# Investigating the Utility of a Bacterial Auxotroph System for Novel Riboswitch Selection

**Nickles Badger**

Department of Biochemistry

April 6, 2023

Thesis Advisor:

Dr. Robert Batey

Department of Biochemistry

Committee Members:

Dr. Jeffrey Cameron

Department of Biochemistry

Dr. Jennifer Martin

Department of Molecular, Cellular, and Developmental Biology

# Table of Contents

*I. Acknowledgements* ..... *2*

*II. Abstract* ..... *3*

*III. Introduction* ..... *4*

The Riboswitch as a Gene Regulatory Element ..... 4

The *pbuE* Riboswitch and Derivatives ..... 8

An Auxotroph System for Selection ..... 10

*IV. Materials and Methods* ..... *13*

Overview of Bacterial Auxotroph System Approach for Riboswitch Selection ..... 13

Plasmid Construction ..... 14

Colony PCR ..... 18

Media Preparation ..... 19

Competent Cell Preparation ..... 20

Transformations ..... 22

Drop Plating ..... 24

Gel Electrophoresis ..... 24

Replica Plating ..... 25

*V. Results* ..... *26*

Strain Behavior and Media Optimization ..... 26

Ligation Assembled J2/3 Library and Evaluation ..... 28

CPEC Assembled J2/3 Library and Evaluation ..... 33

Plasmid modifications for counterselection ..... 34

*VI. Discussion* ..... *37*

Future Directions and Orthogonal Approaches ..... 37

Importance of Novel Riboswitch Development ..... 39

*VII. References* ..... *40*

*Appendix A* ..... *42*

*Appendix B* ..... *43*

# I. Acknowledgements

I would like to thank all members of the Batey lab for their support. Savannah Spradlin, Desmond Hamilton, and Lukasz Olenginski were especially helpful, offering useful insights on many occasions. I express my appreciation to Ankur Sarkar, Andrew Markley, and Nolan O'Connor for engaging in ongoing productive dialogues. I would also like to thank Dr. Jennifer Martin and Dr. Jeffrey Cameron for their support and for serving on my honors committee. Most importantly I would like to thank Dr. Robert Batey for the opportunity to embark on this project and for the latitude to explore different research avenues. Rob inspires with wise words and he leads by example at the lab bench, in the department, and in the broader scientific community. I thank The Biological Sciences Initiative and Lisa Romero de Mendoza for funding as the work herein would not have been possible without this support.

# II. Abstract

Riboswitches serve as metabolite-sensitive RNA regulatory elements in the leader sequences of many prokaryotic transcripts. Synthetic riboswitches selectively responsive to medically or industrially relevant small molecules have potential as useful mechanisms for dosage regulation and biosensing applications. Development of such novel functional RNAs has remained challenging due to large library sizes, throughput limitations of screening approaches, and unreliable *in vivo* function of riboswitches obtained via *in vitro* methods. To overcome these impediments, a bacterial double auxotroph selection system initially employed for transcription factor binding specificity determination was repurposed to investigate the utility of the approach for identifying novel riboswitches responsive to alternate ligands from large libraries (Meng & Wolfe, 2006). Library sequences used in the system possessed a riboswitch scaffold with randomized bases in the aptamer positioned upstream of positive and negative selectable markers, *hisB* and *URA3* in a bicistronic reporter plasmid. In histidine drop-out medium, members with a riboswitch responsive to ligand were conferred a survival advantage by reduced attenuation in expression of *hisB*, an essential enzyme in histidine biosynthesis. Validation experiments using P4C, a known functional derivative of the purine responsive *pbuE* riboswitch, inserted upstream of the reporter genes, demonstrated ligand dependent survival and potential utility of the system for library selection. Furthermore, a P4C scaffolded library influencing reporter genes in a selection strain grown in solid drop-out media with optimized 3-aminotriazole inhibitor and 2-aminopurine ligand was subjected to positive selection. Results suggest library members which inhibit expression even in the presence of purine ligand were successfully selected against. Mutant library members and other constructs conferring a survival advantage and outcompeting properly constructed members despite the lack of a functional riboswitch remain a challenge. Approaches to mitigate against the introduction of these evaders were investigated and robust counterselection appears to be necessary to achieve successful implementation of an auxotroph-based riboswitch selection tool. Although continued efforts are necessary, the approach offers a potential avenue to select for therapeutically applicable biosensors responsive to alternate ligands from large libraries.

# III. Introduction

## The Riboswitch as a Gene Regulatory Element

A large diversity of mechanisms has evolved to regulate gene expression including epigenetic modifications, transcriptional and translational modulation, post translational modifications, and macromolecule degradation. As these mechanisms have become better understood, new opportunities have arisen to manipulate them for medical and technological applications. Among these are noncoding RNAs, once thought to serve few functions beyond central dogma roles of tRNAs and rRNAs. Over the past few decades a number of classes of new noncoding RNAs have been discovered (Patil et al., 2014). An important type of noncoding RNAs are riboswitches, regulatory elements found in the 5' UTRs of some transcripts, which modulate gene expression in response to small molecule ligands (Mandal & Breaker, 2004). Riboswitches are common in bacteria and have also been found in archaea and eukaryotes *(McCown et al., 2017)*. These regulatory RNAs are classified based on the ligands which they bind, and based on their structure, with dozens of classes identified (Montange & Batey, 2008).

Riboswitches modulate expression at the translational or transcriptional level, folding into alternate conformations depending on ligand binding and thereby modulating transcript polymerization or altering binding dynamics of transcript to ribosome (Garst & Batey, 2009). Ligands recognized by these RNAs are often intermediate metabolites or cofactors in regulatory pathways (Garst et al., 2011). Downstream genes in turn, often function to regulate intracellular concentrations of the metabolite. Consequently, in many cases, riboswitches indirectly maintain homeostatic concentrations of the small molecule to which they bind (Porter et al., 2017). Translational riboswitches regulate protein expression by folding into conformations which sterically occlude the ribosome binding site (RBS) on the transcript or leave the site available for efficient mRNA binding and subsequent translation (Lemay et al., 2011). The modulatory outcome is controlled by the ligand dependent riboswitch conformation. Transcriptional riboswitches modulate expression by influencing functional transcript synthesis of the downstream gene. These transcriptional *cis* regulatory elements can

function as ON or OFF riboswitches, up or down regulating gene expression respectively in response to ligand binding. Discussions of riboswitches here forth will refer to transcriptional riboswitches as they are the focus of this investigation.

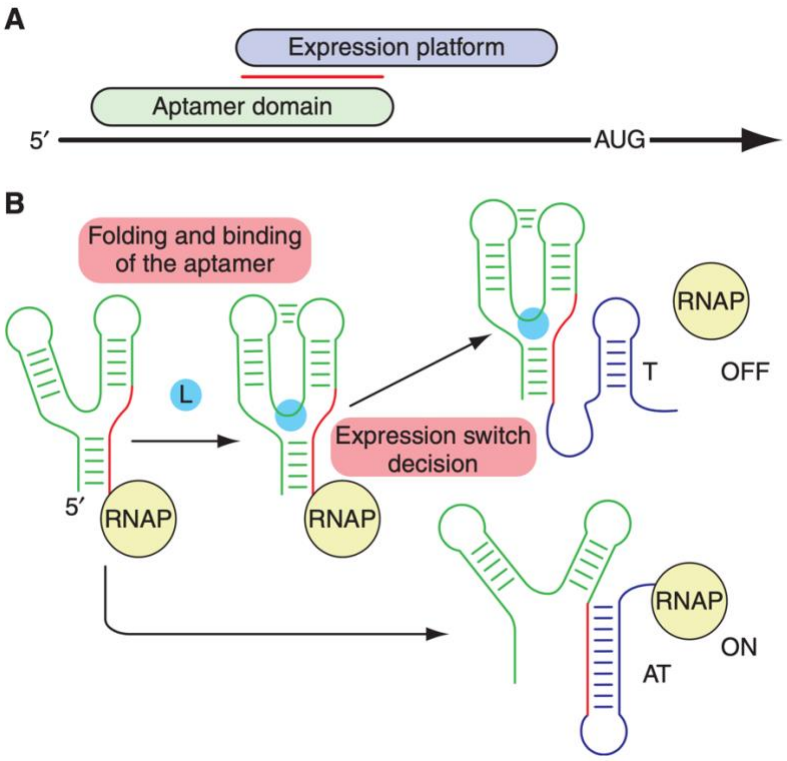


*Figure 1: Schematic of generalized riboswitch. (A) The aptamer domain and the expression platform are constituents of the riboswitch situated in the 5' leader sequence of an mRNA. The two domains overlap in sequence. (B) Regulatory outcome is ligand dependent. In the OFF riboswitch represented, ligand binding stabilizes the aptamer leading to concomitant stabilization of a terminator hairpin in the expression platform. With no ligand bound, strand exchange occurs and an anti-terminator forms affording full transcription. Figure from Garst et al. 2011*

Riboswitches are encoded between the promoter and the Shine Dalgarno sequence in the 5’ UTR of genes they influence (figure 1). These regulatory elements consist of two domains, an aptamer, and an expression platform. The aptamer contains a binding pocket accommodating a small molecule recognized with high specificity which stabilizes the tertiary structure when bound (Batey et al., 2004). The aptamer conformation and the intrinsic terminator structure in the case of ON riboswitches or anti-terminator in the case of OFF riboswitches are mutually exclusive (figure 1). Transitions are mediated by ligand binding or release. Strand exchange occurs in which the strand comprising the 3’ side of the aptamer disassociates and invades the

expression platform to contribute to the establishment of the alternate conformation (Garst et al., 2011).

Stabilization of an intrinsic hairpin disrupts RNAP binding leading to polymerase disassociation from the template DNA, consequent termination of transcription, and a resulting incomplete transcript. The net result is downregulated expression of the downstream gene or genes. Alternatively, in the presence of sufficiently high concentrations of ligand, binding stabilizes the aptamer, dynamically inhibiting timely strand exchange from occurring (Porter et al., 2014). In the case of an ON riboswitch, ligand binding prevents the intrinsic terminator from forming in close proximity to active RNAP, and transcription proceeds into the gene body sequence, resulting in full gene expression. In these riboswitches, the terminator helix is the more thermodynamically stable structure and the ON conformation is dynamically stabilized in a kinetically driven manner (Gilbert et al., 2006). Many riboswitches of this class, bind nucleobases or amino acids and regulate intracellular concentrations of the metabolite via metabolic or efflux regulation mechanisms. Riboswitches functioning in this way, include the well characterized natural guanine responsive *pbuX* and adenine responsive *pbuE* riboswitch.

## Synthetic Aptamer Development Approaches

Novel riboswitch development efforts generally have focused on performance improvement or adaptation for response to alternate small molecules (Porter et al., 2017). Progress has been made in both areas however substantial challenges remain. Successful efforts to increase fold induction, mitigate leaky expression, and otherwise improve performance have focused on modifying the expression platform while preserving overall aptamer structure to maintain ligand binding ability (Marcano-Velazquez & Batey, 2015). Investigations aimed at synthesizing riboswitches responsive to choice, alternate, ligands have involved strategically modifying the aptamer domain while usually making more modest modifications to the expression platform. Others have demonstrated the utility of modular approaches, mixing and matching components from different parent riboswitches (Ceres et al., 2013). In these cases, and in essentially all

successful efforts to develop novel riboswitches, designs begin with scaffolds of naturally evolved riboswitches. To understand previous efforts to synthesize novel riboswitches, examples of aptamer development against alternate targets will be discussed before considering examples of performance improvements achieved via expression platform modifications.

