Genetic Screen of the *B. subtilis pbuE* Adenine-Responsive Riboswitch Expression Platform Reveals Preferences for Base Pairing in the Nucleator Hairpin-Stem

Ava Lorraine Altenbern

Department of Biochemistry

University of Colorado at Boulder

Defended: April 11th, 2023

Thesis Advisor:

Dr. Robert T. Batey, Department of Biochemistry

Committee Members:

Dr. Jeffrey C. Cameron, Department of Biochemistry (Honors Council Representative)

Dr. Robert Buchwald, Honors Residential Academic Program

Acknowledgements

It is necessary that I first acknowledge the participants from the Batey lab who have irreversibly impacted the foundation and development of the *pbuE* Adenine-Responsive Riboswitch research project. Thank you to Parker and Peyton, my colleagues and fellow graduates of 2023, for your unconditional support and openness to collaboration, and to Lisa Hansen, who has been not only an invaluable resource, but my mentor and friend. I will forever be appreciative of your kindness, patience, and generosity.

I would also like to recognize Joan Gallegos, my high school AP Chemistry instructor who, with great tolerance, gave me the confidence to pursue a career in science. Thank you for your lifelong dedication to developing the passions and education of young women in STEM.

Finally, I must extend my deepest gratitude to my committee members for their respect, cooperation, and overall eager encouragement of student growth through their participation in the Honors Thesis program at CU Boulder. Thank you to Dr. Jeffrey Cameron for your commitment in overseeing the undergraduate research program in biochemistry. To Dr. Robert Buchwald, as your enthusiasm and passion for the biological sciences inspired me to seek out research opportunities at the start of my collegiate career. Lastly, my greatest thanks to Dr. Robert Batey for granting me the opportunity to develop invaluable laboratory skills within your cohort over the last 2 years. Your trust and confidence in my ability to practice biochemical analysis has allowed me to grow into the scientist I am today.

Table of Contents

ABSTRACT **1**

DISCLAIMER **2**

I. INTRODUCTION **3**

B. subtilis as a Model Organism 3

Various RNA Regulatory Mechanisms of Genetic Expression 4

Rho-Dependent Transcription Attenuation 5

Feedback Inhibition and Rho-Independent Transcription Attenuation 5

The Riboswitch 7

The *pbuE* Adenine Responsive Riboswitch 10

Conformations and Known Mechanisms of the *NH5* Variant 12

Structure 13

Strand Invasion 13

Genetic Screening of the P4 Hairpin Stem Library 15

II. METHODS **18**

Cloning of the P4 Hairpin Stem-Loop Library 18

Polymerase Chain Reaction 19

Restriction Enzyme Digest 21

Polynucleotide Kinase Treatment (PNK) 22

Ligation of Vector and Insert 22

Transformation of Ligated Plasmids 22

Primary and Secondary Colony Screening 23

Sequencing 25

Activity Assay 25

Data Analysis 26

III. RESULTS **28**

Various Preferences for Base Pairing 32

IV. DISCUSSION **39**

V. CONCLUSION **42**

Subjects of Future Investigations into the P4 Hairpin-Stem Library 42

Applications of the Riboswitch 43

REFERENCES **45**

APPENDIX **48**

Abstract

Located on the 5’ untranslated regions of bacterial messenger RNA, riboswitches are regulatory structures that are responsible for the modulation of genetic expression through ligand-dependent binding. Inclusive of two components, the riboswitch will probe its environment without the aid of an additional protein or DNA structure to sense and attach a specific metabolite to the region known as the upstream aptamer domain. The downstream expression platform then endures changes in its folding pattern to adopt one of two secondary structures, resulting in either the inhibition or continuation of mRNA production. Due to its smaller size, the *B. subtilis pbuE* adenine-responsive riboswitch has been the focus of many previous studies that sought to determine how the tertiary structure of the aptamer domain allows for tight binding with high specificity. The expression platform, however, is similarly interesting, as it participates in strand invasion in order to produce a transcription terminating hairpin that is rho independent.

Through the mutagenic cloning of a novel *pbuE* variant named NH5, the investigation into the reduced nucleator Hairpin-Stem library containing 6 randomized nucleotides revealed a strict preference for genetic base pairing proximal to the L4 loop. Additionally, the data suggests that weak A-U and G-U interactions or even non-canonical coupling between the nucleobases furthest from the polyuridine tract is tolerated if supplemented with three strong Watson-Crick pairs. Paving the way for the creation of additional synthetic riboswitch structures, the robust screening of the P4 region allows for a more thorough understanding of the fundamental requirements that promote the formation of the terminator helix and the subsequent mechanism of strand invasion.

Disclaimer

The following study was, in part, a collaboration between multiple undergraduate students at CU Boulder under the direction of Dr. Robert Batey. Due to the substantial laboratory requirements and the immense volume of work necessary to develop an expansive understanding of the B. *subtilis pbuE* riboswitch, a collective organization was developed. Subsequently, a select group of figures that are highly specific to the foundation of the project are identical to those in previous studies. Each participant fronted an individual investigation into a unique structural division of the principal riboswitch, and as such, were required to operate under similar laboratory conditions. While the P4 Hairpin-Stem library has been previously discussed in theses by Alexandra Brown and Lisa Hansen, the data set was incomplete and thus required a more intensive inspection. Additional students who have taken part in the mission to better integrate undergraduates into research laboratories include Mia Kim, Matthew Payne, Peyton Roeder, and Parker Jules.

I. Introduction

B. subtilis as a Model Organism

The Gram-positive, catalase-positive bacterium *Bacillus subtilis* is a highly responsive model organism popularly utilized for research in bacterial chromosomal replication, genetic regulation, and cellular differentiation. Due to rapid fermentation cycles, genetically stable expression systems, and a single cell membrane that helps to abridge downstream processes and assist protein secretion, the bacterium is extremely adaptable and favored in various laboratory applications[1]. Containing a single circular chromosome, *B. subtilis* is replicated bidirectionally from the origin and develops with overlapping cycles in rich media, much like its Gram-negative counterpart, *Escherichia coli*. Using a growth rate-dependent cell cycle model, replication initiation may occur simultaneously at several origin sites, allowing for a new round to be prompted before the termination of the previous[2]. Although similar in rates of propagation, the Firmicute *B. subtilis* is evolutionarily distant from the γ-Proteobacteria *E. Coli,* with distinctions extending far beyond their relative replication initiation complexes.

According to the tenets of the central dogma of molecular biology, the transmission of genetic information from DNA to RNA to protein is a necessary action for all units of life[3]. Genetic regulation of transcriptional and translational processes is therefore heavily controlled to maintain an organism's homeostatic equilibrium. Within the DNA-focused *Escherichia coli* regulatory system, coupled transcription and translation is a distinguishing characteristic of genetic expression. RNA polymerase (RNAP) is physically and kinetically united with the initiating ribosome, forming a signal-integration complex where messenger RNA transcription is paired with a trailing ribosomal unit. This proximity between RNAP and the pioneering ribosome allows for transcriptional attenuation at operon leaders and Rho selective suspension

within coding regions[4]. Conversely, *Bacillus subtilis* participates in a runaway transcription model in which RNA polymerase is indifferent to translational regulatory systems (Figure 1). Similarly controlled by transcriptional attenuators at the operon, fundamental mechanisms within the Gram-positive bacterium are largely dependent on RNA-binding proteins and riboswitches- not a lagging ribosome[5]. The fundamental division of transcription and translation within *Bacillus subtilis* therefore allows for simplified biochemical manipulation of genetic expression and regulation at the transcriptional RNA level.

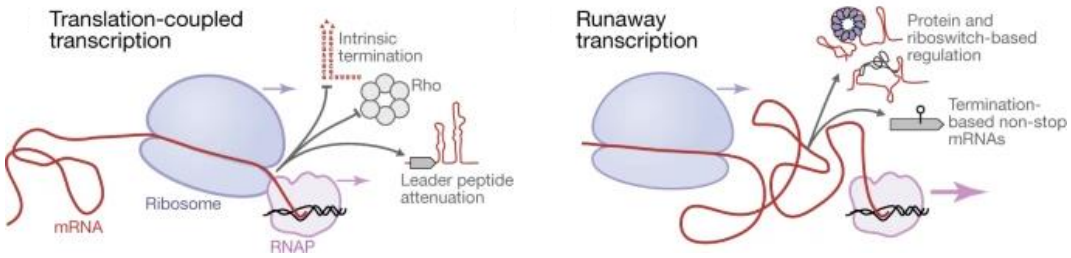


Figure 1. Schematic showing systems of translation-coupled transcription, prevalent in *E. coli*, and runaway transcription, common in *B. subtilis*. In Translation-coupled transcription, the Ribosome physically attatches to RNA Polymerase, allowing translation to initiate without the conclusion of transcription. The emerging mRNA is bound by the ribosome, producing an 'RNAP-nascent mRNA-ribosome' center that promotes various methods of translation-controlled termination of transcription. In the runaway transcription model, RNAP outpaces the ribosome, evading imeadiate translation and allowing for intrinsic transcription terminators to be more widespread. Taken from (4).

Various RNA Regulatory Mechanisms of Genetic Expression

Vital to the transmission of sequence information among nucleic acids, the control of transcription by non-coding regulatory messenger RNAs through both *cis* and *trans* mechanisms modify nearly all procedures encompassed in the central dogma, including mRNA degradation, protein function, and protein stability. Possible methods of RNA regulation consist of, but are not limited to, the targeting of mRNA by antisense oligonucleotides, interactions with proteins through allosteric inhibition, RNA interference, and the formation of complex structures, like Riboswitches, to induce attenuation[6]. The creation of intrinsic terminators is a consequence of

the current conditions and needs of the cell, with RNA acting as a highly specific biosensor and forming various local conformations to enable RNA polymerase dissociation.

Rho-Dependent Transcription Attenuation

Requiring both *cis* and *trans* transcription elements, Rho-dependent termination is a genetic regulatory mechanism in which Rho, a homo-hexameric protein, binds to mRNA at the Rho utilization site (*rut*). Dependent on the presence of C-rich and G-poor sequences, the rut site typically lacks an extensive secondary structure. Transcription is terminated in the distal region at multiple transcription stop points (*tsp*) using a ribosome free sequence of at least 85-97 nucleotides. After loading onto the mRNA at the binding position, Rho activates the ATP dependent RNA-DNA helicase to propel itself in the 5' to 3' direction towards the polymerase pause site. Various Rho-dependent termination factors also interact with Rho, such as NusA, NusB, and NusG, to help release the subsequent terminated transcripts[7]. Due to its pervasive runaway transcriptional model, however, the *B. subtilis* bacterium only allows Rho to selectively terminate operons and remove antisense RNAs without being influenced by translation[8].

Feedback Inhibition and Rho-Independent Transcription Attenuation

Succeeding transcriptional initiation and occurring between the promoter region and operon, attenuation independent of the accessory protein Rho is contingent on the establishment of a stable hairpin secondary structure. The G-C rich symmetrical dyad sequence prior to the site of termination is supplemented by a polyuridine tract at the 3' terminus that induces the interruption of RNAP[9]. In *Bacillus subtilis*, the tryptophan gene cluster *trpEDCFBA* is similarly moderated by the formation of an alternative RNA secondary structure[5]. As an example of a negative feedback loop, the activation or inhibition of the *trp* operon through an attenuation mechanism is a direct response to the quantity of tryptophan in the immediate environment.

