

***Purine Riboswitches: A model system for the study of RNA structural  
biology and a platform for the advancement of synthetic RNA  
technologies***

By

Joan Gabriel Marcano-Velázquez

B.S. University of Puerto Rico, Mayagüez P.R 2009

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirement for the degree of  
Doctor of Philosophy  
Department of Chemistry and Biochemistry

2015

This thesis for the Doctor of Philosophy degree entitled:  
*Purine Riboswitches: A model system for the study of RNA structural biology and a  
platform for the advancement of synthetic RNA technologies*

Written by Joan G. Marcano-Velázquez  
has been approved for the  
Department of Chemistry and Biochemistry

---

Dr. Robert T. Batey

---

Dr. Marcelo Sousa

---

Date

The final copy of this thesis has been examined by the signatories, and we find that both  
the content and the form meet acceptable presentation standards of scholarly work  
in the above mentioned discipline.

Marcano-Velazquez, Joan G. Marcano (Ph.D. Chemistry and Biochemistry)

*Purine Riboswitches: A model system for the study of RNA structural biology and a platform for the advancement of synthetic RNA technologies*

Thesis directed by Professor Robert T. Batey.

The scientific discipline known as synthetic biology aims to develop a set of tools and engineering principles to design artificial biological systems for the production of commodity chemicals like biofuels and therapeutics. A variety of bacterial non-coding RNAs named riboswitches, are particularly attractive for the field of synthetic biology for their ability to create intricate tertiary structures capable of binding a small molecule metabolite and translating this event into the regulation of gene expression.

Riboswitches have been studied with a variety of biophysical and biochemical techniques that have provided a wealth of information useful in the rational design of synthetic RNA systems. However, although the field has been successful in creating artificial riboswitches that work robustly *in vitro*, when these RNAs are expressed in the cellular environment they fail to achieve their purpose. I believe that the reason for these failures originates from the lack of research efforts that aimed to understand riboswitches in the intracellular environment.

To better understand the *in vivo* regulatory activity of riboswitches, I designed an intracellular reporter assay that couples the riboswitch's activity to a quantifiable fluorescence output. With this fluorescence reporter assay, I performed a comprehensive mutagenesis analysis of the *Bacillus subtilis* adenine-binding *pbuE* riboswitch, probing the structural features identified as being responsible for ligand binding, global structure acquisition, and regulation of transcription. The results of this

mutational analysis revealed that some of the structural motifs, for example a conserved loop-loop interaction that is important for ligand binding *in vitro*, display differential effects in their contribution to the *in vivo* regulatory activity. Moreover, the intracellular assay allowed me to evaluate structural elements of the expression platform that direct the regulatory activity, an aspect of riboswitches that has been understudied. My data revealed that the stability of a nucleator stem-loop element at the distal tip of the transcriptional terminator stem has a significant impact on the regulatory activity, most likely affecting the rate of the terminator formation. With this new data, I proposed a mechanism of regulation that suggests that the ligand interaction with the aptamer acts as a kinetic barrier between the interchange of the alternative secondary structures that mediate the riboswitch's regulatory response.

With the new insights of the *in vivo* activity of the purine riboswitch, I describe the creation of innovative techniques towards the development of novel synthetic riboswitches. This new approach attempts to create an artificial evolution system, in which using the tetracycline antibiotic resistance marker, I can select for riboswitches that bind novel ligands and capable of producing a regulatory response. The preliminary results of these experiments indicate that riboswitches are a robust platform for the *in vivo* directed evolution of RNAs, and that with the results generated in the purine riboswitch *in vivo* mutational analysis, I can provide ideas to improve the selection platform. With the development of this technology, I attempt to contribute to the discipline of synthetic biology by providing novel RNA tools that will enable the future of biotechnology.

## CONTENTS

### CHAPTER 1 –Introduction

1.1 Biological solutions to society’s problems.....	1
1.2 An example of how synthetic biology can play a central role in the achievement of long term human goals.....	2
1.3 Synthetic biology products used by our society.....	5
1.4 RNA solutions for the improvement of microbial cell factories.....	6
1.5 An example of how synthetic RNAs will play a role in the development of future therapeutics.....	8
1.6 The use of riboswitches for synthetic biology applications.....	10
1.7 The purine riboswitch: An ideal system for the development of new RNA parts.....	13
1.8 Rational approach towards changing the specificity of the purine riboswitch.....	16
1.9 Using the purine riboswitch as a scaffold RNA in SELEX experiments generates a novel 5-hydroxytryptophan aptamer, but not a novel riboswitch.....	19
1.10 Taking a step back: Understanding the <i>in vivo</i> activity of the purine riboswitches in order to devise strategies for the development of novel riboswitches.....	21

### CHAPTER 2- Structure-guided Mutational Analysis of Gene Regulation by the *Bacillus subtilis pbuE* Adenine-responsive Riboswitch in a Cellular Context

2.1 Design and validation of an adenine riboswitch <i>in vivo</i> activity reporter assay.....	24
2.2 Effect of the leader sequence on the <i>pbuE</i> riboswitch.....	28
2.3 The stability of the P1 “communication sequence” and its effect on the regulatory activity.....	30
2.4 A “nucleator” stem-loop is required for promoting efficient regulatory activity.....	38
2.5 Summary and discussion of results.....	43
2.6 Mechanistic model of the <i>pbuE</i> riboswitch.....	47
2.7 Experimental procedures.....	49

## **CHAPTER 3- RNA regulatory elements and their applications for synthetic biology**

<b>3.1</b>	<i>In vivo</i> evolution of novel riboswitches using tetracycline as a selection marker.....	53
<b>3.2</b>	Chimeric Purine riboswitch as platforms for <i>in vivo</i> directed evolution experiments.....	56
<b>3.3</b>	Optimization of selection platform and preliminary results on selection experiments.....	58
<b>3.4</b>	Experimental Procedures.....	68
	<b>References.....</b>	<b>69</b>

## **Appendix A- Structural characteristics of purine aptamer that lead to an adjusted regulatory response**

<b>A.1</b>	The aptamers of riboswitches classes are highly conserved in their structure, but vary significantly in their regulatory activity.....	75
<b>A.2</b>	A computational phylogenetic analysis reveals a pattern of co-variation for non-structurally interacting nucleotides.....	76
<b>A.3</b>	A region of co-variation was identified at the base of P2 and termed the “Tune” box.....	78
<b>A.4</b>	Mutational analysis tuning box reveals the nucleotide involvement in the binding kinetics.....	79
<b>A.5</b>	The P2 tune box variants impart functional diversity.....	82
<b>A.6</b>	Discussion.....	84
	<b>Appendix B- Supporting Figures.....</b>	<b>87</b>

## FIGURES

### CHAPTER 1

**Figure 1.1** Microbial production of acetaminophen

**Figure 1.2** Combinatorial library of *tunable intergenic regions* (TIGRs)

**Figure 1.3** Ligand dependent proliferation of engineered anti-cancer T-cells

**Figure 1.4** Riboswitches control gene expression by directly binding a small molecule

**Figure 1.5** Riboswitch-based biosensors for the study of the intracellular small-molecules metabolism

**Figure 1.6** Molecular architecture of purine riboswitches

**Figure 1.7** Changing the ligand specificity of purine riboswitches by directed mutations

### CHAPTER 2

**Figure 2.1** *In vivo* activity reporter system for the *pbuE* riboswitch

**Figure 2.2** Ligand-dependent regulatory activity of the *pbuE* riboswitch in *E. coli*.

Figure 2.3 Highly deleterious mutations disrupt the ligand-dependent regulatory activity of the *pbuE* riboswitch in *E. coli*

**Figure 2.4** Effects of the leader sequence length of the *pbuE* riboswitch in the *in vivo* regulatory activity

**Figure 2.5** Architecture of the conserved loop-loop interaction in purine riboswitches

**Figure 2.6** The effect of point mutations in L2 on 2AP-dependent regulatory activity

**Figure 2.7** Effect of mutations that change the length of the P1 helix on regulatory activity

**Figure 2.8** Invasion of the P4b nucleator element by the extended P1 helix

**Figure 2.9** Effects on the *in vivo* regulatory activity by mutations that alter terminator hairpin nucleator P4b

**Figure 2.10** The effect of extending the P4b helix to compensate for its partial disruption by the P1+8 mutation

**Figure 2.11** Energy Landscape of the co-transcriptional folding of the *pbuE* riboswitch

## CHAPTER 3

**Figure 3.1** *In vivo* selections with the tetracycline dual selector system

**Figure 3.2** Chimeric riboswitches used for the *in vivo* selection

**Figure 3.3** Positive *in vivo* selections using the *xpt*(C74U)/*metH* riboswitch chimera

**Figure 3.4** Negative *in vivo* selection using the *xpt*(C74U)/*metH* riboswitch chimera

**Figure 3.5** Positive *in vivo* selections using the *pbuE/pbuE\** modular riboswitch

**Figure 3.6** Negative *in vivo* selection using the pRS-*pbuE/pbuE\** plasmid

## APPENDIX A

**Figure A.1** Evolutionary conservation of the purine aptamer

**Figure A.2** Normalized mutual information analysis on the purine aptamer domain alignment

**Figure A.3** Classification of co-variation sequences and their impact to the ligand binding properties of the purine aptamer.

**Figure A.4** *In vivo* regulation of gene expression by a purine/SAM hybrid riboswitch

## APPENDIX B

**Figure B.1** Sequences used in the in the mutational analysis of the *pbuE* riboswitch in the intracellular context

**Figure B.2.** Sequence of pRR-*pbuE* plasmid construct

**Figure B.3** pRS plasmid (Complete sequence)

**Figure B.4** B.4 pRS-*xpt*(C74U)/*metH*

**Figure B.5** pRS- *pbuE/pbuE\**

## Chapter 1

### Introduction

#### 1.1 Biological solutions to society's problems

The challenges that societies around the world will face this century reside on many fronts. This includes producing renewable and sustainable energy sources that counteract climate change, providing food security to the ever-increasing world population, curing complex diseases for which current medicines fail to treat such as cancer and AIDS, and producing our commodity chemicals in environmentally friendly processes<sup>1</sup>. To solve these complex problems, scientists and engineers have combined efforts to find answers in biological systems<sup>2</sup>. But in order to solve our current societal needs using a new generation of biology-based technologies, we first need to learn how to reliably and cost effectively engineer biology.

Synthetic biology is an emerging scientific discipline that intends to draw parallels between the engineering of electronics for which parts are highly modular and biological systems in order to streamline the design of biology<sup>3</sup>. To achieve this, synthetic biologists aim to separate the roles of developing and manufacturing biological “parts” from a discipline that will solely devise a “code” that will result in a programmed action. The main impediment in this strategy is that genetic elements that have been characterized in a specific biological context will not perform as expected when used in a different genetic context.

## **1.2 An example of how synthetic biology can play a central role in the achievement of long-term human goals**

At the current moment, the National Aeronautics and Space Administration (NASA) is planning several missions to explore nearby asteroids, set colonies on the Moon, as well as to explore our neighbor planet Mars<sup>4-6</sup>. These missions pose many challenges, and the most pressing goal is to reduce the amount of shipped materials in order to support a safe voyage and return of the space crew. The recent advances in synthetic biology are considered realistic solutions to the challenges of these proposed space missions. Specifically, living organisms can be engineered to generate fuels, foods, biopolymers to be used in 3D printing of spaceship parts and research tools, and medicinal compounds<sup>7-9</sup>.

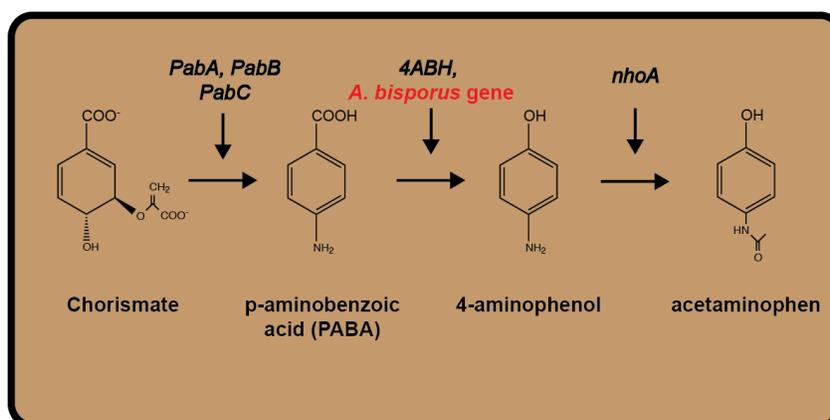
The main approach for producing these chemicals in outer space will be taking advantage of compounds available at the target destination. For example, the surface of Mars contains substantial levels of CO<sub>2</sub> and nitrogen<sup>10</sup>. With these available chemicals, it is feasible to envision a microorganism that can fixate CO<sub>2</sub> and nitrogen for the production of chemical “building blocks” to be used in the further production of compounds needed to support the mission. A proposed method for the production of these important compounds includes the use of an acetogenic microorganism that can fix CO<sub>2</sub> for the production of acetate. In turn, acetate is a viable feedstock for its role in aerobic metabolism through the citric acid cycle, and it is the precursor of biosynthetic pathways that generate biofuels<sup>11</sup>. Another attractive microorganism are cyanobacteria due to their well-established genetic manipulation techniques, their short sequenced and annotated genomes, rapid growth, and ability to use nitrogen gas (N<sub>2</sub>) directly<sup>12</sup>.

The use of microorganisms for the production of compounds is also the proposed approach to solve another challenge in space mission: the short shelf life of medicines due to radiation exposure<sup>13</sup>. This issue is of particular importance since it directly affects the quality of life of the astronauts in the mission. At the moment, one of the approaches to tackle this problem is to send an unmanned spaceship to replenish the supplies for the crew at different stages of the mission, which will be extremely difficult to coordinate and expensive. Optimally, NASA would like to include means for the astronauts to generate medicines in their spaceships when needed. In this scheme, they will send inactive microorganisms in lead-protected vials to prevent the organism from being exposed to radiation, and then to be cultivated when a specific medicine is needed.

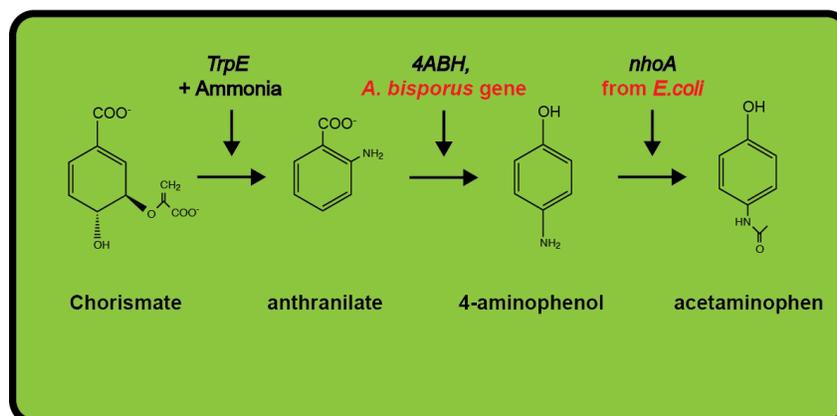
Synthetic biology has already demonstrated how useful it can be for the production of medicines through the genetic manipulation of microorganisms<sup>14</sup>. As an example on how these microorganisms could be used in a space mission, a recent review explained how the production of acetaminophen could be achieved in cyanobacteria<sup>15</sup>. *E. coli*, are capable of producing acetaminophen by introducing a single enzyme; 4-aminobenzoate hydroxylase (*4ABH*) from *Agaricus bisporus* (crimini mushroom) converts para-aminobenzoic acid to 4-aminophenol, which in turn is converted to acetaminophen by a native *E. coli* enzyme (*nhoA*)<sup>16</sup> (Figure 1.1). This pathway can be engineered into the cyanobacteria *Synechocystis* sp. PCC6803 by transposing both the mushroom's and *E. coli*'s enzymes to complement an alternative pathway that utilizes the intermediary anthranilate (Fig. 1.1). This route is more

attractive since the production of anthranilate can be achieved from the available resource ammonia instead of glutamine.

### *E. coli*



### *Synechocystis* (cyanobacteria)



**Fig. 1.1 Microbial production of acetaminophen.** The endogenous enzymes in *E. coli* *pabA*, *pabB* and *pabC* convert chorismate into para-aminobenzoic acid. A heterologous enzyme (*4ABH*) from the mushroom *Agaricus bisporus* is expressed in *E. coli* to convert PABA into 4-aminophenol, which in turn is converted to acetaminophen by a native enzyme *nhoA*. This pathway can be reconstructed in the cyanobacteria *Synechocystis* by shunting anthranilate, which is produced natively from chorismate by the *trpE* gene, into 4-aminophenol using again the mushroom's *4ABH* that also use anthranilate as a substrate. To achieve the final conversion step into acetaminophen, it is necessary to express the *E. coli*'s *nhoA* gene in *Synechocystis*. Figure adapted from Menezes et. al.

While the majority of the ideas described above are theoretical, they are deeply rooted in what synthetic biology has already achieved. In the following sections, I will present examples of the advances in synthetic biology that our current societies already enjoys in addition to projects that are still in progress.

### **1.3 Synthetic biology products used by our society**

There are technologies as a result of synthetic biology that have already reached the consumer. The most notable are genetically engineered foods that contain genes from other organisms that provide the plant with favorable traits such as flavor and appearance, or to increase crop yields in difficult to grow environments<sup>17,18</sup>. In addition, plant systems have been engineered to produced synthetic polymers that can be harvested to produce environmentally friendly fabrics, such as DuPont's SONORA<sup>®</sup>. In the near future, we will likely use plants for the production of many chemicals of interest, as well as other cell types in medical and environmental breakthrough technologies. However, before we can achieve the true potential of synthetic biology, new parts must be developed that expand the catalogue of biological activities in order to produce compounds that are not found in nature.

Synthetic RNA devices can play an important role in the advancement of the field, as there are several advantages to using nucleic acids for synthetic biology applications. First is the cost effectiveness of the chemical nucleic acid synthesis compared to the chemical synthesis of proteins (~ \$0.10 per nucleotide vs. ~ \$2.00 per amino acid). This gives RNA researchers additional tools, like the ability to cheaply test

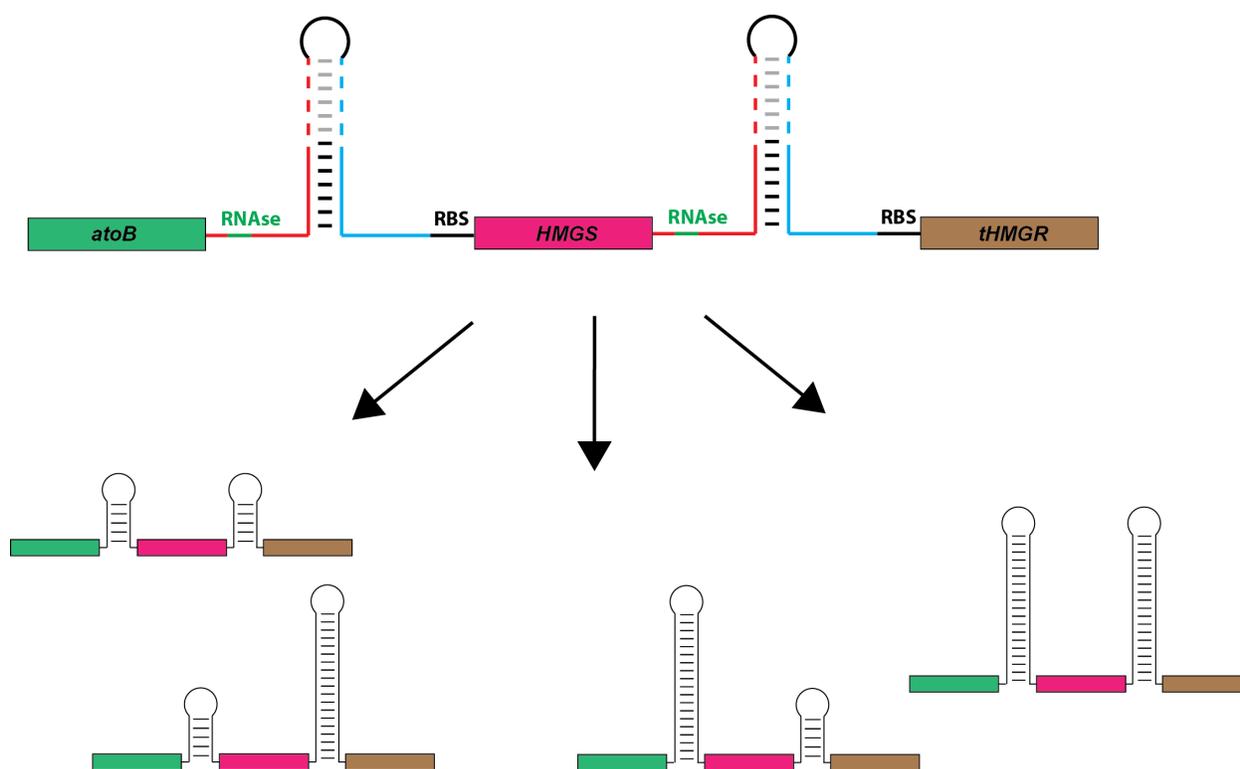
large, chemically-synthesized libraries for different applications. Secondly, biological activities like small molecule binding and regulation can be encoded in small RNAs, which allow the cells to respond faster to changes in the environment<sup>19</sup>. Finally, with RNA, it is possible to use well-established directed evolution techniques, such as SELEX, that allows for the rapid development of novel activities<sup>20</sup>. For these reasons, RNA has been the solution for challenges in the metabolic engineering and medicinal research. In the following sections, various examples of successful synthetic RNA approaches will be highlighted.

#### **1.4 RNA solutions for the improvement of microbial cell factories**

One of the most famous examples of synthetic biology to produce medicines is the development of *E. coli* cells that can produce artemisinic acid, a precursor of the popular malaria drug Artemisinin that is traditionally isolated from the plant *Artemisia annua* L. For this application, multiple genes were inserted into *E. coli* containing the metabolic genes to synthesize artemisinic acid<sup>21</sup>. Initially, the production yields were minimal, mostly because of metabolic bottlenecks in the pathway, suggesting that fine-tuning the expression levels of the genes was necessary to achieve higher yields.

To produce artemisinin in bacteria, the three heterologous genes *atoB*, *HMGS*, and *tHMGR* have been introduced into *E. coli* as an operon (MevT) to synthesize mevalonate, a precursor of the production pathway<sup>22</sup>. A minimal production of mevalonate was identified as a one of the bottlenecks in the pathway, but overexpression of the operon led to detrimental effects on the cell's growth probably due to an accumulation of toxic intermediates. The Smolke research group devised a

strategy to individually tune the genes in the *MevT* operon by introducing *tunable intergenic regions* (TIGRs)<sup>23</sup>.



**Fig 1.2 Combinatorial library of *tunable intergenic regions* (TIGRs).** The library is constructed such that a constant region has the potential to base pair (red pairs with blue) but resulting in hairpins of different lengths. This will generate a series of hairpins with the potential to sequester the RNAse cleavage sites. If the RNAse site is exposed and the mRNA is cleaved, the stability of the transcript as well as the translation initiation is modulated resulting in differential expression of the genes contained in the operon.

To generate the TIGRs, the authors introduced a combinatorial library of RNA secondary structures with the potential to sequester RNAse cleavage sites and a ribosomal binding site (RBS) (Fig 1.2). The resulting library was transformed into *E. coli*, and the cells were screened for the increased production of mevalonate. This approach improved the heterologous pathway production of mevalonate by seven-fold<sup>23</sup>. Further characterization of the optimized operon revealed that the randomized regions of the

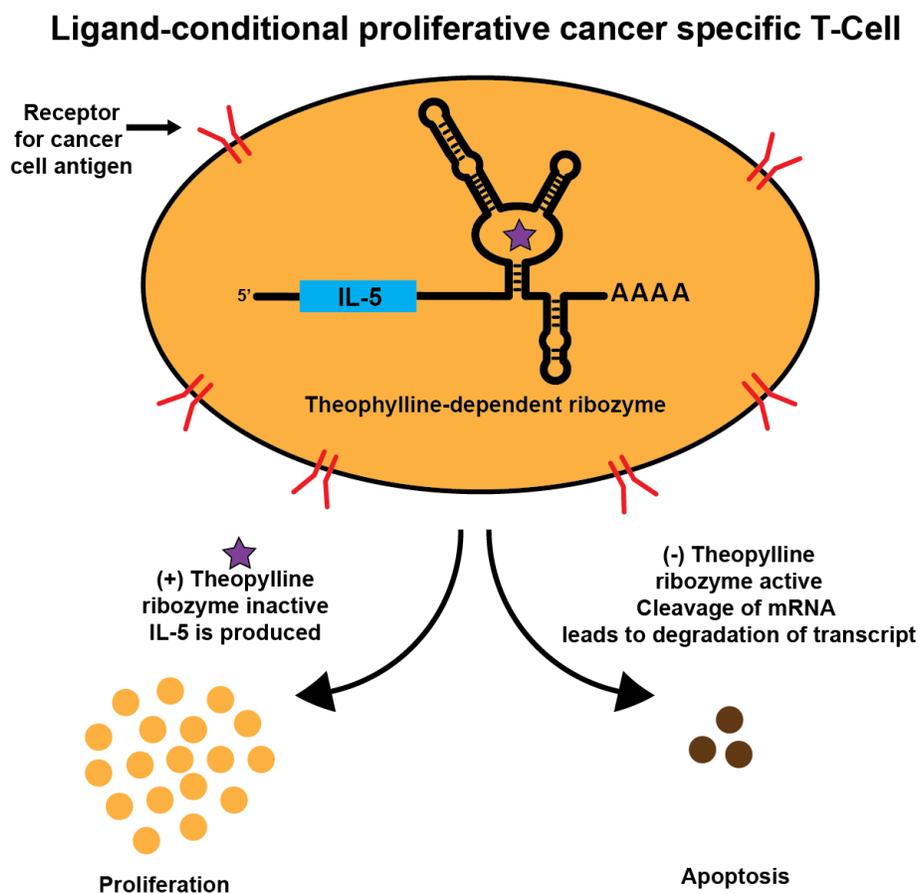
TIGRs formed a variety of RNA hairpin structures that resulted in differences in mRNA processing, transcription termination, and RBS sequestration. This work highlights how the utilization of RNA parts for the improvement of artificial metabolic pathways can expand the tools of synthetic biology towards the development of microbial cell factories.

