibility to improve riboswitch design by optimizing parameters for our prediction algorithm to avoid the formation of stable hairpin structure with SD sequence in ON conformation while maintaining the stability of the OFF conformation.

Although the efficacy of our first TMR riboswitches is limited, given the unique chemical properties of TMR molecule, these riboswitches can lead to various potential downstream applications. The fluorescent molecule with its aptamer has demonstrated their important utilities in unveiling RNA folding and dynamics in various studies.<sup>29,30</sup> More fascinating research with TMR riboswitches awaits to be carried out in both prokaryotes and eukaryotes.

The comparison between *in vitro* and *in vivo* assay results also deserves our attention. Previously, we always attributed the apparently higher functional ligand concentration (comparing to aptamer  $K_D$  value) to the lack of permeability of the cells to ligands (or an active efflux system). Since the permeability of ligands in our *in vitro* cell-free system is not an issue, yet we observed similar *in vitro* and *in vivo* gene activation function for most of our TMR riboswitches. However, there are discrepancies for a number of TMR riboswitches shown in Figure 4.6. These discrepancies suggest the environmental factors can play an important part in riboswitch functions. Given the same composition of biological macromolecules and similar expression level, we speculate that the crowding effect could be the reason to explain the discrepancy. Our collaborators in the Salis lab investigated this crowding effect which is summarized in our submitted paper. In summary, higher riboswitch mRNA levels decrease riboswitch activation ratios inside the cell's crowded environment, but retain the function within the relatively dilute *in vitro* cell-free conditions. When the total volume is limited (a crowded system), linear riboswitch-containing mRNA will occupy a significant fraction of the free space, leading to the exclusion of ligand and a lower fraction of ligand-bound riboswitch. Adding mRNA to a crowded system will increase expression in both ON and OFF states, but will decrease the overall activation ratio due to a higher basal expression level. In contrast, a dilute system provides enough free volume so that adding more mRNA will not exclude ligand from the system, and activation may increase if ligand-binding follows second

order kinetics.

In the meantime, we noticed that our TMR riboswitches can only achieve modest gene activation comparing to our theophylline riboswitch. Most TMR riboswitches can only render 5-7 fold gene activation. We speculate that the limited regulatory ability of TMR riboswitches may result from their different three dimensional structures with stacking interaction. In theophylline riboswitches, the ligand can form multiple hydrogen bonds with its aptamer and provide large free energy to stabilize the complex, whereas in TMR riboswitch, TMR can only recruit  $\pi$ -stacking interaction to stabilize its RNA-ligand structure.

In order to achieve more significant genetic control, a high concentration of TMR, which is cytotoxic, is used in our gene activation experiment. This high concentration of TMR could induce a slow cell growth rate and, thus, a hindered expression of the reporter gene. In our experiment, we normally observed a ~50% reporter signal reduction with 20  $\mu$ M TMR in our *in vivo* assay and a 35% reporter signal reduction with 30  $\mu$ M TMR in our *in vitro* assay. Thus, if more efficient TMR riboswitches that could work in a lower TMR concentration can be designed from our improved prediction algorithm, the TMR riboswitch family will become more useful in the downstream applications.

In conjunction with our data from theophylline riboswitches<sup>35</sup>, our TMR riboswitch *in vitro* decoupling experiment suggests there may be a predominant co-transcriptional kinetic component for many of our TMR riboswitches. It helped to explain why

riboswitches characterized in cells typically respond to several magnitudes higher concentrations of ligand, while aptamers bind ligands at submicromolar concentrations *in vitro*. It also suggested that kinetic riboswitches are more efficient in genetic regulation since the half-life of RNA may not be long enough to establish the equilibrium required for gene activation.

Meanwhile, a similar rule from theophylline riboswitches seemed to apply to the TMR riboswitch family-riboswitches with higher activation ratios tend to adopt kinetically driven mechanism. One of our best *in vitro* TMR riboswitch, TMR S2 construct, exhibits distinct kinetically driven mechanisms, which is comparable to our theophylline D switch. Another two TMR riboswitch, S7 and S8, exhibit a mixed mechanism with the major component from the kinetically driven mechanism. This result suggests our computational screening process is capable of identifying not only equilibrium driven riboswitches, but also kinetically driven riboswitches.