When limited, the transcription and translation of the *trp* genes is completed in order to biosynthesize the necessary enzymes for tryptophan production. However, in tryptophan rich conditions, the 11-subunit RNA-binding protein TRAP (*trp* RNA-binding attenuation protein) binds the growing RNA and inhibits the development of the anti-terminator structure. Instead, the attenuator forms, inducing transcription termination (Figure 2)[10,11].

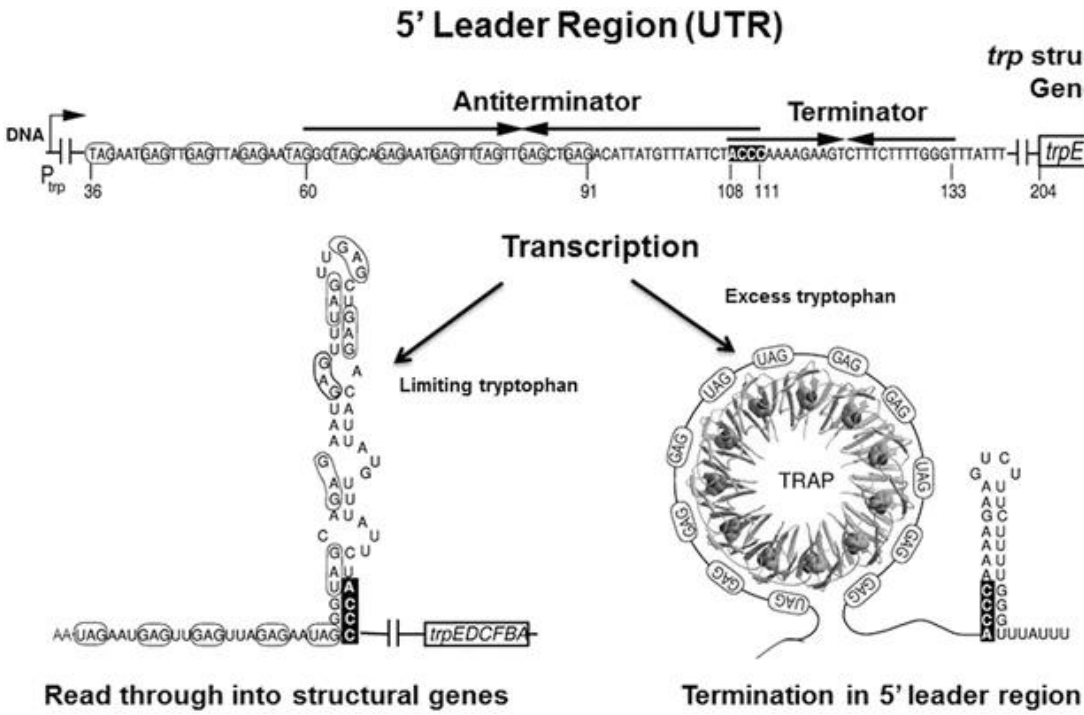


Figure 2. Facilitated regulation of transcription and negative feedback loop by the *trp* RNA-binding attenuation protein (TRAP). When tryptophan is limited in the environment, the anti-terminator is formed and allows the formation of the *trp* operon. In the presence of excessive tryptophan, however, TRAP binds to the emerging RNA and advances the establishment of the terminator. Displayed as a hairpin loop with a rich G-C stem, the 5' region of the terminator highlighted in a dark black box are the residues that overlap between the TRAP binding site and the terminator. Taken from (10).

The Riboswitch

As essential non-coding regulatory segments, riboswitches modulate gene expression through the immediate binding of small, intracellular metabolites. Located in the 5'-untranslated region of bacterial mRNAs, the structure of a riboswitch is inclusive of two primary regions: the aptamer domain (where the ligand-binding pocket resides) and the regulatory domain (also known as the expression platform). The joining of the targeted ligand to the receptor is successful in the absence of additional protein factors, and the RNA alone utilizes a *cis*-mechanism to execute genetic control and chemical recognition with high specificity[12]. Among wide variation, the thiamin pyrophosphate (TPP)-binding riboswitch is the largest class of molecule-sensing RNAs identified in not only archaea and bacteria, but also in plants, fungi, and algae[13]. Defined by the highly specific nature of the receptor, all other classes of riboswitches are restricted to the prokaryotic domain with a broad assortment of ligand preference, aiding in the biosynthesis of various molecules including metabolites like purine, and vitamin derivatives such as cobalamin.

A vital characteristic of the hypothesized 'RNA world', the riboswitch's ability to recognize and respond to the surrounding environment without DNA or protein contribution suggests their evolutionary descension from an ancient regulatory system[12,15]. The mechanism by which riboswitches regulate their respective gene expression further depends on a bacteria's taxonomy, with Gram-positive firmicute riboswitches acting as attenuators at the transcriptional level, and Gram-negative γ-Proteobacteria controlling translation initiation by exposing or obstructing the ribosome binding site (also known as the Shine-Dalgarno sequence)[14]. Transforming into compact RNA conformations, riboswitches can contain a primary base stem, a central loop, and several branching hairpin formations. Therefore, they are categorized both by

the type of ligand they bind, and their subsequent secondary structure[17]. As roughly 2% of the genes in the *Bacillus subtilis* bacterium are modulated by riboswitches, the investigations of genetic expression in the Gram-positive model organism have examined a wide variety of riboswitch classes, beginning with those that aid in the repression of the riboflavin operon[16]. Although originally thought to be the result of regulator proteins RibC and RibR, both substrates modulated the *rib* operon indirectly, as flavokinase activity helped to decrease flavin mononucleotide concentration. Instead, evidence of both transcriptional attenuation at the 5'-untranslated region and the conservation of upstream sequences with the ability to fold into alternative, complex secondary structures were present[14].

With the ability to be exist in either structural region, the incorporation of a switching sequence into the aptamer domain or the expression platform is highly dependent on ligand binding, and subsequently determines the mRNA's expression result (Figure 3). When integrated into the aptamer domain, the expression platform transforms into a Rho-independent transcriptional terminator stem-loop that causes RNA polymerase to abort. Conversely, if the switching sequence is inserted into the expression platform, an anti-terminator stem-loop is produced, allowing transcription to advance. Although some classes are exclusively conditional to ligand concentration and operate thermodynamically, a high percentage of riboswitches also function under kinetic control and fail to reach equilibrium. As such, the ability of the riboswitch to modulate transcription is contingent on the rate of RNA polymerization and overall ligand association[16]. The genetic outcome made by the riboswitch is therefore irreversible.

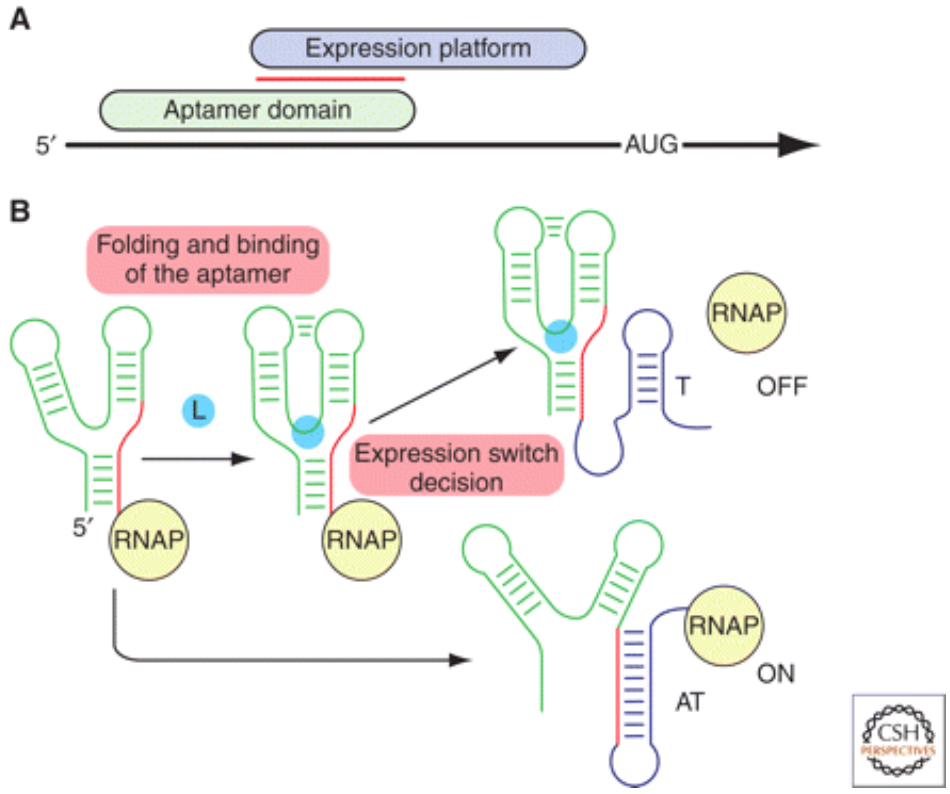


Figure 3. Mechanism of a generalized riboswitch. **(A)** The domains of the riboswitch are arranged as overlapping sequences in the mRNA transcript, with the aptamer domain (green) preceding the expression platform (blue). Nucleotides that may base pair in either structural element are conveyed by the red bar. **(B)** Throughout the transcription of the riboswitch, the binding site may be occupied by the ligand (cyan) or left empty, creating the terminator or anti-terminator in the expression platform. If the switch turns *off*, RNAP disconnects from the mRNA. Taken from (15).

To allow for sufficient binding, the kinetically controlled riboswitch requires a ligand concentration greater than the dissociation constant (K_D)[16,18]. When concerning the order of synthesis throughout transcription, the aptamer domain is the first element of the RNA to be created and folded, producing a ligand binding pocket. The RNA polymerase is then temporarily stalled at the pause site, permitting the riboswitch to recognize and respond to the cellular environment. Dependent on the speed of ligand binding, the switching sequence is integrated into either the aptamer domain or the expression platform, resulting in the respective terminator

or anti-terminator stem-loop structure. As RNA polymerase transcribes at approximately 100 nucleotides per second, the concentration of ligand must sufficiently allow for the rapid binding of the metabolite to the aptamer domain in order to create the Rho-independent attenuating secondary structure[19].

The *pbuE* Adenine Responsive Riboswitch

The purine family is a commonly studied binding class of riboswitch that is associated with various ligands and their analogs, including nucleotide derivatives and coenzymes. Serving as a model by which the mechanism of ligand recognition can be examined, the three-way junction seen within the purine responsive aptamer domain is composed of three paired helices (labeled P1, P2, and P3) that contain scarcely conserved, Watson-Crick base paired sequences[21]. On the other hand, the subsequent joining regions (J1/2, J2/3, and J3/1) and capping hairpin loops (L2 and L3) have a pattern of nucleotide conservation that suggests a functional importance. Facilitated through the coaxial stacking of the P1 and P3 regions to form an unbroken helix, the tertiary structure of the binding pocket is created when the L2 and L3 loops form a pseudoknot. As the structure folds, P2 and P3 are brought together to form the recognition site, and the various junctional regions are used in stabilizing interactions with the ligand (Figure 4)[19].