### **1.5 An example of how synthetic RNAs will play a role in the development of future therapeutics**

The initial advances of synthetic biology focused on programming bacterial cells. However, lessons learned at the microbial level have been applied to mammalian cells, mostly with the intent of producing new therapies for diseases that have no current treatment. Although the techniques for manipulating mammalian cells have lagged behind the developments for bacterial cells, recent advances of genome editing tools such as TALEN and CRISPR have revived the efforts for mammalian synthetic biology<sup>24</sup>.

An example of the promising advances in medicine delivered by synthetic biology is the Chimeric Antigen Receptor-based T-cell therapy (CARs)<sup>25</sup>. CARs are patient derived genetically engineered T-cells repurposed *ex vivo* to express a synthetic receptor that recognizes cancer cells antigens. These engineered T-cells have been shown to be very efficient in targeting cancer cells for destruction, but the display of off target activity (attacking healthy cells) makes them toxic to the host at elevated levels. Recently, a group engineered CARs that would only proliferate in the presence of the small molecule theophylline<sup>26</sup>. To achieve this, the group engineered a ribozyme with a theophylline-dependent activity at the 3' side of the Interleukin-5 (IL-5) mRNA that is

necessary for cell growth and proliferation. Thus, the cells would only proliferate when expressing IL-5 in the presence of theophylline (Fig. 1.3). With such direct control, the therapeutic cells could be removed before their accumulation becomes toxic by eliminating theophylline from the system.



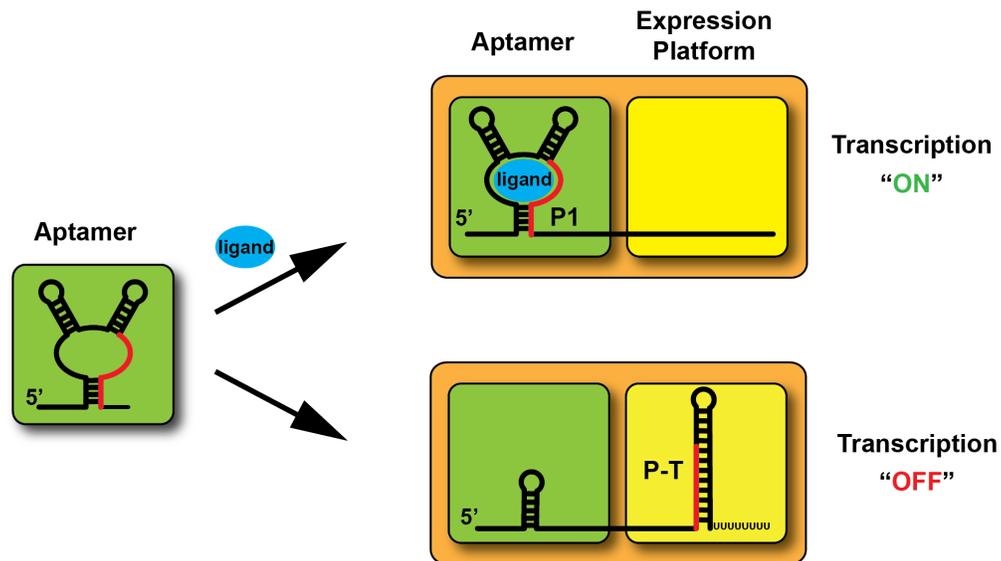
**Figure 1.3 Ligand dependent proliferation of engineered anti-cancer T-cells.** The immunologic cells are designed to express receptors specific to antigens displayed by cancer cells. This permits the cells to circulate around an organism to bind and destroy malignant cancer cells. In addition, these cells possess a catalytic ligand-dependent ribozyme that controls the expression of the Interleukin-5 gene, which is necessary for cell proliferation. When theophylline is present, it will bind the aptamer domain of the ribozyme to repress its activity. When theophylline is not present, the ribozyme is active and upon cleavage it promotes the rapid degradation of the RNA transcript. When the mRNA is degraded, the IL-5 expression is shut down. Adapted from Chen et. al.

Besides cancer-killing T-cells, there has been medical advances that employ gene therapies through lentiviral delivery for lipoprotein deficiency<sup>27</sup>, encapsulated mammalian cells that release uric acid to treat hyperuricemia<sup>28</sup>, and light-activated kidney cells that release glucagon-like peptide-1 that regulates glucose homeostasis<sup>29</sup>. These technologies represent only a glimpse of what the future of medical biotechnology holds for us. At the moment, much optimization needs to be performed before this new breed of therapies reaches the consumers, but as shown with the T-cell example, regulatory RNAs may serve as invaluable tools to advance medical synthetic biology technologies.

### **1.6 The use of riboswitches for synthetic biology applications**

The example described in the previous section used a synthetic RNA that is capable of responding to an external chemical cue that directs its function. These type of input to output RNA devices are extreme useful as it provides the system's user the means to tightly control the genetic system on demand. In bacteria this type of regulatory RNAs is called a *riboswitch*, which naturally interrogates the chemical environment and elicits a genetic response. These devices have already proven useful as synthetic biology tools. Riboswitches are mRNA leader sequences that regulate gene expression via their ability to directly bind a small molecule ligand. They are called riboswitches because the binding event causes the RNA to “switch” from one conformation to another with a different outcome. Riboswitches are composed of two main structural elements widely known as the *aptamer* and the *expression platform* (Fig

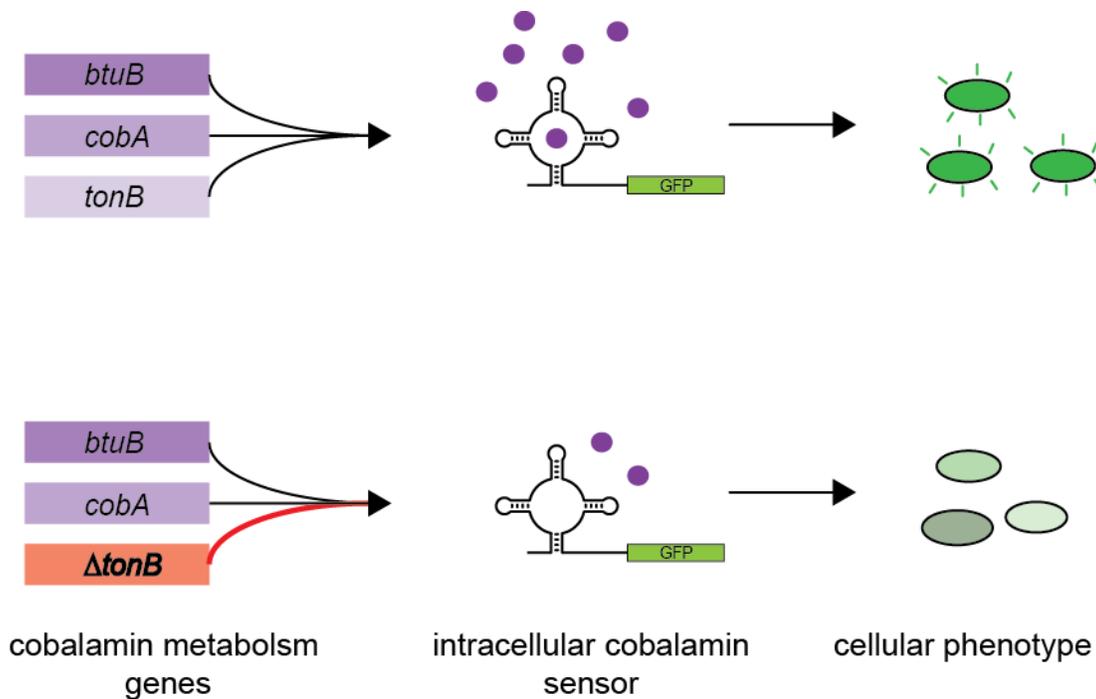
1.4). The aptamer domain acts as a receptor domain that adopts a structure with the ability to bind its cognate ligand with high affinity and specificity. The regulatory effect of the riboswitch is mediated by the expression platform, and in the majority of riboswitches this translates to an activation or repression of the transcription or translation processes.



**Fig 1.4 Riboswitches control gene expression by directly binding a small molecule.** In the simplest scheme one can envision for a transcriptional "on" riboswitch, the ligand binding to the aptamer domain (green box) prevents the folding of the expression platform (yellow box). If the aptamer does not find the ligand in the timescale of transcription, folding of the expression platform will occur, which contains a strong transcriptional terminator that prevents the expression of the downstream gene.

Two very desirable activities are contained within a riboswitch: molecular recognition of a chemical input and regulation of gene expression. These two activities may be repurposed, both for synthetic biology applications and novel research tools. The initial and most straightforward of these applications has coupled the activity of the

riboswitch to an enzymatic or fluorescence assay, creating an intracellular small molecule sensor. As an example of how a riboswitch can be used as a research tool was exemplified by one study that repurposed a cobalamin riboswitch to detect the cobalamin levels inside the cell<sup>30</sup>(Fig. 1.5). In this system, the mRNA leader sequence of the *btuB* riboswitch in *E. coli* is cloned upstream of a fluorescent protein, so that the cell's fluorescence could serve as an output of the riboswitch's activity. With their reporter, they were able to systematically evaluate genes involved in cobalamin metabolism, validating a set of genes thought to be involved in cobalamin biosynthesis, but for which no direct evidence was available. Furthermore, with the cobalamin reporter, the extent of the contribution towards the cobalamin pool can be quantified, as opposed to a qualitative assay that inform whether or not the gene is involved in the pathway. Since there are a number of riboswitches that recognize diverse ligands, it is possible that many different metabolic pathways could be studied by creating riboswitch-based reporters.

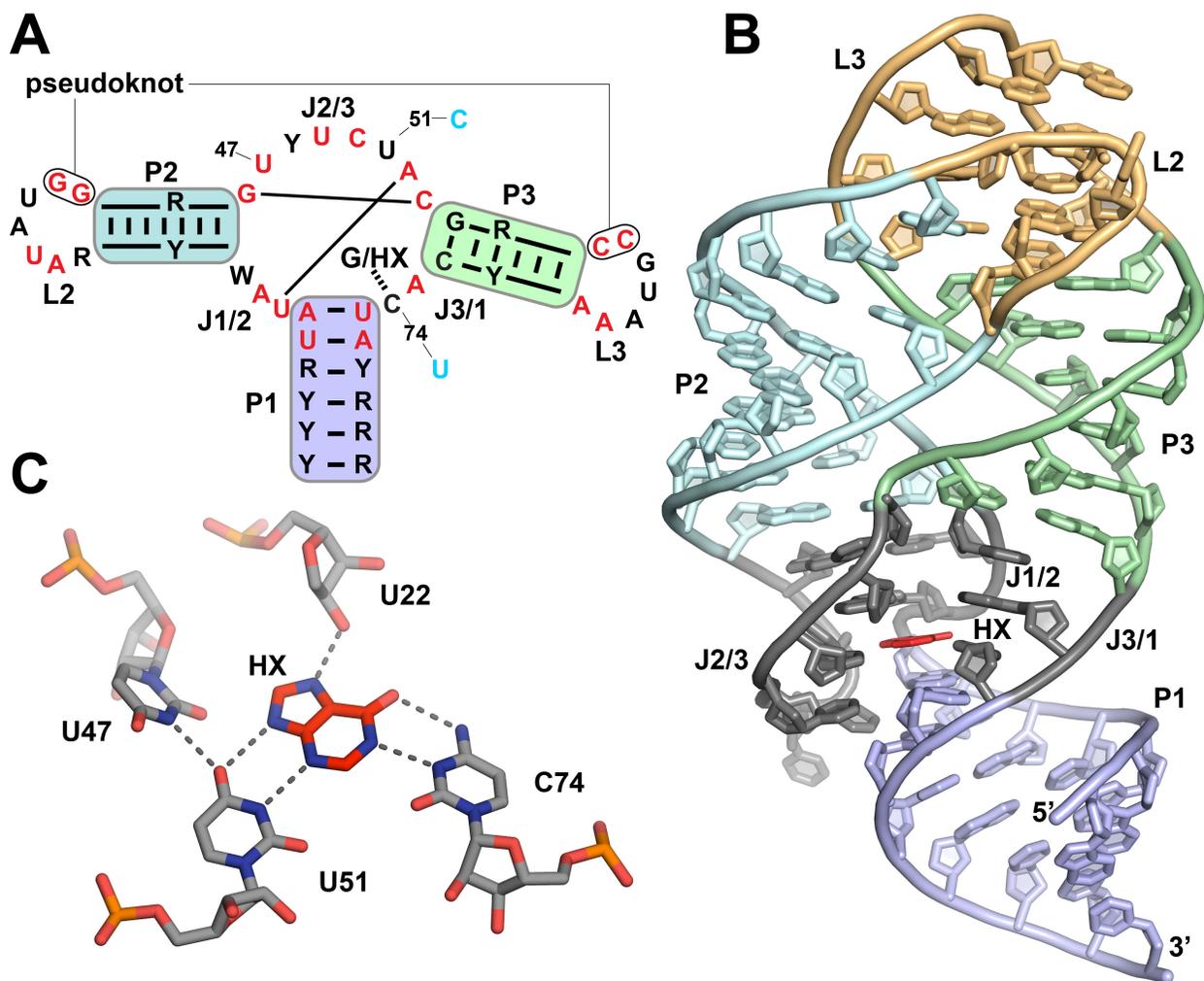


**Fig. 1.5 Riboswitch-based biosensors for the study of the intracellular small-molecules metabolism.** In this example, from the work of Fowler et. al., a cobalamin biosensor was created by fusing the *btuB* riboswitch to the green fluorescence protein gene.

### 1.7 The purine riboswitch: An ideal system for the development of new RNA parts

After their discovery, purine riboswitches rapidly became a popular system for studying RNA structure, folding, and interactions with small molecules<sup>31</sup>. This is in part because the purine riboswitch is a small and simple structure. The structure of the RNA is composed of three paired regions around a three-way junction with tight helical packing and tertiary interactions between two distal loops (Fig. 1.6 A, B). Moreover, this RNA is remarkably well behaved in terms of the stability of its conformation due to the evolutionary pressure to fold rapidly in the dynamic event of transcription. Finally, this architectural arrangement is found through nature in other important biological systems,

like other riboswitch domains, the hammerhead ribozyme, the signal recognition particle (SRP), and various substructures of the ribosomal RNA<sup>32</sup>.



**Fig. 1.6 Molecular architecture of purine riboswitches. A)** Consensus secondary structure of purine riboswitches derived from the structural alignment in the Rfam database. The red nucleotides are at least 90% conserved in the alignment. The background shadow follows the color scheme of the crystal structure. **B)** Crystal structure of the *xpt* guanine sensing aptamer bound to hypoxanthine. **C)** Binding pocket of the *xpt* aptamer bound to hypoxanthine, a guanine analogue. The nucleotides making contacts with the ligand are universally conserved, except for C74 that changes to U74 in adenine binding aptamers. Figure taken from (Porter, Marciano, & Batey, 2014)

Our understanding of how purine riboswitches bind their cognate ligand comes from the crystal structure of four different riboswitches that have been elucidated so far: the guanine-binding *B. subtilis xpt-pbuX*<sup>33,34</sup>, the adenine-binding *V. vulnificus add* riboswitch<sup>34</sup>, the adenine-binding *B. subtilis pbuE*<sup>35</sup>, and the 2' deoxyguanosine binding 1A from *M. florum*<sup>36</sup>. These crystal structures revealed that all purine riboswitches possess a highly conserved architecture independent of the ligand they recognize. The binding pocket is composed by three almost invariable nucleotides (U47, U51 and C74) that form critical hydrogen bonds necessary for ligand recognition, and the hydroxyl group of U22<sup>37-39</sup> (Fig. 1.6 C). Discrimination between guanine and adenine is achieved through the identity of a single nucleotide, C74 for U74, which enables Watson-Crick interactions with guanine and adenine, respectively (Fig. 1.6 A, C). Theoretically, uracil could recognize guanine through a wobble pair, but the architecture of the binding pocket prevents the organization of the adjacent nucleotides to accommodate the shifted U74 necessary for the non-canonical pair to occur<sup>40</sup>.

Recognition of 2' deoxyguanosine is achieved primarily by changing U51 into cytosine (Fig. 1.6, A)<sup>36,41</sup>. In the bound state, C51 shifts towards position 74 relative to U51 in the guanine-bound aptamer, sterically accommodating the 2'-deoxyribose sugar. Accordingly, nucleotide 47 disengages from its interactions with residue 51 and reorients itself out towards the solvent. Thus, it appears that the identities of just two nucleotides, 51 and 74, govern discrimination between different purine nucleobases and nucleosides<sup>41</sup>.

In addition to natural purine compounds, purine riboswitches are capable of binding different chemical analogues, such as 2,6-diaminopurine (DAP) and 2-aminopurine (2-AP)<sup>33,42</sup>. The binding activities towards these compounds have served as an invaluable tool in the study of riboswitch biology contributing to the understanding of ligand binding recognition<sup>33,42</sup>. The purine analogue 2-AP possesses fluorescence properties that can be used to track the localized folding and binding activity of the RNA, as it can be easily incorporated to purine aptamer due to its short length (~60 nt)<sup>43</sup>.

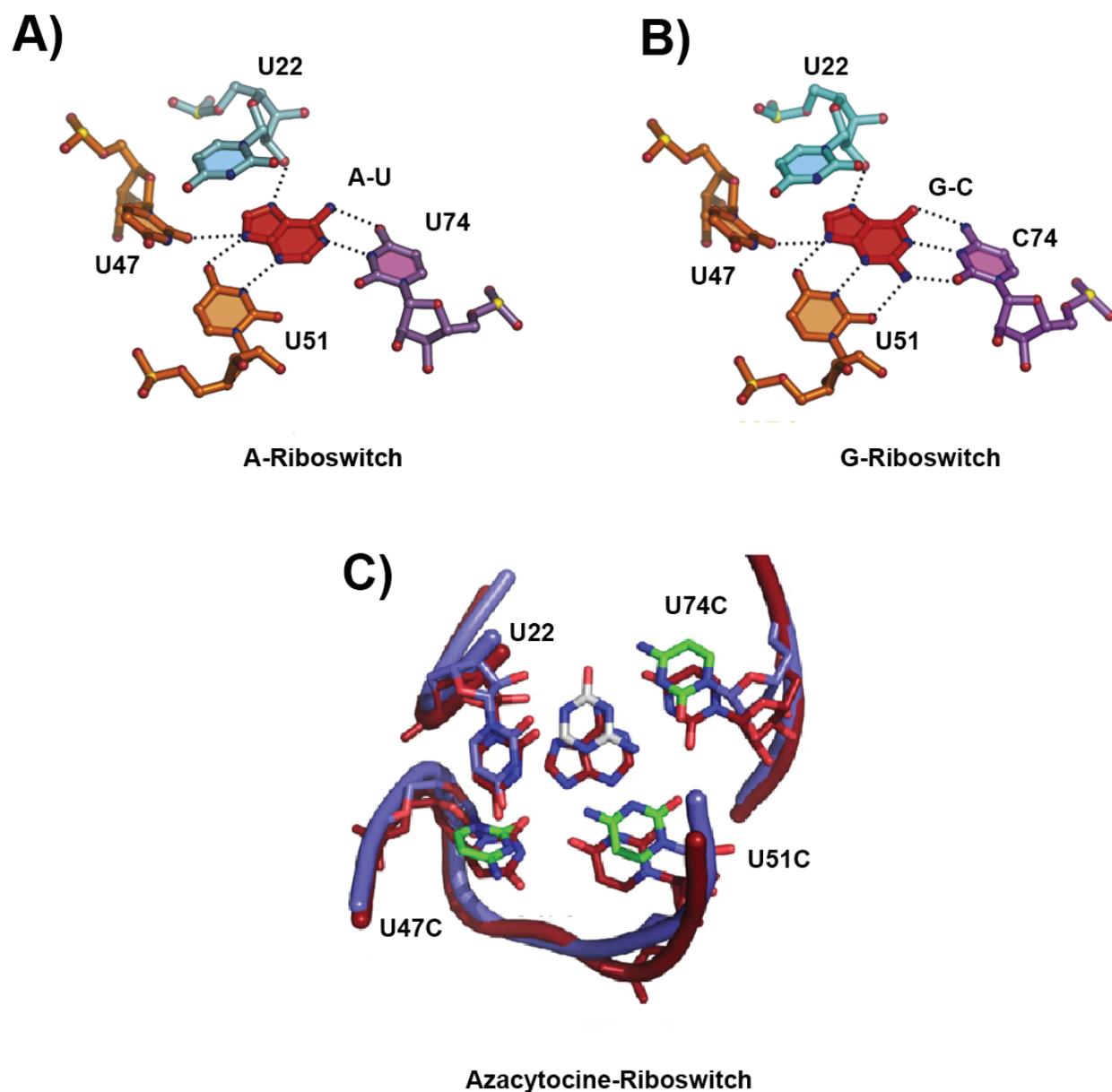
All of these favorable properties of purine riboswitches have placed them at the center of a multitude of biophysical research projects that have substantially advanced our understanding of RNA biology<sup>31</sup>. Moreover, given the wealth of understanding of this RNA and its favorable properties, several research projects in the Batey lab and others have used this riboswitch as a scaffold for the development of novel RNA synthetic parts. These projects have mainly focused in the modulation of the binding pocket towards recognizing orthogonal ligands, while attempting to retain the riboswitch's regulatory activity.

### **1.8 Rational approach towards changing the specificity of the purine riboswitch**

The crystal structures of the purine riboswitches have provided invaluable understanding on how these RNA organize their binding pocket to bind their cognate purine ligands. This information can be used to rationally devise approaches aimed to change the ligand specificity of the purine riboswitch to create novel regulatory devices for synthetic biology. Such is the work presented by the Micklefield group, which re-engineered the adenine sensing *add* riboswitch to recognize purine analogues<sup>44</sup>. In their

work they generated a riboswitch library by randomizing the two conserved nucleotides U47 and U51 that recognize the non-WC faces of adenine as shown by the crystal structure (Fig 1.7, A). The resulting 15 sequences were cloned in plasmid upstream of the chloramphenicol acetyl transferase gene (CAT) and transformed into *E. coli*. The cells were screened against a library of 80 heterocyclic compounds for their ability grow in the presence of chloramphenicol<sup>44</sup>. From this compound screen, the researchers identified a new riboswitch (U47C, U51C) that binds ammeline and permits the translation of a gene downstream. To characterize the regulatory properties of this new riboswitch, the authors exchanged the CAT gene for that of GFP, and transformed the plasmid once again into *E. coli*. With this system, the authors report a 6-fold induction of fluorescence upon adding 500  $\mu$ M of ammeline to the growth media, and no cross activation of the switch upon addition of the natural riboswitch's ligand adenine<sup>44</sup>.

In the context this new ammeline riboswitch, the authors mutated the nucleotide U74, which confers specificity between adenine and guanine in the purine aptamer (Fig. 1.7, A and B), to C74 to create a triple mutant (U47C, U51C, U74C) (Fig.1.7, C). This riboswitch was then screened against the heterocyclic compound library using the CAT system, and it was determined that triple mutant riboswitch displays a 2.8-fold induction upon the addition of azacytosine to the media<sup>44</sup>.



**Fig 1.7 Changing the ligand specificity of purine riboswitches by directed mutations.** The purine riboswitch uses four nucleotides to fully recognize all faces of the purine ligands. In the report by Dixon et al, a very limited library was constructed using the *add* adenine binding riboswitch by changing U47 and U51 for all the possible combinations (15 total sequences) and screening the resulting riboswitches against a library of heterocyclic compounds. A mutant sequences (U47C, U51C, U74C) was able to bind ammeline and activate the expression of the riboswitch *in vivo*. Figure adapted from Nixon et. al.

This work represents a very directed and rational approach to change the specificity of a purine riboswitch. Their library was significantly small enough that they could create each possibility easily. However, due to the limited changes, this library is restricted to screen small purine-like compounds and does not highlight the expanded ligand possibilities for the purine aptamer.

### **1.9. Using the purine riboswitch as a scaffold RNA in SELEX experiments generates a novel 5-hydroxytryptophan aptamer, but not a novel riboswitch**

Theoretically, it is possible to generate an RNA aptamer towards any ligand that can be chemically coupled to a solid matrix using the *Systematic Evolution of Ligands by Exponential Enrichment* (SELEX) approach<sup>45</sup>. The target ligand is attached to a resin and incubated with a nucleic acid randomized library; the “winner” sequences will bind the ligand and then are physically separated from the library due to their interaction with the resin. The winning RNA sequences are eluted from the resin and amplified by reverse transcription and PCR. This step recreates a pool of selected sequences that should contain an enriched population of those RNAs capable of binding the ligand in the artificial conditions just described. After various rounds of these steps, the pool of RNA is a mostly homogenous population of the winner sequence, from which we can obtain a novel aptamer.

Countless DNA and RNA aptamers have been raised against a variety of targets ranging from ions like  $Zn^{2+}$ <sup>46</sup> and  $Ni^{2+}$ <sup>47</sup>, small molecules like cocaine<sup>48</sup> and malachite green<sup>49</sup>, and even proteins like VEGF<sup>50</sup> and streptavidin<sup>51</sup>. Aptamers have even been

raised against fluorophores that emit fluorescence only when bound to the RNA such as *Spinach*<sup>52</sup>. The SELEX generated aptamers are invaluable tools for research<sup>20</sup>, but also these aptamers can be used as therapeutics<sup>53</sup>. Furthermore, it is possible to couple these SELEX-generated aptamers to selected riboswitches expression platforms to obtain an artificial chimeric riboswitch that is responsive to the target ligand for which the aptamer was generated<sup>54</sup>. However, there is only one example of such a chimeric riboswitch that uses a theophylline-binding aptamer<sup>54</sup>, several other SELEX-generated aptamers have failed to produce a functional riboswitch when coupled to a natural expression platform (Ceres and Garst, personal communication).

We believe that the failure in these previous approaches is the lack of structural requirements in the SELEX generated aptamers necessary to communicate the ligand-binding event to the expression platform and elicit a regulatory response. To overcome this limitation, Ely Porter in the Batey lab devised a modification of the SELEX technique that would target the binding pocket of a natural riboswitch aptamer, instead of starting with a completely randomized RNA. This way, the selected RNA would maintain the necessary structural requirements necessary for an appropriate riboswitch function, but would have altered ligand specificity. He decided to enable this approach using the purine riboswitch aptamer domain, particularly due to the wealth of biophysical and biochemical information available for this natural aptamer. Successfully, he changed the specificity of the *xpt* guanine-binding aptamer to bind 5-hydroxytryptophan (5-HT) with an affinity in the low micromolar range (Ely Porter, personal communication). Most importantly, the crystal structure of this 5-HT aptamer revealed that it keeps the global

structural motifs of the original guanine aptamer, which was the main intention of this selection.