Many riboswitches have been developed to respond effectively to alternate ligands often beginning with scaffolds of natural riboswitches. In an early example, the relatively simple guanine binding *pbuX* riboswitch was altered to bind adenine. In the native riboswitch, a canonical Watson-Crick base pairing with a key cytosine residue in the J3/1 loop, forms the core of the binding pocket (Gilbert et al., 2006). The investigators made a simple C74U mutation successfully altering the natural guanine riboswitch to an adenine responsive mechanism. Here, the natural aptamer sequence was retained, excepting only the exchange of the single ligand binding pyrimidine residue in the binding pocket. Altering aptamer recognition to more structurally divergent small molecules has proved substantially more difficult. In the case of purine and other riboswitches, changes to the 3' region of the aptamer often destroys the strand exchange capability of the switch and/or the capability to form an intrinsic terminator (Mandal & Breaker, 2004). To successfully identify sequences that will induce expression upon binding an alternate ligand while also maintaining the ability to form an intrinsic terminator in the absence of ligand, usually requires the evaluation of a large library of sequences. SELEX and other methods have been used for *in vitro* aptamer development which has afforded the ability to generate aptamers responsive to a wide range of ligands however a major limitation is the lack of functionality of many these novel elements when placed in a cellular environment (Darmostuk et al., 2015). SELEX approaches using libraries scaffolded with natural riboswitches have successfully made headway overcoming some of these challenges with aptamers recently raised against neurotransmitters (Porter et al., 2017). Nonetheless, while *in vitro* approaches such as SELEX have advantages, RNAs generated using them still often do not function as expected when implemented in a cellular context.

Other synthetic riboswitch development efforts have focused on improved performance against ligands recognized by natural or synthetic riboswitchwes. Key goals have included improving fold induction and mitigating leaky expression, often focused on expression platform modifications while only making compensatory changes to the aptamer domain (Drogalis & Batey, 2020; Marcano-Velazquez & Batey, 2015). An important method for improving and evaluating performance of natural and artificial riboswitches, is florescence reporter assays, which provide accurate data on fold induction and leaky expression. The method has been used successfully to evaluate scaffolded libraries constructed with degenerate sequences in the expression platform of purine riboswitches. The development and evaluation of rationally designed riboswitches particularly from natural purine riboswitch scaffolds using fluorescence reporter assays has led to vast enhancements in recent years. These screening methods however, have limited throughput, requiring time intensive effort and implemented design principles to restrict library size. The high performance, *pbuE* derived, synthetic P4C riboswitch was developed using these methods.

## The *pbuE* Riboswitch and Derivatives

The *pbuE* riboswitch, natural occurring in *Bacillus subtilis*, controls expression of the PbuE purine efflux pump known to facilitate the removal of inosine, adenosine, guanosine, and hypoxanthine from the bacteria (Johansen et al., 2003). This ON riboswitch functions in an adenine-dependent manner, allowing unattenuated transcription of the efflux pump gene when the ligand is present at elevated concentration (figure 2). At reduced concentrations of adenine, as the nascent transcript is synthesized, the riboswitch aptamer loses stability allowing strand exchange and the formation of an intrinsic terminator effectively downregulating efflux pump expression affording retention of purines (Porter et al., 2014). The *pbuE* riboswitch and related mRNA purine control elements are among the simplest riboswitches, making them important models for understanding riboswitch structure and function. These regulatory elements have been the subject of extensive study and are well understood.

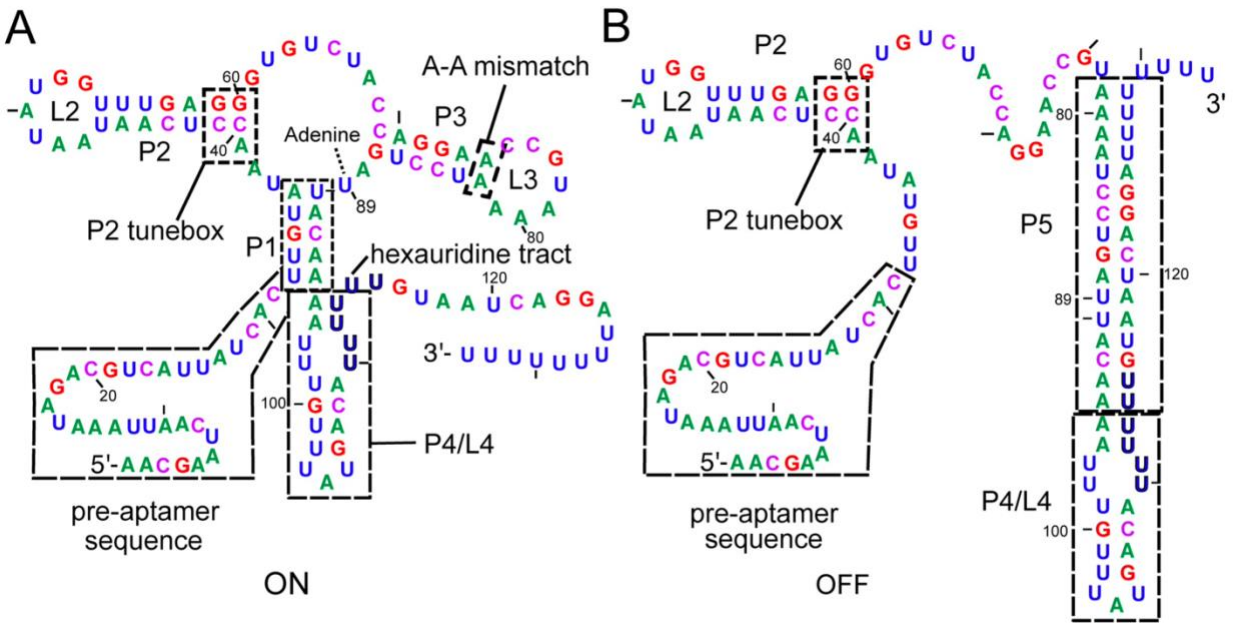


*Figure 2: Secondary structures of the pbuE wild type riboswitch. (A) ON conformation secondary structure. Note U89 which forms a canonical base pair with adenine or 2AP ligand resulting in unattenuated transcription. (B) When the ligand is unbound, an intrinsic terminator hairpin forms leading to polymerase dissociation. Note, modifications to this wild type riboswitch in P4C included removal of the pre-aptamer sequence and P4/L4 alteration. Figure taken from Drogalis & Batey 2020*

The scaffolds of purine riboswitches have proven useful in the development of rationally designed artificial riboswitches. Specifically, the scaffold of *pbuE* has served as the core sequence in successful fluorescence reporter guided efforts for development and validation of a synthetic adenine responsive riboswitch with a remarkable 120-fold induction compared with the wildtype which exhibits a <6 fold induction (Drogalis & Batey, 2020). Key alterations to the wild type scaffold included a 27-nucleotide deletion of the pre-aptamer sequence which eliminated complicating and unnecessary secondary structures, a truncated P4 stem, and the inclusion of a strong pause sequence (figure 3).

The high performing variant P4C was employed for validation experiments in the study at hand. The core of this artificial regulatory sequence also served as the scaffold for a library used in this study for preliminary selection experiments.

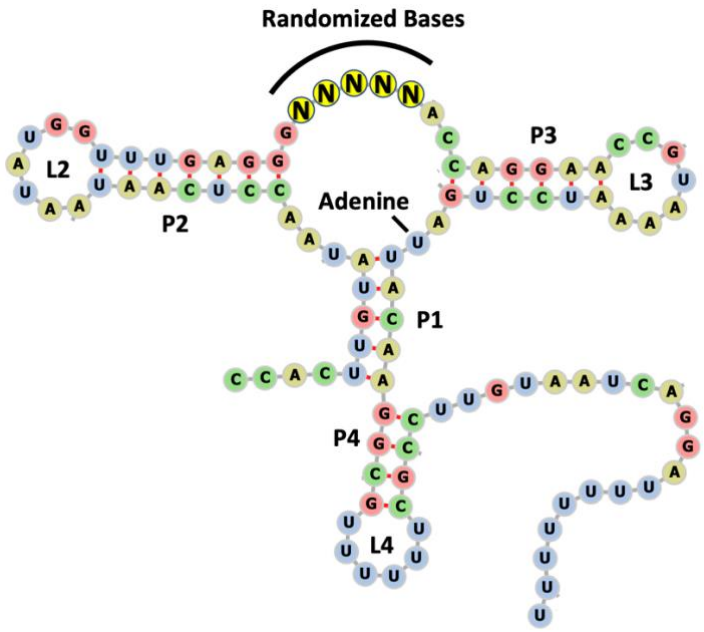


*Figure 3: Secondary structure of the ligand bound pbuE derived P4C riboswitch based on Drogalis & Batey 2020. Alterations to the wild type included the elimination of the pre-aptamer sequence and a truncated P4. A library with five nucleotides in the J2/3 loop randomized was used in this study.*

## An Auxotroph System for Selection

A fundamental challenge of developing synthetic riboswitches responsive to alternate ligands lies in the need to scan vast possible sequence permutations for functional switches that effectively form a suitable binding pocket and communicate with an expression platform for efficient switching (Miller et al., 2022). Given the inability of standard screening approaches to efficiently evaluate large libraries, and the failure of many *in vitro* developed riboswitches to function *in vivo*, identifying functional riboswitches with novel aptamers by selection in an *E. coli* system stands as an attractive alternative.