Due to the highly conserved nature of the RNA sequencing and the stringent requirement of specificity at the binding pocket, the aptamer domain has been the focus of many previous investigations. In *B. subtilis*, the *xpt-pbuX* guanine riboswitch and the *pbuE* adenine responsive riboswitch were the foremost members of the purine family to be discovered and extensively studied[21]. Although similar in sequence and secondary structure when concerning their respective binding pockets, the adenine responsive riboswitch exhibits a cytidine (C74) to

uridine (U74) mutation in the J3/1 joining region that proposes its involvement in ligand recognition[20]. Many elements of the expression platform, however, remain misunderstood, as wide variability within the RNA sequence can lead to similar results in regulation. Consequently, this study focuses specifically on the expression platform of the *pbuE* adenine responsive riboswitch in order to further investigate various structural interactions between either domain that allow for successful transcriptional modulation.

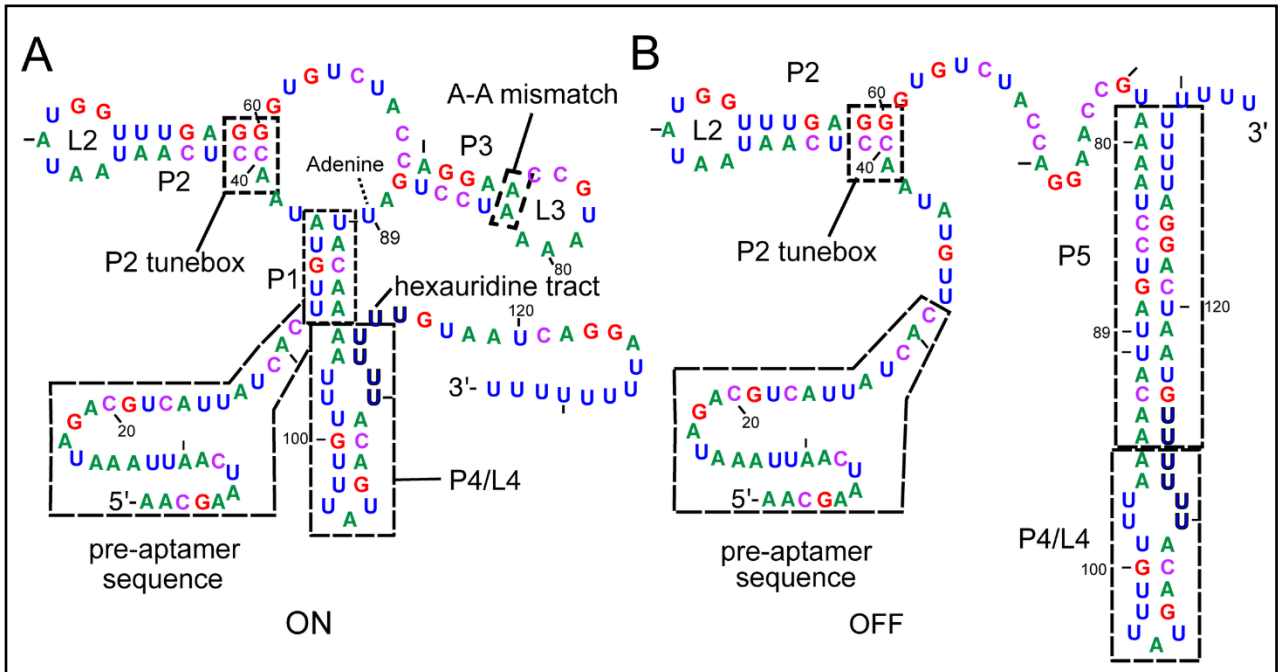


Figure 4. *ON* and *OFF* conformations of the wildtype *pbuE* adenine responsive riboswitch. **(A)** The *ON* secondary structure includes the adenine ligand bound to the aptamer domain to promote the formation of the anti-terminator. **(B)** When there is a lack of adenine in the environment, strand invasion occurs and the terminator hairpin-loop (P5) is formed. RNA polymerase then disengages with the sequence and transcription is repressed. Taken from (22).

Advantages of utilizing the *pbuE* adenine responsive riboswitch as a model system are abundant. Like others in the purine family, *pbuE* contains a simple structural design that results in predictive behavior and a resistance to misfolding[22]. Furthermore, the riboswitch is highly tolerable to adenine analogs, such as 2-aminopurine (2AP), that are less metabolically utilized and not as harmful to an organism in higher intracellular concentrations. Finally, the *pbuE*

riboswitch, in encoding for the removal of excess toxic adenine through the genetic modulation of the purine efflux pump, functions as an *on* switch in the presence of the adenine ligand. Thus, in the unbound state, the creation of the purine efflux pump is repressed by the premature attenuation of the transcriptional sequence[23]. The riboswitch is therefore kinetically contingent on both the concentration of ligand in the cellular environment and the ability for the aptamer domain to rapidly fold throughout transcription.

Conformations and Known Mechanisms of the NH5 Variant

While various riboswitches have been commonly explored since their initial 2002 discovery, numerous questions about what allows successful genetic modulation by the RNA component remains. An effective riboswitch has an increased difference between expression observed in the *on* state versus the *off* state, and although leaky expression is probabilistic, the system's ability to accurately modulate a gene is ultimately restricted by an amplified expression at the basal level. Hence, an insight into the riboswitch's sequence, specifically within the expression platform, is required to understand how the system effectively adjusts transcription with or without the binding of a ligand to the aptamer domain.

This study utilizes a previously engineered riboswitch modeled after the structure of the *pbuE* adenine responsive riboswitch. Obtained through a mutagenic process, the regulatory system, referred to as NH5, was created by Batey and Drogalis with the intent of both simplifying the wildtype and further identifying the regions of the *pbuE* expression platform that directly influence an effective secondary structure (Figure 5)[22]. Exhibiting far greater ligand-control mechanisms than the original model, the shortening of the pre-aptamer sequence to

reduce mis-folding and the alteration of the P1 and P4 regions allowed for an overall improved riboswitch.

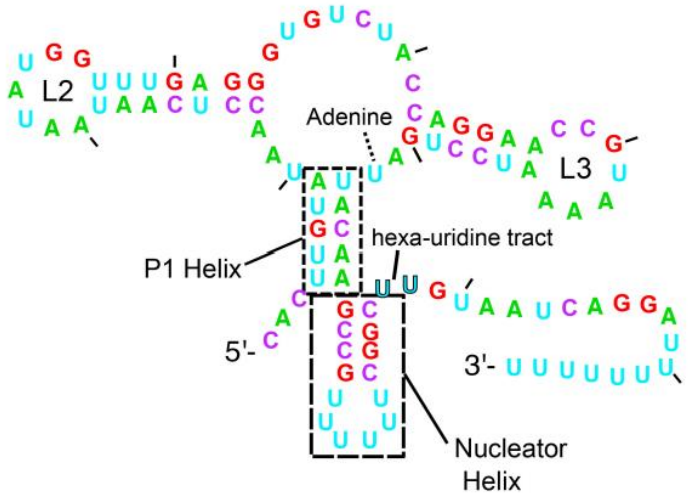


Figure 5. The Anti-Terminator loop of the NH5 Variant. In contrast to the original *pbuE* wildtype, the NH5 variant contains a minimized expression platform and shortened nucleator helix that lacks an internal bulge. As the focus of this study, the Nucleator Helix hairpin stem-loop, P1 Helix, and hexa-uridine tract are labeled. The P4 region also includes four canonical GC base pairs at the stem and six uridines within the loop. Designed in the Batey Lab based on Drogalis and Batey (22).

Structure

In comparison to the 4.5-fold induction value of the *pbuE* wildtype found in *Bacillus subtilis*, the NH5 variant presents a dramatic increase to 120-fold induction when transformed into *Escherichia coli.* Inclusive of an unchanged aptamer domain, NH5 also contains a simplified "nucleator helix" that truncates the original P4/L4 region of the wildtype to only include 4 GC base pairs. The hexa-uridine tract was also decoupled from both P4 and P1 and moved to expose L4, thus eliminating the internal bulge. Finally, the 11 nucleotides of the pre-aptamer sequence were deleted to improve the overall expression of the purine efflux pump in the presence of a 2AP ligand[22].

Strand Invasion

The course by which both the *pbuE* wildtype and the minimized NH5 design disrupts base pairing and forms the competing intrinsic terminator is a process called strand invasion. Throughout transcription, a hexa-uridine tract (found at the 3'-end of the P4/L4 or nucleator

stem-loop region and present within both *on* and *off* conformations) is formed as the aptamer domain simultaneously folds into the recognition site. However, without the binding of the ligand, the sequence on the 3' side of the P4/Nucleator element begins to "invade" the P1 helix. The J3/1 and P3 regions are rapidly penetrated, and as a result, the terminator helix is formed. Conversely, if the ligand is present, strand invasion is blocked by the arrangement of a ligand-dependent "gate" that prevents disruption of J3/1 and P3 by forming base triples. The RNA polymerase is provided additional time to continue transcription and pass the hexa-uridine tract, officially avoiding disengagement by the riboswitch (Figure 6)[22].

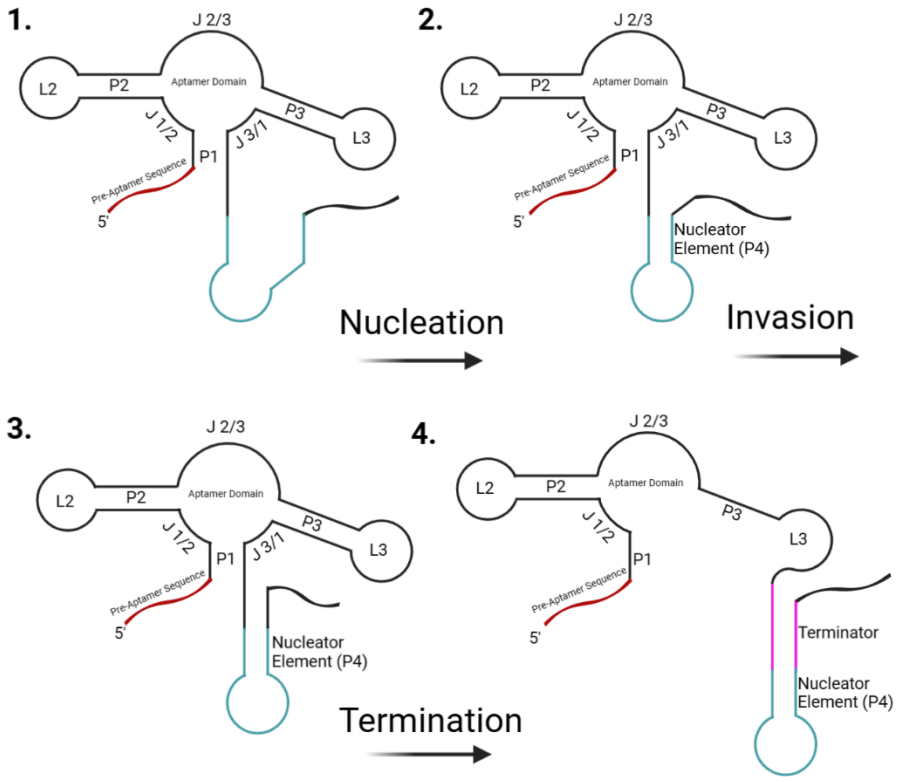


Figure 6. Structural Role of the P4 helix toe-hold in promoting strand invasion in the NH5 variant. The nucleation of the terminator helix is dependent on the base pairing of P4, or the nucleator element. In NH5, four GC pairs create the toe hold that influences successful genetic repression in the absence of the adenine ligand. The 3' end of the switch invades into P1, allowing for both the formation of new base pairs and the proliferation of P4. P1 is completely displaced, the aptamer domain becomes disconnected at the P3/L3 regions, and the terminator helix is formed. Made with BioRender.