While the use of the purine aptamer was a success in evolving a 5-HT aptamer, when coupled to natural expression platform the resulting riboswitch displays marginal *in vitro* activity. Moreover, no activity was observed when this riboswitch was expressed in *E. coli* and the growth medium was supplemented with 5-HT. There are several reasons that could explain this failure; first it could be a ligand issue. It is possible that 5-HT in the media does not enter the cell; either through a passive or active transport mechanism, or once inside the ligand gets degraded. Second, that throughout the selection the aptamer lost the functionality to communicate to with the expression platform, since the selective pressure for regulation was not included in this experiment. Third, and most likely, is that *although we possess an abundance of biochemical and biophysical data on the purine riboswitch, we still do not understand how the ligand binding to the aptamer domain of the riboswitch directs its in vivo regulatory activity.*

### **1.10 Taking a step back: Understanding the *in vivo* activity of the purine riboswitches in order to devise strategies for the development of novel riboswitches**

The approaches discussed in the previous sections use the purine riboswitch as starting point for the development of novel riboswitches. One of the studies used a rational and conservative approach by mutating three nucleotides in the binding pocket, and yielded *in vivo* functional riboswitches that respond to ammeline and azacytosine. The other utilized an *in vitro* selection of the aptamer alone, and the result was a 5-

hydroxytryptophan aptamer, but not a functional riboswitch. Neither of these studies provided the “golden ticket”: a platform that would allow the development of a riboswitch that reacts to any compound of interest and function robustly *in vivo*. I believe that we need to re-evaluate our understanding of purine riboswitches, and in order to do so, it is necessary to investigate the biochemical and biophysical properties of these RNAs in the context where they evolved to function: the intracellular environment.

When purine and other riboswitches were discovered, much effort was directed towards understanding the molecular mechanisms that mediate ligand binding. Rapidly, the crystal structures of the majority of classes of riboswitches were elucidated<sup>55</sup>. However, this reductionist approach for studying riboswitches in the context of only the aptamer domain failed to address the fundamental mechanisms of how ligand binding translates into a regulatory decision. In other words, the majority of riboswitch studies equate ligand binding with regulation. *The full activity of riboswitches can only be captured in the context of both the aptamer and the expression platform*

In addition to determining the crystal structure of the aptamer domain, a wealth of studies on riboswitches have focused in the folding dynamics of the RNA and how this correlates to ligand binding<sup>42</sup>. These studies have not only been fundamental in our understanding of riboswitch function, but have also advanced our general understanding of RNA folding. However, it is critically important to study riboswitches in the context of transcription, but unfortunately, the techniques to co-transcriptionally probe the structure of RNA are not well established. Moreover, inside the cell riboswitches fold co-transcriptionally, in the presence of chaperone proteins like Hfq that may aid in the folding of RNA<sup>56</sup>. This suggests that these studies fail to recapitulate the

RNA folding pathway that occurs *in vivo*. These facts led me to develop an *in vivo* reporter assay to explore and quantify the contributions of the most important structural aspects the purine riboswitch.

The experiments presented in Chapter 2 were designed with a very specific goal: understanding how the *in vitro* observations made for purine riboswitches correlate to the activity of this RNA in the intracellular context. To achieve this, first an *in vivo* reporter assay was developed to track and quantify the activity of the RNA inside the cell. With this system, a comprehensive mutagenic analysis of the most important structural aspects of the *pbuE* riboswitch was performed, and their contribution towards the *in vivo* activity was scored. The results of this work not only allowed us to correlate the activity of the riboswitch from *in vitro* to *in vivo* but also revealed novel insights into the regulatory mechanism of the RNA by exploring the expression platform of the RNA, an aspect of riboswitches commonly overlooked.

In Chapter 3, I will present preliminary results of a different approach to evolve novel riboswitches using the purine riboswitch as a scaffold. My technique combines ideas from the two approaches presented in the previous sections. First, it attempts to select novel riboswitches *in vivo* by coupling the activity of the purine riboswitch to the expression of the antibiotic resistance marker *tetR*. Then, the idea is to design a library around the binding pocket of the aptamer, and challenge the library with the target compound for which we would want to generate a riboswitch. Moreover, I will present how some of the conclusions from the study presented in Chapter 2 can be incorporated in the design of the selection platform. With this work, I aim to reconcile the biochemistry of purine riboswitches towards achieving the future goals of biotechnology.

## Chapter 2

### **Structure-guided Mutational Analysis of Gene Regulation by the *Bacillus subtilis* *pbuE* Adenine-responsive Riboswitch in a Cellular Context**

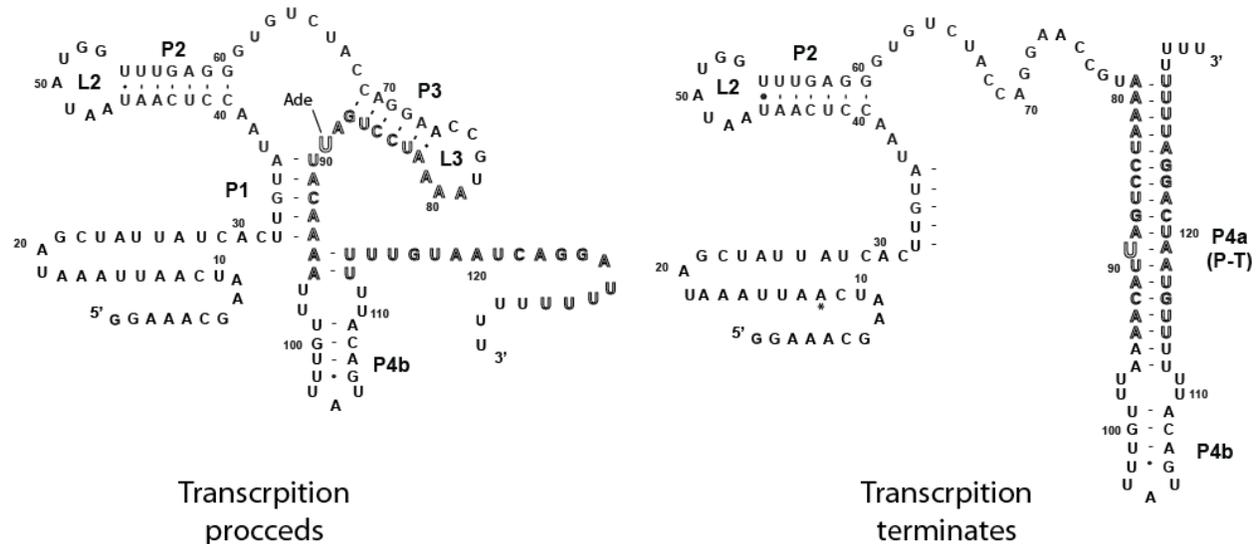
This chapter presents a collection of experiments aimed at understanding the structural basis that mediate the *in vivo* regulatory activity of the adenine-sensing *pbuE* riboswitch. To achieve this goal, a reporter assay was constructed that permitted a comprehensive mutagenesis analysis of the most important structural features of the RNA, and provided novel mechanistic insights of the riboswitch's activity. The results of this project are collected in a recent publication in the *Journal of Biological Chemistry* ISSN1083-351X. PubMed ID: 25550163

#### **2.1 Design and validation of an adenine riboswitch *in vivo* activity reporter assay**

The outcome of adenine binding to the *pbuE* riboswitch is the direct upregulation of transcription of the downstream gene, which in turn leads to increased protein levels (Fig. 2.1, A). In order to track the activity of the riboswitch *in vivo*, I constructed a transcriptional fusion of *pbuE* riboswitch from *Bacillus subtilis* to the Green Fluorescence Protein (GFP). This transcriptional fusion was cloned downstream of a strong transcriptional terminator, that prevents cryptic initiation of transcription from upstream elements, and inserted into a low copy plasmid (pBR322) (Fig. 2.1, B). The plasmid was introduced into *E. coli* (BW25113), the parental strain of the Keio knockout collection. These cells were grown in a chemically defined media, and the fluorescence of the cultures was measured when grown in the presence of different purine ligands.

The fluorescence values were then normalized to culture's cell density, and corrected for background fluorescence using cells carrying an empty plasmid (no GFP).

**A)**



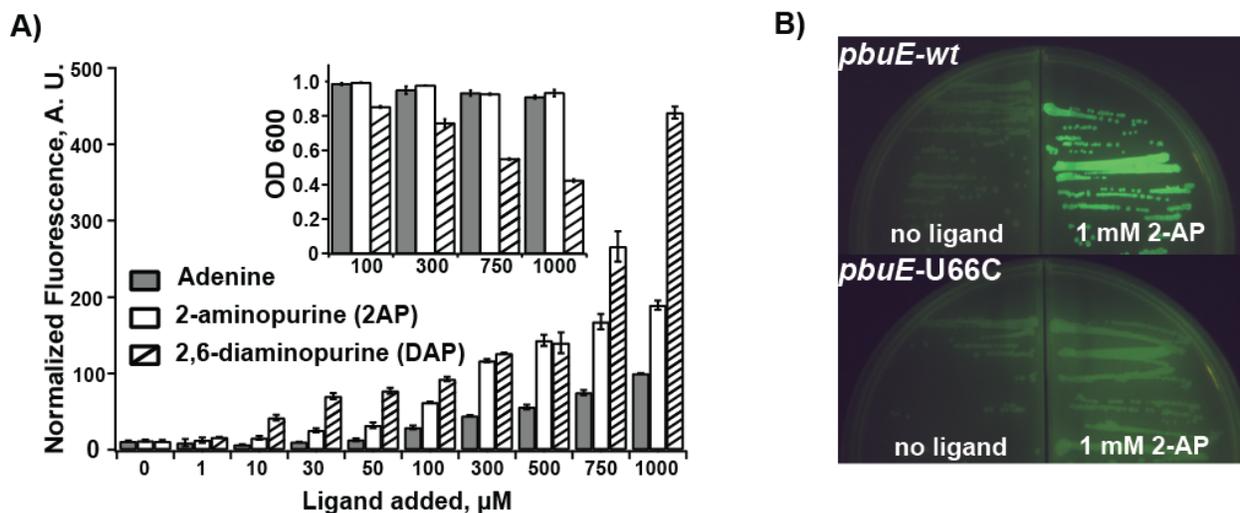
**B)**



**Fig 2.1 *In vivo* activity reporter system for the *pbuE* riboswitch.** **A)** Secondary structure of the *pbuE* riboswitch highlighting the two active conformations that leads to gene regulation. **B)** Diagram of the transcriptional fusion of the *pbuE* riboswitch to the gene fluorescence protein. When adenine activates the riboswitch, it directly affects the expression of the gene that leads to a fluorescent phenotype in the cells.

Adenine (Ade), 2-aminopurine (2-AP) and 2,6-diaminopurine (DAP) are purine ligands demonstrated to bind the adenine riboswitch, while guanine and hypoxanthine display minimal to no binding activity. Addition of adenine, 2-AP and DAP to the media

induces a robust increase of fluorescence, while no effect was observed by the addition of guanine (Fig. 2.2, A). The activation of fluorescence can be directly visualized in cells grown in solid media in the presence and absence of 2-aminopurine (Fig 2.2, B)



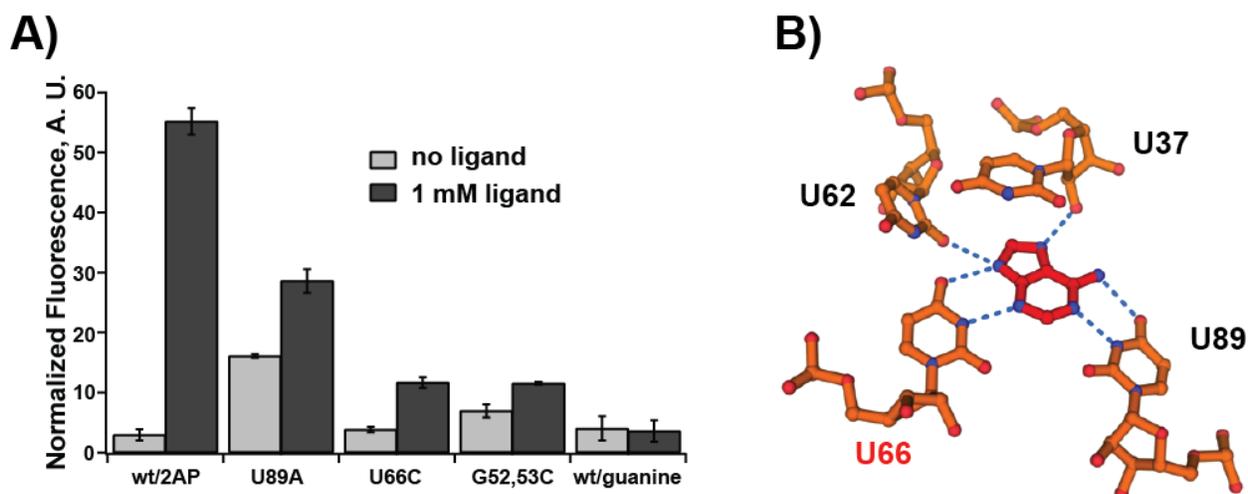
**Fig. 2.2 Ligand-dependent regulatory activity of the *pbuE* riboswitch in *E. coli*.** **A)** Titration of *E. coli* cultured in a rich defined medium with varying concentrations of adenine, 2AP and DAP. At each concentration, the bar height is the cell fluorescence corrected for background fluorescence from cells containing a null reporter vector (no *gfpuv*) and normalized to measured optical densities at 600 nm. The inset shows the measured optical densities for select ligand concentrations. *Error bars* in this figure represent the standard error of the mean (s.e.m.) of three separate experiments. **B)** *E. coli* cells harboring the pRR-*pbuE* wt plasmid were grown in solid CSB media in the presence and absence of 2-aminopurine (1 mM). A mutation to the binding pocket of *pbuE* (U66C) abrogates ligand binding; consequently 2-aminopurine fails to activate the expression of GFP. Imaging of the cells was by UV excitation of the cells at 395 nm and a 480 nm filter for the image collection.

Consistent with previous data that indicate a higher affinity for DAP to the purine aptamer; the induction of fluorescence for this ligand begins at lower concentrations (300  $\mu\text{M}$ ). However, DAP induce toxicity effects on the cells at concentrations higher than 500  $\mu\text{M}$  as shown by the cell density values (Fig. 2.2 A, inset). The lower levels of activation by the addition adenine, could be due to the utilization of the ligand by the

cells to support normal growth. These experiments made evident that 2-AP is the ideal ligand to be used as a proxy for regulation due to its high induction of fluorescence and lack of toxicity effects in concentrations up to 1 mM. Furthermore, 2-AP has been used in a several biophysical studies due to its physiochemical properties, such as emission of fluorescence. This fact will allow me to directly correlate these biophysical studies with *in vivo* experiments that use the same ligand to effect the riboswitch.

To ensure that the observed induction of fluorescence is due to the ligand binding to the riboswitch, we examined a series of point mutations in the in the riboswitch's aptamer. The global fold of the *pbuE* riboswitch is supported by a conserved tertiary interaction, which is strongly mediated by two G-C pairs (explained in detail in section 2.3). A double mutant of (G52C, G53C), which disrupt the L2-L3 interaction<sup>42</sup> fails to mediate ligand dependent fluorescence activation when 2-AP is added to the media (Fig. 2.3, A) In addition to the mutations in L2, we tested two mutants in the binding pocket of *pbuE*. Molecular recognition of the ligand in the binding pocket is supported by four nucleotides located in J2/1(U37), J2/3 (U62 , U66) and J3/1 (U89) (Fig. 2.3,B). Changing the uracil 66 to cytosine constitutes a very conservative mutation shown to maintain the architecture of the binding pocket but takes away the ability of the RNA to recognize the sugar face of the ligand reducing its affinity to undetectable levels<sup>35,39</sup> . In our reporter assay, this mutation is unable to activate fluorescence in a ligand-dependent manner, validating our approach (Fig 2.3, A). The second mutant U89A disrupts ligand recognition by the aptamer, but it also disrupts the terminator by introducing an A-A mismatch in the center of the stem (Fig. 2.1, A). Consistent with previous observation, this mutant displays elevated levels of

fluorescence in the absence of the ligand when compared to the wild type (Fig. 2.3, A)<sup>57</sup>. The results of this mutational analysis confirmed that we developed an assay capable of monitoring the activity of the riboswitch inside a cell. In the next sections, I will present and discuss a comprehensive analysis of different structural aspects of the riboswitch and their importance to the *in vivo* regulatory activity of *pbuE*.

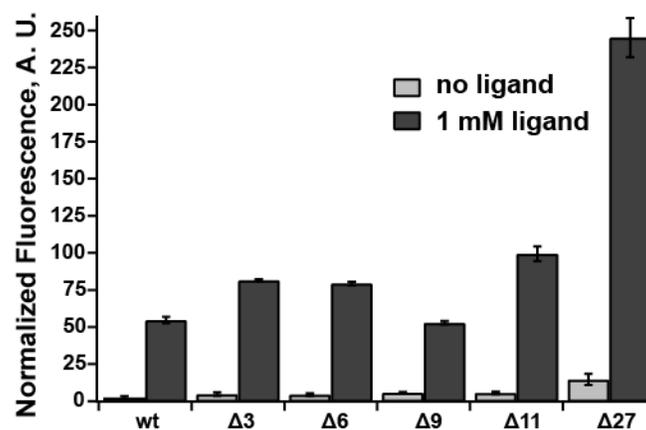


**Fig 2.3 Highly deleterious mutations disrupt the ligand-dependent regulatory activity of the *pbuE* riboswitch in *E. coli*.** **A)** Normalized fluorescence for wild type and select mutants that disrupt the ligand binding site (U89A and U66C) or the L2-L3 interaction (G53C, G54C) in the absence and presence of 2AP. As a further control, the effect of adding 1 mM guanine to the medium was also tested and showed no stimulation of growth. Below the graph is given the fold induction (FI) values. *Error bars* in this figure represent the standard error of the mean (s.e.m.) of three separate experiments. **B)** Nucleotides involved in the molecular recognition of the ligand. A point mutation of the U66 to a cytosine has been shown to disrupt ligand binding while maintaining the overall architecture of the binding pocket.

## 2.2 Effect of the leader sequence on the *pbuE* riboswitch

The original annotation of the *pbuE* riboswitch constituted an mRNA with a leader sequence of 155 nucleotides upstream of the starting AUG codon of the *pbuE* gene coding sequence. More recently, the transcriptional start of the mRNA was experimentally determined to include an additional 11 nt at the 5' side of the leader

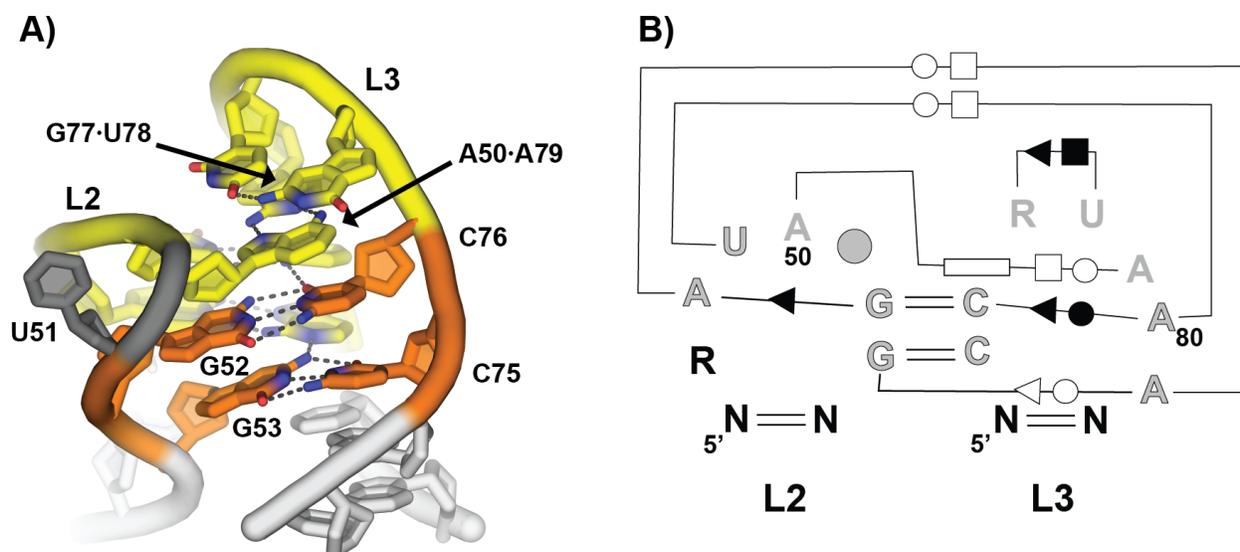
sequence<sup>58</sup>. To evaluate what potential effects this additional sequence has on the *in vivo* regulatory activity, I performed a systematic set of mutations that eroded the 5' side of the riboswitch up to 27 nucleotides, resulting in a minimal sequence that transcriptionally starts next to the 5' side of P1 on the aptamer domain (Fig 2.1, A). Figure 2.4 shows the results of the leader sequence mutants. No significant effects were observed dependent on length of the leader sequence from the 5' side. Interestingly, The  $\Delta 27$  deletion displays higher levels of fluorescence than the other mutants when stimulated 2AP; however, it also shows elevated levels of fluorescence in the absence of the ligand, resulting in induction values comparable to the wild type sequence. The most reproducible results were obtained for the  $\Delta 11$  mutant and the rest of the mutations tested *in vivo* and presented in this study are in the context of this 5' deletion sequence.



**Fig. 2.4 Effects of the leader sequence length of the *pbuE* riboswitch in the *in vivo* regulatory activity. C)** The effect of deleting nucleotides at the extreme 5'-end of the mRNA on regulatory activity in presence and absence of ligand. *Error bars* in this figure represent the standard error of the mean (s.e.m.) of three separate experiments.

### 2.3 *In vivo* sensitivity to perturbations of the tertiary interactions that support the global structure

Purine riboswitches possess a conserved tertiary interaction between the two distal loops P2 and P3. The energy contribution of this interaction has been calculated to be about ~2.9 kcal/mol for the *pbuE* riboswitch<sup>42</sup>. Although this loop-loop interaction is necessary to achieve high affinity ligand binding, it is not required for low affinity interactions between the ligand and the RNA. The overall architecture of this tertiary interaction, as exemplified by the *xpt* riboswitch, is a set of four non-canonical base pairs that scaffold two highly conserved Watson-Crick G-C pairs (Fig. 2.5, orange). The G53-C75 pair interacts with a reverse Watson-Crick/Hoogsteen A48-A81 pair in its minor groove, while the G52-C76 pair interacts with a reverse Watson-Crick/Hoogsteen U49-A80 pair. Distal to these two base quartets are two other pairs: an A50•A79 pair and a side-by-side G77•U78 pair (Fig. 2.5). These pairs make significantly weaker energetic contributions to ligand binding, as reflected in their lower degree of phylogenetic conservation<sup>39</sup>.



**Fig 2.5 Architecture of the conserved loop-loop interaction in purine riboswitches.** **A)** This tertiary interaction has been extensively studied with a myriad of biophysical techniques that have elucidated its role in maintaining the overall architecture of the purine riboswitch aptamer. In addition, this interaction supports the organization of the binding pocket and its disruption affects the affinity of the RNA for the ligand. Figure adapted from (Porter, Marcano, & Batey, 2014). PDB 1U8D **B,** Schematic of the loop-loop interaction highlighting base-base interactions and conservation patterns. Nucleotides in outline/gray are >97% conserved, in gray, >90% conserved and black, 75% conserved. The circle at position 51 denotes that there is a nucleotide present at this position in >97% of sequences. The symbols denoting base-base interaction are the Leontis and Westhof notation (46), reflecting Watson-Crick pairing (double lines) and non-canonical base-base interaction (single lines with symbol denoting the edge of the base used in the interaction: circle, Watson-Crick; square, Hoogsteen; triangle, sugar edge)

Several biophysical and biochemical techniques, such as smFRET<sup>59</sup>, NMR<sup>60</sup> and in line chemical probing<sup>61</sup>, have been employed to study and understand the nature of this highly stable interaction in guanine riboswitches. The collective knowledge gathered from these studies indicates that this interaction can form at low concentration of magnesium (1 mM). Even with no magnesium ions present, the L2-L3 interaction has been observed to form transiently, supporting that this interaction occurs *in vivo* and likely pre-organizes the binding pocket to enable rapid acquisition of a ligand competent

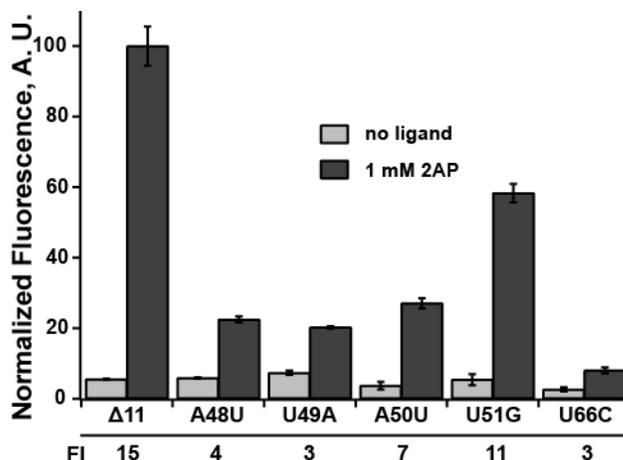
state<sup>42,62</sup>. Strikingly, the L2-L3 interaction is observed to form outside the context of the three-way junction or the P1 helix, further emphasizing its intrinsic stability<sup>60</sup>.

In order to obtain a deeper understanding of the role of the L2-L3 interaction, I re-evaluated its influence to ligand binding using a well-established 2-aminopurine (2-AP) fluorescence quenching assay<sup>42,63</sup>. Binding of 2-AP to the RNA results in a significant decrease in its fluorescence emission, that directly reports about the ligand-RNA interaction<sup>64</sup>. Our assays were performed at (37 °C, 50 mM HEPES pH 7.5, 135 mM KCl, 15 mM NaCl, 10 mM MgCl<sub>2</sub>); which more closely reflect the physiological ionic conditions inside the cell than previous experiments (25 °C, Tris-Cl pH 8.0, 100 mM KCl, 10 mM MgCl<sub>2</sub>)<sup>42</sup>. I determined the K<sub>d,app</sub> values of the wild type aptamer and four destabilizing mutants that resulted in agreement with previously reports for this riboswitch (Table 2.1)<sup>42</sup>. Note that I only introduced mutations in L2, as mutations on L3 would also interfere with the formation of the secondary structural switch (Fig. 2.1, A). Mutants U49A, A50U and U51G display a 2-3 fold reduction in affinity, while the A48U displays a drastic change of about ten fold. The behavior of the A48U mutant is consistent with its role as part of a critical base quartet, the central feature of the L2-L3 interaction (Fig 2.5)<sup>33,34</sup>.