In our *in vitro* DNA template dose-response assay, we observed an increasing gene activation function with higher input of DNA molecules. It is expected that the increasing amount of DNA input will utilize the RNA polymerase in the cell extract to produce more mRNA. In the kinetically driven mechanism, the ratio of ON/OFF conformations for riboswitch is only depended on the concentration of the ligand concentration, however, in the thermodynamic equilibrium, the ratio of ON/OFF conformation of riboswitch is not only depended on the concentration of the ligand concentration but also the concentration

of the RNA molecule itself. Therefore, with an increasing amount of RNA molecules, the general equilibrium is more likely to favor the ON conformation and provide extra genetic activation for the riboswitch.

However, in our *in vivo* assays with different plasmids, we observed a contrary response when we tried to increase the plasmid copy numbers in cell. Based on their origins of replication, the pBAC vector will have 1 copy/cell, the p15A vector will have ~10 copies/cell, the pColE1 vector will have ~50 copies/cell, and the pUC19 vector will have 100+ copies/cell in *E. coli* TOP10 strain. Among these four vector systems, the lowest copy number plasmid pBAC achieved the highest regulatory performances (12.3-fold activation for S2 switch) and other vectors performed poorly (5.7-fold activation for S2).

The discrepancy between *in vitro* and *in vivo* DNA template dose-response assays coincides with our finding in the varying performances of the same TMR riboswitch under cell-free and cell-based assays. The results made our argument about the crowding effect more convincing. That is, riboswitches tend to perform better in a diluted environment but poorly in a crowded environment. However, the limited results from a single riboswitch family can only provide a rough and qualitative explanation and more data are required to construct a predictive model for engineering purposes.

## **4.4 Experimental**

### ***General Consideration***

All plasmid manipulations were conducted with the standard cloning techniques and all constructs were verified by DNA sequencing (MWG Operon, Huntsville, AL). Purifications of plasmid DNA, PCR products, and enzyme digestions were facilitated by using kits from Qiagen (Germantown, MD), New England Biolabs (Ipswich, MA). 5-bromo-4-chloro-indolyl- $\beta$ -D-glucuronide X-gal was purchased from US Biological (Swampscott, MA). Theophylline, *o*-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG), ampicillin, and all other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Bacto™ Agar was purchased from BD (Sparks, MD). Synthetic oligonucleotides primers were purchased from IDT (Coralville, IA). All electrophoresis supplies were purchased from Bio-Rad Laboratories (Hercules, CA). All experiments were performed in *E. coli* TOP10 F' strain or TOP10 cells (Invitrogen, Carlsbad, CA) cultured in media obtained from EMD Bioscience (Merck KGaA, Darmstadt, Germany).

### ***Secondary Structure Modeling***

The estimated  $\Delta G$  for riboswitch sequences were generated using mFold<sup>20</sup> and cross-validated with Vienna RNA webserver<sup>21</sup>. The entire 5'-UTR along with the 5'-fragment of the luciferase reporter was set as the input for the free energy calculation. Inclusion of additional codons did not appear to disrupt the structures of the functional riboswitches. The ligand binding energy was set to -10.2 kJ/mol based on the dissociation constant of the aptamer sequence determined by experimental characterization.<sup>26</sup> Briefly, the molar Gibbs free energy of binding is calculated via  $\Delta G_{binding} = RT \ln \frac{K_d}{c^\emptyset}$ , where  $K_d$

is the dissociation constant measured by experiment and  $c^\theta$  is the standard reference concentration (1 mol/L). A fixed cut-off strategy was set to include the first 68 nt of the luciferase encoding sequence, whereas an adaptive cutoff strategy was set to accommodate the optimal sequences based on folding which varies from 75 to 95 nt of the luciferase sequence.

### ***In vitro Luciferase Expression Assay***

*In vitro* expression reactions were performed using S30 *E. coli* extract (Promega, item numbers L1020 and L1030). Reaction conditions followed by the manufacturer's instructions, and included S30 premix, S30 extract, amino acids, template DNA, and either ligand or water. An amount of 40 ng/ $\mu$ L of plasmid DNA template was used for each reaction. TMR concentrations were typically 30  $\mu$ M, except for the dose response curves. Reaction volumes were typically 20  $\mu$ L, but results were reproducible regardless of total volume. Reactions were incubated at 37°C for 30 min after mixing all the components, unless otherwise noted. Reactions were quenched by placing them on ice with dilution buffer provided by the manufacturer.