Survival based selection is a promising high throughput approach successfully applied to evaluation of similarly high complexity libraries in the context of transcription factors (Serebriiskii & Joung, 2001). Yeast auxotroph hybrid systems have been in use for three decades to investigate macromolecular interactions with the evaluation of complex libraries by selection on drop out media (Wilson et al., 1991). In early implementations, investigators employed a

*HIS3* and *URA3* knockout *Cerevisiae Streptomyces* strain. In conjunction with a reporter plasmid containing these genes, DNA binding specificity was determined for the transcription factor NGFI-B expressed as a chimera in a second plasmid. Variations of this system have been applied in yeast to meet a number of similar synthetic biology challenges where selection against large libraries is necessary (Vidal, 1999). More recently, the approach was adapted for use in *E. coli.*

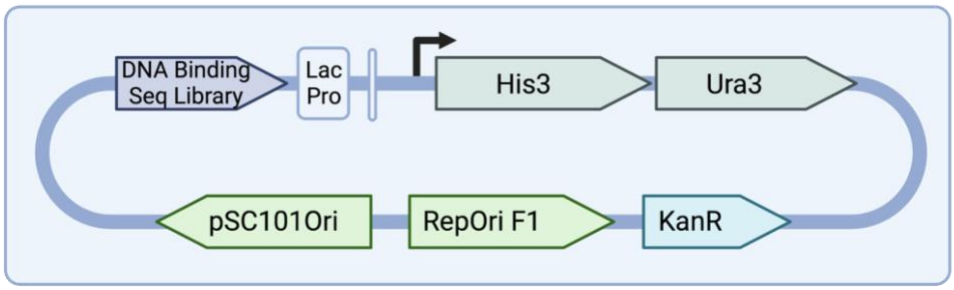


*Figure 4: Reporter plasmid constructed by Meng & Wolfe 2005 for determining transcription factor binding specificity in conjunction with an E. coli auxotrophic strain and B1H plasmid. HIS3 and URA3 are heterologous reporters used in this implementation.*

A key advantage to an *E. coli* based implementation of the platform is a 3-4 order of magnitude higher transformation efficiency obtainable in the bacteria (Singh & Weil, 2002). Investigators employed a bacterial one hybrid system consisting of a plasmid expressing a transcription factor of interest linked to the α-subunit of RNAP (Meng et al., 2005). A second reporter plasmid contained selectable marker genes *HIS3* and *URA3* in a bicistronic construct under control of a weak lac promoter (figure 4). An insert was placed up stream of the promoter, constituting a library of potential DNA binding sequences. The investigators made an accompanying knockout using a DH5α parental strain, removing *pyrF* and *hisB* from the *E. coli* genome yielding a double auxotroph strain. Using the histidine/uracil auxotroph, a survival advantage was conferred to members with enhancer sequences recognized by the transcription factor DNA binding domain as a result of RNAP recruitment by the linked α-subunit, and consequent elevated transcription (Meng et al., 2005).

In this system, the *HIS3* marker served as the positive selection gene and *URA3* afforded negative selection. The HIS3 enzyme in yeast and hisB enzyme in *E. coli* are dehydratases in the

histidine biosynthesis pathway. Binding of the α-subunit linked transcription factor fusion protein to a recognized member of the sequence library yields upregulation of *HIS3* transcription and thus cells harboring plasmids with the favorable library member were able to survive on histidine dropout media. The competitive *HIS3*/*hisB* inhibitor, 3-aminotriazole was used to modulate stringency to mitigate undesired survival due to expression in the absence of transcription factor binding (Meng et al., 2006). The use of 5-fluoroorotic acid (5-FOA) afforded the option for counter selection. The yeast URA3 enzyme and *E. coli* pyrF enzyme are decarboxylases functioning in the pyrimidine biosynthesis pathway. URA3 enzyme accepts 5-FOA as substrate, decarboxylating it into toxic fluorouracil. Counterselection was performed by growing library transformants in media containing 5-FOA without inducing expression of the fusion protein. Consequently, any cells containing reporter plasmids with self-activating sequences were eliminated as a result of URA3 expression in the absence of transcription factor binding, yielded fluorouracil. Any cells containing plasmids with non-self-activating sequences theoretically survived negative selection.

# IV. Materials and Methods

## Overview of Bacterial Auxotroph System Approach for Riboswitch Selection

To investigate the utility of a bacterial auxotroph selection platform to select for novel riboswitches responsive to alternate small molecules, the approach of Meng et al., 2006 was repurposed. Targeting functional RNAs in this approach instead of DNA binding proteins obviated the need for chimera expression from a second plasmid, allowing the deployment of a single plasmid library. Initially, the pH3U3 parent plasmid (Meng & Wolfe, 2005) was used to construct pHBU3, followed by inserting the P4C riboswitch between the promoter and the Shine Dalgarno sequence influencing reporter genes *hisB* and *URA3*. This plasmid was used to evaluate the behavior of USOΔHisBΔpyrF auxotrophic cells with a reporter plasmid containing a known functional riboswitch. A 1024-member aptamer library, scaffolded by the P4C core sequence, consisting of an aptamer degenerate at five consecutive bases in the J2/3 loop, was used as a test library to for assay development (figure 3, figure 5).

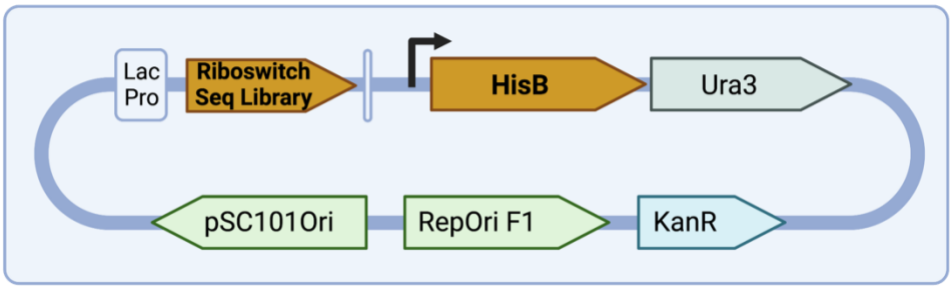


*Figure 5: Reporter plasmid with 1024-member J2/3 library for riboswitch selection validation. Highlighted components are modifications from the parent plasmid. Note the position of the library and hisB gene substituting HIS3.*

Constructed plasmid libraries were transformed into TOP10 cells for expansion (figure 6). Cultures were then grown on nonselective solid media and recovered. Library plasmids were extracted and then purified. The complete purified library was then subjected to positive selection to obtain plasmids possessing functional riboswitches responsive to 2AP ligand, an analog of adenine and known activator of P4C.

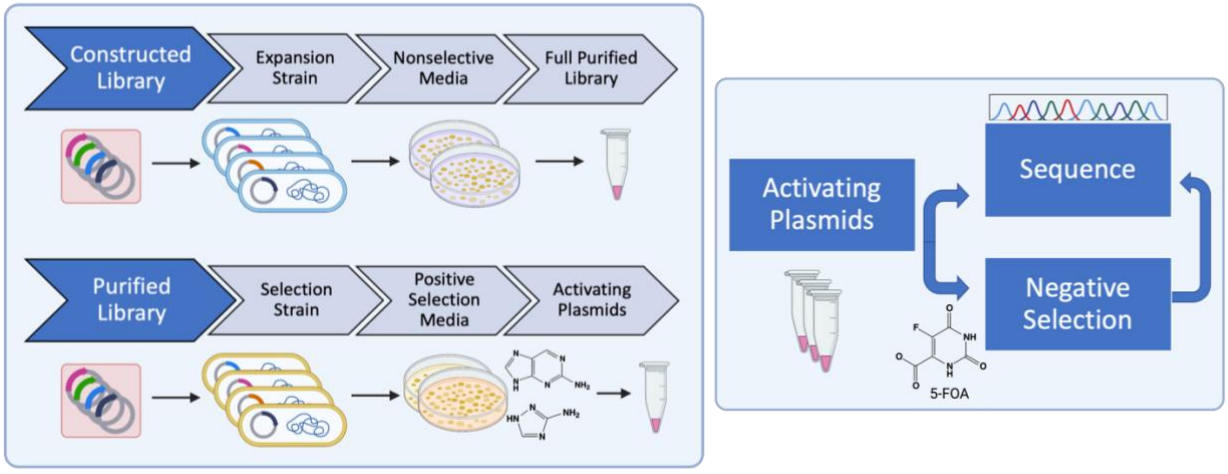


*Figure 6: Workflow for riboswitch selection platform.* ***Left:*** *Constructed unpurified library is first transformed into the expansion strain followed by growth and purification. The purified library is then transformed into the selection strain and grown on positive selection histidine dropout media with 2AP and 3AT. Activating plasmids are purified.* ***Right****: Activating plasmids are individually sequenced or subjected to another round of transformation and negative selection before sequencing.*

To accomplish this, the purified plasmid library was transformed into the selection strain and plated on solid positive selection media without histidine (figure 6). Positive selection media contained 2AP to activate functional riboswitches and upregulate *hisB* expression. Selection media also contained 3AT to inhibit catalysis by hisB enzyme expressed at basal levels due to leaky expression by nonfunctional members. Subsequently, purified activating plasmids were individually sequenced. Alternatively, negative selection on histidine containing media with 5-FOA and without 2AP could be performed with another round of transformation, selection, and purification prior to individually sequencing plasmids. Negative selection would theoretically eliminate evasive members as URA3 enzyme expressed in the absence of ligand would catalyze the conversion of 5-FOA into toxic fluorouracil.

## Plasmid Construction

All plasmids assembled and used in this study possess a vector backbone from parent plasmid pH3U3-mcs (Addgene plasmid # 12609; http://n2t.net/addgene:12609 ; RRID:Addgene_12609) constructed by Scot Wolfe and Xiangdong Meng (Meng et al., 2005). Selectable markers in the

parent plasmid include *HIS3* and *URA3*, heterologous genes from *S. cerevisiae* for positive and negative selection respectively. A weak lac promoter drives expression of reporter genes. The vector possesses kanamycin resistance gene *kanR* and replicates at low copy number with a PSC101 origin of replication. Two approaches were employed to assemble plasmids and plasmid libraries, ligation assembly and CPEC.

*Ligation Assembly*

Inserts were amplified from genomic DNA, existing plasmids, or from ultramers synthesized by Integrated DNA Technologies (IDT), and then mini-column purified using the Omega Bio-Tek E.Z.N.A. Cycle Pure Kit. Inserts were double digested using a pair of NEB High Fidelity restriction enzymes of the following: KpnI-HF, SalI-HF, and BamHI-HF. Reactions were prepared using 43 µL of purified DNA at a concentration of 50-150 ng/µL in 5 µL 10X NEB Cutsmart buffer, and 1 µL of each restriction enzyme. Digestions were run for 1 hour at 37 ˚C followed by mini-column purification. To improve ligation efficiency, inserts were first phosphorylated at 5' hydroxyls using homemade T4 polynucleotide kinase (PNK). These PNK reactions were prepared using 300-2000 ng of purified insert eluted in 15 µL $ddH_2O$, 1 µL PNK (1 U), 2 µL 10 mM ATP, and 2 µL 10X PNK buffer. These phosphorylation reactions were run for 1 hour at 37 ˚C followed by a 20-minute PNK deactivation at 65 ˚C.

Backbone DNA was prepared by first extracting plasmids from TOP10 cell overnight cultures using Omega Bio-Tek E.Z.N.A. DNA Miniprep Kit I. Digestions were executed in 100 µL reactions using 2-12 µg of parent plasmid in 86 µL of $ddH_2O$, 10 µL of 10X NEB Cutsmart buffer, and 1.5 µL of each restriction enzyme, in addition to 1 µL (1 U) of calf intestinal phosphatase (CIP). CIP was used to remove 5' phosphates to reduce undesired re-ligation of backbone ends.

After preparation of insert and backbone, ligation reactions were carried out using 90-100 ng of total DNA in a 3:1 molar ratio of insert to backbone. Total DNA mass was increased to ~100-150 ng for libraries and insert proportion increased to 5:1 or 7:1. Total reaction volumes for

ligations were 20 µL using 1 µL of NEB T4 DNA ligase and 2 µL of 10X NEB ligase buffer with $ddH_2O$ used to reach total volume. Controls were run using backbone DNA and $ddH_2O$ instead of insert. Ligation reactions were run for 1 hour at 25 ˚C before transforming into competent cells.