RNA	Kd, nM	Kd, Rel
wt	300 ± 28	1.0
A48U	2700 ± 250	9.2
U49A	880 ± 90	3.0
A50U	720 ± 120	2.4
U51G	870 ± 90	2.9
G52C,G53C	N.D.	N.D.
U66C	200000 ± 186	660

**Table 2.1 Affinity of the L2 mutants determined by the 2-aminopurine fluorescence-quenching assay**

In a previous section of this chapter (section 2.1), I began to feature the influence of L2-L3 kissing loop motif with the double mutant of two universally conserved G-C pairs (G52-C75, G53-C76) (Fig. 2.2, A). This mutant completely lost its regulatory activity, in agreement with previous biochemical results on individual mutations of G52 and G53. At 1 mM 2AP, mutants A48U, U49A and A50U all exhibit about a five fold reduction in activity (observed fold induction of measured GFP fluorescence) as compared to the wild type riboswitch, despite significant differences in their affinities (Fig. 2.6). For all these mutants, the observed effect seems to arise with an inability of the riboswitch to anti-terminate transcription at high 2-AP concentrations, as show by the fluorescent levels in the absence of the ligand. The U51G mutant retains substantial regulatory activity, but it is reduced relative to wild type. This is consistent with its low phylogenetic conservation and its lack of a structural role as indicated by the crystal structure that shows U51G flipped out into solvent (Fig. 2.5, gray). The discrepancy between  $K_d$  affinity values and the *in vivo* regulatory activity for the conserved nucleotides in L2 may reflect their importance in the kinetics of structure acquisition of the aptamer. This limited dataset highlights the importance of examining the binding properties in a regulatory context rather than using the aptamer in isolation.



**Fig 2.6 The effect of point mutations in L2 on 2AP-dependent regulatory activity.** Below the graph is given the fold induction (FI) values. *Error bars* represent the s.e.m. of three separate experiments.

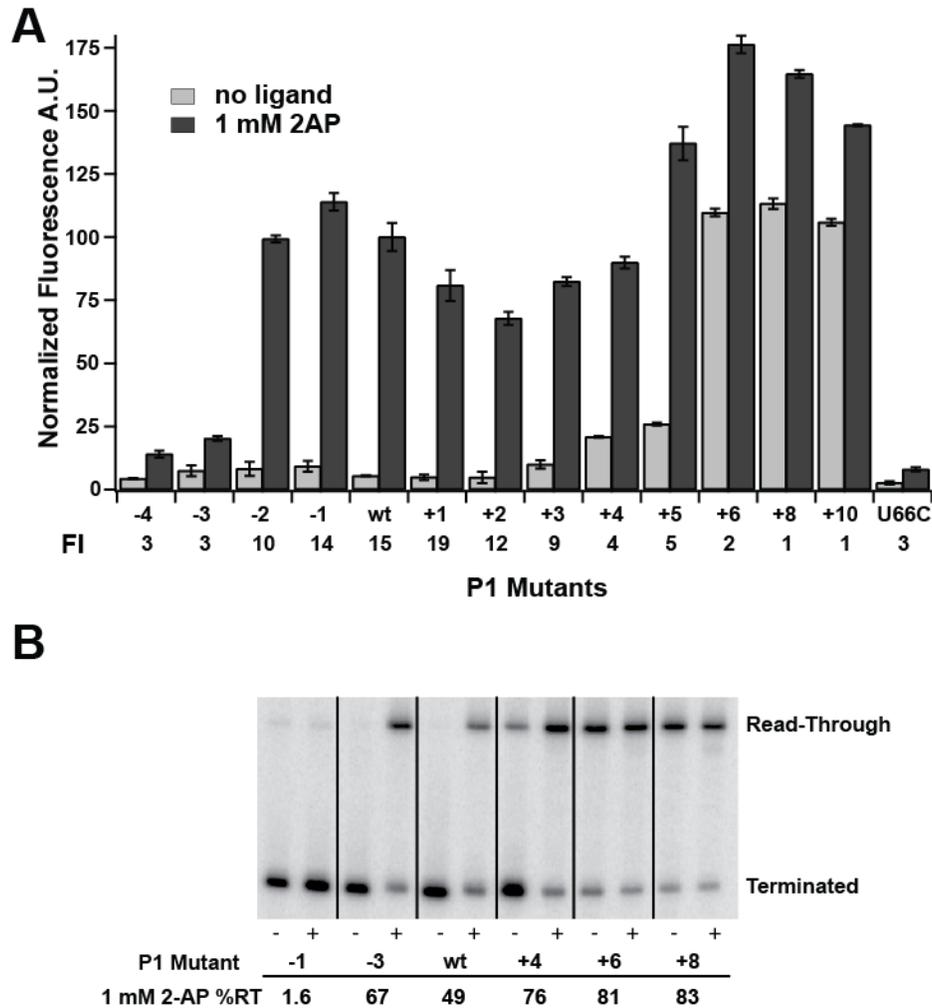
### 2.3 The stability of the P1 “communication sequence” and its effect on the regulatory activity

The *pbuE* riboswitch represents the simplest secondary structural switch that can be envisioned. It adopts two mutually exclusive secondary structural conformations, each with a specific function, that leads to either termination or anti-termination of transcription (Fig. 2.1, A)<sup>57</sup>. Transcription for all RNAs occurs from 5' to 3', and naturally, the nucleotides that form the aptamer are transcribed first. When the terminator sequence is transcribed, 16 nucleotides contained in the aptamer need to “switch” and become part of the more energetically stable expression platform. As this switching occurs, the first structural element from the aptamer to be abolished is the P1 helix and thus presents the first energetic barrier for the folding of the expression platform. Given its central position at the core of the riboswitch, it was hypothesized that the thermodynamic stability of the P1 helix is finely tuned to mediate ligand-dependent regulation.

To define the influence of the length and stability of the P1 helix on the regulatory activity, a systematic set of mutations were introduced at the 5' side of the P1 helix. These mutations were engineered to add or reduce the number Watson-Crick base pairs of P1 while maintaining the ability to form base pairs with P4a (Fig. 2.1, A). The ability of the P1 mutants to mediate ligand dependent gene expression was evaluated with the intracellular fluorescent reporter assay. Remarkably, I observed that a wide range of P1 lengths still support ligand dependent regulation (Fig. 2.7, A). The minimal length of P1 that is necessary to support ligand binding is three nucleotides, further erosion of the stem presumably ablates ligand binding<sup>65</sup>. Two of these base pairs, A36-U90 and U35-A91 (Fig. 2.1, A), are required to form ligand-dependent triplets with J2/3, while the third is weakly conserved as an R-Y pair with no obvious role in ligand binding<sup>33,34</sup>. This is consistent with the observations that a four base pair P1 helix could support near wild-type affinity while only two base pairs in P1 reduced affinity by ~50-fold<sup>65</sup>. However, it should be reminded that the affinity measurements do not completely correlate to the intracellular activity, suggesting that other parameters such as the folding kinetics may play an important role in regulation.

Stabilizing the P1 helix by sequential elongation of the stem shows that this element can be extended up to 10 base pairs before losing the ability to regulate gene expression upon the addition of 2-AP (Fig. 2.7, A). With the additional base pairs corresponding to P1+4 and P1+5 the expression of the GFP begins to increase in the absence of 2-AP, suggesting that the terminator element formation is slightly impaired. Further extension of P1 by one base pair (11nt total) strongly reduces the regulatory activity by 5-fold due to a significant increase in expression in the absence of the ligand.

All together, this data indicates that the terminator element of the *pbuE* riboswitch can effectively form over a broad range of P1 lengths, challenging an establish knowledge that the stability of the P1 helix directs the regulatory activity.



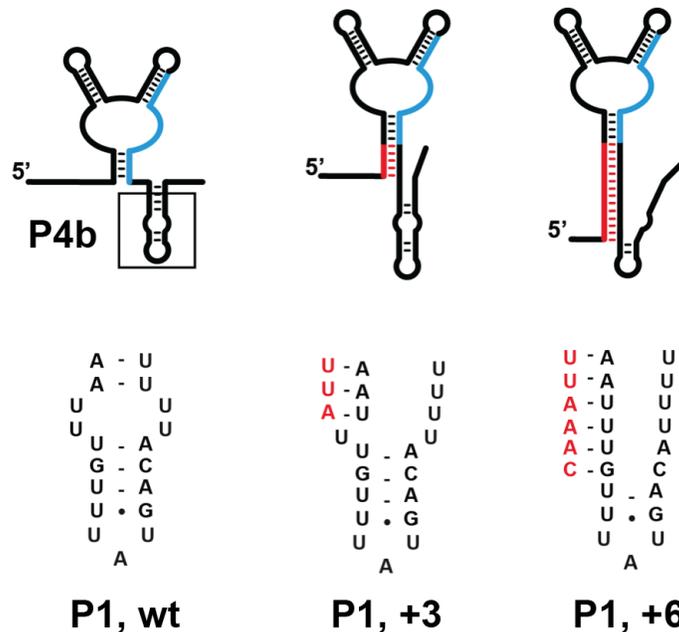
**Fig. 2.7 Effect of mutations that change the length of the P1 helix on regulatory activity.** **A)** The effect of point mutations on the 5'-side of P1 that either remove (-4 through -1) or add Watson-Crick pairs (+1 through +10) on 2AP-dependent regulatory activity. A binding-incompetent mutant (U66C) is shown for comparison. Below the graph the fold induction (FI) values are stated. **B)** Single turnover transcription assays of a subset of P1 length variants tested in panel A. *Error bars* represent the s.e.m. of three separate experiments.

The experiments described above indirectly measure the activity of the riboswitch by quantifying the fluorescence of GFP, which is regulated by the RNA. In addition, in the intracellular context there may be additional protein factors contributing to the switching mechanism of the RNA. Furthermore, changes in the expression or the mRNA stability induced by the engineered mutations are not tested by the GFP assay. In order to directly observe the intrinsic regulatory activity of the riboswitch, we examined the P1 mutants using an *in vitro* single turnover transcription assay. This assay consists of a minimal system containing *E. coli* RNA polymerase holoenzyme, 75  $\mu$ M of each NTP,  $\alpha$ -<sup>32</sup>-P labeled ATP, and a DNA template encoding the sequence of the riboswitch. After the transcription assay, products are separated by polyacrylamide gel electrophoresis and the band quantified in a phosphorimager.

The transcription assay revealed an almost identical pattern of regulation to that observed in the intracellular assay (Fig 2.7, B). Again, I observed that a P1 stem of 3 nucleotides can robustly regulate transcription when 1 mM 2-AP is added to the reaction; accordingly, a P1 length of ten nucleotides is the longest mutant that can terminate transcription in the absence of the ligand. If the data is examined closer, one can observe that the percent read-through (%RT) do not faithfully replicate the activation observed for some of the mutants in the intracellular assay. These subtle differences are to be expected, since I do not possess complete control of the cellular factors that can constantly vary through the experiment. However, most read-through percent values do align with the *in vivo* fluorescence, and the results clearly indicate that the regulatory functions of the riboswitch are solely encoded in the RNA.

## 2.4 A “nucleator” stem-loop is required for promoting efficient regulatory activity

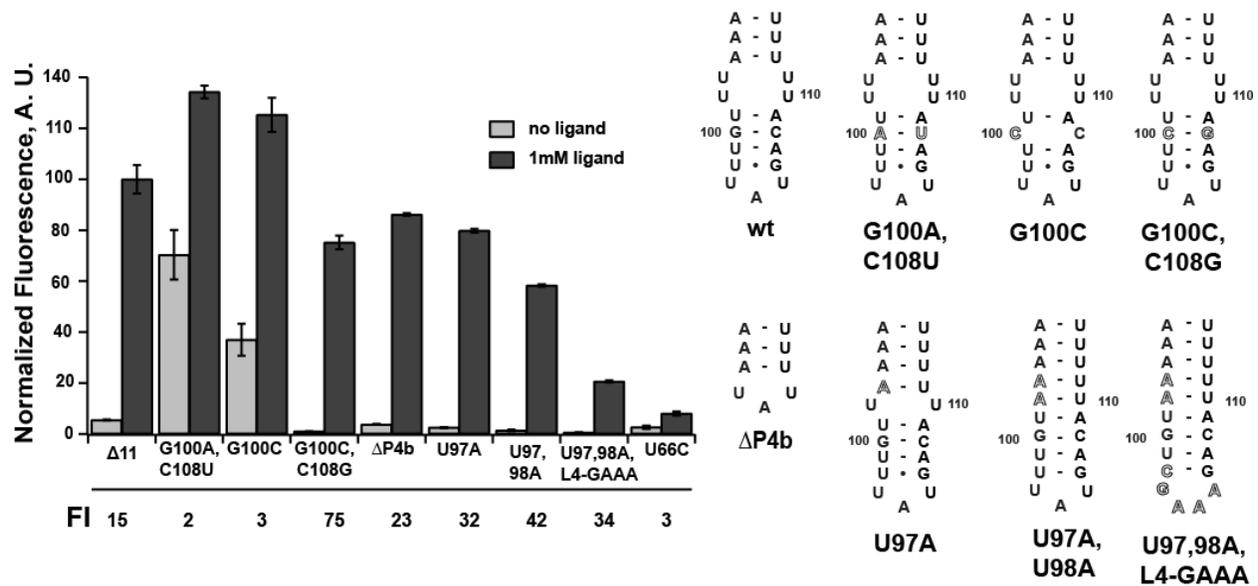
The mutagenesis analysis of the P1 helix directed me to ask questions regarding the functionality of the terminator element. I observed that an extended P1 helix length of 11nt (P1+6), added nucleotides on the 5' side that began to sequester nucleotides of the P4b stem-loop (Fig. 2.8). Particularly, the (P1+6) mutant enables the formation of a base pair that competes with the formation of the G100-C108 base pair in P4b (Fig 2.8). It is important to note that in the wild-type riboswitch, this element forms independently of the folding of the aptamer, so I proposed that it could serve as a nucleation element that promoted a strand invasion of the terminator helix into the aptamer domain<sup>57</sup>. A similar type of stem-loop capping motif exists in the lysine riboswitch *lysC* from *B. subtilis*, which has been proven to be necessary for efficient termination *in vitro*. These stem-loop capping motifs are not universally conserved among all expression platforms, but their common occurrence suggests that they may be an important feature for the regulatory activity of many riboswitches.



**FIG 2.8. Invasion of the P4b nucleator element by the extended P1 helix.** The added nucleotides in the 5' side of P1 engage in base pairs with nucleotides that also form part of the terminator helix P-4. In the wild type sequence the P4b stem is formed by nucleotides that never base pair with P1. The extension of P1 by 6 nucleotides begins to sequester the nucleotides of P4b such that now will have to compete with P1 for its formation

To investigate the role of P4b in the regulatory activity of the *pbuE* riboswitch, I tested a series of mutations predicted to stabilize or destabilized the stem-cap motif. Analogous to the P1 helix mutagenesis, these mutations were designed to minimize the formation of alternative structures in the P1/P4a switch. A mutation in P4b converting the G100-C108 base pair into an A-U pair is predicted to destabilize P4b by 1.9 kcal/mol, and the introduction of a C-C mismatch (G100C) is predicted to completely eliminate P4b. These mutations retain weak 2AP-dependent stimulation of reporter expression, but in the absence of ligand there is significantly elevated expression of GFP (Fig. 2.9). Introducing a compensatory mutation to G100C (G100C, C109G) rescues the riboswitch's activity by allowing the system to terminate in the absence of

the ligand. Interestingly, another mutation that eliminates P4b and the 2x2 internal loop separating P4a and P4b by deletion of nucleotides 97-108 ( $\Delta$ P4b; Fig. 2.9) shows wild-type levels of regulatory activity. In this RNA, a UAU tri-loop directly capping P4a is sufficient to promote regulatory activity (note that the terminal two base pairs of this stem do not participate in the regulatory switch). Together, these data reveal that while P4b is not essential for regulatory activity, it positively influences the efficiency of termination.



**Fig 2.9 Effects on the *in vivo* regulatory activity by mutations that alter terminator hairpin nucleator P4b.** The effect of point mutations in P4b or the internal loop between P4a and P4b to the 2AP-dependent regulatory activity. To the right is the secondary structure of each mutant tested. Below the graph is given the fold induction (FI) values. *Error bars* represent the s.e.m. of three separate experiments.

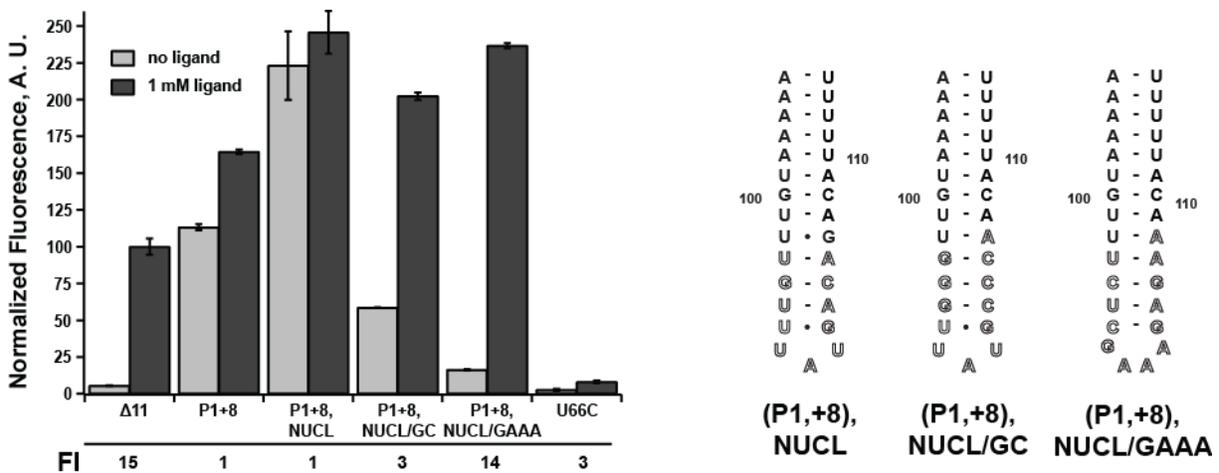
The stabilizing mutations on P4b display various levels of regulatory activity, as compared to wild type, by decreasing the expression of GFP in both the absence and presence of 2-aminopurine. Stabilization of P4b near the junction to P4a, by a single (U98A) or double (U98, U99A) mutation that closes the internal 2x2 bulge, both support

significant levels of 2AP-stimulated expression (Fig. 2.9). Additionally, these mutants display reduced expression levels in the absence of ligand comparable to that of the empty vector background. Further stabilization of P4b, in which the terminal loop is converted to an ultra-stable GAAA tetraloop<sup>66</sup> along with ablating the internal loop, (Fig. 2.9) yields a functional riboswitch with reduced level of GFP expression in the presence and absence of 2AP. Notably, all these mutant riboswitches possess greater fold induction levels (F.I.) as compared to wild type due to the augmented repression of fluorescence in the absence of the ligand.

This data made evident that the stability of the P4b influences the regulatory activity of the RNA. However, the observed decrease in read-through RNA could be attributed to the mutations interfering with the intrinsic pausing of RNAP. The *pbuE* riboswitch possesses a short poly uridine stretch (nucleotides 112-117) that has been hypothesized to serve as a pause for the RNA polymerase that gives extra time to the nascent RNA to properly fold during the transcription process<sup>58</sup>. However, none of our mutations directly alter this tract, and thus would not be expected to affect pausing. Therefore these data, especially those in the absence of 2AP, can be interpreted as P4b promoting efficient formation of the terminator element likely through rapid nucleation and subsequent propagation of the terminator stem via invasion of the P1 helix.

To further test this hypothesis, I constructed a set of terminator stem-loop variants that aimed to rescue the highly deleterious effect of the P1+8 mutant (13 base pairs in P1) that is predicted to disrupt formation of P4b (Fig. 2.10). To accomplish this, I extended P4b with additional base pairs and/or stabilizing tetraloops. Extension of P4b

with a weak stem-loop reflecting the sequence of the wild type P4b did not rescue the loss of termination at low 2AP concentrations ((P1+8), NUCL; Fig. 2.10). However, the addition of stem-loops that either strengthen base pairing, or add an ultra-stable C(GAAA)G tetraloop, restored efficient 2AP dependent regulatory activity ((P1+8), NUCL/GC and (P1+8), NUCL/GAAA; Fig. 2.10). It should be noted that the tetraloop itself is not expected to contribute to the rate of hairpin nucleation, but rather the stability of the closing G-C base pairs<sup>67</sup>. The observed rescue by the addition of G-C pairs in both the NUCL/G-C and NUCL/GAAA that is predicted to increase the rate of stem-loop nucleation, suggests that efficient nucleation of the terminator is critical for regulatory activity.



**FIG 2.10** The effect of extending the P4b helix to compensate for its partial disruption by the P1+8 mutation. *Error bars* represent the s.e.m. of three separate experiments.

It is interesting to note that the NUCL/GAAA variant has a 2AP-dependent regulatory response nearly identical to wild-type, with a P1 helix of 13 base pairs in length. This contrasts with a recent report suggesting that a *pbuE* riboswitch variant

with an elongated P1 helix (9 base pairs stabilized with four additional G-C pairs) and that stabilizing P1 has a negative impact on RNA folding and ligand affinity<sup>68</sup>. I observe that the concentration of 2AP in the growth media required to elicit a half-maximal regulatory response is also very similar between the two RNAs (~300  $\mu$ M for wild type and ~100  $\mu$ M for (P1+8), NUCL/GAAA) (data not shown), suggesting that P1 stability is not important for regulatory function. This further reinforces the idea that riboswitches may behave differently *in vitro* and in the context of transcription and the cellular environment.

## 2.5 Summary and discussion of results

The experiments presented in the previous sections provide new information regarding the biology of gene regulation in bacteria through the action of transcriptional riboswitches. The aptamer domain of purine riboswitches has been extensively studied, and our results were a necessary step to close the gap between the biophysical understanding of riboswitches and their activity at the whole organism level. In this study, I made an attempt to study riboswitches under conditions that match the speed of transcription, salt concentration and molecular crowdedness where riboswitch carry their activity. Still, my riboswitch reporter system is heterologous (riboswitch from *B. subtilis*, tested in *E. coli*), and it's expressed from a plasmid instead of being integrated in the genome. These conditions are different from the natural system in the aspects of copy number and promoter strength. For these reasons, we like to stress that our assay is not truly *in vivo*, but rather a molecular assessment of RNA activity in the cellular context. Nonetheless there is at least *in vitro* evidence indicating that riboswitches,

including *pbuE*, behave similarly when transcribed with *E. coli* or *B. subtilis* polymerase<sup>58,69</sup>.

We now have a comprehensive study of a riboswitch in the cellular context that provides critical information into the structural elements that mediate regulation. First, we have observed that *in vitro* ligand binding does not fully account for the patterns of phylogeny. In the L2-L3 interaction there are highly conserved nucleotides, but contribute modestly to the ligand binding activity. These nucleotides proved to be essential for a robust regulatory response inside the cell. It is likely that that these nucleotides are necessary for rapid structure acquisition or to promote a highly folded population in the *apo* state. However, our cell-based assay cannot discriminate between these alternatives, as they would both manifest as a low fluorescence activation of the switch as seen in (Fig. 2.6). Overall, contrary to previous studies that indicated that the identities of nucleotides at L2 and L3 were not important for the *in vitro* binding activity<sup>39,42</sup>, their contributions to the intracellular activity better recapitulate their evolutionary conservation when tested in my intracellular assay.

One of the most interesting results of this study is the tolerance of the *pbuE* riboswitch to wide range of P1 helix lengths. In the lower end, three nucleotides (P1-2) are sufficient to support a regulatory response. At the other end, ten nucleotides yield a functional riboswitch with a small increase in fluorescence in the absence of ligand. This indicates that the terminator does not form in the majority of the transcription events despite being, in the P1+6 mutant, significantly more thermodynamically stable (24.6 kcal/mol) than the aptamer (13.3 kcal/mol).

These observations are best understood under a co-transcriptional folding model, where secondary structures fold sequentially but exchange rapidly through branch migration<sup>70</sup>. In this mechanism the P1 helix folds first, but is quickly disrupted by the 5'-side of the terminator helix. In the riboswitches with the longest functional P1 helices, increased background expression in the absence of 2AP is observed, likely due to the terminator not being able to fully displace P1 and P3 before the polymerase can clear the termination poly-U tract. In the majority of the transcription events, regardless of the P1 length, the terminator invades the less stable P1 helix until it reaches base pairs adjacent to the junction. Further branch migration is stalled by ligand-dependent structure around P1 and J3/1. In particular, ligand-induced formation of base triples between J2/3 (U65 and C66) and the two Watson-Crick pairs proximal to the junction (U35-A91 and A36-U90, respectively)<sup>33,34</sup> likely function as the primary barrier to strand exchange in the bound state, consistent with the results of Frieda and Block that investigated the folding of the *pbuE* riboswitch under *in vitro* co-transcriptional conditions<sup>71</sup>.