Strand invasion is heavily dependent on the successful nucleation of the P4 region, or the toehold, that permits the continued propagation into P3. Both occurrences require complex, unknown conditions of base pair stability throughout various regions of the riboswitch's secondary *off* structure, prompting an investigation into the physical and kinetic necessities that promote successful repression with minimized leaky expression. Previous studies investigating whether or not the expression platform would tolerate changes in base pairing showed that specific sequences of nucleotides- like the site of stand invasion initiation (three base pairs of P1 distal to the junction)- display meaningful impacts on the effectiveness of a switch. Structures with weak canonical base pairs (-AU or -GU), within either P1 or the invading strand, fail to form the terminator, while those with stabilizing pairs (-GC) demonstrated expression measurements similar to that of the wildtype. However, the riboswitch's acceptance of variations in the length of the P1 helix also produced two suggestions: that the mechanism of invasion was governed by kinetics rather than thermodynamic stability, and that the P4 Hairpin-Stem and L4 Loop structures are significant in dictating the regulatory response of the expression platform[22,24].

Genetic Screening of the P4 Hairpin-Stem Library

Originally divided into 4 nucleotide regions of interest, the NH5 expression platform has been the focus of an extensive collaborative project within the Batey laboratory. This study, however, will be specifically analyzing the modified Hairpin-Stem library (located in the P4 region) under the assumption of previous findings from the P4 Hairpin-Loop. The rapid initiation of strand invasion in all riboswitches is usually coupled with a section of the aptamer domain and includes a hairpin element that is formed in both the *on* and *off* conformational states. In the *pbuE* adenine-responsive wildtype and the NH5 variant, this hairpin element is the P4/L4 region.

Diverging from the structure of *pbuE*, the Hairpin-Stem of the NH5 variant contains four G-C pairs that, due to a substantial increase in fold induction, were thought to aid in nucleating the P4 helix and generally stabilizing strand invasion. While an investigation on base pairing in the P4 Hairpin-Stem was previously conducted by a fellow Batey lab member, Lisa Hansen, the size of the sequence tested (8 nucleotides) prevented a clear result. Utilizing an equation quantifying the percent completion of each library (detailed in section II of this study), the randomization of 8 nucleotides would require approximately 350,000 colonies to be screened to approach a 95% confidence that all unique sequences were observed. Though the investigation only reached 13.2% (9,245 colonies), flexibility of the region was suggested due to the variety of successful riboswitches found with unpredictable base pairs. However, because of the diverse nature of the selected switchers, the most effective sequences failed be identified despite the stem region displaying a preference for stability.

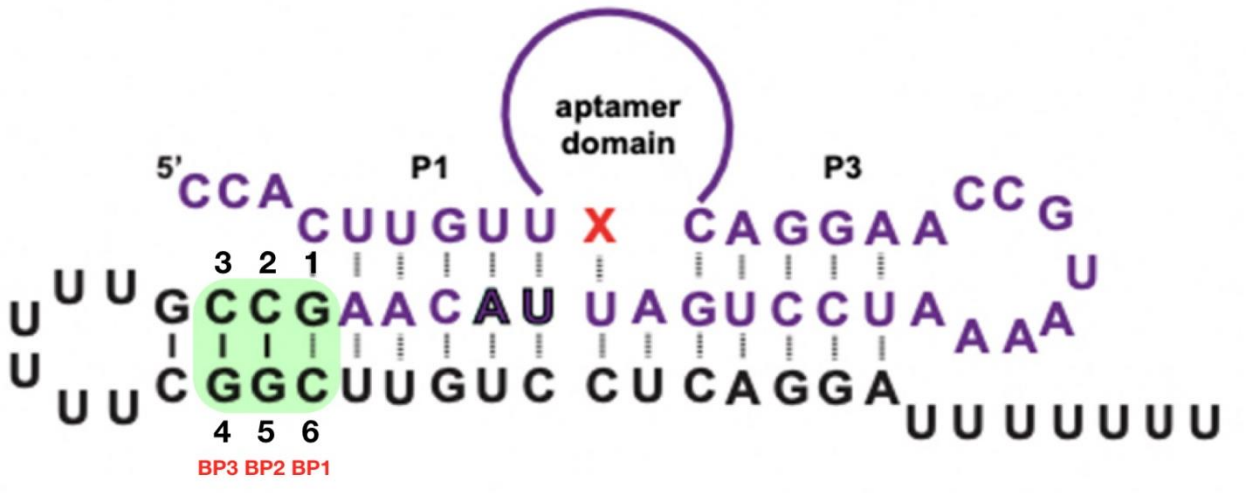


Figure 7. Base pairing within the P4 Hairpin Stem Library of the NH5 variant riboswitch. The secondary structure of the NH5 riboswitch expression platform includes various organizational elements, such as the P1 and P3 regions (purple) that are invaded during transcriptional termination. The stem of the nucleator helix contains four strong Watson-Crick G-C base pairs, although only the three most distal to the polyuridine loop (labeled BP for base pair and numbered 1 through 3) are selectively examined throughout this study. Highlighted in green, the six nucleotides of focus are numbered 1 to 6 starting at the 5' end of the sequence. Adapted from the NH5 structure created by Robert T. Batey PhD.

To obtain a further understanding of how specific base pairs encourage genetic repression in the absence of ligand, the following mutagenic screening of the NH5 sequence was performed. Involving the randomization of 6 nucleotides within the hairpin stem, the stabilizing GC pair proximal to the polyuridine loop was maintained in order to truncate the selected library, thus reducing the amount of colonies necessary to screen in order to reach a 95% confidence of observation (Figure 7). Prior reports have determined that the functional ability of the riboswitch is not contingent on the sequence specificity of the polyuridine loop. Therefore, the effectiveness of the region relies heavily on canonical base pairing within the P4 helical stem[25,26]. Coupled with a fluorescence based assay, the following muti-step examination revealed sequences that repressed expression in the absence of ligand in similar manner to or even better than the NH5 variant due to a preference for weaker stability in the randomized base pair closest to both the P1 region and the hexa-uridine tract. Additionally, further evidence is provided towards the hypothesis that canonical base pairing is a required element of the P4 library within *Bacillus subtilis's pbuE* adenine-responsive riboswitch, allowing for the formation of a toehold that promotes strand invasion.

II. Methods

Cloning of the P4 Hairpin Stem-Loop Library

The schematic illustrated below (Figure 8) establishes the protocol in which the various libraries of the *B. subtilis* pbuE riboswitch, including the P4 Hairpin Stem-Loop, were initially cloned, screened, sequenced, and activity assayed. Each step following the initial Polymerase Chain reaction, Restriction Enzyme Digest, Polynucleotide Kinase Treatment, and ligation into the pRR5-gfpUV plasmid was repeatedly performed in 10 separate rounds to maximize the amount of colonies observed. The percent completion of the Hairpin Stem-Loop was quantified by the following equation, where n is the number of nucleotides randomized (in this case, 6) and *P* is the probability of observing every possible colony:

$$\textit{Colonies screened} = -4^{n} \ln (1-P)$$

For this study, a 96.83% probability was met, with a total of 14,145 colonies screened on primary plates.

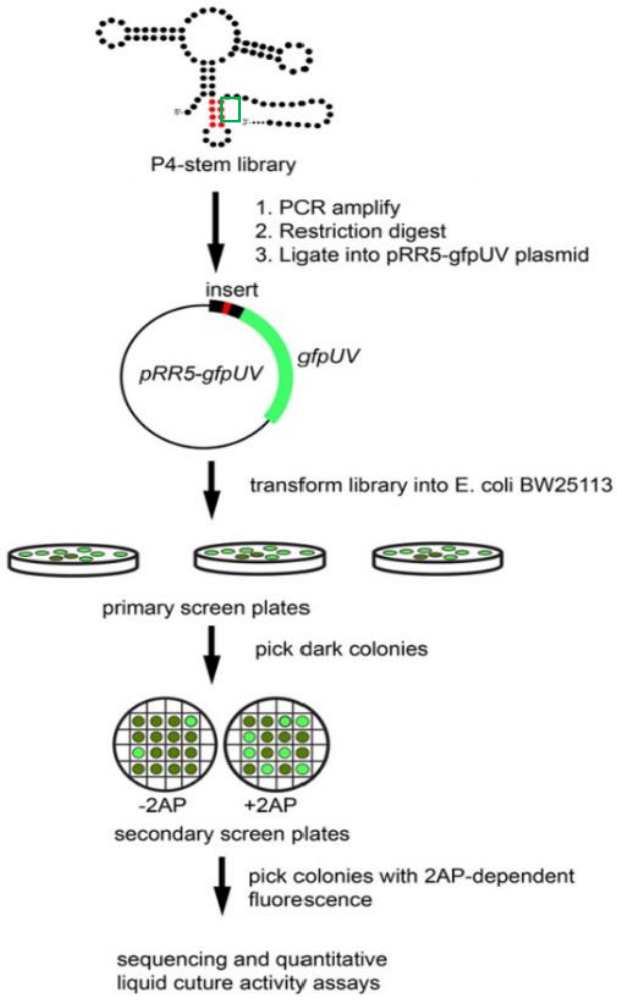


Figure 8. Graphic detailing the workflow of cloning and screening the P4 Hairpin-Stem library. Synthesized through a Polymerase Chain Reaction, the nucleator element of the NH5 riboswitch was then annealed and ligated into the pRR5-gfpUV reporter plasmid. The library was transformed into *E. coli* (BW25113) to allow for a primary screen on CSB carbicellin plates without 2-Aminopurine. Colonies were then chosen based on visible fluorescence, with the dimmest "winners" moving forward to a secondary gridded plate screen in both the prescence and absence of ligand. Finally, those elected from the secondary screen based on successful repression and expression (once again, through the visual analysis of fluorescence) moved on to sequencing and the liquid culture activity assay. Colonies with mutations outside of the randomized library (the 6 nucleotides surrounded by a green box distal to the Hairpin-Loop) were not considered viable and removed from the study.

Polymerase Chain Reaction

The double-stranded DNA used within the genetic screen was constructed, copied, and amplified through a standard in vitro polymerase chain reaction, or PCR (Figure 9)[28]. Manufactured by Integrated DNA Technologies (IDT), the single stranded oligonucleotide sequences and various primers utilized throughout the PCR process included the isolated Hairpin-Stem library, the specially designed outer DNA primers "5'Gblock_Nsil" and "3'Gblock_HindIII", and two inner Ultramers (detailed within Table 1). In preparation, the oligonucleotide was suspended in ddH_2O to create a concentration of 100 µM. Each insert was then synthesized by a 50 µL reaction consisting of 36.5 µL ddH_2O, 10 µL 5X Q5 buffer, 1 µL

dNTPs, 0.5 µL NsiI 5' primer, 0.5 µL HindIII 3' primer, 1 µL of a 1:100 solution of adaptor and Hairpin-Stem library insert (1 µL 3'_SpeI_HindIII_pRR adaptor, 1 µL insert, and 98 µL of ddH-$_2$O) and 0.5 µL Q5 polymerase. The mixture was then subjected to 12 rounds of amplification in a PCR thermocycler, with each round including 30 seconds at 95 °C for denaturing, 30 seconds at 60 °C for the primer to anneal, and 45 seconds at 72 °C for the primer to extend. A 10 minute period at 72 °C was then conducted to confirm that the resulting product was double stranded. Altogether, the final dsDNA consisted of 228 base pairs. The length of the product was confirmed by running the reaction in a 2% agarose gel, as seen in figure 10. Once verified, the PCR product was cleaned with the Omega Biotek EZNA Cycle Pure Kit according to the manufacturer's recommendations[27].