*Circular Polymerase Extension Cloning (CPEC)*

CPEC was employed as an additional approach for plasmid assembly, library assembly, and site directed mutagenesis. CPEC assemblies were performed using the standard protocol with minor deviations (Quan & Tian, 2011). Initially 100 µL PCR reactions were run for each DNA component with primers designed to yield high Tm double stranded overlaps. PCRs were purified, quantified, and used at a 2:1 molar ratio of insert to backbone with a total DNA mass of 120-160 ng for CPEC reactions. Volumes of 0.2 µL high fidelity Q5 polymerase, 1 µL DMSO, 0.4 µL dNTPs (10 mM per nucleotide triphosphate), and 4 µL 5X Q5 buffer were used, with $ddH_2O$ to bring the total reaction volume to 20 µL. The reactions were run on a Bio-Rad T100 Thermocycler, initially denaturing for 30 seconds at 98 ˚C. The following steps were repeated for 25 cycles: denaturation for 10 seconds, slow ramp annealing from 70 ˚C to 56 ˚C in 1 ˚C steps over 3 minutes, annealing at 55 ˚C for 30 seconds, and extension at 72 ˚C for 2-3 minutes. An extension for 5 minutes at 72 ˚C served as the final step of these reactions.

*Site Directed Mutagenesis (SDM)*

SDM was employed to remove or to introduce KpnI sites from the gene body or the 5’ UTR respectively, of pH3U3. Codon degeneracy was exploited to accomplish restriction site removal while maintaining amino acid sequence. Primers were designed using the Agilent Quick Change Primer Design Tool (Novoradovsky et al., 2005) to optimize oligonucleotides. The CPEC protocol described herein was performed to make these mutations.

*pHBU3 Construction*

Extracted and purified genomic DNA from a TOP10 *E. coli* cell line was used as template to obtain the *hisB* gene by PCR amplification. Primers were designed with overhangs to yield amplicon with KpnI and BamHI restriction sites at 3' and 5' ends respectively. Following PCR amplification and gel electrophoresis verification of the ~1150 bp DNA product, a double restriction enzyme digest reaction was run to introduce the desired sticky end overhangs. A KpnI and BamHI reaction was concurrently run to digest purified pH3U3-mcs plasmid followed by concomitant gel verification. Subsequently, the amplified HisB gene and linearized backbone vector each with complementary four nucleotide restriction site overhangs were mini-column purified. A ligation reaction was run using a 3:1 mole ratio of insert to plasmid and transformed into RbCl competent TOP10 cells. A complementary reaction run without *HisB* served as a negative control. The primary impetus for building the pHBU3 plasmid was to introduce a key KpnI restriction site between the Lac promoter and the Shine-Dalgarno site upstream of reporter genes in preparation for insertion of functional riboswitches and libraries.

*P4C-pHBU3 Construction*

The P4C riboswitch sequence was PCR amplified from plasmid template using primers designed with overhangs to yield DNA amplicon with SalI and KpnI restriction sites at 3' and 5' ends respectively. The pair of corresponding restriction enzymes was employed to digest the resulting DNA insert in parallel with digestion of purified newly assembled pHBU3 plasmid for subsequent ligation. The 120 bp digested insert was gel verified before assembly and subsequent sequence verification.

*Intergenic Modifications*

In efforts to improve expression of the second reporter gene, *URA3,* in the bicistronic elements of pHBU3 and pH3U3 for counterselection, plasmid sequences were evaluated to identify sequence modifications to improve translation (Appendix B). To compute intergenic sequence modifications expected to achieve this, the operon Salis calculator was employed (Cetnar & Salis, 2021). The algorithm utilizes thermodynamic and accompanying structural modeling to estimate and predict translation initiation rate, translation coupling rate and elongation rate for bacterial operons. Parameters were chosen in aim to elevate translation of URA3 while minimally modifying expression of the first gene in the operon, HisB. Three primer pairs were designed to modify the intragenic sequence predicted to elevate gene two expression by increments. CPEC was employed to assemble the modified plasmids. New constructs were evaluated in liquid 2xYT media with 5-FOA to assess improvements in response to the compound due to URA3 expression.

## Colony PCR

Colony PCRs were performed for initial assessment of plasmid library construction to indicate the presence or absence of an appropriately sized insert or presence of spurious constructs. After transformation of ligation assemblies or CPEC library assemblies, a subset of colonies outgrown on solid media were used as template for amplification of a target sequence extending from ~114 bp upstream of the lac promoter through the HisB gene, to the intragenic operon region of the backbone. The amplicon product for plasmid library members containing a correctly sized insert were expected to be of length 1352 bp. Based on availability and target region coverage, primers pBRforEco (designed for sequencing) and RBS5e_strong_rev (designed for RBS modification) were chosen for these PCRs. Although better resolution could be obtained using specially designed primers covering a more constrained region around the 5' UTR, use of existing primers proved effective at initially discerning plasmid library quality and the presence of evasive constructs prior to sequencing for several library assemblies. Colony

PCRs were prepared by making a master mix and aliquoting 8-10 µL into 100 µL PCR tubes and adding a picked colony or control template to each well. Control reactions were run using a final template concentration of 400-600 pg/uL.

## Media Preparation

For TOP10 cells, 2xYT rich media was prepared using standard protocols. LB liquid media was used for outgrowth in preparation of chemically competent and electrocompetent TOP10 cells. Plates and liquid 2xYT media were prepared with 50 µg/mL kanamycin and/or 10 µg/mL tetracycline.

*Drop out amino acid media*

Utilization of the histidine auxotroph strain for selection required preparation and use of amino acid drop out media (Serebriiskii & Joung, 2001). To achieve the necessary media composition, Sigma Aldrich reagent SA Y2021 was used containing all essential amino acids and nucleobases except histidine, adenine, tryptophan, and leucine. Dropout solutions for use in media were prepared at 10X concentration in 100 mL batches. Reagents were prepared at concentrations of 13.9 mg/mL SA Y2021 DOAAs, 0.76 mg/mL tryptophan, and 3.8 mg/mL leucine in a volumetric flask in MiliQ water. Preparations were stirred on a hotplate at 60 ˚C for ~1 hour to completely solvate all components. The solutions were cooled to 25˚C and then syringe filter sterilized through 0.22 µm pores before storing at 4˚C.

CSB salts were diluted to standard concentration and autoclave sterilized. Complete liquid drop out amino acid media was prepared by combining 10X DOAA solution with media supplement solution, filter sterilizing, and adding to autoclaved salts solution. DOAA solid media was prepared by adding agar to the salt solution prior to autoclave sterilization to yield a final agar concentration of 1.8%.

## Competent Cell Preparation

Two *E. coli* strains were used in this investigation - TOP10 cells for general transformations and USOΔhisBΔpyrF cells for selection. TOP10 cells were propagated from commercial Thermo Fischer ancestors. The USO knockout strain was made by Scot Wolfe from a DH5α parental strain and contain the *tetR* gene conferring tetracycline resistance. Several types of competent cells were prepared and used in this study. For initial experiments to assess plasmid and strain behavior, calcium chloride competent cells were prepared for immediate use. Rubidium chloride competent cells were made to attain higher transformation efficiency. To achieve maximum transformation efficiency electrocompetent TOP10 and USO cells were also prepared.

*Calcium Chloride Competent Cells*

Calcium chloride competent USOΔhisBΔpyrF DH5α cells were prepared for initial experiments to assess new plasmid constructs and make glycerol stock. First, cells were streaked out from agar stab (Add Gene Strain #12614) onto a tetracycline 2xYT plate and grown overnight. A single colony was selected to inoculate a 5 mL 2xYT liquid starter culture containing tetracycline which was incubated overnight to yield a culture suitable for expansion. Solutions of 100 mM $CaCl_2$, 100 mM $MgCl_2$, and 85 mM $CaCl_2$ plus 15% glycerol were made and chilled at 4 ˚C (common stocks of these reagents were used). A baffled flask with 100 mL 2xYT with tetracycline served as the expansion media. A starter culture of 1 mL was transferred into the expansion culture and grown at 37 ˚C shaking at 220 RPM. $OD_{600}$ measurements were taken every 15 minutes after 2.5 hours of growth. Due to required absorption of both uracil and histidine by the double auxotroph strain in the absence of rescue plasmid, approximately 4-5 hours of growth was necessary to reach a desired $OD_{600}$ of ~0.35. Following outgrowth, the culture was transferred to falcon tubes, incubated on ice for 20 minutes and then centrifuged in prechilled rotor buckets for 18 minutes at 3400 RPM at 4 ˚C in a tabletop centrifuge. The supernatant from each tube was decanted and pellets were resuspended in 20 mL of 100 mM $MgCl_2$ by pipetting up and down and inversion. The suspensions were subjected to a second

centrifugation for 15 minutes at 3000 RPM at 4 ˚C. The supernatant was decanted, and the pellet resuspended in 20 mL of 100 mM $CaCl_2$. The suspension was allowed to chill on ice before a third centrifugation for 15 minutes at 4 ˚C. After decanting the supernatant, the pellets were resuspended in 10 mL of 85 mM $CaCl_2$, 15% glycerol solution and centrifuged a final time for 15 minutes at 4 ˚C. The supernatant was decanted after the final wash, resuspended in 400 µL of 85 mM $CaCl_2$, 15% glycerol solution and aliquoted into chilled microcentrifuge tubes. Prepared $CaCl_2$ competent ΔhisBΔpyrF DH5α cells were immediately used for transformations of pHBU3 and pH3U3 plasmids and ultimately glycerol stocks were made.

*Electrocompetent cells*

Electrocompetent TOP10 cells and ΔhisBΔpyrF DH5α cells were prepared for high efficiency transformation of libraries. Initially, overnight cultures of each strain were grown by inoculating each 2xYT liquid culture with a single colony. Existing chemically competent cell stock and glycerol stock was used as the source for TOP10 and selection strains respectively. For this protocol, growth media for TOP10 cells was made without antibiotic while tetracycline was used for selection strain growth. Care was taken to avoid cross contamination while preparing two strains simultaneously and all procedures were performed aseptically. Baffled 1 L flasks of sterile 500 mL 2xYT (with tetracycline) and 500 mL LB were used for expansion. For ΔhisBΔpyrF DH5α culture expansion, 5 mL of overnight was added to the 2xYT growth media. For TOP10 cell expansion, 2 mL overnight culture was added to the LB containing no antibiotic. The baffled flasks were placed in a shaker preheated to 37 ˚C and agitated at 220 rpm for 3 hours at which time $OD_{600}$ measurements were taken every ~15 minutes thereafter. For the selection strain and TOP10 cells respectively, suitable $OD_{600}$ values of ~0.40 and ~0.36 were reached after ~3.5 hours and ~5 hours indicating that both cultures were in logarithmic growth phase.

Subsequently, two washes were performed. All centrifugations in this protocol were run for 10 minutes at 4 ˚C and 4000 rpm. The cultures were each transferred into ten 50 mL falcon tubes, placed in chilled rotor buckets and centrifuged. Supernatants were decanted and pellets were

resuspended in 35 mL of chilled sterile, MiliQ $H_2O$ for the first wash. Suspensions were subjected to a second centrifugation before decanting and resuspending each in 20 mL ice cold water. Suspensions were combined and subjected to a third centrifugation. Pellets were then resuspended in 5 mL of ice cold 15% glycerol solution each and then suspensions of respective strains were combined into a single tube per strain. A final centrifugation was performed followed by dencanting and resuspension in 500 µL of 15% glycerol for each strain. Highly concentrated competent cells were then aliquoted into chilled microcentrifuge tubes in a 4˚ C cold room, flash frozen in liquid nitrogen, and place in -80 ˚C for storage.