While the *pbuE* riboswitch is able to accommodate a spectrum of P1 lengths, there is a clear upper boundary (P1+6; Fig. 2.8, A). One interpretation of this model is that the terminator cannot invade through a P1 helix of greater than ten base pairs in the timescale of transcription. In the model I propose, the activity loss is due to disruption of a small stem-loop (P4b; Fig. 2.9) that forms independently of the occupancy status of the aptamer domain and likely serves to promote efficient terminator formation. Mutations that destabilize this element negatively impact transcriptional termination in the absence of ligand, and one of these mutations

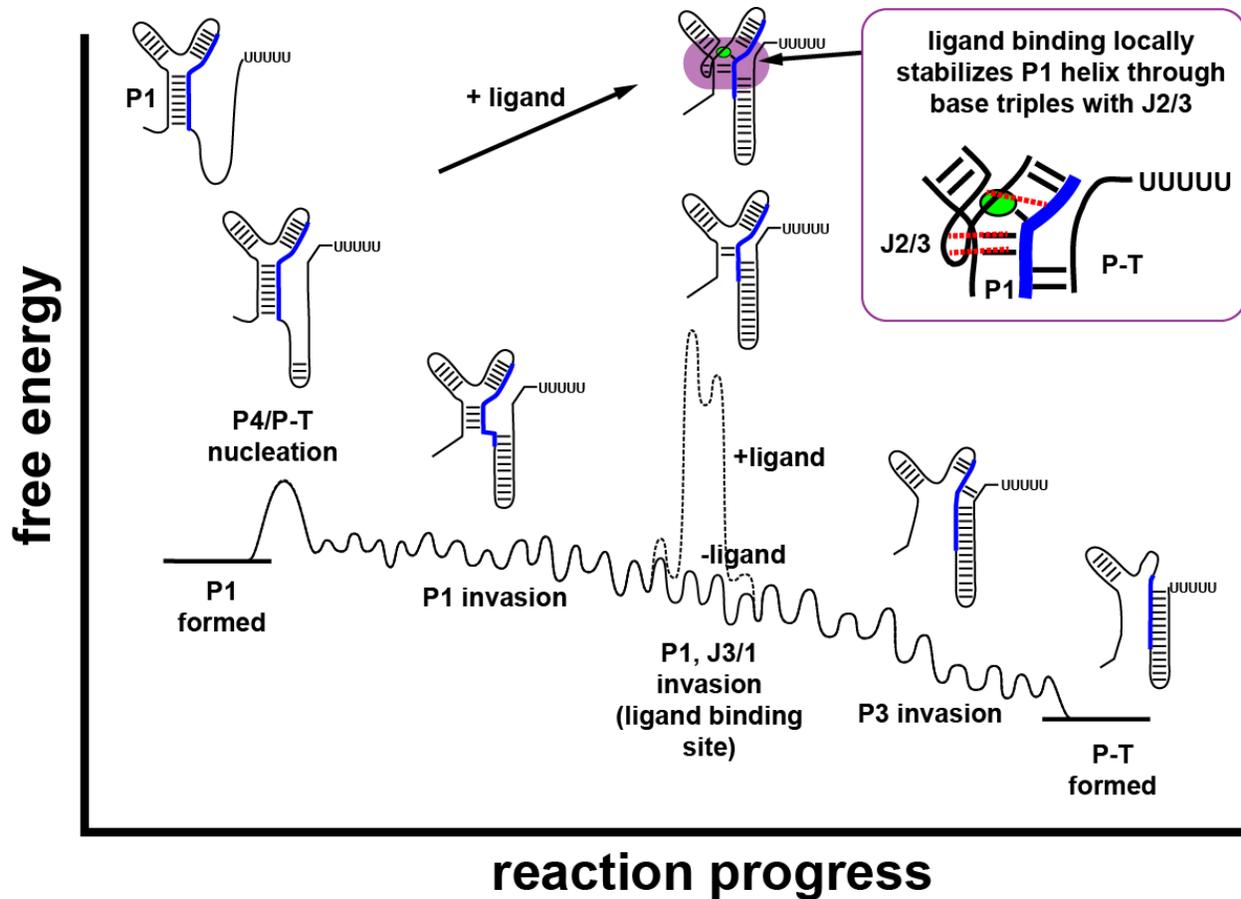
(G100C), can be rescued to wild type levels through a compensatory C108G mutation. Conversely, mutations that stabilize P4b substantially reduce the reporter expression regardless of the presence of 2-AP. These variants exhibit the largest fold induction, primarily because the amount of reporter expression in the absence of 2AP approaches that of the negative control. As direct support that the length of P1 alone is not the cause of loss of regulatory activity, I was able to rescue the regulatory activity of P1+8 (13 base pairs) by extending the length of the terminator helix and restoring a nucleation site disrupted by the extended P1. Thus, I proposed that another key determinant in the regulatory activity of the *pbuE* riboswitch is the rate of terminator nucleation, which is related to sequences that are not part of the structural switch. For many riboswitches, the rate of nucleation of the alternative helix to P1 may be a way of tuning the dynamic range to meet the needs of the transcriptional unit<sup>72</sup>. Unfortunately, very little quantitative information is available regarding how stem-loop sequences and their structure determine the rate at which it forms<sup>67</sup>

It is worth noting that the *pbuE* riboswitch relies upon the ligand directly blocking branch migration step necessary for the functional terminator element. However, there are a number of riboswitches for which this is not the case. In these RNAs, the minimal sequence required for high-affinity ligand binding is distinct from sequences that directly participate in the secondary structural switch—that is, in these RNAs the helix alternative to P1 does not invade into the aptamer domain. These expression platforms and associated regulatory switches can host a variety of distinct aptamers and retain efficient ligand-dependent regulatory activity<sup>54,73</sup>. An interesting difference is that these riboswitches (*e.g.*, the *B. subtilis metE* SAM-I riboswitch) can be exquisitely

sensitive to P1 length, and that an efficient regulatory activity is disrupted by the addition or deletion of a single P1 base pair<sup>54</sup>. How those observations relate to the observations of this study is not clear and merits further investigation. It is likely that elements of the above model relate to these riboswitches, but that the details of how ligand binding prevents formation of the competing helix may differ.

## **2.6 Mechanistic model of the *pbuE* riboswitch**

The results discussed in the previous section present a novel insight into the regulatory mechanism of the *pbuE* riboswitch. These new perspectives are likely representative of numerous riboswitches that control transcription or translation. Furthermore, the results highlight the importance of investigating riboswitches, and RNAs in general, in the context of the cell. To finalize this chapter, I would like to present a summary of what I have learned of the mechanism of gene regulation by the *pbuE* riboswitch (Fig. 2.11).



**Fig. 2.11 Energy Landscape of the co-transcriptional folding of the *pbuE* riboswitch.** The left axis represents the free energy of the folding of the riboswitch. The pathway follows a steadily decreasing rugged landscape culminating in the folding of the expression platform that is the most energetically stable conformation.

When the leader sequence of the *pbuE* gene in *B. subtilis* is transcribed, the sequence of the aptamer is transcribed first, as it sits on the 5' end. As the RNA is synthesized, it rapidly acquires the conformation of the aptamer and the P4 nucleator. At this stage of the process, the aptamer is capable of binding the ligand, but it is likely that a process of invasion of the aptamer start to occur as soon the nucleator forms. I hypothesize that the terminator invasion represents a net energetically favorable process, as the terminator is more thermodynamically stable. However, if the ligand is

present, now the energy required for the switching process is substantially higher since now the terminator needs to disrupt additional direct interactions to P1 that do not occur in the absence of the ligand. Presumably, this presents a kinetic barrier that stalls the terminator invasion and allows the polymerase to clear the expression platform before the structural transcriptional termination signal is able to form. Notably, these two scenarios eventually lead to the same thermodynamic state, in which the terminator helix dominates the folding equilibrium. I presume that once the regulatory decision is made, the ligand will eventually dissociate from the aptamer and the terminator will form to dominate the final folding state.

## 2.7 Experimental procedures

**Generation of riboswitches in a reporter vector** — Each riboswitch was cloned into a parental reporter vector upstream of to regulate its expression in a ligand-dependent fashion derived from a low copy pBR327 plasmid<sup>54</sup>. The parental vector contains a moderately strong synthetic insulated transcriptional promoter “proD”<sup>74</sup> downstream of the strong *rrnB* terminator to limit transcriptional read through from upstream genes. Construction of each plasmid vector was accomplished using standard molecular cloning methods and sequence verified. Table 1 gives the full and annotated sequence for each riboswitch starting from the transcriptional initiation site (+1) to the translation initiation codon.

**Cell-based fluorescence assays** — All experiments were performed using the *E. coli* cell strain K-12 BW25113 (parental strain of the Keio knockout collection<sup>75</sup>). Cells were transformed with a reporter plasmid containing wild type *pbuE* riboswitch, or

mutant variants thereof, using standard molecular biological protocols. Transformants were grown on Luria Broth plates containing 100 µg/mL carbenicillin for selection for the *bla* resistance marker. Three individual colonies were picked and used to inoculate individual cultures in 5 mL of CSB growth medium (a rich chemically defined medium<sup>54</sup>) containing 100 µg/mL ampicillin and incubated overnight at 37 °C under constant agitation. Saturated cultures were used to inoculate fresh CSB media as a 1/100 dilution to a final optical density at 600 nm wavelength light ( $OD_{600}$ ) of 0.05. At this stage, effector ligand was added to the medium at a defined concentration and the cells incubated at 37 °C for six hours, at which time the cells are in late-log phase growth.

Ligand-induced expression was determined by measuring the  $OD_{600}$  and fluorescence intensity of the culture using a plate reader Infinite M200 Pro (Tecan). For all fluorescence measurements the cells were excited at 395 nm and fluorescence read at 510 nm. The raw fluorescence values were divided by the  $OD_{600}$  to calculate the normalized fluorescence per cell density. To this value, the background fluorescence of cells was subtracted calculating the normalized fluorescence of cells carrying the parental pBR327 vector that does not contain the *gfpuv* reporter and were subjected to the same conditions.

***In vitro transcription assays*** — DNA transcription templates were amplified by PCR from the appropriate plasmid constructed to test activity of the riboswitch *in vivo* (see above). The two primers used were 5'-GTGGTTGCTGGATAACTTTACGGGC and 5'-CCCGGGGATCCTCTAGAGTCGAC at 1 µM concentration in a standard PCR amplification reaction. These DNA templates were transcribed as described previously<sup>40,54</sup>. Polymerase was bound to the promoter in a first reaction step by

incubating 50 ng of DNA template at 37 °C for 10 min in 12.5 µL of 2X transcription buffer (140 mM Tris-Cl pH 8.0, 140 mM NaCl, 0.2 mM EDTA pH 8.0, 28 mM β-mercaptoethanol, 70 µg/mL BSA), 2.5 µL of 25 mM MgCl<sub>2</sub>, 0.5 µCi of α-[<sup>32</sup>P]-ATP, and 0.25 units of *E. coli* RNA polymerase σ<sup>70</sup> holoenzyme (Epicentre). Transcription was initiated by adding 7.5 µL of NTP mix (75 µM final concentration of each NTP, 0.2 mg/mL heparin and 2-aminopurine (2AP) at 3.3 mM) and incubated for an additional 20 min at 37 °C. Transcription was quenched by adding 25 µL of RNA loading buffer (95% (v/v) deionized formamide, 10 mM EDTA pH 8.0, 0.25% (w/v) bromophenol blue, 0.25% xylene cyanol) and incubation at 65 °C for three minutes. The RNA products were electrophoretically separated on a denaturing 29:1 (acrylamide: bisacrylamide) gel. Quantification of the intensities for the bands was performed with Image J (<http://imagej.nih.gov/ij/>).

*2AP fluorescence binding assay* — RNA aptamers encompassing the nucleotides from 32 to 94 from the *pbuE* riboswitch with two guanine residues appended to the 5'-end for efficient transcription were generated by T7 *in vitro* transcription and purified by denaturing PAGE. 2AP binding assays were performed in 50 mM HEPES pH 7.5, 135 mM KCl, 15 mM NaCl, 10 mM MgCl<sub>2</sub> and 50 nM 2AP with a protocol used previously to monitor 2AP binding to the purine riboswitch<sup>76</sup>. Increasing concentrations of RNA were added to the reactions and incubated for 10 minutes prior taking fluorescent measurements in a plate reader Infinite M200 Pro (Tecan). Excitation of 2AP was performed at 300 nm and emission was collected from 360 – 370 nm. The change in fluorescence produced by RNA binding to 2AP was fit to the single site binding equation

$$F = F_{\min} + \Delta F \left( \frac{[RNA]}{[RNA] + K_{D,app}} \right)$$

where  $F$  is the observed fluorescence,  $\Delta F$  is fluorescence change,  $F_{\min}$  is the fluorescence in the absence of RNA and  $K_{D,app}$  is the apparent dissociation constant.

The error in the fit values represents the standard error of the mean for three independent titrations.

## Chapter 3

### Development of novel RNA devices using the purine riboswitch as a scaffold for directed evolution

The discipline of synthetic biology already implements RNA devices as tools for the engineering of biological technologies. This chapter attempts to devise a platform for the development of novel RNA control devices that respond to any compound of interest.

#### 3.1 *In vivo* evolution of novel riboswitches using tetracycline as a selection marker

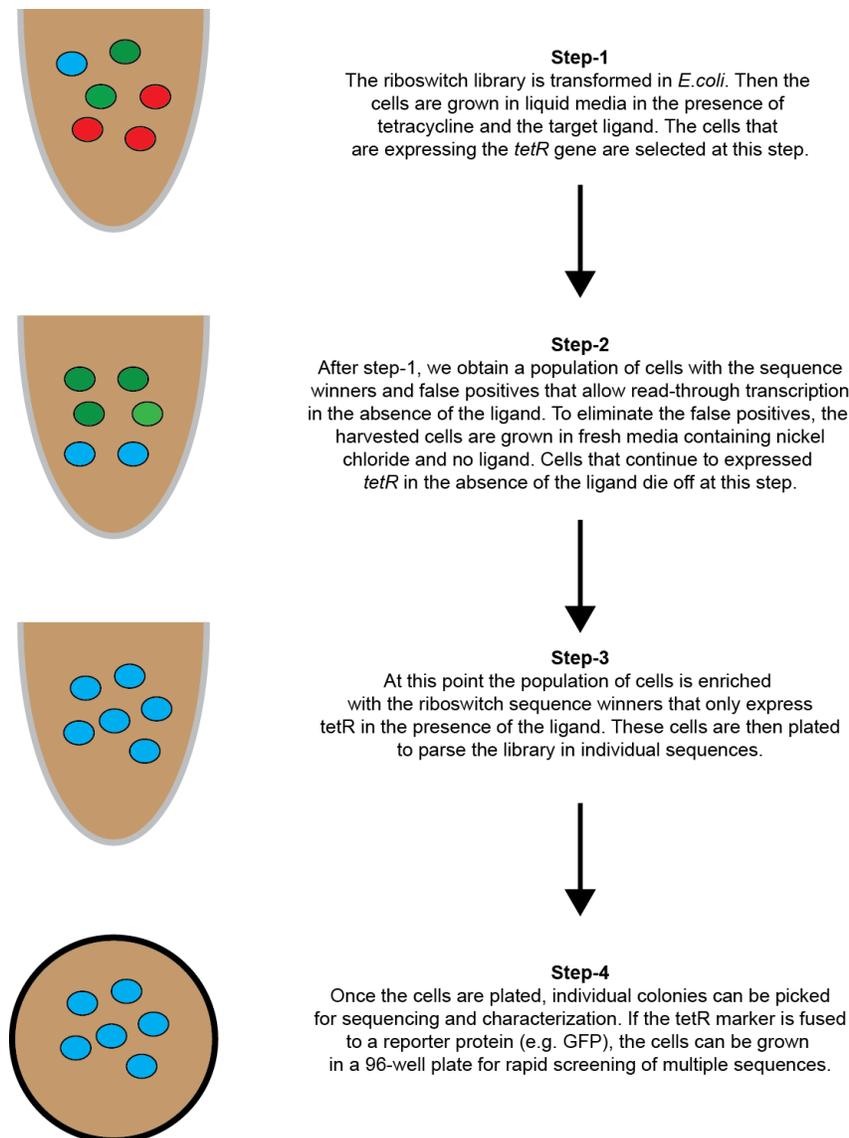
As mentioned in the introductory chapter of this thesis, one of the advantages of using RNA in genetically encoded devices is that it is relatively easy to evolve new activities or functions using *in vitro* approaches such as SELEX. However, the majority of the *in vitro* evolved RNAs have poor to no activity in a cellular context. I reasoned that if newly evolved riboswitches were to work in *E. coli* cells, it would be best to carry out the evolution experiment in the cellular context. The challenge then becomes choosing an ideal riboswitch system that robustly controls the expression of a selection marker.

Recently, the Yokobayashi research group devised a strategy that uses the tetracycline resistant gene (*tetR*) to select for riboswitches responsive to thiamine in *E. coli*<sup>77-79</sup>. The advantage of using *tetR* as a selection marker is that positive and negative selections can be achieved using a single marker, a requirement for selection experiments to eliminate false positives. The product of the *tetR* gene is an H<sup>+</sup> antiporter protein that pumps out the tetracycline from the cells making them resistant to the

antibiotic<sup>80</sup>. Moreover, this antibiotic pump confers the cells metal sensitivity due to the inner membrane alterations that occur when the protein is inserted in the membrane, such that cells overexpressing the *tetR* gene will not proliferate when nickel or cadmium salts are added to the media<sup>80,81</sup>. The *tetR* gene in the Yokobayashi reporter was fused to the GFP protein through a flexible peptide linker. This fusion allowed the researchers to follow and quantify the selection steps by means of fluorescence<sup>77</sup>

In the Yokobayashi report, they cloned a natural thiamine aptamer upstream of a fifteen randomized nucleotide sequence. Downstream of the randomized sequence they inserted a *Shine-Dalgarno* (SD) sequence and a translation initiation codon AUG. After one round of *in vivo* positive and negative selection, they obtained a “communication sequence” capable of relaying the binding event into a regulation of gene expression at the translational level<sup>79</sup>.

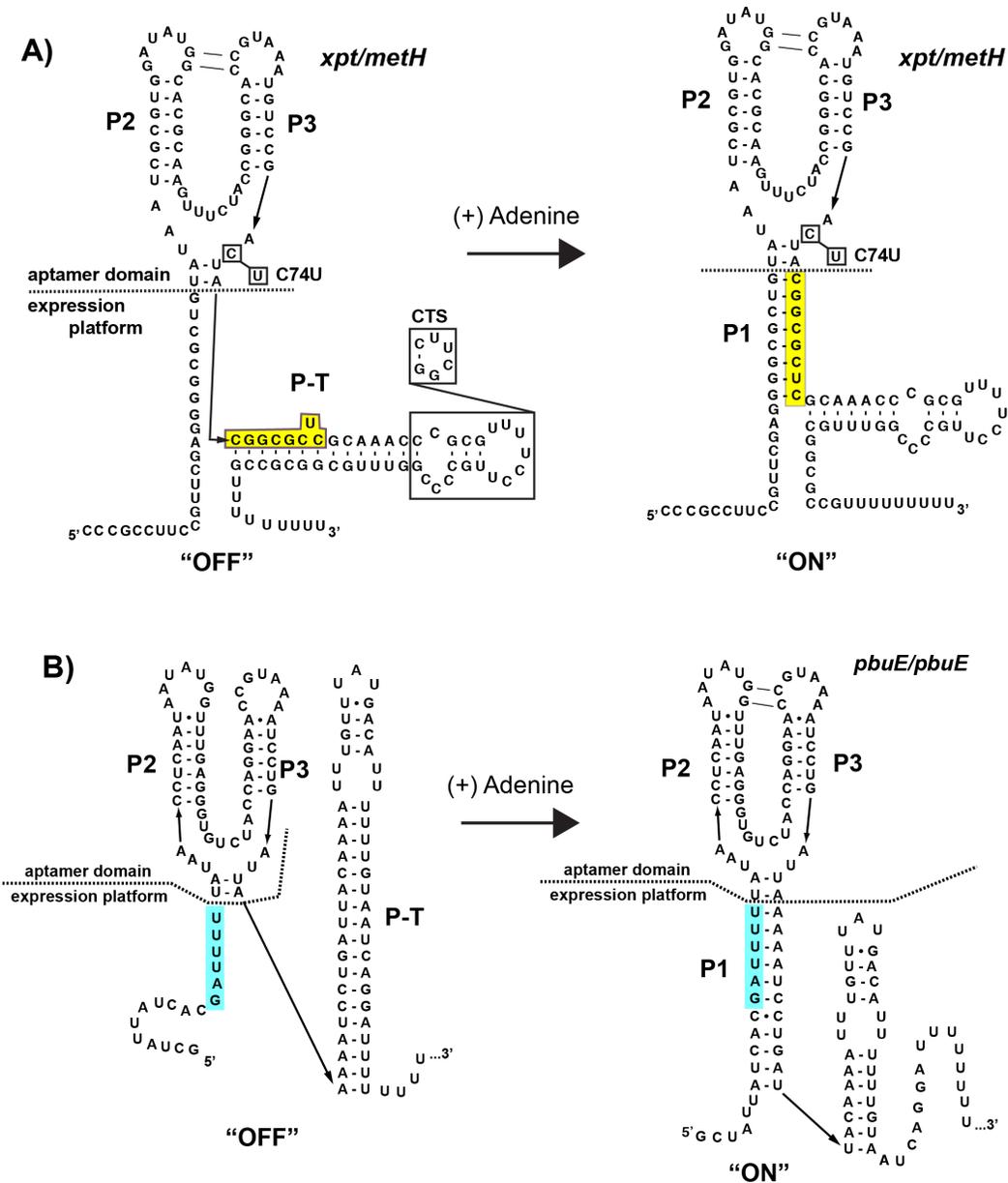
Building on the Yokobayashi group’s work, I hypothesized that it could be possible to generate a novel riboswitch that responds to an orthogonal ligand using a platform coupled to the *tetR* system. In this strategy, I would design a minimal library around the binding pocket of a riboswitch and clone it upstream of *tetR*. This library would be transformed into *E. coli* cells and subjected to various rounds of positive and negative selection in liquid media. At the end of each round, surviving cells will be plated and individual clones will be screened (Fig. 3.1). Initially, the clones will be screened for their ability to display ligand-dependent growth, but also if GFP is included, I can perform a quantitative assessment of the extent of regulation. Then, the individual clones will be sequenced to obtain the evolved “winner” sequence.



**Fig. 3.1 *In vivo* selection with the tetracycline dual selector system.** The power of this system relies in the ability of the tetracycline resistance (*tetR*) gene to be used simultaneously as a positive and negative selector. Cells displaying ligand-dependent expression of *tetR* are represented in blue. The false positives, or cells expressing *tetR* in the absence of the ligand are shown in green. The cells that never express *tetR* are represented in red.

### 3.2 Chimeric Purine riboswitch as platforms for *in vivo* directed evolution experiments

In order to preliminarily assess if this approach is feasible, I constructed a selection plasmid that employs the *tetR* and *gfpuv* genes. Similar to the Yokobayashi work, my system also contains the GFP protein but it is transcriptionally, instead of translationally, fused and contains its own start codon and ribosomal binding site. I inserted this construct in the pRR vector backbone presented in Chapter 2, which is a low copy plasmid derived from pBR322, and called it pRS (riboswitch selection). To direct the expression of this polycistronic mRNA to be used for the evolution experiments, I chose two artificial purine riboswitches described by Ceres *et al*<sup>54,73</sup>. This work presents a strategy to generate artificial transcriptional “on” riboswitches by combining different natural aptamers to different expression platforms. The first one, the *xpt/meth* “chimera” contains the *xpt* (C74U, mutant) adenine-binding aptamer connected to the S-adenosylhomocysteine-binding *meth* riboswitch expression platform, a combination that produces an adenine sensing transcriptional “on” switch (Fig. 3.2, A). The second one, is a modified version of the *pbuE* riboswitch named *pbue/pbuE\**, in which a sequence was engineered in the 5' side of the P1 helix that decouples the aptamer from the expression platform (Fig. 3.2, B). My rationale for using these modular riboswitches is based on an engineering perspective suggesting that whatever new RNA “part” we evolve can be transplanted to another context and still be active. In other words, I did not want to generate a new part whose function is dependent on another module adjacent in the RNA.



**Fig 3.2 Chimeric riboswitches used for the *in vivo* selection.** **A)** *xpt(C74U)/methH* chimeric riboswitch. The splice sites between the aptamer and the expression platform are indicated in the secondary structural map of the chimeric riboswitch. The nucleotides highlighted in yellow participate in the structural switch. The expression platform of the *methH* riboswitch contains an engineered mutation in the transcriptional termination (CTS) that stabilizes this structural element. **B)** The *pbuE/pbuE\** modular riboswitch. The nucleotides highlighted in blue were engineered in the 5' side of the P1 helix in order to structurally separate the aptamer and the expression platform.

Both chimeric riboswitches used in this study contain an aptamer that binds adenine but also 2-aminopurine, which is a better effector ligand for *in vivo* experiments as described in Chapter 2 of this thesis. Both constructs were cloned in pRS plasmid as a transcriptional fusion of the *tetR-gfpv* mRNA, such that it will regulate the expression of the gene fusion. These plasmids were transformed into Top10 *E. coli* cells and were subjected to growth assays in the presence and absence of 2-aminopurine.

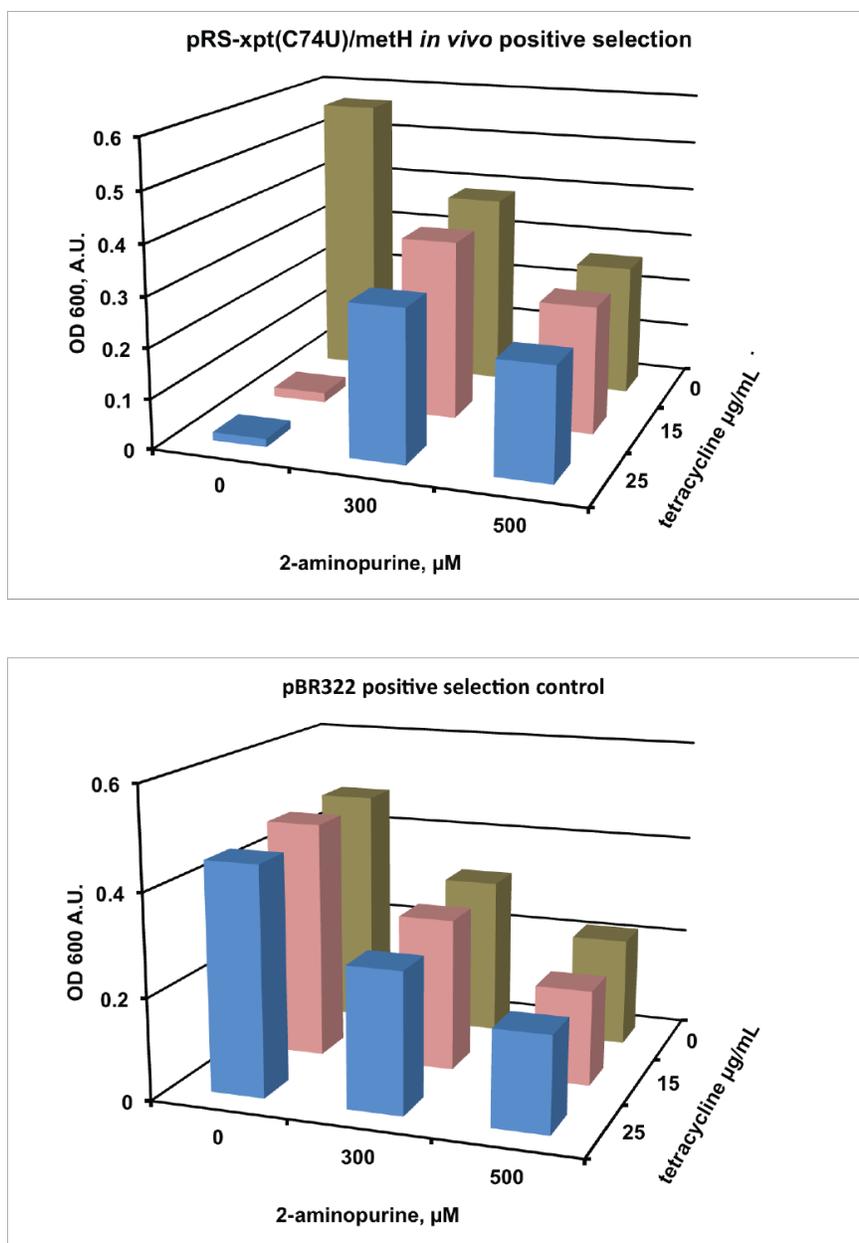
Initially, this system did not performed as expected. The cells would either proliferate when challenged with tetracycline or not at all, independently of the presence of 2-AP. This highlights a recurrent pitfall that synthetic biology attempts to address regarding the modularity of genetic parts. While the Yokobayashi report described a robust platform for *E. coli* selection, when I used the same genes in a similar but not equal context the system did not perform.