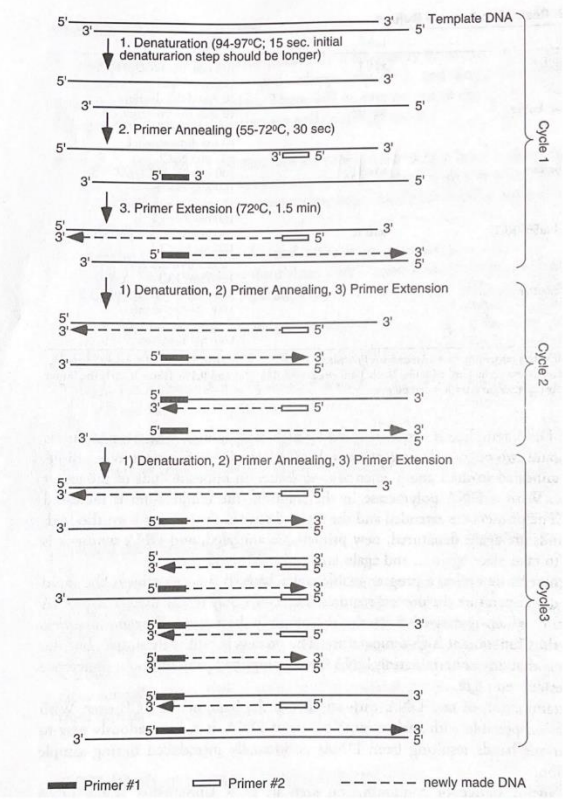


Figure 9: Generalized process of a Polymerase Chain Reaction. While a typical PCR utilizes a series of 20-40 rounds, the Hairpin-Stem library was synthesized from 12 thermocycles to reduce biases that occurred throughout the amplification process. After denaturation, the temperature of the reaction is lowered, and two oligonucleotide primers flank the DNA on opposite ends. The primers are annealed to their complementary sequences, and once temperature increases, the DNA polymerase in the solution begins extending the region between each primer. This process is then repeated in each individual cycle. In comparison to the figure, primer extension within the study was shortened from 1.5 minutes to 45 seconds as the precise time required for elongation is highly dependent on the DNA polymerase utilized and the length of the target region. After the finishing cycle, a 10 minute period at 72 °C confirmed that any remaining single-stranded DNA was fully elongated. Taken from (28).

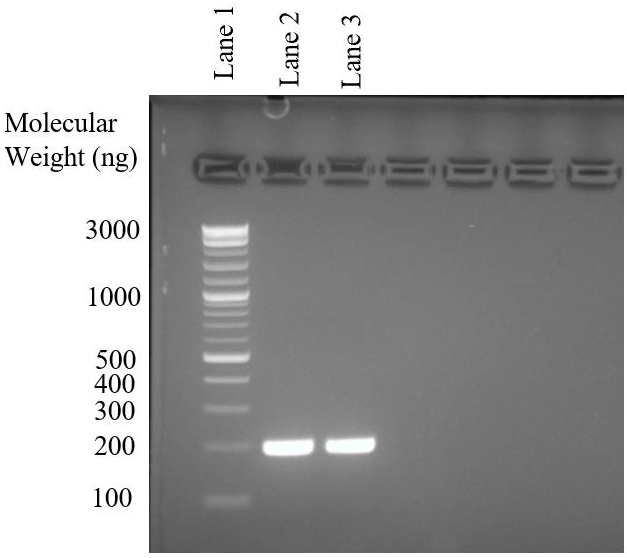


Figure 10: 2% agarose gel containing the PCR product. After 12 rounds of amplification, the length of the resulting dsDNA was verified through processing 200 ng of DNA in an agarose gel. Lanes 2 and 3 contain 200 ng of the product mixed with a 6X purple loading dye from New England Biolabs (NEB). Lane 1, conversely, contains the 5 μL of the NEB Quick-Load 2-Log DNA ladder (0.1-10kb). The gel was run for 50 minutes at 110 V/400 mA.

Table 1: Oligonucleotides utilized for the creation of the Hairpin-Stem dsDNA

The randomized nucleotides in the NH5 Hairpin-Stem Library are denoted by the letter N below.

Name	Sequence (5' to 3')
Del27_NH5_lib_stem	TTTACGGGCATGCATAAGGCTCGTATAATATATTCCAC TTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACC AGGAACCGTAAAATCCTGATTACAANNNNTTTTTTNNN NTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTA AGTAAAGGAGTT
3'_SpeI_HindIII_pRRadaptor	GCATGCAAGCTTGGCGTAATCATGGTCATAACAAACTC CTTTACTTAAATGTACTAGTA
5'GEN-Gblock_NsiI	TTTACGGGCATGCATAAGGCTCGTATA
3'GEN-Gblock_HindIII	AGGCATGCAAGCTTGGCGTAATCATGG
pRR_Forward	GCGCTAGCCACAGCTAACAC

Restriction Enzyme Digest

A digest with 1 μL NsiI-HF and 1 μL HindIII-HF restriction enzymes (constructed by New England Biolabs) was performed in order to anneal the insert with the reporter plasmid. The 50 μL reaction, also consisting of 5 μL of the 1X CutSmart buffer and 43 μL of the PCR product,

was conducted through a 1 hour incubation at 37 °C. Moreover, the reporter vector was similarly digested in a 100 μL reaction which contained 85 uL of the vector (pRR-gfpUV), 10 μL of 1X CutSmart buffer, 2 μL of NsiI-HF, 2 μL of HindIII-HF, and 1 μL calf intestinal alkaline phosphatase (CIP). The inclusion of CIP permitted the dephosphorylation of any cleaved terminal phosphate groups within the plasmid prior to the polynucleotide kinase treatment. Both products were once again purified with the Omega Biotek EZNA Cycle Pure Kit following the instructions given by the manufacturer[27].

Polynucleotide Kinase (PNK) Treatment

The restriction enzyme insert (15 μL) was treated with 1 μL of T4 polynucleotide kinase, 2 μL PNK buffer, and 2 μL 10mM ATP for 30 minutes at 37 °C and heat activated for 20 minutes at 65 °C. Originally utilized to remove phosphatase contamination present in the HindIII-HF enzyme, the PNK treatment also allows for improved efficiency and a higher yield of the following ligation procedure.

Ligation of Vector and Insert

According to molecular cloning protocol, the PNK treated PCR Hairpin-Stem insert products were ligated with the digested vector. The 20 μL reaction, containing 14.5 μL ddH_2O, 2 μL of the pRR-gfpUVvector (close to 80 ng/μL), 0.5 μL of the insert (about 40 ng/ μL), 2 μL of 10X T4 ligase buffer, and 1 μL ligase T4, sat at room temperature for 45 minutes.

Transformation of Ligated Plasmids

To complete the replication process, both the ligated plasmids and various control inserts (pBR322, NH5, and gfpUV) were transformed into KEIO parental cells that were descended from E. coli K-12 BW25113 and made chemically competent by the rubidium chloride method. 3 μL of the nucleator ligation plasmid and inserts were added to a 100 μL KEIO cell tube, while

1 μL of each control were pipetted into separate collections of 30 μL KEIO cells. Each sample was subsequently allowed to rest on ice for 15 minutes, heat shocked in a 37 °C water bath for 90 seconds, placed back on ice for another 5 minutes, and injected into 700 μL (350 μL for the controls) of 2xYT media following a sterile technique. The cells were then incubated in the 37 °C water bath for a shortened length of 30 minutes to allow for proliferation but prevent over replication. Once removed from the bath, the samples were centrifuged at 5000 rpm for 5 minutes and resuspended in 400 μL (100 μL for controls) fresh 2xYT media. Finally, the randomized library and transformed pBR322, NH5, and gfpUV were plated onto separate CSB carbenicillin agar plates and left to grow overnight at 37 °C.

The procedure of plating the Hairpin-Stem library onto media without the addition of 500 μM 2-Aminopurine ligand diverges from other examinations of the NH5 riboswitch variant as it reduces the selection of false positive colonies within the primary screening process. In order to determine what type of primary plate was necessary for each library, the cells were initially transformed onto both CSB+ and CSB- plates (in which + refers the presence of 2AP and – denotes its absence). As the Hairpin-Stem library prevents the riboswitch from forming a terminator helix in environments with high concentrations of the ligand, colonies that were able to effectively repress transcription in the absence of 2AP were more easily selected (dim colonies on CSB – plates).

Primary and Secondary Colony Screening

The transformed colonies on the CSB carbenicillin plates were analyzed under a long-wave 365 nm *UV* light in order to elect the most effective riboswitches based on noticeable fluorescence (Figure 11). For the Hairpin-Stem library, colonies that appeared the dimmest were

chosen for a secondary gridded plate screening under the assumption that they were successful in repressing expression.

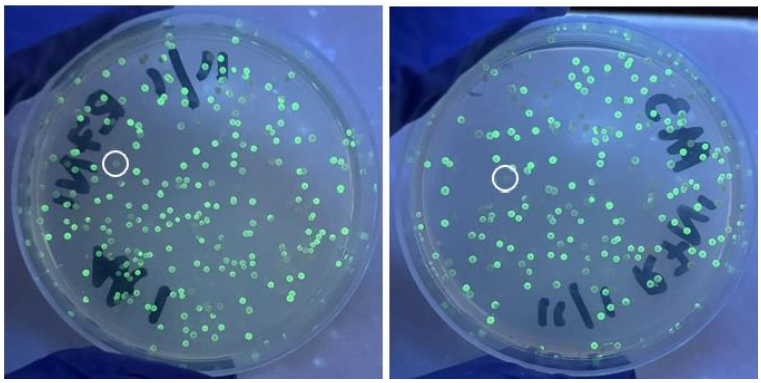
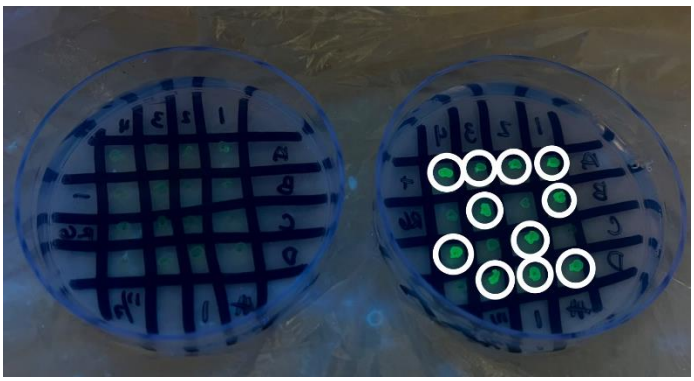


Figure 11. Primary screening of the NH5 Hairpin-Stem Library. After being transformed onto CSB- plates, dim colonies (circled in white) were chosen to proceed based on visual fluorescence and assumed to form a successful terminator helix.

Those selected for the secondary screening were plated in both the presence and absence of the 2AP ligand and, once again, incubated overnight at 37 °C. As seen in figure 12, the pairs were subsequently observed under the long-wave UV light to isolate efficient riboswitches that were capable of turning *off* when placed on the negative plate (dim) and *on* when placed on the positive (bright). Colonies that were considered "winners", or functional, were then permitted to be sequenced and activity assayed.