Luciferase assays were conducted using samples from *in vitro* expression reactions. In each assay, 20  $\mu$ l of the quenched reaction mixture was then added to a well on a black opaque 96-well plate. An equivalent amount of luciferase assay reagent (Promega) was added to each well and mixed right before the measurement. Luciferase activity was measured by reading the luminescent signals with a Biotek Synergy HT plate reader,

using the luminescence procedure with a sensitivity of 150.

For *in vitro* expression reactions, the data were normalized relative to the amount of DNA used. These values were then normalized relative to a single control reaction or time point within that experiment. In most cases, an aptamer-deletion construct with the luciferase gene was used as the control trial for normalization. All experiments were repeated three times on different days. Standard deviations were determined and graphed as error bars. From these average expression values, estimated activation ratios (ARs) were determined. Specifically, an AR is calculated by dividing the amount of gene expression in the presence of ligand by the amount of gene expression in the absence of ligand.

#### ***Ligand Dose-response Assay***

Ligand dose response curves were conducted *in vivo* as described in earlier chapters, but with varying amounts of ligand as shown in the figures. More specifically, concentrations of 0.1, 1, 10, 20, 100  $\mu\text{M}$  of TMR are used to induce gene activation. Also, the induction time is shortened to 20 minutes to reduce the cytotoxicity of high concentration of TMR. The same luciferase signal measurement method and normalization strategy are used.

#### ***DNA Template Dose-response Assay***

DNA template dose response curves were conducted *in vitro* as described above, but with varying amounts of DNA template as input. More specifically, plasmid DNA is

harvested from cell culture and the concentration of plasmid is measured by Nanodrop. A 500 ng/ $\mu$ L DNA stock solution is made by vacuum drying the excess water. Concentrations of 0.1, 1, 10, 20, 100 ng DNA per  $\mu$ L *in vitro* expression reaction as final DNA concentration are used to measure the gene activation effect. The same luciferase signal measurement method and normalization strategy are used.

### ***Gene Activation Assay with Various Vector Systems***

TMR riboswitches and luciferase reporter gene fusion were constructed and inserted into vectors derived from plasmid pBAC (mini-F' origin, Kan<sup>R</sup>, addgene.org), p15A (p15A origin, Amp<sup>R</sup>, Salis Lab), pFTV1 (also known as pColE1, ColE1 origin, Cm<sup>R</sup>, Salis Lab), and pUC19 (modified pMB1 origin, Amp<sup>R</sup>, Invitrogen) using standard molecular cloning. Briefly, DNA fragments were computationally designed, synthesized, and assembled using either annealing of oligonucleotides, PCR assembly of oligonucleotides, or PCR amplification of DNA fragments (Integrated DNA Technologies). DNA fragments were then digested by BamHI and XbaI restriction enzymes, followed by ligation with digested plasmid, transformation, plating on selective media, and verification of purified plasmid by sequencing. *In vivo* luciferase assay is conducted as described previously.

### ***In vitro Transcription-Translation Decoupling Experiment***

Decoupling experiments were modified from our previous publication<sup>35</sup>. Briefly, 40 ng of DNA plasmid template per  $\mu$ L reaction was used for a normal cell-free reaction

with Promega *E. coli* S30 cell extract system. Amino acids were omitted during the initial 10 minute incubation. Rifampicin was then added at 250 ug/ml final concentration in the reaction to halt transcription. After a 1-minute incubation, amino acids were added (0.1mM per each amino acid as recommended by the manufacturer) and/or TMR, yielding a final concentration of 40  $\mu$ M TMR. This mixture was then incubated for up to 10 minutes. At 0, 1, 2, 4, 6, and 10 minutes after adding amino acid, an aliquot of reaction mixture is quenched and measured with our standard luminescence assay. Each experiment is conducted with a control where TMR is omitted in any stage of the reaction. Experiments are repeated three times on different days. Averages of three trials are reported and standard deviations are shown as error bars.