## Transformations

All transformations were performed under sterile conditions with cells kept on ice. Following addition of recovery media, cultures were place in a 37 ˚C shaking incubator for one hour. An alternate protocol, incubating statically at 37 ˚C yielded slightly lower transformation efficiency and therefore the shaking incubation method was favored. After outgrowth, transformations were plated on solid media with the use of glass beads.

*Chemically Competent cell transformations*

Heat shock was employed to transform plasmids into calcium chloride competent cells and rubidium chloride competent cells. Initially, aliquots of chemically competent cells were removed from -80C storage and thawed on ice. Cells were inoculated with 0.5 - 1.0 µL of purified plasmid, 3 µL ligation assembly reaction, or 5 µL of CPEC assembly reaction and then incubated on ice for 30 minutes. Cells were subjected to heat shock in a 37 ˚C water bath for 90 seconds followed by a 2-minute ice incubation to achieve transformation. Subsequently, 900 µL of 2xYT media was added to cells for recovery. Transformants were outgrown at 37 ˚C with agitation at 220 rpm for one hour. Culture volumes of 20-120 µL were plated on 60 mm 2xYT solid media plates and distributed with glass beads.

*Electroporation*

Plasmid libraries were transformed into electrocompetent cells using a Bio-Rad Gene Pulser with an auxiliary Bio-Rad Pulse Controller Unit. The instrument was set to a capacitance of 25 µFD and resistance of 200 Ω for all transformations. Due to availability, electroporation cuvettes with an electrode gap of 1 mm were used in initial electroporation experiments and cuvettes with a 2 mm gap were used for later transformation batches. An electric potential of 1800 V was used for electroporations when using 1 mm electrode gap cuvettes and an electric potential of 2500 V was used for electroporations when using 2 mm gap cuvette electrodes. CPEC assemblies were desalted by mini-column prior to electroporation to eliminate magnesium from the polymerase buffer to reduce conductivity and avoid arcing. Ligation assembled constructs were used in electroporations without further purification. Observed time constants ranged from 3.5 ms to 5 ms, with smaller values recorded for transformations of ligation assemblies. Following electroporation, cells transformed in 1 mm gap cuvettes or 2 mm gap cuvettes were recovered in 700 µL or 900 µL of 2xYT media respectively.

*Library Transformations*

The plasmid transformation protocols for chemically competent and electrocompetent cells described above were employed for library transformations with deviations and additional steps. To maximize transformation efficiency, a 6 % v/v concentration of ligation reaction libraries or a 10 % v/v concentration of CPEC reaction libraries in competent cells was used for transformations. It was found that transformation efficiency directly into the ΔhisBΔpyrF DH5α selection strain was low despite high recorded efficiency when transforming purified DNA. Consequently, unpurified library assemblies were transformed into electrocompetent or chemically competent cells first followed by resuspension into liquid media from solid media and purification by the plasmid extraction method described. Direct expansion in liquid media was attempted as an alternative however this approach was not used ultimately due to likely introduction of bias, skewing library member representation. A volume of 2-3 µL of purified

library was then transformed into the selection strain utilizing the standard procedure. Following recovery incubation, stage two library transformations were subjected to centrifugation at 4000 rpm for 5 minutes. Pellets were gently resuspended in 400 µL of liquid drop out amino acid media with or without modifier compounds and plated on drop out media selection plates with kanamycin, 3AT, and 2AP.

## Drop Plating

Drop plating was performed on selective media to evaluate the behavior of the selection strain harboring specific plasmids. Histidine/adenine dropout plates were prepared with or without 2AP and with varying concentrations of 3AT. Fresh transformations were performed for all drop plating experiments. A single colony for the selection strain harboring each plasmid of interest was then used to inoculate 2 mL drop out amino acid liquid cultures. This was repeated for each strain to acquire biological replicates. After 24 hours of growth, an $OD_{600}$ measurement was taken for each culture. Liquid cultures were then normalized, adjusting concentrations by individually diluting each to an $OD_{600}$ of 0.06. A precision multichannel pipette was then used to drop 7 µL of OD normalized culture onto selection plates. Cultures were allowed to grow on solid media for 24 hours at 37 ˚C. Plates were then photographed and processed to attain drop series for each plasmid or replicate at each 3AT concentration.

## Gel Electrophoresis

To verify amplicon sizes and assess PCR quality, gel electrophoresis was employed. For DNA >200 bp in size, an NEB 1 kb Plus DNA ladder was used for reference on a 1 % agarose gel. For small inserts and libraries, an NEB low molecular weight ladder was used for reference on a 2 % agarose gel. Gels were run at 120 V for 25-30 minutes or 45-60 minutes for large or small PCR products respectively. TAE buffer was used for all agarose gel preparations and for running buffer. Gels were loaded with 5 µL of PCR reaction.

## Replica Plating

To initially assess response to ligand dependent survival advantage of colonies yielded by selection, replica plating was conducted. Histidine/adenine dropout media with 2 mM 3AT plates were prepared with and without 1 mM 2AP. Individual colonies were picked from original selection plates and used first for inoculation of 2AP negative agar and then of 2AP positive agar. Plates were incubated at 37 ˚C for ~20 hours and evaluated for the presence of growth. Cells harboring plasmids with functional purine responsive riboswitches were expected to grow on 2AP+ plates and not survive on 2AP- plates.

# V. Results

## Strain Behavior and Media Optimization

Initial development of the auxotroph-based riboswitch selection platform consisted of plasmid construction, media optimization, and selection experiments. After completing construction of the pHBU3 and P4C-pHBU3 plasmids followed by sequence verification, experiments were conducted to determine function of the plasmids in the selection strain. Primary experiments sought to understand survival behavior on histidine dropout media, associated leaky expression, and responsiveness to ligand. As reported in previous studies using the auxotroph platform for transcription factor binding specificity and in evaluations of the designed P4C riboswitch, leaky expression was observed. To address the effects of leaky expression and select against library members expressing low levels of *hisB* in the absence of ligand, the competitive inhibitor 3AT was used to improve positive selection. Two separate time trial experiments aimed to identifying optimal growth time, tuning 2AP, and tuning 3AT were performed. The experiments were conducted over 14-20 hours in liquid dropout media with variable concentrations of modifier compounds.

In the first time series experiment, two identical 3AT series were prepared, one containing no 2AP, the second with 1 mM 2AP (figure 7). To prepare cultures for these experiments, 2xYT medium was inoculated with a single colony of USO selection strain harboring the P4C-pHBU3 plasmid – both kanamycin and tetracycline were used for overnights. Culture tubes were prepared with 2 mL of histidine/adenine dropout media. A dilution series ranging from 0.05 to 3.5 mM 3AT was prepared. Cultures were grown at 37 ˚C in agitating conditions and $OD_{600}$ measurements were then taken every 1-2 hours from 13 to 20 hours of growth in technical duplicate measuring 200 µL of culture in a flat bottom plate. A clear increase in growth for all cultures containing 2AP was observed, consistent with ligand binding to the P4C riboswitch, upregulating expression of *hisB* and thereby conferring a survival advantage (figure 7).

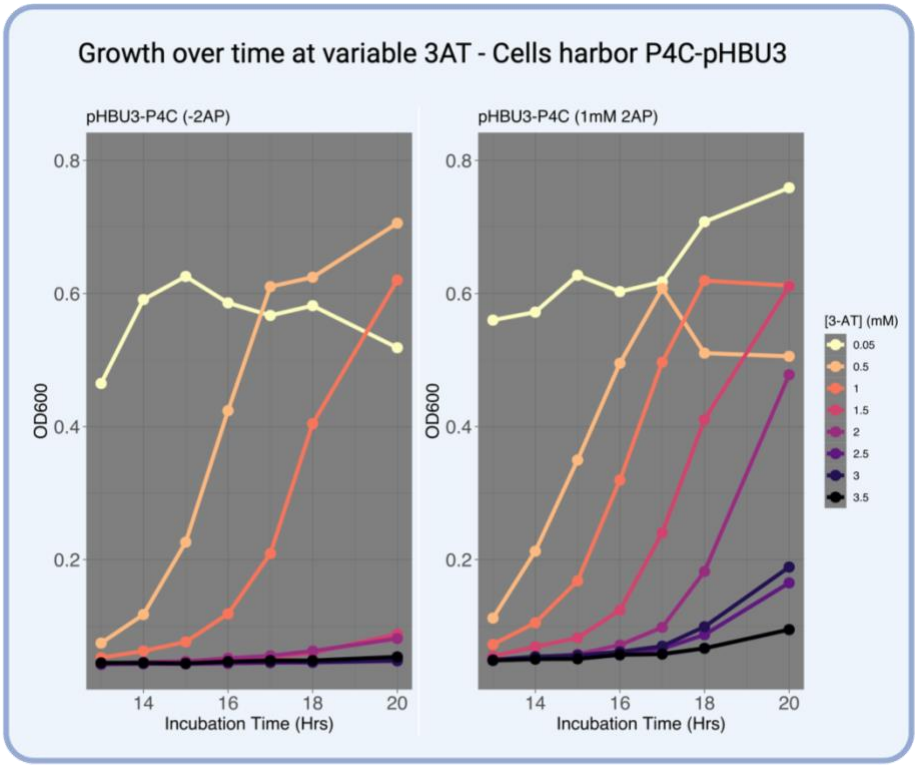


*Figure 7: Growth over time of liquid cultures. Selection strain harbored the P4C-pHBU3 plasmid and was grown at variable 3AT inhibitor concentration. Left panel (-2AP), right panel (+ 1mM 2AP). Note difference in $OD_{600}$ of ~0.4 at a 1.5 mM 3AT concentration when grown with ligand.*

This increase in growth rate with 2AP present was observed for all concentrations of 3AT. Cultures grown without 2AP and with 0.05-1.0 mM 3AT all reached the logarithmic phase before 18 hours of growth. This observation indicates in the absence of ligand, some *hisB* is transcribed. Leaky expression of ON transcriptional riboswitches has been widely reported and thus this observation is consistent with previous investigations (Barrick & Breaker, 2007). Because P4C is an adenine binding riboswitch, the nucleobase endogenously synthesized may account for basal levels of activation. Alternatively, local conditions may stochastically afford basal levels of aptamer stabilization in the absence of ligand. At increased concentrations of 1.5 and 2 mM 3AT, cultures with 2AP reached logarithmic growth phase by 17 and 19 hours respectively, while cultures without ligand showed little to no growth by the end of the trial at 20 hours. These observations demonstrate concentrations of 3AT in the 1.5-2.0 mM range are suitable to effectively inhibit *hisB* to suppress growth in cultures without ligand for up to 20 hours in liquid media.

To determine the optimal 3AT concentration on solid media, a similar experiment was performed by drop plating (figure 8). Plates were prepared in two series, with and without 1mM 2AP, both at concentrations of 3AT ranging from 0-3 mM in 0.5 mM increments. Observations were similar to those obtained from liquid media experiments corroborating previous data suggesting that *hisB* is expressed in a leaky fashion in the absence of ligand. These results provided guidance for positive selection conditions in library evaluation.