Figure 12. Secondary Screen Grid Plates for the NH5 Hairpin-Stem Library. Colonies were plated on both CSB- (left) and CSB+ (right) plates and left to grow overnight. Based on the repression and expression ability of the riboswitch, functionality was determined by observations of fluorescence. If a switch was dim in the absence and bright in the presence of 2AP, they were considered "winners" (circled in white) and proceeded on to both sequencing and the activity assay.



Sequencing

The colonies selected from the secondary screen were grown overnight in 3 mL of 2xYT and 3 μL of ampicillin. Utilizing the Omega Biotek EZNA Plasmid DNA Mini Kit I, the plasmids were subsequently isolated and eluted in 50 μL of ddH_2O. After being archived, 10 μL of the plasmid was combined with 5 μL of the pRR_Forward primer (detailed in Table 1) so that a final primer concentration of 1.67 μM was reached. These primers were then sent to the QuintaraBio Lab for Sanger sequencing and finally compiled into a FASTA file (located in the Appendix). However, sequences with mutations outside of the 6 nucleotides randomized were ultimately removed from the screen.

Activity Assay

To determine the fold induction of each riboswitch, an activity assay was performed and consequently compared with the respective sequences to measure effectiveness. Each assay utilized three separate controls that are detailed as follows: pBR322 (a negative control that does not contain the *gfpUV* gene and fails to exhibit activity in the presence of 2AP), pRR5_gfpUV (the positive control that continuously expresses the *gfpUV* gene at an increased level), and the NH5 riboswitch plasmid. The colonies that were able to pass both secondary screening and the sequencing analysis (as well as the pBR322, NH5, and gfpUV controls) were once again grown overnight (about 14-18 hours) at 37 °C in a rotating drum using 3 mL of 2xYT and 3 μL of ampicillin. The following morning, each culture was then transferred into two separate tubes: a negative sample (containing 3 mL of defined CSB media, 3 μL of the cell culture, and 3 μL of ampicillin) and a positive sample (containing 3 mL of defined CSB media, 3 μL of the cell culture, 3 μL of ampicillin, and 15 μL, or 500 of μM, 2AP).

The selected cultures were allowed to mature for 6 hours to reach an optical density (OD_{600nm}) range of 0.4-0.6. Once achieved, 200 µL of each culture was pipetted in triplicate into a Costar® 96 well plate- along with a single sample of 5 µg/mL fluorescein to set gain- and analyzed by a Tecan Infinite M200® PRO plate reader. The program, using an excitation wavelength of 395 nm and an emission wavelength of 510 nm, determined both the fluorescence and OD_{600nm} of each culture. Each assay was also performed at least three times in biological conditions (resulting in a total of 9 data points), with some samples needing additional screens due to either large variations in fold induction value or increased interest.

Data Analysis

In order to quantify fold induction, measurements of optical density (OD) corrected fluorescence of each colony was calculated by the following formula:

$$OD\ corrected\ fluorescence = \frac{total\ fluorescence}{OD\ (600nm)}$$

Using the subsequent normalized value, the background corrected fluorescence was found by subtracting the median of the three pBR322 corrected fluorescence from the OD corrected fluorescence:

$$background\ corrected\ fluorescence = (OD\ corrected\ fluorescence) - (median\ background\ fluorescence)$$

Finally, the fold induction for each colony was acquired by dividing the background corrected fluorescence in the presence of ligand by the background corrected fluorescence in the absence of ligand.

$$fold\ induction = \frac{background\ corrected\ fluorescence\ in\ the\ presence\ of\ ligand}{background\ corrected\ fluorescence\ in\ the\ absense\ of\ ligand}$$

It is important to once again note that since each culture was observed in both technical and biological triplicate, *at least* nine data points exist for all colonies that expressed a fold induction above the value of 2 and had no mutations outside of the randomized library. After compiling all of the calculated data into an organized excel (Microsoft) document, the file was formatted into a csv and analyzed within R studio so that the library would be consistent with the standards set by previous studies within the Batey collaborative project. Colonies were removed if they exceeded a factor of 1.5 with respect to the interquartile range, and various plots were created to observe the relationship between repression value and fold induction. The code utilized within the study was created by Lisa Hansen and is freely available at the following link: https://github.com/bateyLab/ExpressionPlatform_purineRiboswitch/blob/main/expression_platform_files/analysis/n6_p4stem_analysis.Rmd. A t-test was also performed on select colonies to determine if a significant difference existed between their -2AP OD corrected fluorescence and pBR322.

Although the calculations used to quantify switching ability successfully provide a ratio between maximal and minimal measures of expression, the correction of background cellular fluorescence has its constraints. Certain variants within the study, in the absence of 2AP, were able to exhibit a smaller amount activity than that of pBR322. As a result, colonies with less fluorescence per cell density than the control had a negative fold induction. In order to include highly repressive colonies in the analysis and correct this limitation, switches that presented nonsensical fold inductions were manually adjusted so that the value of fluorescence in the absence of ligand is equal to 50. This quantity remains lower than the 340 average of the parental NH5, but ultimately allows for a reasonable comparison between variants. The reported standard error from the original data set was also retained.

III. Results

Investigations into the P4 Hairpin-Stem library has, in the past, failed to reach a strong conclusion due to the original size and inability to observe over 95% of all possible variants within a shortened time period[25, 26]. By reducing the focus to 6 randomized nucleotides rather than 8, 96.83% of all possible sequences (14,145 colonies) were able to be detected on primary plates. After the initial screen, 395 colonies were subsequently surveyed on a secondary grid plate, and of those, 191 were collectively assayed and sent to the QuintaraBio Lab for Sanger sequencing. Of the final selection, 127 functional switches (about 66.5% of the samples nominated from the grid plate) exhibited a fold induction greater than 2 and conserved sequencing outside of the selected randomized nucleotides (Table 2).

Table 2: Hairpin-Stem library Colony Selection

Number of colonies observed on primary screen plate	14145
Number of colonies picked for secondary screen	395
Number of colonies picked from secondary screen	191
Number of colonies that exhibited at least two-fold switching	127
Fraction of variants observed	96.83%

With the intent of determining the structural requirements of an effective switch, each colony was categorized into three separate bins based on both repression value and fold induction (Figure 13). Those that have a repression value greater than 9 are considered inefficient and were sorted into the “Bottom” class. The “Middle” bin is inclusive of both measurements, containing colonies that have a repressions less than or equal to 9 and a fold

induction below 0.5. Finally, those in the “Top” division retained a fold induction above 0.5 and are considered the most successful riboswitches.

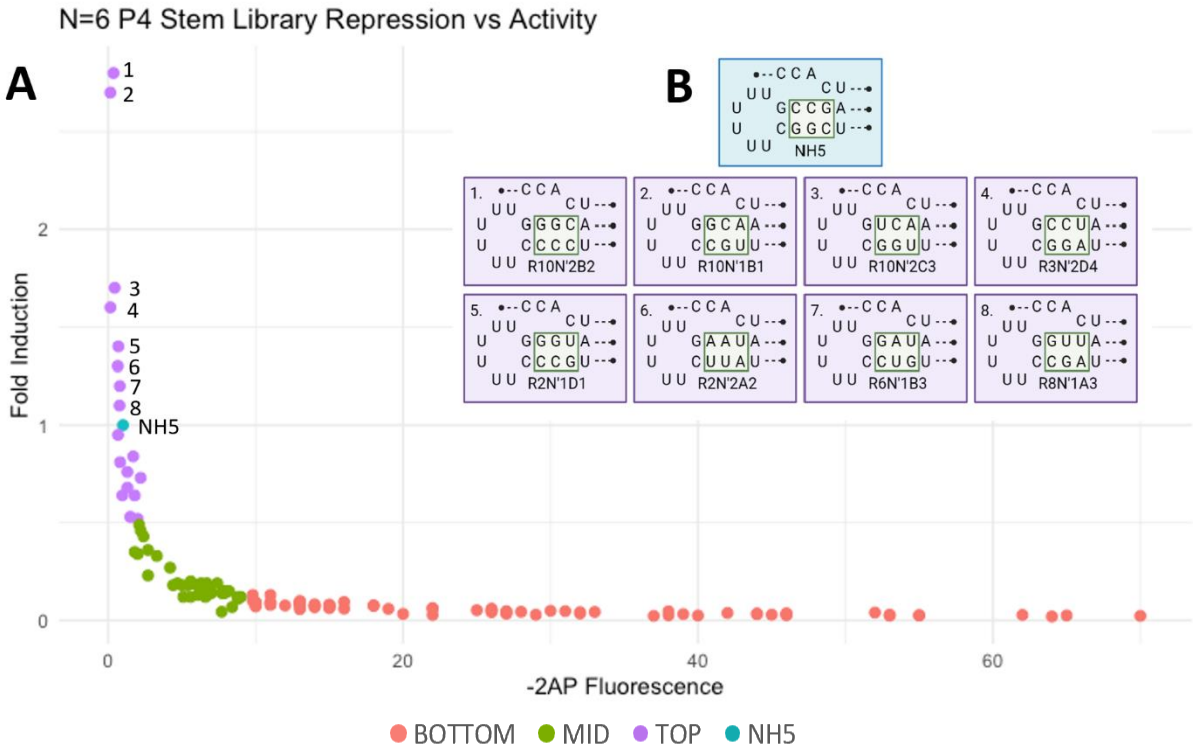


Figure 13. Graph, seperated by bins, depicticng the P4 Hairpin-Stem Library Background Corrected Flourescence in the Absence of 2AP vs Fold Induction. (A) Each item on the graph corresponds to a unique sequence that has at least 9 data points and a corrected fold induction above the value of 0.20. The NH5 variant is highlighted in blue, and each bin is separated by color, as detailed in the legend. The “Top” division (purple, 19 total colonies) requires a fold induction value above 0.5, while the “Bottom” division (coral, 71 total colonies) contains cultures with a repression value above 9. The “Middle” bin (green, 37 total colonies) is, therefore, inclusive of the remaining switches falling outside of either parameter. It is important to note that if evaluated against comparable graphs of the other libraries of the NH5 variant, the P4 Hairpin-Stem exhibits a similar negative exponential function. **(B)** Corresponding to the data points numbered on supplement graph A, the various sequences detailed are those that exhibited a Fold Induction higher than that of the NH5 variant. Adapted from the original graph made by Lisa Hansen.

As the original *pbuE* riboswitch lacks a GC-rich base within the P4 region, the NH5 parental sequence of the Hairpin-Stem was thought to prefer strong canonical base pairing. Grounded in both the improved fold induction of NH5 from the initial wildtype and the common

inclusion of a GC abundant hairpin element within intrinsic transcriptional attenuators, the nucleation of the P4 helix was assumed to be primarily dependent on increased helical stability[22, 25, 26]. However, initial observations of successful switchers within the screened pool of variants displayed several deviations from the usual NH5 base pair arrangement. Although a considerable majority of the sequences selected demonstrated both a repression value above and fold induction below the NH5 switch, 8 colonies were able to out perform the parental variant in both measurements (Figure 13). Out of this highly select group, all 8 switches contained either a weak or non-canonical couple within the first base pair position (furthest from the polyuridine loop). Furthermore, Base Pair 2 (located in the center of the randomized region) exhibited 3 weak AU or GU combinations, while Base Pair 3 (proximal to L4) only displayed 2. Further scrutiny therefore became warented.