### **3.3 Optimization of selection platform and preliminary results on selection experiments**

To obtain a functional platform, I focused in optimizing two parameters: the strength of the promoter that drives the expression of *tetR-gfpuv* gene and the growth conditions of the experiments. A report from the Sauer group presented a collection of promoters for which their relative strength is quantified relative to that of the *tac* promoter utilizing a standard genetic context<sup>74</sup>. This set of promoters was also used to optimize the *in vivo* reporter system presented in Chapter 2. Several of these promoters were cloned in the pRS plasmid and the best results were obtained with the “*proD*” promoter, one of the strongest promoters in the collection. As for the media conditions,

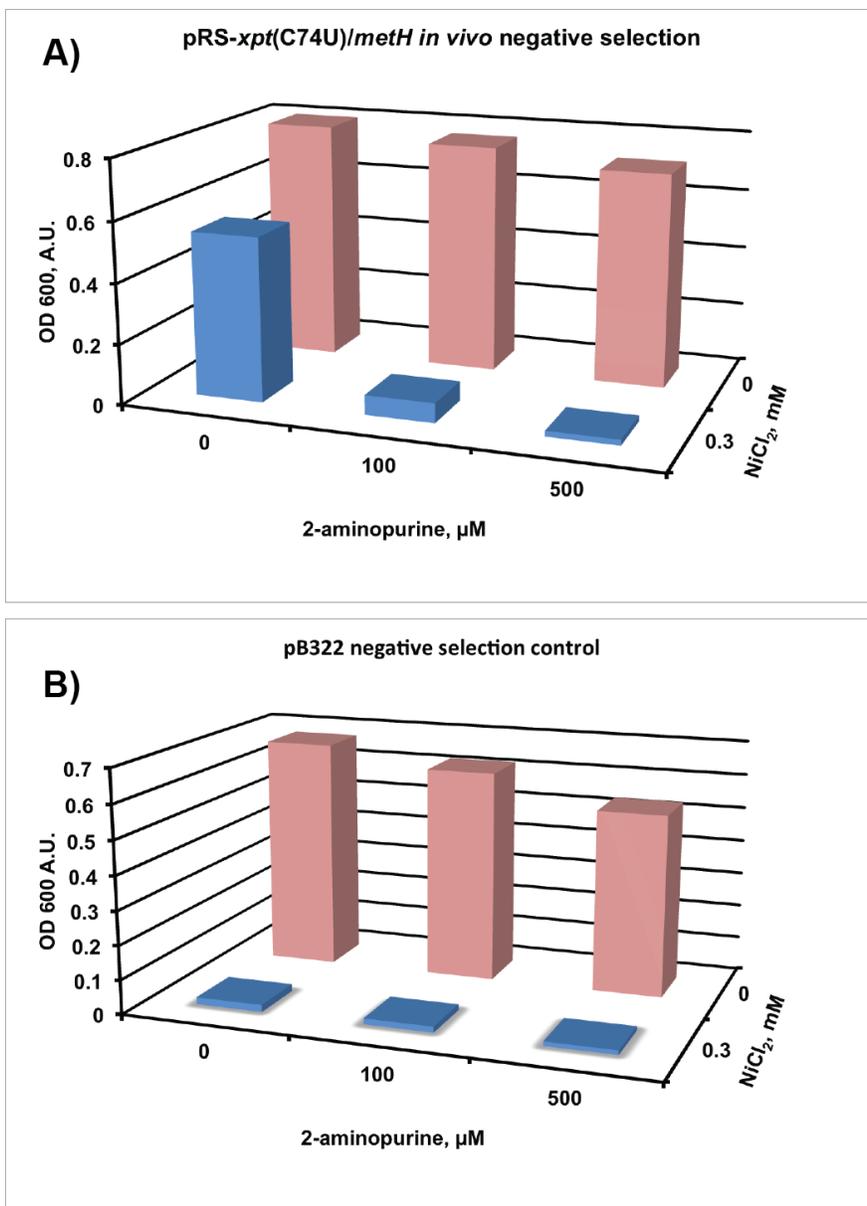
the best results were obtained the minimal media M9<sup>82</sup>, supplemented with 0.1% casamino acids, 0.08% glycerol as a carbon source, and 0.01 mg/mL of thiamine. The experiments presented in the following sections used these conditions to assess the ligand-depend growth of the pRS constructs.

Top10 *E. coli* cells carrying the pRS-*xpt/metH* plasmid can survive concentrations of 15-30 mg/mL of tetracycline, when at least 500  $\mu$ M of 2-AP is added to the media (Fig. 3.3, A). Cells transformed with the pBR322 plasmid, which constitutively expresses *tetR*, can grow in these concentrations of tetracycline even when 2-AP is not added (Fig. 3.3, B). This suggests that the cells containing the chimeric riboswitch proliferate due to the ligand activation of *tetR* expression by 2-AP.



**Fig 3.3 Positive *in vivo* selections using the *xpt*(C74U)/*methH* riboswitch chimera. A)** Selection of cells carrying the pRS-*xpt*(C74U)/*methH* plasmid. No growth is observed without the addition of 2-aminopurine when 15-25 mg/mL of tetracycline is added to the media. **B)** *E. coli* cells carrying the pBR322 plasmid grown on the same conditions as pRS-XS as a control. In the absence of the ligand, these cells can grow at all the tetracycline concentrations tested. Interestingly, under this media conditions the addition of 2-aminopurine seems to have an adverse effect in the cell's growth. Nevertheless, these experiments indicate that this chimeric riboswitch possess the regulatory power to select cells that activate the expression of TetR in a ligand dependent manner.

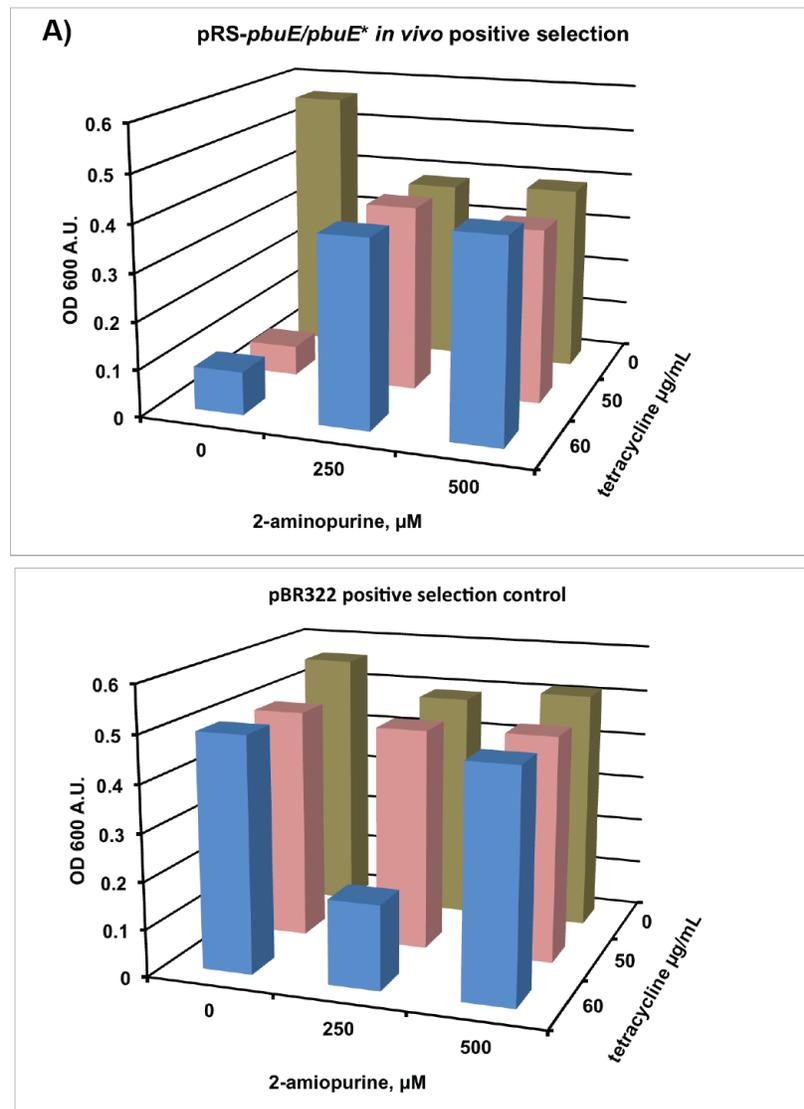
The pRS-*xpt/methH* construct also proved to be efficient in the negative selection step. Cells carrying the pRS-*xpt/methH* plasmid can only grow in the absence of the ligand when 0.3 mM of nickel chloride added to the media, suggesting that they can only express *tetR* when 2-aminopurine bind the riboswitch (Fig.3.4, A). As expected, this is not the case for the cells carrying the pBR322 plasmid that could not grown at concentrations of 0.3 mM nickel chloride (Fig 3.4, B). With this data, I can conclude that the pRS-*xpt/methH* is a suitable system to perform *in vivo* selections of novel riboswitch aptamers. However, this system is not ideal in the sense that we do not fully understand the structural requirements that mediate regulation in the expression platform of the *methH* riboswitch. Ideally we would like to have as much information from the riboswitch as possible, as this will facilitate the downstream rational optimization of the selected RNA. The work presented in Chapter 2 of this thesis dissected the structural features of the *pbuE* riboswitch's expression platform making it a preferable choice.



**Fig 3.4 Negative *in vivo* selection using the *xpt(C74U)/meth* riboswitch chimera.** **A)** Addition of 2-AP to the media activates the expression of the *tetR* gene making the cells susceptible to concentrations of 0.3 mM nickel chloride. **B)** Since pB322 constitutively expresses *tetR*, the cells cannot survive 0.3 mM of nickel chloride even when no 2-AP is added to the media, a condition that permits the proliferation of pRS-XS

To evaluate the potential of the *pbuE* riboswitch to be used as selection platform, I replaced the *xpt(C74U)/meth* chimera with the *pbuE/pbuE\** modular riboswitch in the pRS plasmid to create pRS-*pbuE/pbuE\**. Cells carrying this plasmid should only

proliferate in the presence of tetracycline, if they are provided with 2-aminopurine. When challenged in the tetracycline growth assay, cells carrying the pRS-*pbuE/pbuE\** only grow when at least 250  $\mu\text{M}$  of 2-AP is present in the media at concentrations of 50 mg/mL of tetracycline (Fig. 3.5, A). Accordingly, cells carry the pBR322 plasmid cannot survive these conditions even when 2-AP is added (Fig. 3.5, B).



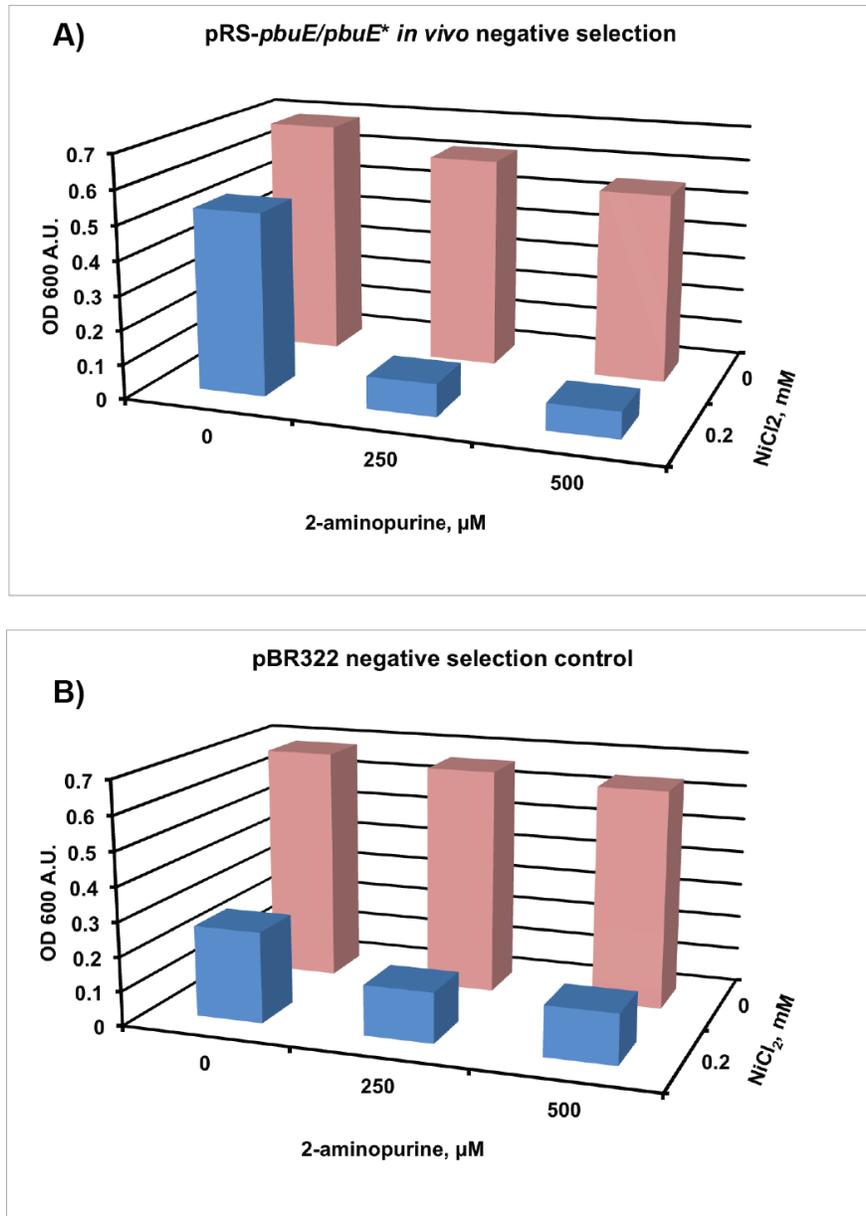
**Fig 3.5 Positive *in vivo* selections using the *pbuE/pbuE\** modular riboswitch. A)** Cells transformed with the pRS-*pbuE/pbuE\** plasmid can only grow in the presence of tetracycline when 2-AP is supplemented to the media to permit transcription of *tetR*. **B)** Cells carrying the pBR322 constitutively expressing *tetR* can survive even when the 2-AP is not supplemented to the media.

In addition, this data shows that the *pbuE/pbuE\** riboswitch does not allow for enough leaky expression in the absence of the ligand to proliferate in media conditions containing 200 mM NiCl<sub>2</sub>. Addition of 250 μM of 2-AP to the media seems to activate the riboswitch to produce enough *tetR* to be susceptible to osmotic stress by the metal salt (Fig 3.6, A). Again, cells carrying the pBR322 control were not able to grow when exposed to nickel chloride (Fig 3.6, B).

Comparing the results of the pRS-*pbuE/pbuE\** with those of the pRS-*xpt/methH* system, I conclude that the pRS-*xpt/methH* system is more robust, as the cells scarcely grow if the 2-AP is not supplemented. Addition of 2-AP seems to induce expression to optimal levels that cells grow as well as when no antibiotic is added. The same digital response is observed when the cells are counter selected with nickel chloride, indicating that the pRS-*xpt/methH* system is a superior choice than pRS-*pbuE/pbuE\**. To perform an *in vivo* selection we would optimally like our system to display the least amount of leaky expression when no target ligand present, but the maximal amount from the “winner” sequence once it binds the target ligand.

The effort of the mutagenesis analysis of the *pbuE* riboswitch in Chapter 2 is already rewarding me with ideas on how to improve the *pbuE/pbuE\** system. For example, I constructed several *pbuE* mutant expression platform that hyper stabilized the transcriptional terminator, one of them U97A/U98A displays a 42-fold dynamic range (Fig. 2.7, A). If this mutation is introduced to the *pbuE/pbuE\** system it has the potential to improve this platform to be even more robust than pRS-*xpt/methH*. Another attractive mutant is the 5'-Δ27 mutant that eroded the 5' side of the *pbuE* riboswitch up to the

beginning of P1 helix. This minimal *pbuE* riboswitch displays substantially higher expression when 2-AP is added to the media as compared to the wild type. A higher level of ligand dependent activation is a desirable property for my *in vivo* selection, but in addition the shorter riboswitch sequence will facilitate the construction of the randomized library. These ideas elevate the relevance of why riboswitches should be investigated in the cellular environment; especially if we want to build synthetic RNA technologies based on natural riboswitch systems.



**Fig. 3.6 Negative *in vivo* selection using the pRS-*pbuE/pbuE\** plasmid.** A) Cells transformed with the pRS-*pbuE/pbuE\** plasmid display ligand-dependent expression of *tetR* gene. When challenged with nickel chloride, only the cells that were not exposed to 2-AP proliferated. B) Cells transformed with the control plasmid pBR322 are significantly crippled when exposed to 0.3 mM nickel chloride a no 2-AP, as the constitutively express *tetR*.

Neither pRS-*xpt/metH* nor pRS-*pbuE/pbuE\**, displayed any cell fluorescence in any of the growth conditions. At this moment, I do not understand why the cells are not

fluorescing; it is possible that some unwanted mutations were introduced in the genetic design or that the cells are silencing the GFP gene for other reasons, such as to prevent toxicity. However, it is important to note that the fluorescence signal is not necessary to perform a riboswitch selection. To analyze the winning sequences, the riboswitches could be sub cloned into another plasmid with a fluorescence marker to perform an *in vivo* characterization. It will be advantageous to also have a fluorescence signal to track the progress of the selection, as it can provide a quick method for the evaluation of individual colonies without the need for additional sub cloning steps, meriting the consideration of utilizing exactly the same *tetR-gfpuv* fusion used by Yokobayashi.

Once this platform is optimized, the library needs to be constructed, for which there are several established methods<sup>78,83</sup>. Initially, I envision it will be a very limited (N=3-6) library encompassing the most highly conserved nucleotides that the RNA utilizes to recognize the purine ligand. As a proof of concept, the library will be challenged with purine-like compounds, such as 2-aminopurine, 2,6-diaminopurine and that permeate into the cell and for which binding sequence solutions exist. After this initial test, larger libraries and additional compounds of interest can be utilized for selection experiments.

There are other approaches that can be incorporated to this selection pipeline. Other research groups have been successful in evolving new RNA switches by employing Fluorescently Activated Cell Sorting (FACS)<sup>83</sup>. I can envision a method where the tetracycline selection is performed followed by a FACS step that can identify the best switches by means of fluorescence. These follow-up experiments will enlighten

the synthetic biology field with additional tools for the development of custom-made, ligand-dependent RNA switches.

### 3.4 Experimental Procedures

**Construction of *in vivo* selection tetracycline system-** I utilized the *pRR* (riboswitch reporter) vector utilized in Chapter 2 as a backbone for the insertion riboswitch-tetracycline-GFP construct. This is a low copy plasmid derived from pBR322, in which the coding sequence of GFPuv from pGFPuv (clonetechn) and the TetR gene from pBR322 itself, were inserted as a transcriptional fusion (A complete map of the plasmid is presented in Appendix B. Upstream of the two genes different riboswitch were cloned to control the transcription of the polycistronic containing GFP and TetR. Flanking the complete construct, the *rrnB* strong transcriptional terminator was inserted to prevent cryptic transcription of upstream elements.

**Ligand dependent growth experiments-** The plasmid aforementioned was introduced into the *E. coli* Top-10 cells. The cells were grown overnight to saturation in M9 media supplemented with (0.1 % casaminoacids, 0.08% glycerol, 0.01 mg/mL thiamine). To begin the growth experiment, the cells were diluted 1/100 into fresh media in a volume of 3 mL in a plastic tube and incubated at 37 Celsius for 12hrs with different concentrations of 2-aminopurine, tetracycline or NiCl<sub>2</sub>. After reaching the time point of 12 hrs, 300 µL of the cultures were transferred to a 96-well plate and the absorbance was measured at 600 nm as an indicator of cell growth in a Tecan Safire II plate reader.

## References

1. Martin, J. *The Meaning Of The 21st Century*. (Random House, 2012).
2. Hara, K. Y. *et al.* Development of bio-based fine chemical production through synthetic bioengineering. *Microb Cell Fact* **13**, 173 (2014).
3. Cheng, A. A. & Lu, T. K. Synthetic Biology: An Emerging Engineering Discipline. *Annu. Rev. Biomed. Eng.* **14**, 155–178 (2012).
4. Friedensen, V. P. & Mazanek, D. D. NASA's plans for human exploration of near-earth asteroids. *Earth and Space* **10**, 9780784412190.162 (2012).
5. Mazanek, D. D., Troutman, P. A., Culbert, C. J., Leonard, M. J. & Spexarth, G. R. Surface buildup scenarios and outpost architectures for lunar exploration. 1–23 (2009).
6. Drake, B. G., Hoffman, S. J. & Beaty, D. W. Human exploration of mars, design reference architecture 5.0. 1–24 (2010).
7. Verlinden, R. A. J., Hill, D. J., Kenward, M. A., Williams, C. D. & Radecka, I. Bacterial synthesis of biodegradable polyhydroxyalkanoates. *J Appl Microbiol* **102**, 1437–1449 (2007).
8. Post, M. J. Cultured beef: medical technology to produce food. *J. Sci. Food Agric.* **94**, 1039–1041 (2013).
9. Abedon, S. T. & Thomas-Abedon, C. Phage therapy pharmacology. *Curr Pharm Biotechnol* **11**, 28–47 (2010).
10. Hanford, A. J. Advanced life support baseline values and assumptions document. (2006).
11. Schiel-Bengelsdorf, B. & Dürre, P. Pathway engineering and synthetic biology using acetogens. *FEBS Letters* **586**, 2191–2198 (2012).
12. Lovley, D. R. & Nevin, K. P. Electrobiocommodities: powering microbial production of fuels and commodity chemicals from carbon dioxide with electricity. *Current Opinion in Biotechnology* **24**, 385–390 (2013).
13. Du, B. *et al.* Evaluation of Physical and Chemical Changes in Pharmaceuticals Flown on Space Missions. *AAPS J* **13**, 299–308 (2011).
14. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact* **8**, 17 (2009).
15. Menezes, A. A., Cumbers, J., Hogan, J. A. & Arkin, A. P. Towards synthetic biological approaches to resource utilization on space missions. *Journal of The Royal Society Interface* **12**, 20140715–20140715 (2014).
16. Tsuji, H., Ogawa, T., Bando, N. & Sasaoka, K. Purification and properties of 4-aminobenzoate hydroxylase, a new monooxygenase from *Agaricus bisporus*. *J. Biol. Chem.* **261**, 13203–13209 (1986).
17. Todaka, D., Shinozaki, K. & Yamaguchi-Shinozaki, K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front Plant Sci* **6**, 84 (2015).
18. Acharjee, S. & Sarmah, B. K. Plant Science. *Plant Science* **207**, 108–116 (2013).
19. Altuvia, S. & Wagner, E. G. Switching on and off with RNA. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9824–9826 (2000).
20. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24**, 381–

- 403 (2007).
21. Keasling, J. D. Metabolic Engineering. *Metabolic Engineering* **14**, 189–195 (2012).
  22. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology* **21**, 796–802 (2003).
  23. Pfleger, B. F., Pitera, D. J., Smolke, C. D. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology* **24**, 1027–1032 (2006).
  24. Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature Publishing Group* **15**, 95–107 (2014).
  25. Kloss, C. C., Condomines, M., Cartellieri, M., Bachmann, M. & Sadelain, M. Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells. *Nature Biotechnology* **31**, 71–75 (2012).
  26. Chen, Y. Y., Jensen, M. C. & Smolke, C. D. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proceedings of the National Academy of Sciences* **107**, 8531–8536 (2010).
  27. Gaudet, D. *et al.* Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPL). *Gene Therapy* **20**, 361–369 (2012).
  28. Kemmer, C. *et al.* nbt.1617. *Nature Biotechnology* **28**, 355–360 (2010).
  29. Ye, H., Daoud-EI Baba, M., Peng, R.-W. & Fussenegger, M. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science* **332**, 1565–1568 (2011).
  30. Fowler, C. C., Brown, E. D. & Li, Y. Using a Riboswitch Sensor to Examine Coenzyme B. *Chemistry & Biology* **17**, 756–765 (2010).
  31. Porter, E. B., Marcano-Velázquez, J. G. & Batey, R. T. Biochimica et Biophysica Acta. *BBA - Gene Regulatory Mechanisms* **1839**, 919–930 (2014).
  32. Pena, M. D. L., Dufour, D. & Gallego, J. Three-way RNA junctions with remote tertiary contacts: A recurrent and highly versatile fold. *RNA* **15**, 1949–1964 (2009).
  33. Batey, R. T., Gilbert, S. D. & Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411–415 (2004).
  34. SERGANOV, A. *et al.* Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chemistry & Biology* **11**, 1729–1741 (2004).
  35. Delfosse, V. *et al.* Riboswitch structure: an internal residue mimicking the purine ligand. *Nucleic Acids Research* **38**, 2057–2068 (2010).
  36. Pikovskaya, O., Polonskaia, A., Patel, D. J. & Serganov, A. structural principles of nucleoside selectivity in a 29-deoxyguanosine riboswitch. *Nature Chemical Biology* **7**, 748–755 (2011).
  37. Burge, S. W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research* **41**, D226–D232 (2012).
  38. Griffiths-Jones, S. Rfam: an RNA family database. *Nucleic Acids Research* **31**, 439–441 (2003).
  39. Gilbert, S. D., Love, C. E., Edwards, A. L. & Batey, R. T. Mutational Analysis of the Purine Riboswitch Aptamer Domain †. *Biochemistry* **46**, 13297–13309 (2007).