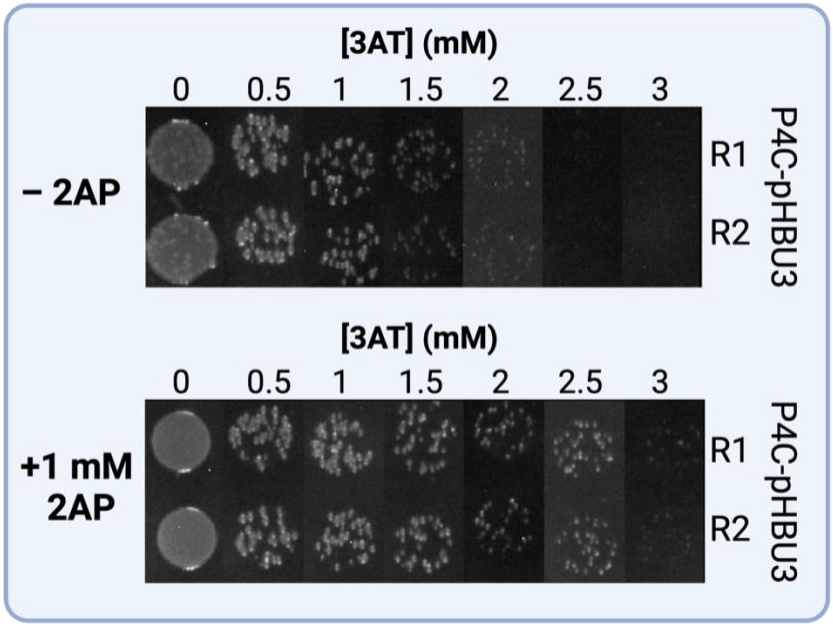


*Figure 8: Drop plating results for P4C-pHBU3 harboring cells in culture on solid dropout media with variable 3AT, with or without 2AP. Note inhibited growth above 1.5 mM 3AT in top panel.*

## Ligation Assembled J2/3 Library and Evaluation

After determining optimal concentrations of 3AT and 2AP for enriching cells containing plasmids with ligand responsive riboswitches, a plasmid library was constructed using the P4C scaffolded J2/3 sequence suite. Prior to digestion and ligation, amplified libraries were gel verified (figure 9). Strong bands at the expected length of 120 bp indicated successful and specific amplification. Weak bands at ~75 bp were possibly due to nonspecific amplification.

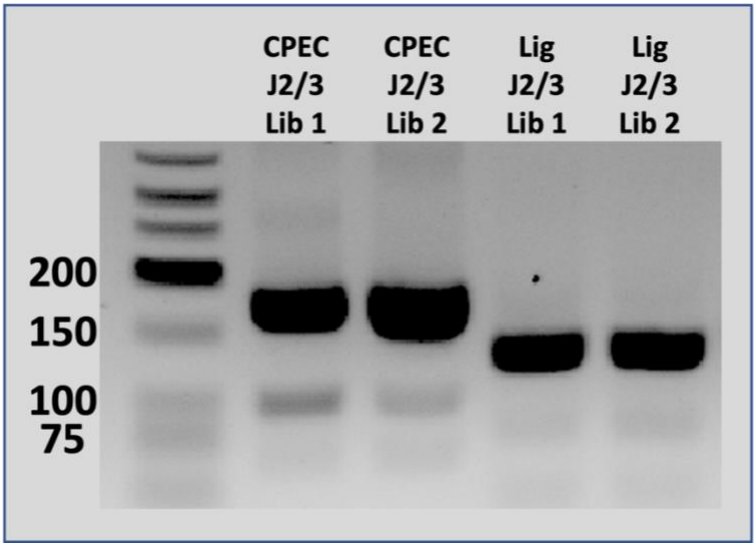


*Figure 9: PCR products yielded from amplification of the J2/3 library run on a 2 % agarose gel. Lanes 1-2 are products from amplification with long CPEC primers. Lanes 3-4 are products from amplification using medium length primers with SalI and KpnI restriction enzymes.*

The unpurified library was then assembled and transformed directly into rubidium chloride competent USOΔhisBΔpyrF cells. Poor transformation efficiency was obtained with fewer than 50 colonies observed when plated on selective media. These chemically competent cells were previously found to possess a sufficient transformation efficiency of $>10^6$ CFU/µg when evaluated using purified plasmids. The double auxotroph strain required non-supercoiled ligated plasmids be taken up by cells and to then express *hisB* and biosynthesize histidine while also absorbing uracil in order to grow on drop out media. The low transformation efficiency was attributed to these factors in addition to the ligand binding requirement, metabolic burden of synthesizing *kanR* and additional metabolic pressure due to 3AT inhibition of *hisB.* The chemically competent selection strain was evaluated again by transforming supercoiled plasmid and by repeating transformation of ligated unpurified plasmid library. The initial observations were corroborated, confirming high transformation efficiency with supercoiled plasmids and insufficient transformation efficiency when using the freshly assembled library.

To address this, the unpurified, ligated library was transformed into commercial TOP10 cells engineered for high transformation efficiency. Acceptable transformation was obtained, and colonies were recovered and purified to yield a supercoiled library. The purified library was

then transformed into the selection strain achieving high efficiency. Following the developed protocol, outgrown and pelleted cells harboring the supercoiled library were resuspended in dropout media and plated on selection media with ligand. After 24 hours of growth, ~150 colonies were observed on media without 3AT, 10-15 colonies were observed on media with 0.5 - 1.0 mM 3AT, and 5-10 colonies were observed on media 1.5 - 2.5 mM 3AT, while 3 colonies were observed on 3 mM 3AT media. A total of ten colonies were selected for replica plating. Replica plating yielded growth of all ten colonies on both 2AP+ and 2AP- plates indicating the plasmids did not contain functional riboswitches able to mitigate expression in the absence of ligand.


*Figure 10: Top sequence is the J2/3 library with P4C scaffold annotated with riboswitch features. Lower sequence was obtained from selection and contains an undesired 73 bp digestion product incorporated into ligation assembly. Orange boxes show undesired repeated sequences.*

To gain insight into the source of the evasive constructs, these ten colonies were outgrown and purified for sequencing. Assessment of sequences revealed a 73 bp insertion constituting a repeat of a portion of the riboswitch scaffold downstream of the J2/3 randomized region on the 3' side of the sequence (figure 10). Inspection of the aberrant construct showed that a member of the J2/3 library possessing an interior additional KpnI restriction site within the riboswitch was cleaved during the restriction digest reaction. Undesired cleavage yielded dsDNA fragments possessing KpnI overhangs on both ends. The aberrant and evasive construct was assembled by undesired ligation of the 73 bp fragment between library insert and backbone KpnI sites. The construct afforded the maintenance of high levels of transcription irrespective of ligand presence and thereby outcompeted cells harboring functional constructs.

To address the issue of evasive constructs generated from undesired cleavage and recombination, DNA assembly components were isolated prior to ligation and the experiment was repeated. The primary goal of this approach was to eliminate the ~73 bp DNA fragments possessing KpnI sites at both ends to curtail the construction of aberrant evasive assemblies. To accomplish this, polyacrylamide extraction and agarose gel extraction were employed to isolate DNA fragments of the correct size and exclude problematic fragments. DNA extraction from a native polyacrylamide gel was performed to isolate the digested J2/3 library with an expected size of 112 bp, adjusted for overhangs. The digested backbone was similarly excised and agarose gel extracted to achieve isolation from any undigested backbone or undesired digestion products. After isolation of sticky ended DNA components, a standard ligation reaction was performed and the assembly was transformed into electrocompetent TOP10 cells.

A relatively low abundance of transformants obtained was attributed to low purity polyacrylamide extraction eluent. A spectrophotometric peak below 230 nm was observed during purity determination, likely attributable to acrylamide not completely removed during column purification. Nonetheless, relative to colonies yielded by the assembly transformation, a proportion of only ~1 % appeared on the vector-only control plate, indicating desired low background. Colony PCR was performed to determine presence or absence of the riboswitch scaffold in the colonies obtained (figure 11). Based on amplification product size compared with controls, colonies from the vector-only transformation did not contain the insert while all colonies evaluated among ligation reaction transformants contained the library insert. After confirming high ligation efficiency, using the same ligation reactions, four more electroporations were performed. Outgrown transformation cultures were concentrated to maximize transformant recovery before plating. A total of approximately 600 colonies were obtained on the solid nonselective media representing ~45 % of the 1024-member library. The population was recovered from plates by suspension in liquid LB media and the plasmid library was mini-column extracted. Six electroporations of the library into the selection strain were performed and selection was carried out on histidine/adenine dropout plates containing 2 mM

3AT inhibitor and 1 mM 2AP ligand. The batch of electrocompetent USOΔhisBΔpyrF selection strain cells was previously determined to have an efficiency of >$10^7$ CFU/µg when transformed with purified plasmid. A control plate on 2xYT media with tetracycline and kanamycin

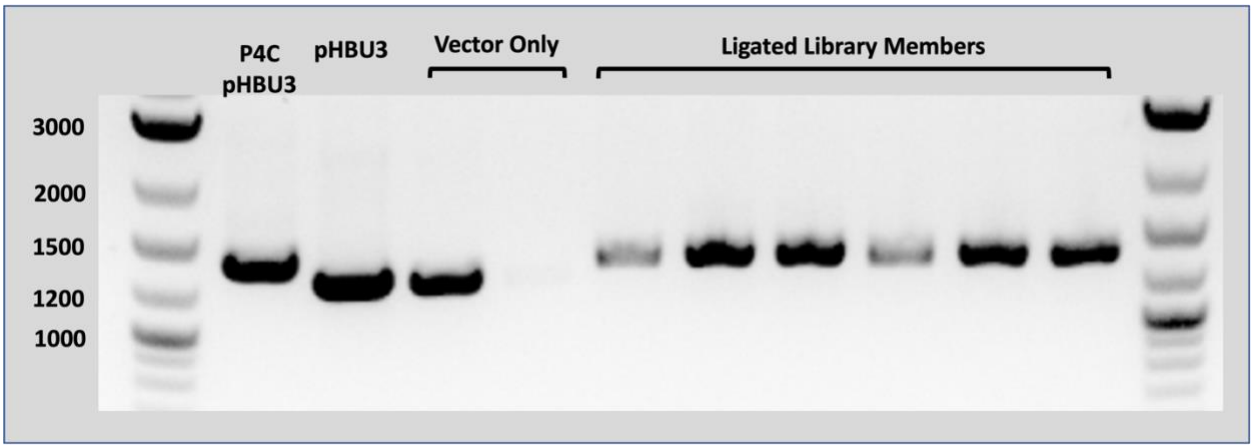


*Figure 11: Colony PCR to determine presence of riboswitch scaffold insert. The first colony from the vector only control plate contains cells with plasmids missing the insert indicated by a band at ~1215-1230 bp.*

confirmed comparable transformation efficiency of the extracted library. A total of 20 colonies were picked from selection plates and used to inoculate overnight cultures for subsequent extraction and sequencing.