#### 4.5 References

1. Desai, S.K. and Gallivan, J.P., *J Am Chem Soc*, **2004**. *126*: 13247-54.
2. Sharma, V., Nomura, Y. and Yokobayashi, Y., *J Am Chem Soc*, **2008**. *130*: 16310-5.
3. Muranaka, N. and Yokobayashi, Y., *Chem Commun (Camb)*, **2010**. *46*: 6825-7.
4. Endo, K. and Saito, H., *Methods Mol Biol*, **2014**. *1111*: 183-96.
5. Hanson, S., Bauer, G., Fink, B. and Suess, B., *RNA*, **2005**. *11*: 503-11.
6. Wieland, M., Benz, A., Klauser, B. and Hartig, J.S., *Angew Chem Int Ed Engl*, **2009**. *48*: 2715-8.
7. Dohno, C., Kohyama, I., Kimura, M., Hagihara, M. and Nakatani, K., *Angew Chem Int Ed Engl*, **2013**. *52*: 9976-9.
8. Davidson, M.E., Harbaugh, S.V., Chushak, Y.G., Stone, M.O. and Kelley-Loughnane, N., *ACS Chem Biol*, **2013**. *8*: 234-41.
9. Lynch, S.A., Desai, S.K., Sajja, H.K. and Gallivan, J.P., *Chem Biol*, **2007**. *14*: 173-84.
10. Lynch, S.A. and Gallivan, J.P., *Nucleic Acids Res*, **2009**. *37*: 184-92.
11. Topp, S. and Gallivan, J.P., *Chembiochem*, **2008**. *9*: 210-3.
12. Nomura, Y. and Yokobayashi, Y., *J Am Chem Soc*, **2007**. *129*: 13814-5.
13. Sharma, V., Nomura, Y. and Yokobayashi, Y., *J Am Chem Soc*, **2008**. *130*: 16310-5.
14. Ceres, P., Garst, A.D., Marcano-Velazquez, J.G. and Batey, R.T., *ACS Synth Biol*, **2013**. *2*: 463-72.

15. Ghazi, Z., Fowler, C.C. and Li, Y., *Methods Mol Biol*, **2014**. *1111*: 57-75.
16. Barrick, J.E. and Breaker, R.R., *Genome Biol*, **2007**. *8*: R239.
17. Meyer, M.M., Roth, A., Chervin, S.M., Garcia, G.A. and Breaker, R.R., *RNA*, **2008**. *14*: 685-95.
18. Kikin, O., D'Antonio, L. and Bagga, P.S., *Nucleic Acids Res*, **2006**. *34*: W676-82.
19. Perutka, J., Wang, W., Goerlitz, D. and Lambowitz, A.M., *J Mol Biol*, **2004**. *336*: 421-39.
20. Zuker, M., *Nucleic Acids Res*, **2003**. *31*: 3406-15.
21. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L., *Nucleic Acids Res*, **2008**. *36*: W70-4.
22. Whitaker, J.E., Moore, P.L., Haugland, R.P. and Haugland, R.P., *Biochem Biophys Res Commun*, **1991**. *175*: 387-93.
23. Detty, M.R., Prasad, P.N., Donnelly, D.J., Ohulchansky, T., Gibson, S.L. and Hilf, R., *Bioorg Med Chem*, **2004**. *12*: 2537-44.
24. Gibson, S.L., Holt, J.J., Ye, M., Donnelly, D.J., Ohulchansky, T.Y., You, Y. and Detty, M.R., *Bioorg Med Chem*, **2005**. *13*: 6394-403.
25. Grate, D. and Wilson, C., *Proc Natl Acad Sci U S A*, **1999**. *96*: 6131-6.
26. Baugh, C., Grate, D. and Wilson, C., *J Mol Biol*, **2000**. *301*: 117-28.
27. Flinders, J., DeFina, S.C., Brackett, D.M., Baugh, C., Wilson, C. and Dieckmann, T., *Chembiochem*, **2004**. *5*: 62-72.

28. Bernard Da Costa, J. and Dieckmann, T., *Mol Biosyst*, **2011**. 7: 2156-63.
29. Reif, R., Haque, F. and Guo, P., *Nucleic Acid Ther*, **2012**. 22: 428-37.
30. Sokoloski, J.E., Dombrowski, S.E. and Bevilacqua, P.C., *Biochemistry*, **2012**. 51: 565-72.
31. Wu, Y., Zhan, S., Xu, L., Shi, W., Xi, T., Zhan, X. and Zhou, P., *Chem Commun (Camb)*, **2011**. 47: 6027-9.
32. Stead, S.L., Ashwin, H., Johnston, B., Dallas, A., Kazakov, S.A., Tarbin, J.A., Sharman, M., Kay, J. and Keely, B.J., *Anal Chem*, **2010**. 82: 2652-60.
33. Gibson, S.L., Hilf, R., Donnelly, D.J. and Detty, M.R., *Bioorg Med Chem*, **2004**. 12: 4625-31.
34. Grate, D. and Wilson, C., *Bioorg Med Chem*, **2001**. 9: 2565-70.
35. Mishler, D.M. and Gallivan, J.P., *Nucleic Acids Res*, **2014**. 42: 6753-61.