40. Gilbert, S. D., Reyes, F. E., Edwards, A. L. & Batey, R. T. Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. *Structure/Folding and Design* **17**, 857–868 (2009).
41. Edwards, A. L. & Batey, R. T. A Structural Basis for the Recognition of 2'-Deoxyguanosine by the Purine Riboswitch. *Journal of Molecular Biology* **385**, 938–948 (2009).
42. Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. A. Folding of the Adenine Riboswitch. *Chemistry & Biology* **13**, 857–868 (2006).
43. Lang, K. & Micura, R. The preparation of site-specifically modified riboswitch domains as an example for enzymatic ligation of chemically synthesized RNA fragments. *Nat Protoc* **3**, 1457–1466 (2008).
44. Dixon, N. *et al.* Reengineering orthogonally selective riboswitches. *Proceedings of the National Academy of Sciences* **107**, 2830–2835 (2010).
45. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
46. Ciesiolka, J., Gorski, J. & Yarus, M. Selection of an RNA domain that binds Zn<sup>2+</sup>. *RNA* **1**, 538–550 (1995).
47. Hofmann, H. P., Limmer, S., Hornung, V. & Sprinzl, M. Ni<sup>2+</sup>-binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair. *RNA* **3**, 1289–1300 (1997).
48. Stojanovic, M. N., de Prada, P. & Landry, D. W. Fluorescent Sensors Based on Aptamer Self-Assembly. *J. Am. Chem. Soc.* **122**, 11547–11548 (2000).
49. Grate, D. & Wilson, C. Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer-ligand complex. *Bioorg. Med. Chem.* **9**, 2565–2570 (2001).
50. Ruckman, J. *et al.* 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.* **273**, 20556–20567 (1998).
51. Srisawat, C. & Engelke, D. R. Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7**, 632–641 (2001).
52. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642–646 (2011).
53. Lee, J.-H. *et al.* A therapeutic aptamer inhibits angiogenesis by specifically targeting the heparin binding domain of VEGF165. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18902–18907 (2005).
54. Ceres, P., Garst, A. D., Marcano-Velázquez, J. G. & Batey, R. T. Modularity of Select Riboswitch Expression Platforms Enables Facile Engineering of Novel Genetic Regulatory Devices. *ACS Synth. Biol.* 130328094931000 (2013). doi:10.1021/sb4000096
55. Serganov, A. & Nudler, E. A Decade of Riboswitches. *Cell* **152**, 17–24 (2013).
56. Hajnsdorf, E. & Boni, I. V. Multiple activities of RNA-binding proteins S1 and Hfq. *Biochimie* **94**, 1544–1553 (2012).
57. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**, 29–35 (2003).
58. Lemay, J.-F. *et al.* Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in

- Riboswitch Regulatory Mechanisms. *PLoS Genet* **7**, e1001278 (2011).
59. Brenner, M. D., Scanlan, M. S., Nahas, M. K., Ha, T. & Silverman, S. K. Multivector Fluorescence Analysis of the xptGuanine Riboswitch Aptamer Domain and the Conformational Role of Guanine. *Biochemistry* **49**, 1596–1605 (2010).
  60. Noeske, J. *et al.* Interplay of ‘induced fit’ and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Research* **35**, 572–583 (2006).
  61. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586 (2003).
  62. Gilbert, S. D. & Batey, R. T. Riboswitches: Fold and Function. *Chemistry & Biology* **13**, 805–807 (2006).
  63. Stoddard, C. D., Gilbert, S. D. & Batey, R. T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **14**, 675–684 (2008).
  64. Gilbert, S. D., Stoddard, C. D., Wise, S. J. & Batey, R. T. Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain. *Journal of Molecular Biology* **359**, 754–768 (2006).
  65. Lemay, J. F. & Lafontaine, D. A. Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA* **13**, 339–350 (2007).
  66. Blose, J. M., Proctor, D. J., Veeraraghavan, N., Misra, V. K. & Bevilacqua, P. C. Contribution of the Closing Base Pair to Exceptional Stability in RNA Tetraloops: Roles for Molecular Mimicry and Electrostatic Factors. *J. Am. Chem. Soc.* **131**, 8474–8484 (2009).
  67. Nagel, J. H. A. *et al.* Structural parameters affecting the kinetics of RNA hairpin formation. *Nucleic Acids Research* **34**, 3568–3576 (2006).
  68. Nozinovic, S. *et al.* The importance of helix P1 stability for structural pre-organization and ligand binding affinity of the adenine riboswitch aptamer domain. *rnabiology* **11**, 655–666 (2014).
  69. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The Speed of RNA Transcription and Metabolite Binding Kinetics Operate an FMN Riboswitch. *Molecular Cell* **18**, 49–60 (2005).
  70. Mahen, E. M., Watson, P. Y., Cottrell, J. W. & Fedor, M. J. mRNA Secondary Structures Fold Sequentially But Exchange Rapidly In Vivo. *PLoS Biol* **8**, e1000307 (2010).
  71. Frieda, K. L. & Block, S. M. Direct Observation of Cotranscriptional Folding in an Adenine Riboswitch. *Science* **338**, 397–400 (2012).
  72. Tomsic, J., McDaniel, B. A., Grundy, F. J. & Henkin, T. M. Natural Variability in S-Adenosylmethionine (SAM)-Dependent Riboswitches: S-Box Elements in *Bacillus subtilis* Exhibit Differential Sensitivity to SAM In Vivo and In Vitro. *Journal of Bacteriology* **190**, 823–833 (2008).
  73. Ceres, P., Trausch, J. J. & Batey, R. T. Engineering modular ‘ON’ RNA switches using biological components. *Nucleic Acids Research* (2013). doi:10.1093/nar/gkt787
  74. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* **39**, 1131–1141 (2011).

75. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, (2006).
76. Stoddard, C. D. *et al.* Nucleotides Adjacent to the Ligand-Binding Pocket are Linked to Activity Tuning in the Purine Riboswitch. *Journal of Molecular Biology* **425**, 1596–1611 (2013).
77. Nomura, Y. & Yokobayashi, Y. Reengineering a Natural Riboswitch by Dual Genetic Selection. *J. Am. Chem. Soc.* **129**, 13814–13815 (2007).
78. Yokobayashi, Y. & Arnold, F. H. A Dual Selection Module for Directed Evolution of Genetic Circuits. *Nat Comput* **4**, 245–254 (2005).
79. Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Research* **37**, e39–e39 (2009).
80. Stavropoulos, T. A. & Strathdee, C. A. Expression of the tetA(C) tetracycline efflux pump in Escherichia coli confers osmotic sensitivity. *FEMS Microbiology Letters* **190**, 147–150 (2000).
81. Podolsky, T., Fong, S. T. & Lee, B. T. Direct selection of tetracycline-sensitive Escherichia coli cells using nickel salts. *Plasmid* **36**, 112–115 (1996).
82. Harwood, C., and S. Cutting. "Chemically Defined Growth Media and Supplements." *Molecular Biology Methods for Bacillus* (1990): 548. Print.
83. Lynch, S. A. & Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Research* **37**, 184–192 (2009).
84. Grundy, F. J. & Henkin, T. M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.* **30**, 737–749 (1998).
85. McDaniel, B. A. M., Grundy, F. J., Artsimovitch, I. & Henkin, T. M. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3083–3088 (2003).
86. Mulhbachter, J. & Lafontaine, D. A. Ligand recognition determinants of guanine riboswitches. *Nucleic Acids Research* **35**, 5568–5580 (2007).
87. Lu, C. *et al.* SAM Recognition and Conformational Switching Mechanism in the Bacillus subtilis yitJS Box/SAM-I Riboswitch. *Journal of Molecular Biology* **404**, 803–818 (2010).
88. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**, 1172–1175 (2006).
89. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* **10**, 701–707 (2003).
90. Lindgreen, S., Gardner, P. P. & Krogh, A. Measuring covariation in RNA alignments: physical realism improves information measures. *Bioinformatics* **22**, 2988–2995 (2006).
91. Martin, L. C., Gloor, G. B., Dunn, S. D. & Wahl, L. M. Using information theory to search for co-evolving residues in proteins. *Bioinformatics* **21**, 4116–4124 (2005).
92. Caporaso, J. G. *et al.* Detecting coevolution without phylogenetic trees? Tree-ignorant metrics of coevolution perform as well as tree-aware metrics. *BMC Evol Biol* **8**, 327 (2008).

93. Wickiser, J. K., Cheah, M. T., Breaker, R. R. & Crothers, D. M. The Kinetics of Ligand Binding by an Adenine-Sensing Riboswitch. *Biochemistry* **44**, 13404–13414 (2005).
94. Vicens, Q., Mondragon, E. & Batey, R. T. Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic Acids Research* **39**, 8586–8598 (2011).
95. Chen, B., Zuo, X., Wang, Y. X. & Dayie, T. K. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Research* **40**, 3117–3130 (2012).
96. Feng, J., Walter, N. G. & Brooks, C. L., III. Cooperative and Directional Folding of the preQ 1 Riboswitch Aptamer Domain. *J. Am. Chem. Soc.* **133**, 4196–4199 (2011).

## Appendix A

### Structural characteristics of purine aptamer that lead to an adjusted regulatory response

#### A.1 The aptamers of riboswitches classes are highly conserved in their structure, but vary significantly in their regulatory activity

In some bacteria, there are multiple copies of a particular type of riboswitch, each regulating its own transcriptional unit. It is hypothesized that the regulation properties of each riboswitch is “tuned.” As an example, in *B. subtilis*, there are 11 different S-adenosylmethionine (SAM) riboswitches, that regulate the genes associated with methionine biosynthesis and transport<sup>84,85</sup>. Similarly, there are five different purine riboswitches have been identified, in *B. subtilis* as well, suggesting that it could be a commonality of metabolic pathways controlled by riboswitches<sup>61,86</sup>. Since these riboswitches belong to the same family, they display high global structural conservation, with their binding pockets being nearly invariant, as demonstrated by the various crystal structures of SAM riboswitches<sup>87,88</sup>. However, these SAM aptamers exhibit a wide range of affinities (1000 fold range in  $K_D$  values), as well as differences in their transcription assay output and *in vivo* regulatory response<sup>72,89</sup>. Particularly, it was noted that the genes involved in biosynthetic processes were more sensitive to changes in the intracellular levels of SAM than those involved in the transport of the compound. As mentioned previously in this dissertation, purine aptamers also display a very high structural conservation, and also display differential affinities for their ligands.

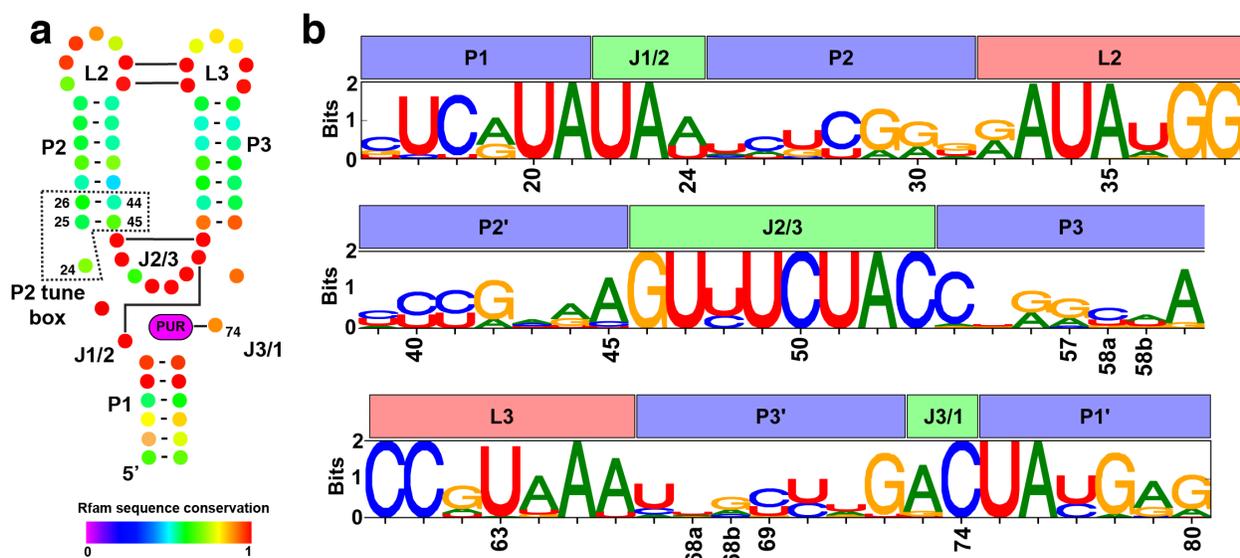
The observations regarding the different amplitude regulatory responses can be attributed to the expression platform of the riboswitches, but also can be rationalized as being the result of different affinities of the aptamers. Since we have studied purine riboswitches extensively, we decided to investigate if there was a structural mechanism mediating a “tuned” response in purine riboswitches using the guanine-binding *xpt* riboswitch from *B. subtilis*. The results of the study emerge from collaborations between biochemists and computational biologists, making of this work an example on how these fields can be brought together. Notably, this study also utilized an *in vivo* reporter system that I designed.

## **A.2 A computational phylogenetic analysis reveals a pattern of co-variation for non-structurally interacting nucleotides**

It is difficult to identify changes in an RNA sequence that can lead to different activities based solely on sequence alignments. The patterns of co-variation are dominated by the tertiary and secondary structural elements in the aptamer. However a statistical technique named Normalized Mutual Information (NMI) is capable of dissecting the patterns of co-variation to identify relevant mutations that are not bound to the structure<sup>90</sup>. This technique has proven useful in identifying relevant regions of co-variation in proteins that lead to different activities<sup>91,92</sup>.

Before an NMI analysis can be performed, it is necessary to obtain a high quality homogenous alignment. The Rfam database contains the RNA alignment for the purine riboswitch aptamer, but a significant amount of the sequences contain insertions or deletions that confound the analysis<sup>37</sup>. For that reason, a curated alignment was generated containing 302 of the 484 sequences available in Rfam at the time. The

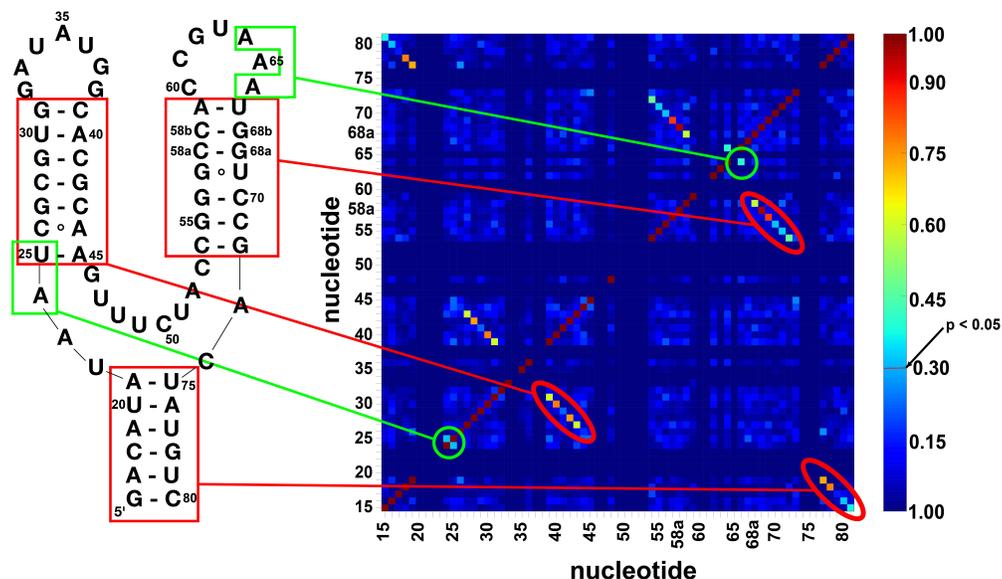
selected sequences were chosen to contain 7 base pairs in the paired regions P1, P2, and P3. The consensus sequences indicates a 7 base pairs helix for P2 and P3, and for the P1, it has been shown that extending the helix beyond 4 base pairs does not have an impact in the binding activity<sup>65</sup>. Figure A.1 shows the secondary structure of the purine aptamer along with the patterns of nucleotide conservation of each nucleotide position. This data shows high-degree of conservation around the three-way junctions where the binding pocket lies, along with conservation at the distal loops that support the tertiary L2-L3 interaction.



**Fig. A.1 Evolutionary conservation of the purine aptamer. A)** Secondary structure with overlaid conservation heat map. Watson-Crick interactions are denoted by bars/lines; red positions are universally conserved (Adapted from the Rfam database, accession number RF00167). **B)** Sequence logo representing conservation of the nucleotides positions, as well as the identity information of the structure-based alignment of 302 sequences. For each position the height of the nucleotide represents the frequency at that position, while the overall height represents the information content at that position.

### A.3 A region of co-variation was identified at the base of P2 and termed the “Tune” box

The NMI analysis yields a score for all the possible pairwise interaction, demonstrating significant co-variation patterns corresponding to the paired regions P1, P2, and P3 (Fig. A.2). In addition to the expected patterns of co-variation in the paired regions, a significant NMI score (NMI=.320,  $p=0.0197$ ) between positions 24 and 25 was observed (Fig A.2). Co-variation between positions 24 and 25 is particularly interesting due to the nearby position to the binding pocket. Position 24 is located in J1/2 and anchors the junction to J1/3 by stacking between positions 72 and 73. Nucleotide 25 displays a low level of conservation, while position 24 is either an adenine (66%) or a uracil (34%).

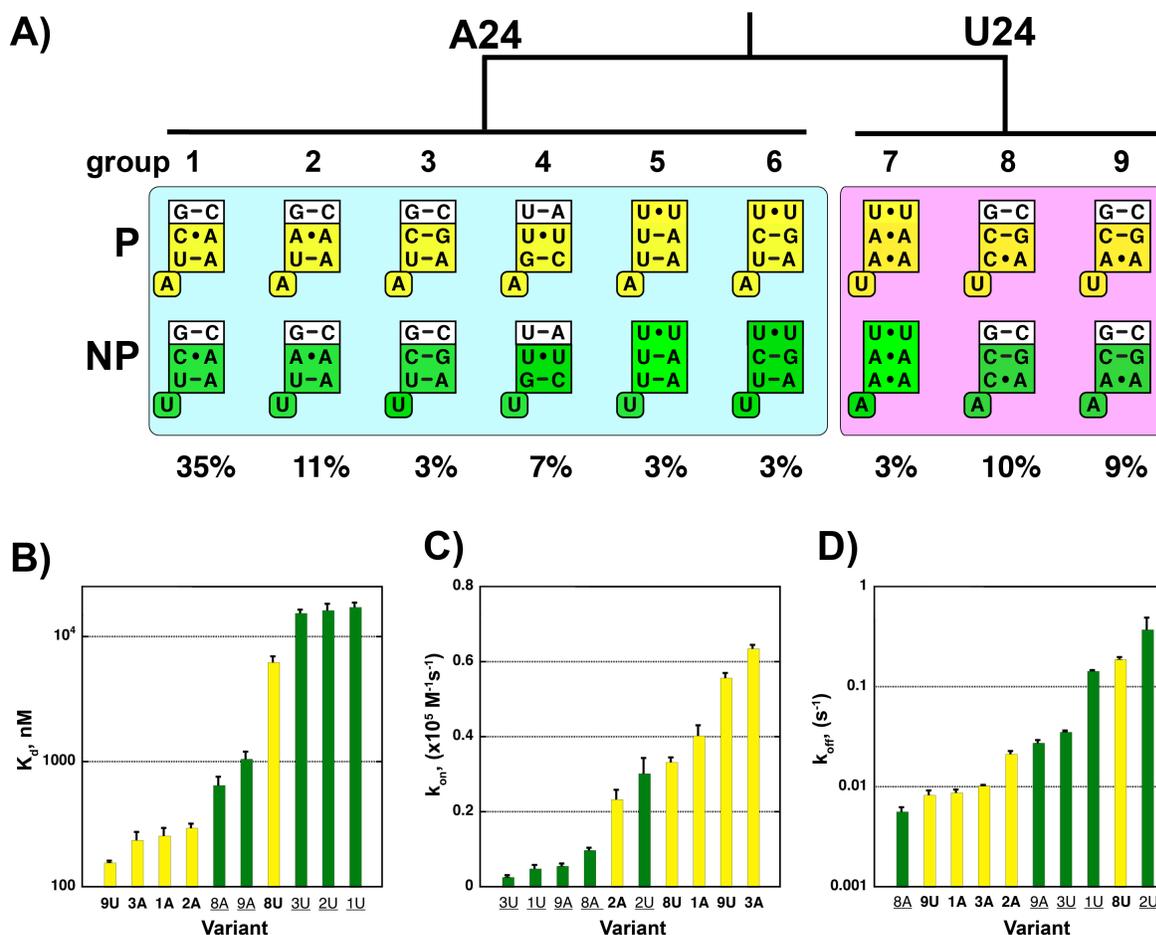


**Fig. A.2 Normalized mutual information analysis on the purine aptamer domain alignment.** NMI scores indicate covariation patterns in the aptamer domain. Secondary structural elements (red) and novel covariation interaction (green) are correlated to a two-dimensional heat map with NMI scoring represented on heat map colored by NMI values. The color bar to the right of the heat map indicates that an NMI value of 0.3 or greater is a significant value ( $p < 0.05$ )

Another section of the aptamer that showed significant values of co-variation of non-interacting nucleotides are nucleotides 64 and 66 in L3 (NMI 0.401,  $p=0.00579$ ). In 80% of the sequences both of these nucleotides are adenosines that form non-canonical base-pairing interactions with L2. As a pair, there is low level of conservation in these nucleotides; (U64/U66, 10%), (G64/A66, 8.5%) and (U64/A66, 1.5%). Moreover, a previous report had shown that the U64/A66 variant binds 2-aminopurine as well as wild type (A64/A66). In addition, it has been shown that the A35-A64 pair in the L2-3 interaction is tolerant to changes at the 35 position, suggesting that the loop-loop interaction can accommodate various pairs at the site. For these reasons, subsequent biochemical analysis of the co-variation patterns focused on nucleotides 24 and 25.

#### **A.4 Mutational analysis tuning box reveals the nucleotide involvement in the binding kinetics**

Further evaluation of the sequence alignments classified the observed patterns of co-variation based on their phylogenetic representation based on the position of nucleotide 24 being and adenine or uracil (Fig. A.3, A). Five of the nine phylogenetically represented sequences embodying 68% of the sequence in the alignment were chosen for further evaluation. In addition a non-phylogenetically represented artificial sequence was created for each of the variants. All the sequences of the P2 tune box were generated in the context of the *xpt-pbuX* aptamer domain to allow the comparison of directly to the other variants. The resulting aptamers were biochemically characterized based on their equilibrium and kinetic ligand binding properties using a fluorescence enabled by the fluorescent properties of 2-aminopurine<sup>42</sup>.



**Fig. A.3 Classification of co-variation sequences and their impact to the ligand binding properties of the purine aptamer. A)** P2 sequence variants representing the majority of all purine riboswitch aptamer domains. Aptamer variants from phylogenetically represented sequences (P) and non-phylogenetic variants (NP) with single base substitutions at position 24 are represented in yellow and green, respectively. The boxed pairs represent the P2.1, P2.2, and P2.3 pairs. The abundance of each variant from the sequence alignment is indicated. **B)** Equilibrium dissociation constants of the representative tuning sequences measured by the 2-AP fluorescence assay. **C and D)** Association and Dissociation rate constants ( $k_{on}$  and  $k_{off}$ ) representative tuning sequences.

It was generally observed that the naturally occurring sequences tend to have higher affinities than their non-natural counterparts (Fig A.3, B). The equilibrium constants were determined by monitoring the fluorescence of a constant concentration

of 2-aminopurine with increasing concentrations of the aptamer RNA. It was observed over 100-fold range in  $K_d$  (160 nM to 17  $\mu$ M), with most of the natural variant displaying values in the low nanomolar range. There was one exception to the natural sequences trend with the 8U variant, which displayed a  $K_d$  of 6.2  $\mu$ M. This sequence is represented in ~10% of the alignment, but it displays ~10 fold lower affinity than its non-natural counterpart 8A.

The kinetic properties of the naturally occurring variants grouped with both higher association rates ( $k_{on}$ ) and low dissociation rates ( $k_{off}$ ) (Fig A.3, C, D). For the kinetic measurements, the fluorescence of 2-aminopurine was monitored with a stopped flow fluorometer using established methods<sup>64</sup>. An exception to the trend was 2A, which displays a slightly lower association rate than its non-natural counterpart 2U. Furthermore, neither the  $k_{on}$  or  $k_{off}$  measurements resolve why the natural variant 8U is selected over 8A. The  $k_{on}$  of 8U is only ~3.4-fold lower than that for 8A, in addition to displaying a  $k_{off}$  ~33-fold greater than the non-phylogenetic variant 8A.

Purine riboswitches have been pointed to be under “kinetic control,” denoting that the aptamer does not reach thermodynamic ligand bind equilibrium before the RNA polymerase finish transcribing the 3' side of the riboswitch ( $\Delta t_{RNAP}$ ). The kinetic parameters determined for the tuning variants can be utilized to assess the mode of action of a riboswitch containing that particular tuning aptamer sequence using a model originally developed for the *pbuE* riboswitch<sup>93</sup>. In this model, it is assumed that a 90% of the RNA needs to be in the ligand bound state to induce a full regulatory response; and that the time it takes for the aptamer to achieve ligand binding equilibrium,  $\Delta t_B$ , is known. The time interval can be estimated by  $\Delta t_B = 1.15/k_{off}$  and compared to  $\Delta t_{RNAP}$ . If  $\Delta t_B$  is of

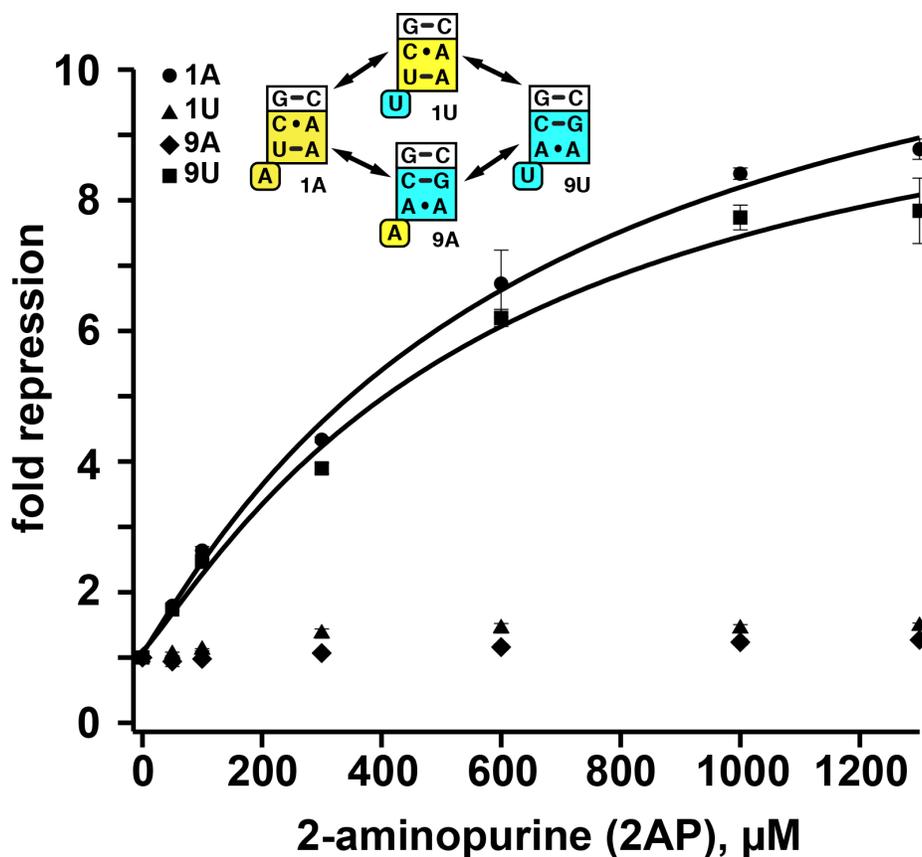
equal or lower value than  $\Delta t_{\text{RNAP}}$ , is presumed that the switch is under “thermodynamic control”; if the  $\Delta t_{\text{B}}$  higher  $\Delta t_{\text{RNAP}}$ , it is described as being under “kinetic control.”