Unique sequences were aligned against the P4C-pHBU3 plasmid from the lac promoter through the 3' end of P4C (figure 12). Mutations were observed in the promoters of all sequences. Although near the SalI cloning site, these mutations were downstream and independent the ligation site. A single nucleotide deletion in J3/1 was observed adjacent and upstream of the canonical thymine ligand binding partner for all recovered sequences. For all sequences, a two-nucleotide deletion in the L4 loop was also observed. All sequences possessed the same J2/3 sequence, ATTAT, a departure from the TGTCT sequence found in this portion of the J2/3 in P4C. These results indicate probable PCR induced mutation which conferred a survival advantage. Notably, deletions in the P4 of all members likely disrupted intrinsic terminator formation leading to constitutive maximum expression of the *hisB* reporter gene. The ubiquitous deletion in J3/1 likely also contributed to the inability of the nascent transcript to

undergo efficient strand exchange to form a stable terminating helix. The recurrence of these deletions may be due to PCR bias (Joyce, 2004).

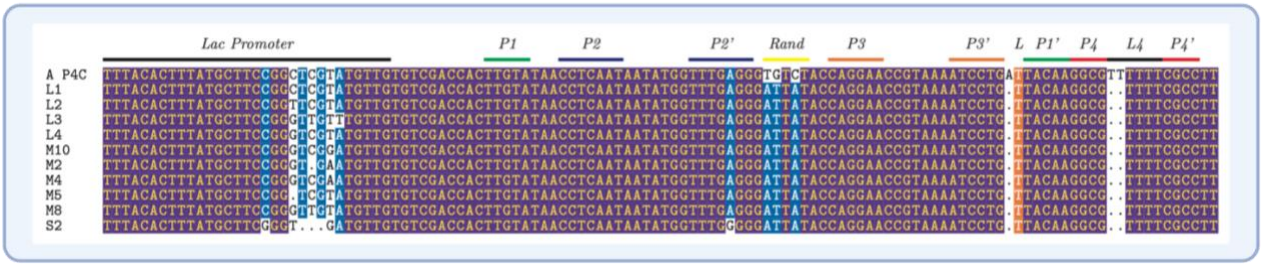


*Figure 12: Alignment of unique sequences obtained from selection with P4C as reference. Note deletions in key loops in the aptamer likely responsible for an unstable hairpin structure.*

The sequencing results provided data further demonstrating the necessity for effective counterselection in the development of this platform. A total of 20 PCR cycles was used for amplification of the riboswitch scaffold library and the high fidelity Q5 polymerase was utilized. Although fewer PCR cycles and a yet higher fidelity polymerase could be employed, it is unlikely all PCR mutations could be reliably avoided. Selection pressure in this approach favors any construct unable to form a stable terminator hairpin.

## CPEC Assembled J2/3 Library and Evaluation

As one approach to address the issue of aberrant, evasive constructs assembled during ligation, an alternative construction method was investigated. CPEC assembly involves preparing double stranded overlapping DNA fragments and running a modified PCR reaction to generate circularly polymerized plasmids. The remaining nicks in resulting plasmid constructs are repaired endogenously by the *E. coli* (Quan & Tian, 2011). This alternate avenue to library assembly allowed construction without the need for restriction digestion and thereby afforded circumvention of the problematic ligation constructs. The riboswitch scaffold library was amplified using primers designed with overlaps corresponding to a second pair designed for backbone amplification (figure 9, Appendix A).

Initial attempts to execute this procedure resulted in high background likely due to undigested or otherwise circularized pHBU3 plasmids without riboswitch scaffold inserts. DpnI is a methylation dependent restriction enzyme used in the protocol to degrade template DNA after amplification of backbone. To mitigate incomplete digestion of parent template plasmid, the experiment was repeated using doubled enzyme concentration and digested for an additional hour. A control experiment was also performed to verify effective digestion. Proper digestion was confirmed after running the control reaction on a gel, indicating complete digestion of methylated DNA, precluding the possibility of deactivated restriction. After preparation of the backbone and library, the components were assembled by CPEC.

The assembled plasmid library and a backbone-only control were transformed into TOP10 cells. The relative number of colonies appearing on the control plate represented ~10 % of the number of colonies obtained on corresponding CPEC plates. These observations indicated the presence of a functioning plasmid without desired insert transformed from a backbone-only reaction. The presence of undigested, methylated parent plasmid was determined to be unlikely after verifying effective and complete DpnI digestion with the previously discussed methods. Furthermore, these transformations were plated on double antibiotic plates with both tetracycline and kanamycin largely ruling out contamination as a possible explanation. Although not conclusive, one possible explanation for these colonies is the polymerization of functional plasmids during the CPEC reaction by self-priming of backbone ends. Since this backbone contains all necessary plasmid components for a functional extrachromosomal DNA unit, self-priming by annealing of backbone tails to one another may have yielded circular DNA products without the desired riboswitch scaffold insert.

## Plasmid modifications for counterselection

To resolve the issue of mutant constructs outcompeting correctly assembled library members during positive selection, counterselection using 5-FOA was investigated. Early pilot experiments in liquid selection media showed little to no growth in media without uracil even

with histidine and adenine present. This behavior was observed for cultures with cells harboring pH3U3, pHBU3, or P4C-pHBU3 and interpreted as a lack of expression of functional URA3. Previous investigators reported effective use of 2 mM 5-FOA for counterselection to eliminate self-activating constructs in transcription factor studies in conjunction with a B2H (Meng et al., 2005). When these plasmids were tested under this counterselection condition, no significant change in growth was observed. This was likely due to the absence of RNAP α-subunit linked protein expressed in the riboswitch system. When grown in liquid media containing 4 mM 5-FOA, an inhibition of growth was observed in cultures with cells harboring the pH3U3 plasmid when compared with pHBU3. After 19 hours of growth, mean $OD_{600}$ values of 0.86 and 1.08 were observed respectively, indicating the presence of URA3 enzyme in pH3U3 cultures and consequent metabolism of 5-FOA into toxic fluorouracil. Surprisingly, this indicated that URA3 expression was a result of increased readthrough in translation of the transcript containing the heterologous *HIS3* gene over that containing the *hisB* gene. This interpretation was corroborated using the Salis RBS operon calculator (Cetnar & Salis, 2021) to evaluate the respective sequences. Based on these results, the pH3U3 plasmid was modified for direct use in the riboswitch platform under development.

To adapt the pH3U3 plasmid while preserving the *HIS3* gene, site directed mutagenesis was employed using CPEC to remove a KpnI site from the *HIS3* gene body and to introduce a necessary KpnI site into the 5' UTR. Comparable growth inhibition was observed in cultures with cells harboring the altered plasmid in 4 mM 5-FOA media. Subsequently, the P4C-pH3U3-Kp construct was made and compared with P4C-pHBU3 with 5-FOA and 2AP ligand, however no difference in growth was observed. In this and other experiments, the presence of the riboswitch in plasmids yielded suppressed expression of reporter genes when compared to plasmids without a riboswitch and thus URA3 expression was reduced below sufficient levels for effective counter selection. These results suggested higher basal URA3 expression would be necessary for counterselection use in this riboswitch platform relative to previous applications for transcription factor binding specificity.

In order to further improve expression of URA3, the Salis calculator was employed to compute intragenic RBS sequences predicted to increase translation of the gene. Using these predictions, three primer pairs were designed to build a suite of modified constructs with RBSs of increasing strength: pH3U3-R9, pH3U3-mid, pH3U3-5E (Appendix B). These plasmids were transformed into the selection strain and evaluated at concentrations of 5-FOA ranging from 4 to 15 mM. Only the pH3U3-5E showed a modest improved response to 5-FOA over pH3U3-Kp. Due to limited and still insufficient improvement in URA3 expression, this approach was not pursued further. The weak lac promoter used in this platform is likely not strong enough to yield sufficient second gene expression while under riboswitch influence regardless of ligand presence. Consequently, to further pursue counterselection, a stronger promoter should be tested to first achieve greater expression of the second reporter gene.

# VI. Discussion

## Future Directions and Orthogonal Approaches

The investigation at hand represents an initial exploration of the utility of one selection approach for developing novel riboswitches. Results indicate that modulating ligand concentration can control cell survival with a known riboswitch placed to influence reporter genes an *E. coli* system. Furthermore, data suggest when using a purine riboswitch scaffolded libarary, constitutively off members can be eliminated using positive selection. Nonetheless, evasive constructs, arising as a result of PCR induced mutation, unfiltered parent template, or other means, present a challenge. These constructs are self-activating regardless of ligand presence. Efforts to eliminate these evaders in this investigation included, CPEC construction approaches, reducing PCR cycles, and agarose and polyacrylamide gel extraction. Selection pressure affords a survival advantage to these constructs, and they persistently outcompete proper library members.

Attempts to avoid the introduction of these evaders have been unsuccessful and do not appear to be readily surmountable. A conclusion of this study is that counterselection is essential for eliminating these evaders. The use of 5-FOA for counterselection presented difficulties with limited success. Reports of successful use of 5-FOA for counterselection primarily appear in yeast studies, while those in *E. coli* studies indicate inconsistent results. Nonetheless, efforts to achieve effective counterselection in this study with 5-FOA were not exhaustive. Three plasmids were constructed to elevate URA3 expression however response to 5-FOA was not substantially improved. The lack of sufficient expression of URA3 is likely due to low maximum levels of operon transcription coupled with reduced expression of the second gene in the cistron. In the Meng et. Al, 2005 study, self-activation increased reporter gene expression above levels produced with lone pH3U3 plasmid in the absence of a B1H plasmid. In contrast, this riboswitch selection system does not involve a means for elevated recruitment of the α-subunit. Consequently, the weak lac promoter optimized for the previously implemented transcription

factor binding specificity system may not be suitable for the riboswitch development system at hand. To achieve sufficient expression of URA3 for successful counterselection with 5-FOA, stronger promoters should next be tested. Nonetheless, inconsistent reported results of this counter selection method in *E. coli*, low solubility, photosensitivity, pH sensitivity, and other limitation of 5-FOA present challenges to the counterselection method. Several other approaches to achieve the desired result should be considered. These include using an alternative negative selectable marker or implementing a selection/screening hybrid approach.

*SacB* is a *Bacillus subtills* levansucrase exoenzyme which converts sucrose into levan polysaccharide. The levan product is toxic to a number of bacteria species including *E. coli,* therefore making it useful for counter selection. As one alternative to 5-FOA selection, the pHBU3 plasmid could be easily modified to replace *URA3* with *sacB.* Using this assembly, counterselection could be performed on sucrose containing media to eliminate evasive, mutant, or otherwise spurious constructs. Under this approach, the robust and commercially available Keio *hisB* knockout would be used as the selection strain. Sucrose is favorable to 5-FOA with better stability and a lower cost than 5-FOA. The use of the single auxotroph Keio strain would also afford direct transformation of ligated or CPEC assembled, unpurified library. This would be possible, as the strain, possessing a genomic copy of *pyrF*, would not be burdened by the necessity to absorb uracil from the media. The approach would streamline the selection workflow saving time and materials.