## Chapter 5. Conclusions

### 5.1 Summaries and Conclusions

More than a decade has passed since the initial discovery of the first riboswitch in nature.<sup>1</sup> The study of both natural and synthetic riboswitches is still an intriguing and fascinating problem waiting for more efforts from molecular biologists. As these simple but versatile genetic regulatory elements extend their influence in more and more downstream areas, riboswitch researchers still lack the proper knowledge of riboswitch mechanism and efficient ways of creating new riboswitches.

In this dissertation, we illustrated the mechanism of our theophylline riboswitches on a molecular level (Chapter 2). Through our analysis, we revealed the detailed ligand-riboswitch interactions and refreshed the biophysical model for our theophylline riboswitches. We found a strong correlation between the free energy level and the gene expression level of the riboswitch complex. Thus, we advanced our knowledge about riboswitch functions onto a quantitative level. Based on this quantitative model, we successfully predicted theophylline riboswitch variants that outperform the existing ones.

We also verified the kinetic component of riboswitch function in a cell-based approach (Chapter 3). Our results indicated that the life span of mRNA molecules may not sustain the proposed equilibrium control by riboswitch. Meanwhile, the attempt to decouple transcription and translation *in vivo* yielded similar results comparing to our *in vitro* characterization<sup>2</sup> which further support our speculation of the predominant kinetic

components in riboswitch functions.

Moreover, we created a new family of TMR riboswitches from our computational predictions (Chapter 4). The *in vitro* and *in vivo* assays on the predicted TMR riboswitches proved the versatility of our biophysical riboswitch model. The results from *in vitro* transcription translation decoupling experiments provided insights to the mechanism of our new riboswitch family. The different performance of *in vitro* and *in vivo* DNA dose-response assays hinted at the influence of crowding effect in riboswitch functions.

Comparison of the functionality of the theophylline riboswitch family and the TMR riboswitch family led us to the conclusion that most best-performing riboswitches are kinetically driven due to the limited time window for riboswitch to function. Knowledge of the biophysical aspect of riboswitch functions revealed the important correlation between the structure and function of riboswitches and helped us to build and improve our computational prediction tools.

To our knowledge, our new TMR riboswitch family is one of the most efficient riboswitches that can work under sub-micro-molar ligand concentration and is the only synthetic gene-activating riboswitch to achieve that efficiency.<sup>3,4,5</sup> Even though the performance of TMR riboswitches deserves further optimization, their discovery demonstrated a novel approach to create functional riboswitches with computational assistance. However, the main disadvantage of our synthetic TMR riboswitches in

comparison to other inducible systems<sup>2,6</sup> is that the toxicity of the ligand is high and the ligand tolerance limits the potential of the riboswitch.

As we are crossing the hurdle of understanding the relationship between the structure and the function of riboswitches, we believe that the accumulated biochemical and structural information will be useful for predicting novel riboswitches for other candidate ligands. Despite incomplete knowledge about riboswitch mechanism, we are beginning to understand the principles that confer efficient ligand-inducible genetic regulatory function by riboswitches and have taken the first steps toward the rational design of riboswitches that target promising ligands.

Future work would concentrate on improving our TMR riboswitches and rapidly expanding the repertoire of the riboswitch family. As we have done some preliminary screening for promising target molecules, we envision that the success rate of creating novel riboswitches is significantly enhanced by the usage of computational screening tools. A step further would be exploring the full potential of riboswitches for medicinal and biotechnological applications which will lead to reveal a whole spectrum of synthetic molecular devices with our synthetic RNA sensors.

## 5.2 References

1. Serganov, A. and Nudler, E., *Cell*, **2013**. *152*: 17-24.
2. Mishler, D.M. and Gallivan, J.P., *Nucleic Acids Res*, **2014**. *42*: 6753-61.
3. Werstuck, G. and Green, M.R., *Science*, **1998**. *282*: 296-8.
4. Suess, B., Hanson, S., Berens, C., Fink, B., Schroeder, R. and Hillen, W., *Nucleic Acids Res*, **2003**. *31*: 1853-8.
5. Weigand, J.E., Sanchez, M., Gunnesch, E.B., Zeiher, S., Schroeder, R. and Suess, B., *RNA*, **2008**. *14*: 89-97.
6. Desai, S.K. and Gallivan, J.P., *J Am Chem Soc*, **2004**. *126*: 13247-54.