To determine if the OD corrected repression values without the subtraction of the background measurement were significantly different ($P < 0.05$) than that of pBR322, a t-test was preformed with the subjects of the "Top" bin (Figure 14). The original motivation of this analysis was to examine the two colonies (R10N'1B1 and R3N'2D4) that, over multiple rounds of screeing, produced negative OD corrected flourescent values in the absence of ligand and thus, a negative fold induction. As a result, over half of the 19 sequences within the upper division were deemed significant, including R10N'1B1 and the NH5 variant. While some colonies, despite enduring additional rounds of screening, retained a large standard error value, they failed to be identified as an outlier within the R-Studio assessment and remained in the pool of viable colonies.

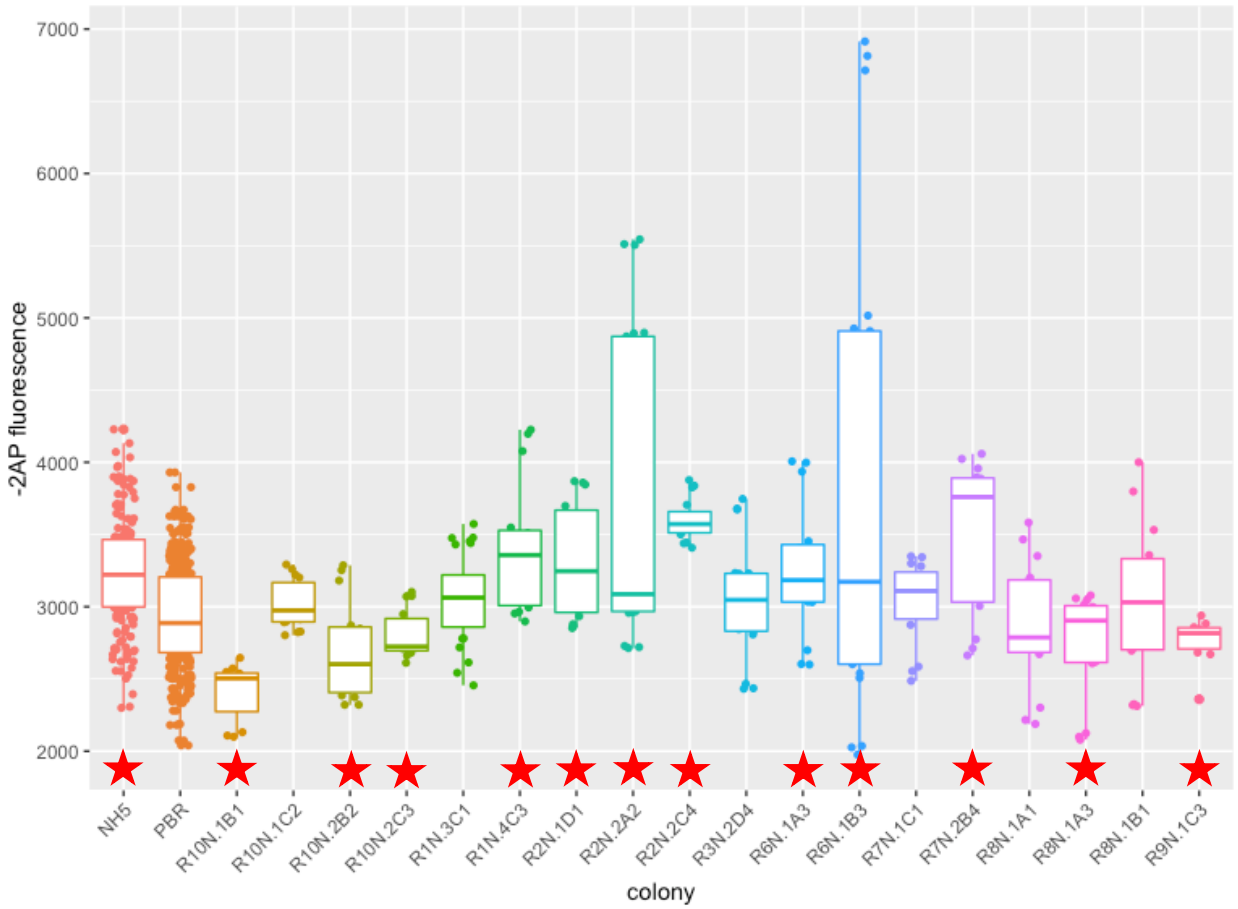


Figure 14. Graph representing the OD corrected repression values of the P4 Hairpin-Stem library "Top" bin (without the subtraction of pBR322). The various box and whisker plots each denote the specific sequences of the "Top" bin. The NH5 variant and pBR322 negative control are also included, and a t-test was executed for each colony (including NH5) against the pBR322 values. Switches denoted by a red star are those that have a repression that is considered significantly different (P-value < 0.05) from the control. A total of 13 colonies in the "Top" bin exhibited OD corrected repression measurements below that of NH5, and two colonies (R10N'1B1, R3N'2D4) continuously expressed negative fold inductions due to repression values below pBR322. To correct this limitation of the data analysis calculation, any switch that exhibited these negative measurements had its repression changed to a value of 50. However, the parameters seen in the figure above are created by the original OD corrected repression measurements. Adapted from the orginal graph made by Lisa Hansen.

Preferences for Base Pairing

The proposed mechanism for the creation of the terminator helix originally called for the assumed polyuridine loop "pause site" at the end of the P4 stem to allow for the aptamer domain to the rapidly inspect the concentration of ligand within the environment. However, the results of preceding papers proposed that the switch does not directly rely on the L4 region and physical stopping of the RNA polymerase due to a lack of sequence preference, and thus the ability to form base pairs within the stem was deemed more significant[25,26]. With a total of 127 unique sequences, the large pool of subjects derived from mutagenic cloning exhibited varied switching efficiency that allowed for a comprehensive analysis of the Hairpin-Stem structure. Each base within the randomized library was numbered 1 through 6 in the 5' to 3' direction, and the nucleotide identities of the respective positions were examined through three distinctive percent composition diagrams specified by bin type (Figure 15).

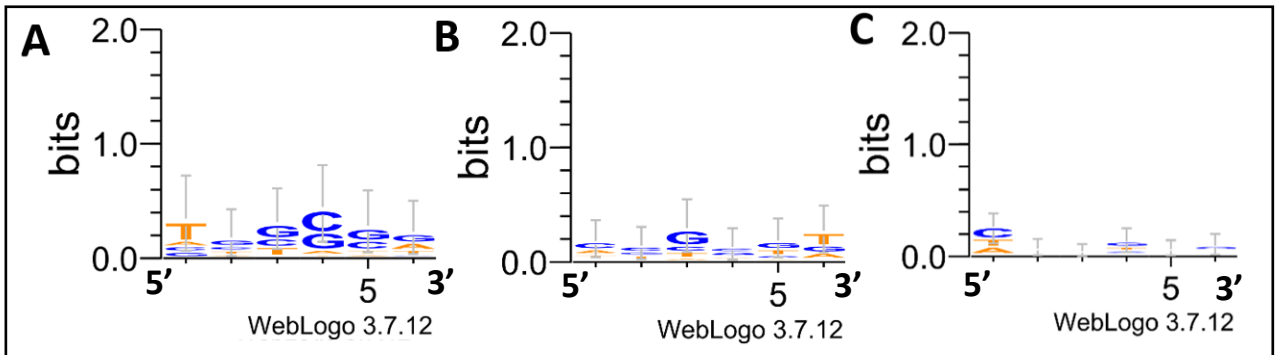


Figure 15. Diagram illustrating the percent composition of each nucleotide in the P4 Hairpin-Stem library, divided by each bin. Extending in the 5' to 3' direction. **(A)** The "Top" bin of the data set is inclusive of 18 unique sequences and the NH5 variant. The N1 position is highly preferential for thymine, or uracil in RNA. Furthermore, the N4 position is shown to equally favor either a guanine or cytosine nucleobase. **(B)** The "Middle" bin displays a reduction of preference for a specific nucleotide with the decline of performace of the riboswitch. **(C)** The plot exhibiting the "Bottom" division demonstrates a lack of sequence conservation and an overall absence of preference for specific nucleobases as efficiency decreases. Created using the WebLogo 3 server (http://weblogo.threeplusone.com/create.cgi).

Within the “Top” class, a preference for either guanine or cytosine is present throughout positions 2 through 4 (or the nucleobases most proximal to the polyuridine loop). In position 6, however, there is an equal partiality for the purine derivatives guanine and adenine, while position 1 heavily favors the pyrimidine thymine (or uracil in the RNA sequence). It is also important to note that all position 1 nucleobases have a competing pairing interaction with a cytosine located in the P1 region .When considering the “Middle” division of riboswitches, preferences decrease in all nucleotide sites other than 3 and 6, where thymine then rises in popularity. Finally, in the “Bottom” bin, a lack of sequence conservation is evident, with positions 1, 4, and 6 changing allegiances, and 2, 3, and 5 dramatically reducing alltogether.

Furthermore, preferences for nucleobase identities become more apparent when comparing the base pair attributes themselves. Categorized by both the nucelotide couple and their specific arrangement, Figures 16 through 18 detail the structure of the hairpin-stem element as well as the various percent compositions at each pair site. Figure 16 focuses specicifically on Base Pair 1, located near the hexa-uridine tract and closest to the initiation site of strand invasion. As a strucutal component integral to the formation of the nucleator site, the stabilization of the the hairpin element was proposed to serve as propellant in the terminator mechanism. While support for this suggestion is evident within the Base Pair 2 and 3 locations, Base Pair 1 unequivocally holds a strong preference for unstable combinations of nucleotides despite originally composed of a G-C unit in the NH5 variant. With uracil favoring the N1 position, the UG wobble and UA canonical combinations equally dominate the “Top” bin of the data set, with non-canonical, GC, and AU arrangements fighting for a majority within the “Middle” and “Bottom” classifications.