An alternative approach involves combining FACS based screening with the selection system at hand. This method would leverage the high throughput capabilities of selection while benefiting from the tunability and flexibility afforded by cell sorting. This system would involve placing *gfp* and a lambda repressor in an operon under control of a stronger lac promoter. The *hisB* gene would be placed under control of a second promoter adjacent to an operator sequence recognized by the lambda repressor. In the histidine auxotroph strain negative selection would be performed first on drop out media in the absence of ligand. Self-activating

members would be eliminated. Subsequent positive selection with ligand present would be performed by FACS sorting to identify members with highest fluorescence intensity.

Although more work remains to complete construction of a functional auxotroph-based riboswitch selection platform, meaningful progress has been made, laying a foundation for ultimate development of a useful tool.

## Importance of Novel Riboswitch Development

A tool for efficiently and reliably developing riboswitches responsive to alternate small molecules remains an unattained synthetic biology goal. A system to produce these regulatory elements with synthetic aptamers selectively responsive to a ligand of choice would have important implications for a range of applications. Notably, a platform of this type would yield *cis* modulatory elements for use in RNA therapeutics for dosage regulation, improving the safety profile of these drugs and affording avenues to new therapeutics. A tool for synthesizing riboswitches has additional applications in industry for rapid identification of compounds of interest in many potential settings. Synthetic biology systems continue to become increasing important for solving problems facing humanity ranging from drug development to addressing energy challenges. As these systems are being engineered to address these challenges in increasingly complex way, the need for fine tuning behavior of individual elements is concomitantly becoming more important. A reliable tool for rapid development of specialized riboswitches to tune these systems would constitute a major contribution to accelerating the advancement of synthetic biology opening new avenues for advancement of many new biotechnologies.

Despite incentives for developing this type of tool, considerable hurdles have made progress slow. Screening approaches have insufficient throughput to evaluate necessarily large libraries. The promise of machine learning, AI, and other computational approaches have not yet delivered in this arena. Bacterial auxotroph selection remains an attractive approach.

# VII. References


Barrick, J. E., & Breaker, R. R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*, *8*(11), R239. https://doi.org/10.1186/gb-2007-8-11-r239

Batey, R. T., Gilbert, S. D., & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, *432*(7015), 411-415. https://doi.org/10.1038/nature03037

Ceres, P., Garst, A. D., Marcano-Velazquez, J. G., & Batey, R. T. (2013). Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. *ACS Synth Biol*, *2*(8), 463-472. https://doi.org/10.1021/sb4000096

Cetnar, D. P., & Salis, H. M. (2021). Systematic Quantification of Sequence and Structural Determinants Controlling mRNA stability in Bacterial Operons. *ACS Synth Biol*, *10*(2), 318-332. https://doi.org/10.1021/acssynbio.0c00471

Darmostuk, M., Rimpelova, S., Gbelcova, H., & Ruml, T. (2015). Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol Adv*, *33*(6 Pt 2), 1141-1161. https://doi.org/10.1016/j.biotechadv.2015.02.008

Drogalis, L. K., & Batey, R. T. (2020). Requirements for efficient ligand-gated co-transcriptional switching in designed variants of the B. subtilis pbuE adenine-responsive riboswitch in E. coli. *PLoS One*, *15*(12), e0243155. https://doi.org/10.1371/journal.pone.0243155

Garst, A. D., & Batey, R. T. (2009). A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta*, *1789*(9-10), 584-591. https://doi.org/10.1016/j.bbagrm.2009.06.004

Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, *3*(6). https://doi.org/10.1101/cshperspect.a003533

Gilbert, S. D., Stoddard, C. D., Wise, S. J., & Batey, R. T. (2006). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol*, *359*(3), 754-768. https://doi.org/10.1016/j.jmb.2006.04.003

Johansen, L. E., Nygaard, P., Lassen, C., Agerso, Y., & Saxild, H. H. (2003). Definition of a second Bacillus subtilis pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL). *J Bacteriol*, *185*(17), 5200-5209. https://doi.org/10.1128/JB.185.17.5200-5209.2003

Joyce, G. F. (2004). Directed evolution of nucleic acid enzymes. *Annu Rev Biochem*, *73*, 791-836. https://doi.org/10.1146/annurev.biochem.73.011303.073717

Lemay, J. F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Masse, E., & Lafontaine, D. A. (2011). Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet*, *7*(1), e1001278. https://doi.org/10.1371/journal.pgen.1001278

Mandal, M., & Breaker, R. R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol*, *11*(1), 29-35. https://doi.org/10.1038/nsmb710

Marcano-Velazquez, J. G., & Batey, R. T. (2015). Structure-guided mutational analysis of gene regulation by the Bacillus subtilis pbuE adenine-responsive riboswitch in a cellular context. *J Biol Chem*, *290*(7), 4464-4475. https://doi.org/10.1074/jbc.M114.613497

McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., & Breaker, R. R. (2017). Riboswitch diversity and distribution. *RNA*, *23*(7), 995-1011. https://doi.org/10.1261/rna.061234.117

Meng, X., Brodsky, M. H., & Wolfe, S. A. (2005). A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol*, *23*(8), 988-994. https://doi.org/10.1038/nbt1120

Meng, X., Smith, R. M., Giesecke, A. V., Joung, J. K., & Wolfe, S. A. (2006). Counter-selectable marker for bacterial-based interaction trap systems. *Biotechniques*, *40*(2), 179-184. https://doi.org/10.2144/000112049

Meng, X., & Wolfe, S. A. (2006). Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nature Protocols*, *1*(1), 30-45. https://doi.org/10.1038/nprot.2006.6

Miller, C. A., Ho, J. M. L., & Bennett, M. R. (2022). Strategies for Improving Small-Molecule Biosensors in Bacteria. *Biosensors (Basel)*, *12*(2). https://doi.org/10.3390/bios12020064

Montange, R. K., & Batey, R. T. (2008). Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys*, *37*, 117-133. https://doi.org/10.1146/annurev.biophys.37.032807.130000

Novoradovsky, A., Zhang, V., Ghosh, M., Hogrefe, H., Sorge, J., & Gaasterland, T. (2005). Computational Principles of Primer Design for Site Directed Mutagenesis. 2005 NSTI Nanotechnology Conference and Trade Show - NSTI Nanotech 2005 Technical Proceedings, Anaheim, CA.

Patil, V. S., Zhou, R., & Rana, T. M. (2014). Gene regulation by non-coding RNAs. *Crit Rev Biochem Mol Biol*, *49*(1), 16-32. https://doi.org/10.3109/10409238.2013.844092

Porter, E. B., Marcano-Velazquez, J. G., & Batey, R. T. (2014). The purine riboswitch as a model system for exploring RNA biology and chemistry. *Biochim Biophys Acta*, *1839*(10), 919-930. https://doi.org/10.1016/j.bbagrm.2014.02.014

Porter, E. B., Polaski, J. T., Morck, M. M., & Batey, R. T. (2017). Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol*, *13*(3), 295-301. https://doi.org/10.1038/nchembio.2278

Quan, J., & Tian, J. (2011). Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat Protoc*, *6*(2), 242-251. https://doi.org/10.1038/nprot.2010.181

Serebriiskii, I., & Joung, J. (2001). Yeast and Bacterial Two-hybrid Selection Systems for Studying Protein-Protein Interactions. In E. Golemis (Ed.), *Protein-Protein Interactions: A Molecular Cloning Manual* (1st ed., pp. 93-142). Cold Spring Harbor Laboratory Press.

Singh, M. V., & Weil, P. A. (2002). A method for plasmid purification directly from yeast. *Anal Biochem*, *307*(1), 13-17. https://doi.org/10.1016/s0003-2697(02)00018-0

Vidal, A. J. M. W. a. M. (1999). A Genetic Strategy to Eliminate Self-Activator Baits Prior to High-Throughput Yeast Two-Hybrid Screens.

Wilson, T. E., Fahrner, T. J., Johnston, M., & Milbrandt, J. (1991). Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science*, *252*, 1296–1300.

# Appendix A

**Oligonucleotides used in PCR reactions**

| Primer Name | Primer Sequence (5' to 3') |
|---|---|
| PHBU3_INTRBS9_FWD | TTAGCGGCTTAACTGTGCCCTC |
| PHBU3_INTRBS9_REV | TGTGCATTCGTAATGTCTGCCC |
| HBU3_P4C_FWD | TGTTGTGTCGACCACTTGTATAACCTCAATAAT |
| RS_UNIV_BB_HU_FWD | CAGGATTTTTTTTATTTACTAGTGGTACCAATTTCACACAGGAAACAGCTATGAGTCAG |
| RBS5E_STRONG_FWD | GTAACCTGTGGAGAGGAAGGGGAGTATCTGTATGTCGAAAGCTACATATAAGGAACGTGC |
| RS_UNIV_BB_HU_REV | CCATATTATTGAGGTTATACAAGTGGTCGACACAACATACGAGCCGGAAGCATAAAGTG |
| RBS9_INT_H3U3_FWD | GAGATGGATATGGGGAGTTATCTGTGATACATGTCGAAAGCTACATATAAGGAACGTGC |
| RS_UNIV_LIB_HU_FWD | CACTTTATGCTTCCGGCTCGTATGTTGTGTCGACCACTTGTATAACCTCAATAATATGG |
| QCh_PH3U3_RMKPN1_REV | GGAGGGAACATCGTTTGTTCCATTGGGCGAGGTG |
| QCh_PH3U3_KPN1_REV | TGTTTCCTGTGTGAAATTGGTACCCGCTCGGTCGACACAAC |
| QCh_PH3U3_KPN1_FWD | GTTGTGTCGACCGAGCGGGTACCAATTTCACACAGGAAACA |
| QCh_PH3U3_RMKPN1_FWD | CACCTCGCCCAATGGAACAAACGATGTTCCCTCC |
| HBU3_P4C_REV | GTGAAATTGGTACCACTAGTAAATAAAAAAAATCCTGATTACAAGG |
| Spacer_HisB_URA3 | GATCCCTGCAGCTTTAAATAATCGGTGTCATTACAGCACTCCTTTCGACGAGGGCAGGG |
| RBS9_Int_HBU3_Rev | GTATCACAGATAACTCCCCATATCCATCTCCCACACAGGTTACAGCACTCCTTTCGAC |
| Ext_J2lib_fwd | GTCGACCACTTGTATAACCTC |
| RS_univ_lib_HU_rev | CTGACTCATAGCTGTTTCCTGTGTGAAATTGGTACCACTAGTAAATAAAAAAAATCCTG |
| pBR_Eco_Fwd | AATAGGCGTATCACGAGGC |
| RBS5e_strong_rev | CAGATACTCCCCTTCCTCTCCACAGGTTACAGCACTCCTTTCGACG |

# Appendix B

**Intergenic regions in modified pHBU3 constructs**

| Sequence Name | Sequence (5’ to 3’) |
|---|---|
| pHBU3-RBS5e Intergenic Region | CCTGTGGAGAGGAAGGGGAGTATCTGT |
| pHBU3-midRBS Intergenic Region | TGACACCGATTATTTAAAGCTGCAGGGATCCCACACAGGAAACAGCT |
| pHBU3-RBS9 Intergenic Region | CCTGTGTGGGAGATGGATATGGGGAGTTATCTGTGATAC |