$\Delta t_{\text{B}}$  was calculated for the all the variants, for which kinetic values were determined, to be compared with a theoretical  $\Delta t_{\text{RNAP}}$ . A theoretical  $\Delta t_{\text{RNAP}}$  was devised by assuming a transcription rate of ~50 nt/s and average riboswitch of 150 nt, yield a value for  $\Delta t_{\text{RNAP}}$  equal to 3 seconds. Notably, the naturally occurring sequences display significantly slower dissociation rates that lead to high values of  $\Delta t_{\text{B}}$ , greater than 50 s. These RNA are then thought to be under kinetic control under the conditions just mentioned. Conversely, the non-phylogenetically represented sequences tend to have lower  $\Delta t_{\text{B}}$  values, suggesting that they might be under thermodynamic control.

#### **A.5 The P2 tune box variants impart functional diversity**

The previous section presented contributions of the nucleotides in P2 tune box towards the ligand binding activity. To determine the effect of these tune variants, the *in vivo* regulatory activity of a small subset of variants (1A, 1U, 9A, and 9U) was tested using a fluorescence assay analogous to the one described in Chapter 2 of this dissertation. To perform this analysis, the aptamer sequences were fused to the expression platform of the *B. subtilis metE* SAM-I riboswitch. This set up would decouple any potential sequence specific interactions between the aptamer and the expression platform, and the differences between regulatory activities can be directly attributed by differences in the ligand binding properties of the aptamer.

The chimeric tuning aptamers were evaluated for their ability to repress the expression of GFP when a defined concentration of 2-aminipurine is added to the media. Only the two naturally represented sequences were able to repress GFP expression of about 8-fold (Fig. A.4). In addition both sequences display an EC-50 value, which defines the concentration of ligand to elicit a half-maximal response of  $\sim 600 \mu\text{M}$ . Conversely, the non-phylogenetically represented sequences are highly detrimental to the regulatory response. This data shows, that a deleterious mutation on nucleotide 24 can be rescued by changes in the first 2 base pairs of the P2 helix. Moreover, these results support the coupling of the identities of these nucleotides, even when they are non-interacting, validating the statistical coupling analysis.



**Fig. A.4** *In vivo* regulation of gene expression by a purine/SAM hybrid riboswitch. Aptamer variants 1A, 1U, 9A, and 9U were fused to the *Bacillus subtilis metE* expression platform and were assayed for the ability to terminate transcription in response to increasing 2AP concentrations. Transcriptional termination was measured by a decrease in GFP signal corrected by optical density.

## A.6 Discussion

This study demonstrates that it is possible to identify regions of covariation in nucleotides sequence alignment that are relevant to the activity of the RNA. Utilizing the NMI method, functionally coupled nucleotides that affect ligand binding were identified for the purine aptamer. Evaluation of the ligand binding kinetic properties revealed that the P2 tune box has a large effect in the ligand dissociation constant ( $k_{\text{off}}$ ) and thus in the ligand dissociation constant ( $K_d$ ). Although the ligand association

constant was also affected, compared to the values of  $\Delta t_{\text{RNAP}}$ , the effects on the regulatory activity by changes in ( $k_{\text{on}}$ ) should be minimal. These *in vitro* observations were supported by the *in vivo* activity experiments that clearly demonstrate co-variation of nucleotide 24 with the first two pairs of the P2 helix. Finally, the P2 tune box site appears to be an important site for affinity tuning, since other binding pocket adjacent regions, like the bottom of P3 and the top of P1, are highly conserved.

A study that evaluated the eleven SAM-I riboswitches indicated that the activity of the riboswitch was tuned based on the functionality of the gene that is being regulated; for this study, such correlations could not be established<sup>72</sup>. A difference in this work is that sequences evaluated originate from different organisms that live in different environments, and it is reasonable to think that changes in the extracellular conditions would have a significant impact on the evolution of the aptamer.

The role of the P2 box seems to be to impart stability to the ligand bound aptamer state. For this project, the crystal structures of three tuning variants were determined, revealing that the structure of the binding pockets was nearly identical to that of the wild type *xtp-pbuX* (data not shown). Furthermore, chemical probing experiments (SHAPE) show very similar patterns of reactivity suggesting that the folding dynamics are not affected (data not shown). Nevertheless, the RNA-ligand stabilization becomes evident due to the observed greater values of  $k_{\text{off}}$ . Additionally, we still don't know what effects these nucleotides might have in the co-transcriptional folding pathway.

A finely tuned genetic regulatory response is an important biological feature that allows organism to respond to the ever-changing cellular environment. For riboswitches,

similar modulated responses, as the one exemplified by the purine riboswitch, have been observed in SAM-I, FMN and preQ1 riboswitch aptamers<sup>94-96</sup>. For purine aptamers, we now have identified the nucleotides mediating this mechanism, and we have provided the bioinformatics tools to extend the analysis to other nucleic acid systems.

## Appendix B

### Supporting Methods and Figures

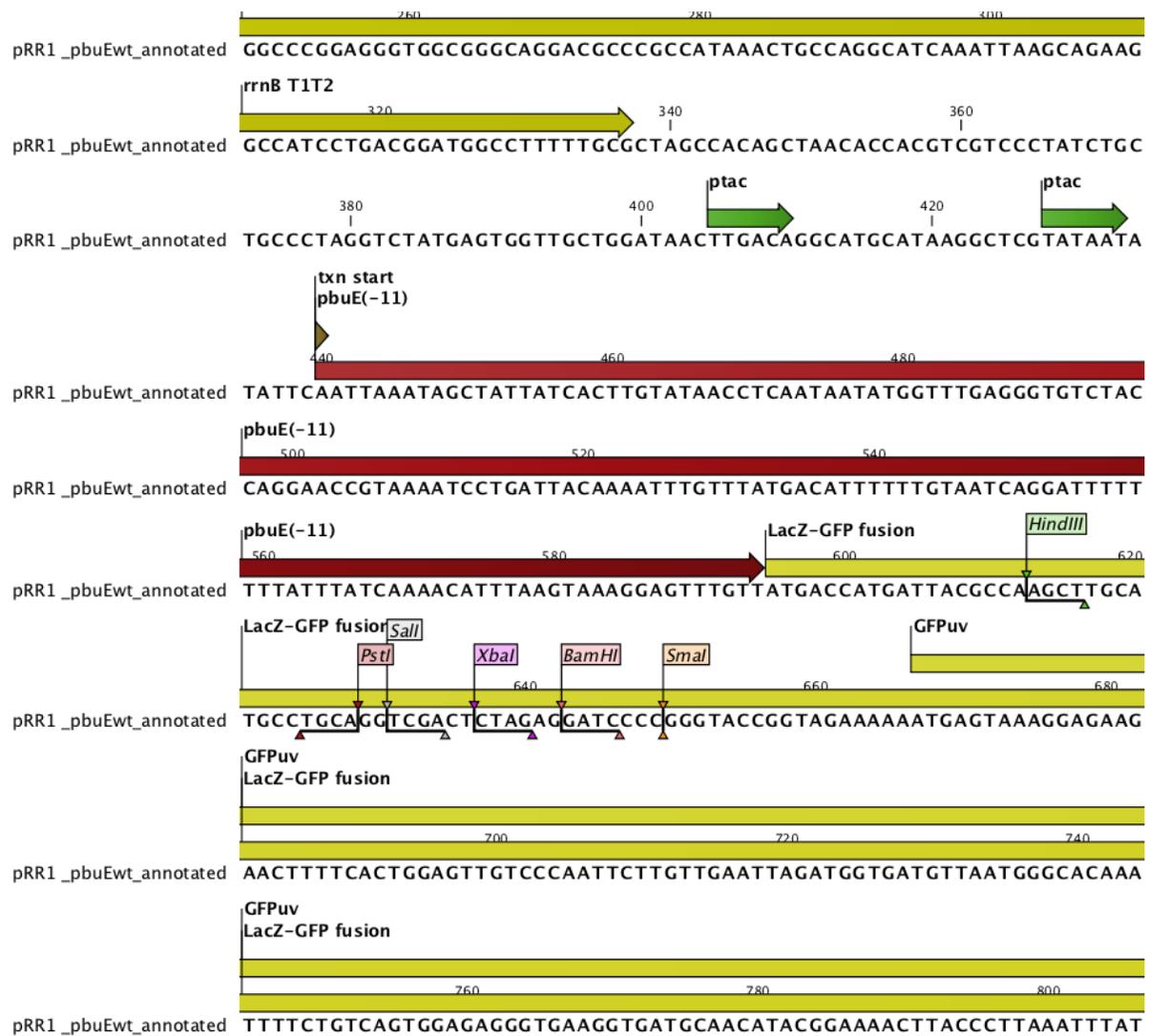
#### B.1 Sequences used in the in the mutational analysis of the pbuE riboswitch in the intracellular context

Riboswitch	Sequence
wild type <sup>a</sup>	<u>TTTACGGGCATGCATAAGGCTCGTATAATATATTCGGAAACGAATCAATTAA</u> ATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTA CCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATTTTTTTGT AATCAGGATTTTTTTTATTTATCAAAACATTTAAGTAAAGGAGTTTGT <u>AT</u> <u>G</u>
U66C <sup>b</sup>	GGAAACGAATCAATTAAATAGCTATTATCACTTGTATAACCTCAATAATA TGGTTTGAGGGTGTCCACCAGGAACCGTAAAATCCTGATTACAAAATTTG TTTATGACATTTTTTTGTAATCAGGATTTTTTTT
U89A	GGAAACGAATCAATTAAATAGCTATTATCACTTGTATAACCTCAATAATA TGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGA <u>AT</u> TACAAAATTTG TTTATGACATTTTTTTGTAATCAGGATTTTTTTT
5' Δ3	AACGAATCAATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGG TTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTT ATGACATTTTTTTGTAATCAGGATTTTTTTT
5' Δ6	GAATCAATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTT GAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATG ACATTTTTTTGTAATCAGGATTTTTTTT
5' Δ9	TCAATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAG GGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACA TTTTTTGTAATCAGGATTTTTTTT
5' Δ11	AATTAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
5' Δ27	CTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAA AATCCTGATTACAAAATTTGTTTATGACATTTTTTTGTAATCAGGATTTTTT TT
Δ11-A48U	AATTAATAGCTATTATCACTTGTATAACCTCAATA <u>T</u> TATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
Δ11-U49A	AATTAATAGCTATTATCACTTGTATAACCTCAATA <u>AA</u> ATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
Δ11-A50U	AATTAATAGCTATTATCACTTGTATAACCTCAATAAT <u>T</u> TGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
Δ11-U51G	AATTAATAGCTATTATCACTTGTATAACCTCAATAATAG <u>G</u> GGTTTGAGGG

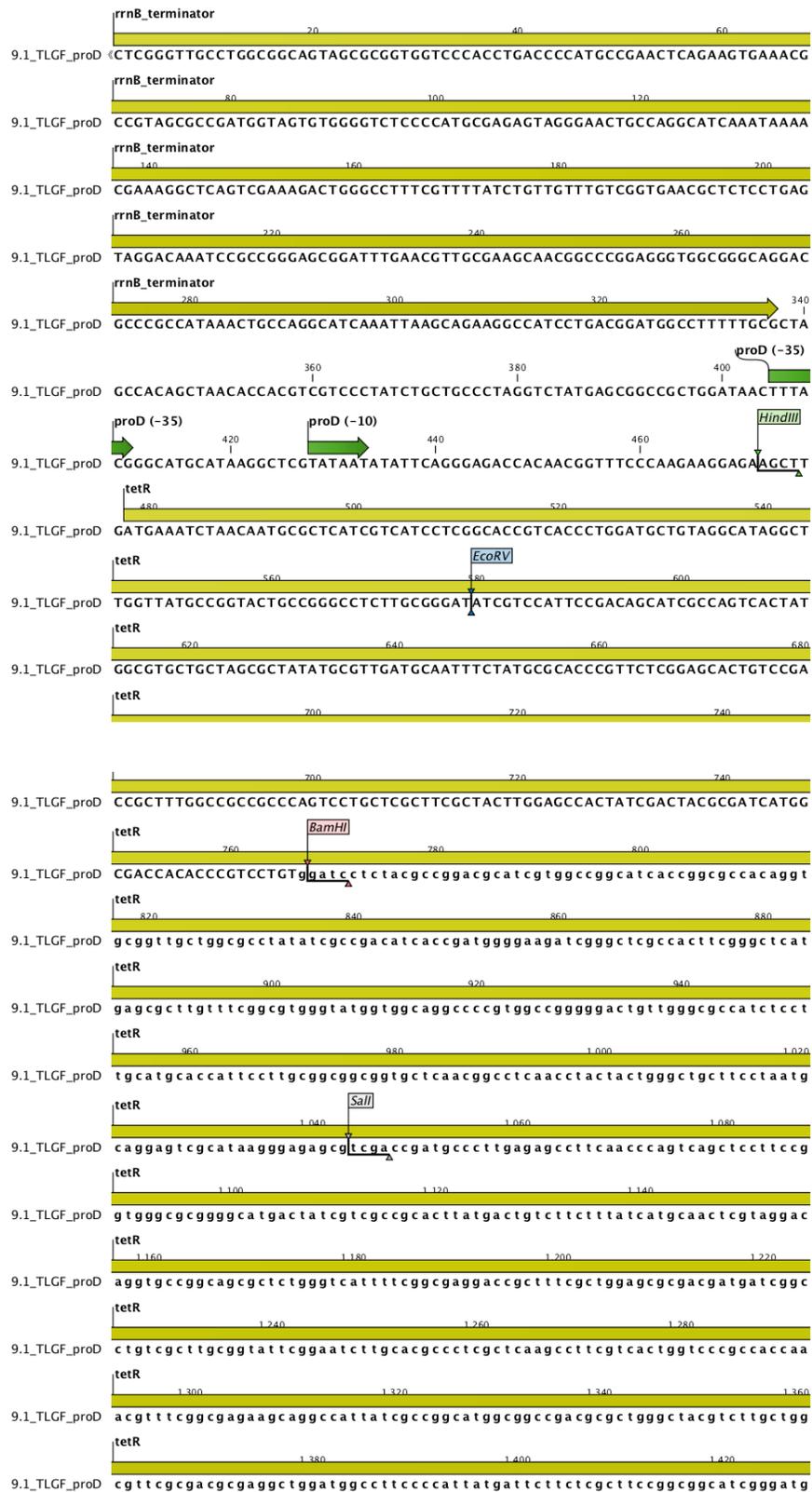
	TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11- G52C,G53C	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATCCTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,-4)	AATTAAATAGCTATTATCACAAACAATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,-3)	AATTAAATAGCTATTATCACAACTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,-2)	AATTAAATAGCTATTATCACAAAGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,-1)	AATTAAATAGCTATTATCACATGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+1)	AATTAAATAGCTATTATCATTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+2)	AATTAAATAGCTATTATCTTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+3)	AATTAAATAGCTATTATATTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+4)	AATTAAATAGCTATTCAATTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+5)	AATTAAATAGCTATTAAATTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+6)	AATTAAATAGCTATCAAATTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+8)	AATTAAATAGCTAAACAATTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-(P1,+10)	AATTAAATAGCAAACAATTTTGTATAACCTCAATAATATGGTTTGAGGG GTGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTGTTTATGACATT TTTTTGTAAATCAGGATTTTTTTT
$\Delta$ 11- G100A,C108U	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTATTTTATGATATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11-G100C	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTCTTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta$ 11- G100C,C108G	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAATTTCTTTTATGAGATT TTTTGTAATCAGGATTTTTTTT

$\Delta 11$ -( $\Delta P4b$ )	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAAATATTTTTTTGTAA TCAGGATTTTTTTT
$\Delta 11$ -U97A	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAATTGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta 11$ -U97,98A	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAAATGTTTATGACATT TTTTGTAATCAGGATTTTTTTT
$\Delta 11$ -U97,98A, (L4-GAAA)	AATTAAATAGCTATTATCACTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAAATGTCGAAAGACA TTTTTTGTAATCAGGATTTTTTTT
$\Delta 11$ - (P1,+8),(NUCL)	AATTAAATAGCT <u>AACAAAT</u> TTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAATGTT <u>TGTTTATGA</u> <u>CAGACATTTTTT</u> GTAATCAGGATTTTTTTT
$\Delta 11$ - (P1,+8),(NUCL/ GC)	AATTAAATAGCT <u>AACAAAT</u> TTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAATGTT <u>GGGTTATGC</u> <u>CCAACATTTTTT</u> GTAATCAGGATTTTTTTT
$\Delta 11$ - (P1,+8),(NUCL/ GAAA)	AATTAAATAGCT <u>AACAAAT</u> TTTTGTATAACCTCAATAATATGGTTTGAGGG TGTCTACCAGGAACCGTAAAATCCTGATTACAAAAATGTT <u>TCTCGAAAG</u> <u>AGAAACATTTTTT</u> GTAATCAGGATTTTTTTT

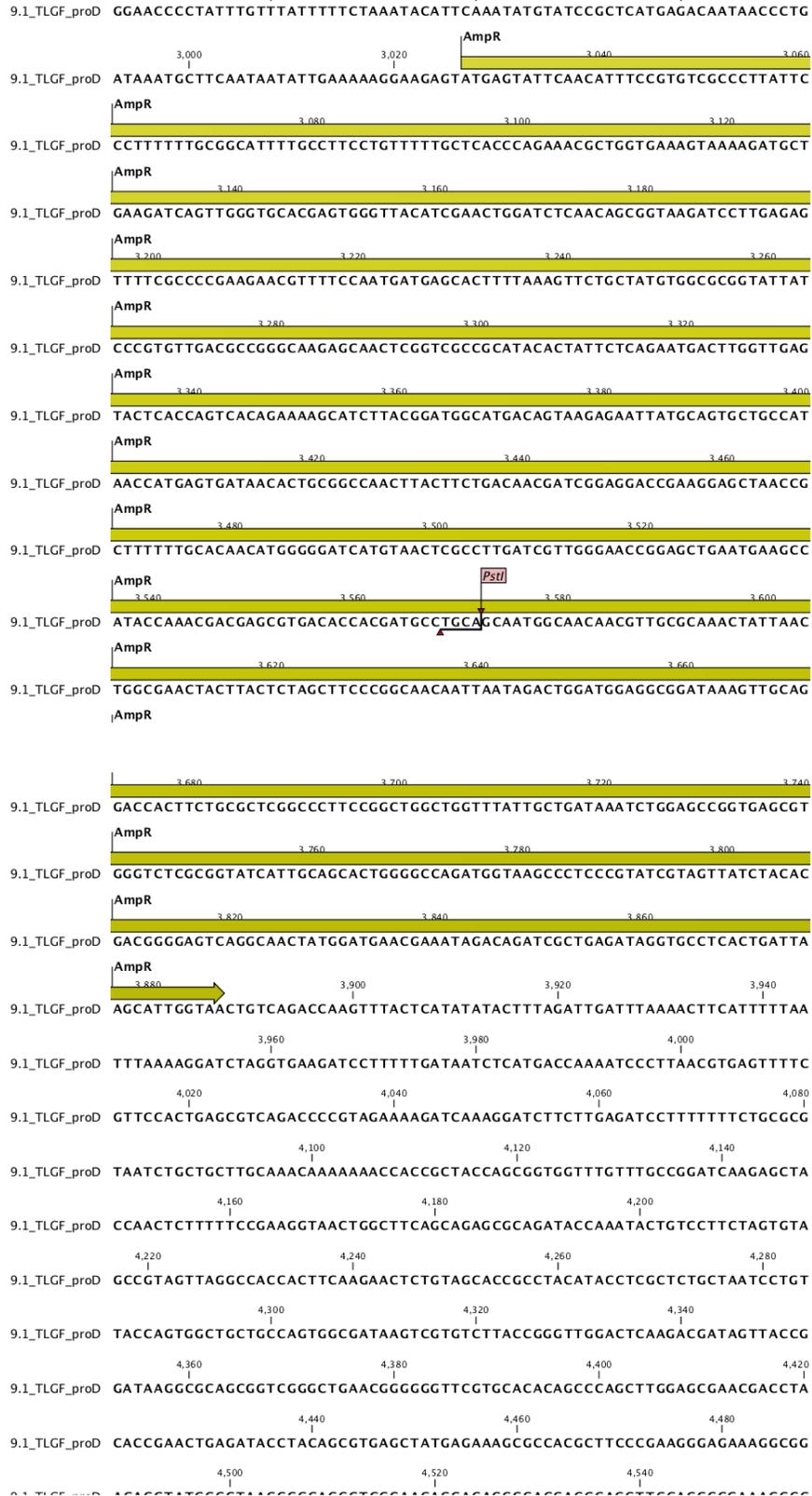
## B.2. Sequence of pRR-pbuE plasmid construct



### B.3 pRS plasmid (Complete sequence)







9.1\_TLGF\_proD <sup>4,200</sup> TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC  
<sup>4,200</sup> <sup>4,640</sup> <sup>4,660</sup> <sup>4,680</sup>  
 9.1\_TLGF\_proD AGGGGGGGGAGCCTATGGAAAAACGCCAGCAACCGGCCCTTTTACGGTTCCTGGCCTTTTGTGCGC  
<sup>4,700</sup> <sup>4,720</sup> <sup>4,740</sup> <sup>4,760</sup>  
 9.1\_TLGF\_proD CTTTGTCTCACATGTTCTTCTCGCTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTGG  
<sup>4,780</sup> <sup>4,800</sup> <sup>4,820</sup>  
 9.1\_TLGF\_proD TGAGCTGATACCGCTCGCCGACGCCGAACGCCGAGCGAGTCAGTGAGCGAGGAAGCGGAAGA  
<sup>4,840</sup> <sup>4,860</sup> <sup>4,880</sup>  
 9.1\_TLGF\_proD GCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTTACACCCGCATATGGTGCACTCTCA  
<sup>4,900</sup> <sup>4,920</sup> <sup>4,940</sup> <sup>4,960</sup>  
 9.1\_TLGF\_proD GTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCA  
<sup>4,980</sup> <sup>5,000</sup> <sup>5,020</sup>  
 9.1\_TLGF\_proD TGGCTGCGCCCCGACCCGCCAACCCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCC  
<sup>5,040</sup> <sup>5,060</sup> <sup>5,080</sup> <sup>5,100</sup>  
 9.1\_TLGF\_proD GCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGACAGGTTTTTACCCTCATCACCGAA  
<sup>5,120</sup> <sup>5,140</sup> <sup>5,160</sup>  
 9.1\_TLGF\_proD ACGCGCGAGGCAGCTGCGGTAAGCTCATCAGCGTGGTCTGTAAGCGATTACAGATGCTGCCTGTT  
<sup>5,180</sup> <sup>5,200</sup> <sup>5,220</sup>  
 9.1\_TLGF\_proD CATCCCGCTCCAGCTCGTTGAGTTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATG  
<sup>5,240</sup> <sup>5,260</sup> <sup>5,280</sup> <sup>5,300</sup>  
 9.1\_TLGF\_proD TTAAGGGCGGTTTTTCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTTACGGGGT  
<sup>5,320</sup> <sup>5,340</sup> <sup>5,360</sup>  
 9.1\_TLGF\_proD AATGATACCGATGAAACGAGAGAGGATGCTCAGGATACGGGTTACTGATGATGAACATGCCCGTTAC  
<sup>5,380</sup> <sup>5,400</sup> <sup>5,420</sup> <sup>5,440</sup>  
 9.1\_TLGF\_proD TGGAACGTTGTGAGGGTAAACAACCTGGCGGTA TGGATGCGGGGGACCAGAGAAAAATCACTCAGGGT  
<sup>5,460</sup> <sup>5,480</sup> <sup>5,500</sup>  
  
 9.1\_TLGF\_proD selection <sup>5,320</sup> <sup>5,340</sup> <sup>5,360</sup>  
 AATGATACCGATGAAACGAGAGAGGATGCTCAGGATACGGGTTACTGATGATGAACATGCC  
<sup>5,380</sup> <sup>5,400</sup> <sup>5,420</sup>  
 9.1\_TLGF\_proD selection GGTACTGGAACGTTGTGAGGGTAAACAACCTGGCGGATGGATGCGGGGGACCAGAGAAAA  
<sup>5,440</sup> <sup>5,460</sup> <sup>5,480</sup>  
 9.1\_TLGF\_proD selection ATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCA  
<sup>5,500</sup> <sup>5,520</sup> <sup>5,540</sup>  
 9.1\_TLGF\_proD selection GCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGACGGCGCTGACTTCCCGTTTCCA  
<sup>5,560</sup> <sup>5,580</sup> <sup>5,600</sup>  
 9.1\_TLGF\_proD selection GACTTTACGAAACACGGAACCGAAGACCATTTCATGTTGTTGCTCAGGTCGCAGACGTTTTG  
<sup>5,620</sup> <sup>5,640</sup> <sup>5,660</sup>  
 9.1\_TLGF\_proD selection CAGCAGCAGTCGCTTACGTTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCA  
<sup>5,680</sup> <sup>5,700</sup> <sup>5,720</sup>  
 9.1\_TLGF\_proD selection ACCCCGCCAGCCTAGCCGGTCTCAACGACAGGAGCACGATCATGCGCACCCGTGCCAGG  
<sup>5,740</sup>  
 9.1\_TLGF\_proD selection ACCCAACGCTG



### B.5. pRS- *pbuE*/*pbuE*\*