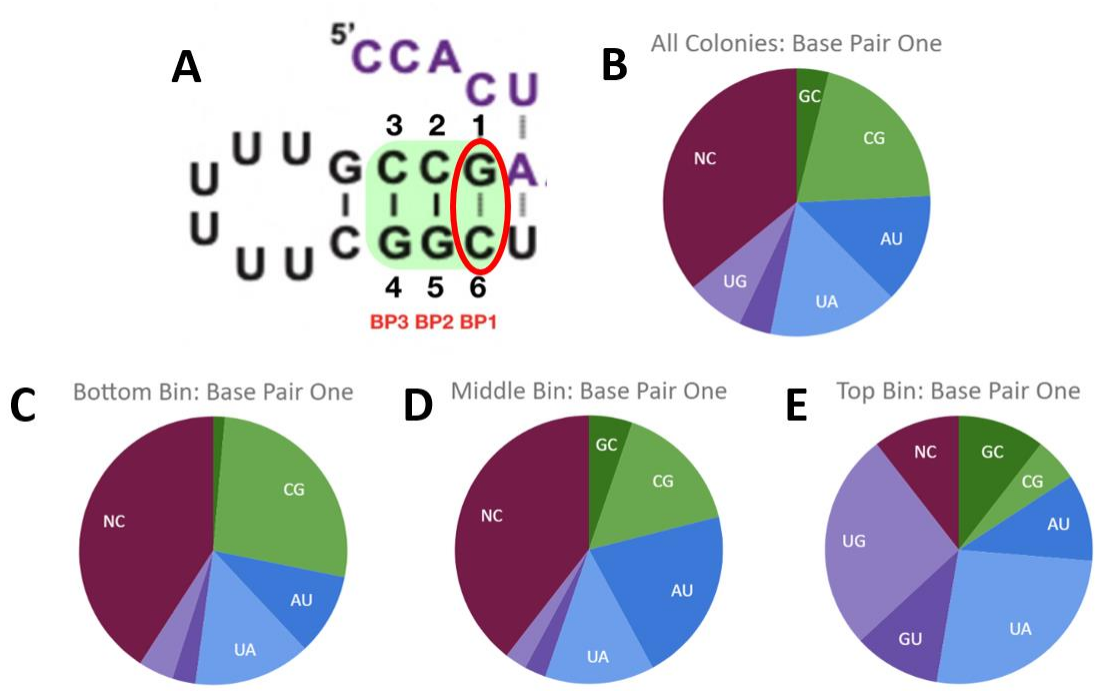


Figure 16. Structure of the NH5 riboswitch Hairpin-Stem library and pie charts displaying preference within the Base Pair 1 (BP1) position across bins. Within each pie chart, **"**NC" is the classification for non-canonical, representing any base pairs outside of GC, AU, or GU interactions. **(A)** Most distal from the polyuridine loop, Base Pair 1 encompasses the N1 and N6 randomized nucleotide positions. BP1 also neighbors the hexa-uridine tract and has competing interactions with cytosine in the P1 region. **(B)** Pie chart displaying Base Pair 1 compositions of every colony within the data set. **(C)** Pie chart displaying Base Pair 1 compositions of every colony within the "Bottom" bin data set. Non-canonical base pairs are highly preferred within inefficient riboswitches. **(D)** Pie chart displaying Base Pair 1 compositions of every colony within the "Middle" bin data set. Although an inclination towards a UA arrangement within BP1 increases, non-canonical base pairs remain the majority. **(E)** Pie chart displaying Base Pair 1 compositions of every colony within the "Top" bin data set. Between the "Middle" and "Top" bins, weak canonical and wobble base pair arrangements (UA/AU and UG/GU) become highly preferred, with a proclivity towards having a uracil nucleotide in the N1 position.

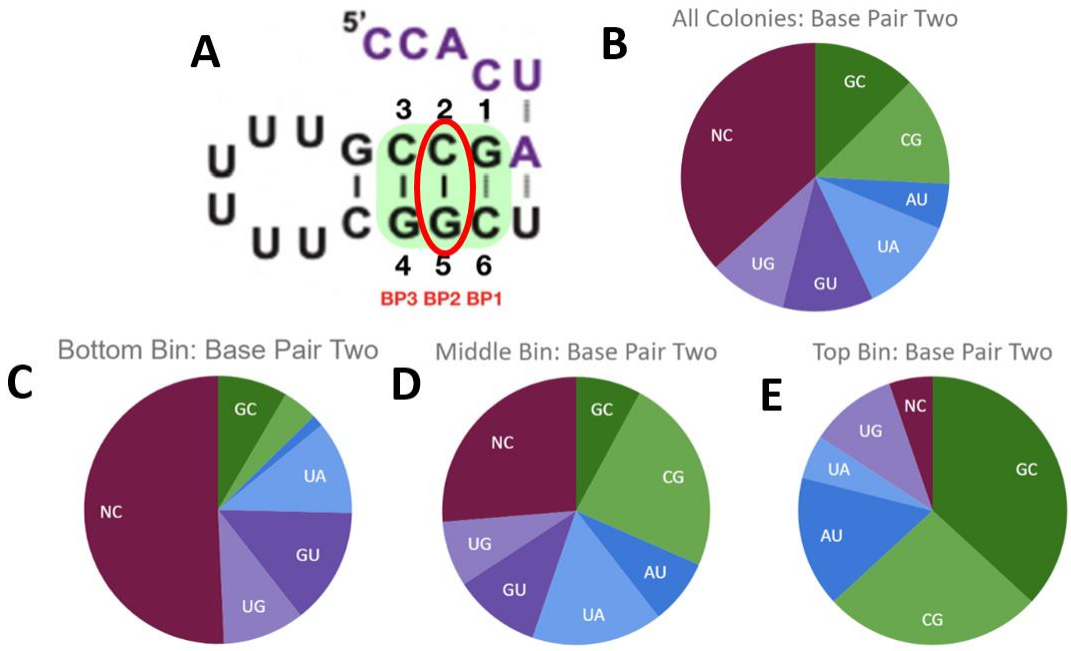


Figure 17. Structure of the NH5 riboswitch Hairpin-Stem library and pie charts displaying preference within the Base Pair 2 (BP2) position across bins. Within each pie chart, **"**NC" is the classification for non-canonical, representing any base pairs outside of GC, AU, or GU interactions. **(A)** As the central base pair in the randomized library, Base Pair 2 encompasses the N2 and N5 positions. BP2 does not exhibit competing interactions within the P1 region. (**B)** Pie chart displaying Base Pair 2 compositions of every colony within the data set. **(C)** Pie chart displaying Base Pair 2 compositions of every colony within the "Bottom" bin data set. A strong affinity for non-canonical base pairing is exhibited. **(D)** Pie chart displaying Base Pair 2 compositions of every colony within the "Middle" bin data set. Non-canonical base pairing begins to reduce, while CG base interactions increase. **(E)** Pie chart displaying Base Pair 2 compositions of every colony within the "Top" bin data set. Non-canonical base pairing, once again, decreases, and GC base interactions expand.

Base Pair 2, central within the randomized region, displayed a more expected result, with an almost equal disposition for GC and CG coupling in the upper division of colonies (Figure 17). In the lowest tier, non-canonical base pairing is evident within over 50% of the sequences, and with increased efficiency, strong preference for strong Watson-Crick couples grows dramatically.

Finally, Base Pair 3, in the closest position to the conserved GC stem and L4 region, once again show a decline in repression values and an increase in fold induction with stable canonical pairs. When linked with cytosine, a slight preference towards guanine in the N3 location also exists. However, this characteristic is not required (Figure 18). Overall, the percentage compositions of weak canonical and wobble couples within Base Pair 2 and 3 are relatively consistent across all bins, with only non-canonical and GC arrangements changing in dilution.

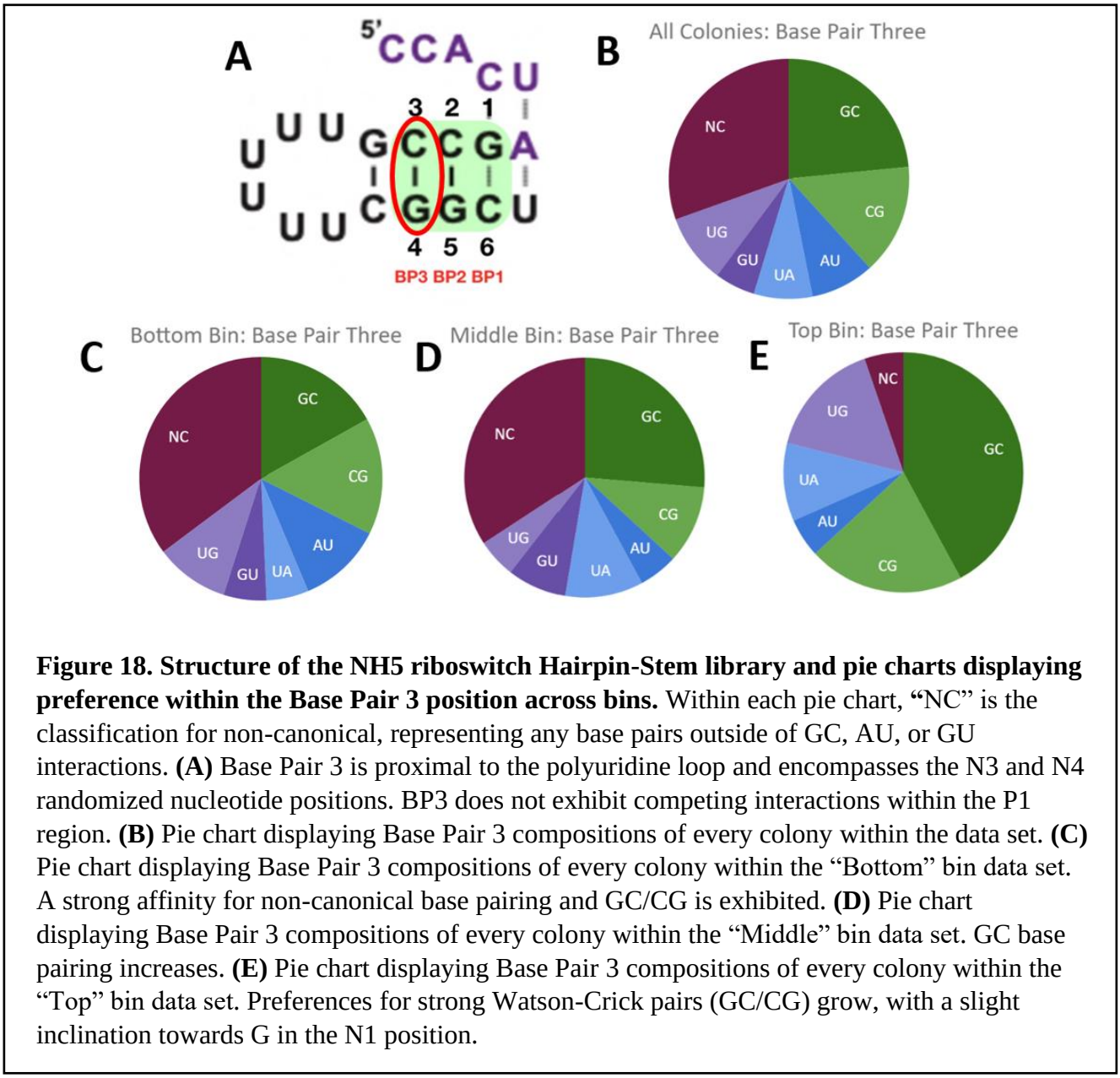


Figure 18. Structure of the NH5 riboswitch Hairpin-Stem library and pie charts displaying preference within the Base Pair 3 position across bins. Within each pie chart, **"**NC" is the classification for non-canonical, representing any base pairs outside of GC, AU, or GU interactions. **(A)** Base Pair 3 is proximal to the polyuridine loop and encompasses the N3 and N4 randomized nucleotide positions. BP3 does not exhibit competing interactions within the P1 region. **(B)** Pie chart displaying Base Pair 3 compositions of every colony within the data set. **(C)** Pie chart displaying Base Pair 3 compositions of every colony within the "Bottom" bin data set. A strong affinity for non-canonical base pairing and GC/CG is exhibited. **(D)** Pie chart displaying Base Pair 3 compositions of every colony within the "Middle" bin data set. GC base pairing increases. **(E)** Pie chart displaying Base Pair 3 compositions of every colony within the "Top" bin data set. Preferences for strong Watson-Crick pairs (GC/CG) grow, with a slight inclination towards G in the N1 position.

With the intent of further relating stability to the minimal expression values in the absence of the 2-Aminopurine ligand, the stacking energies throughout the randomized region were examined and contrasted to NH5 (Figure 19). Although base pairing is often considered the dominant force in the nucleic acid structure, coaxial base stacking interactions among adjacent nucleotide pairs can either restrict or allow folding mechanisms[29]. Obtained by utilizing the UNAFold Web Server's Two State Melting Hybridization calculator (http://www.unafold.org/Dinamelt/applications/two-state-melting-hybridization.php), the changes in Gibbs free energy between each nucleotide unit of all 127 sequences were found.

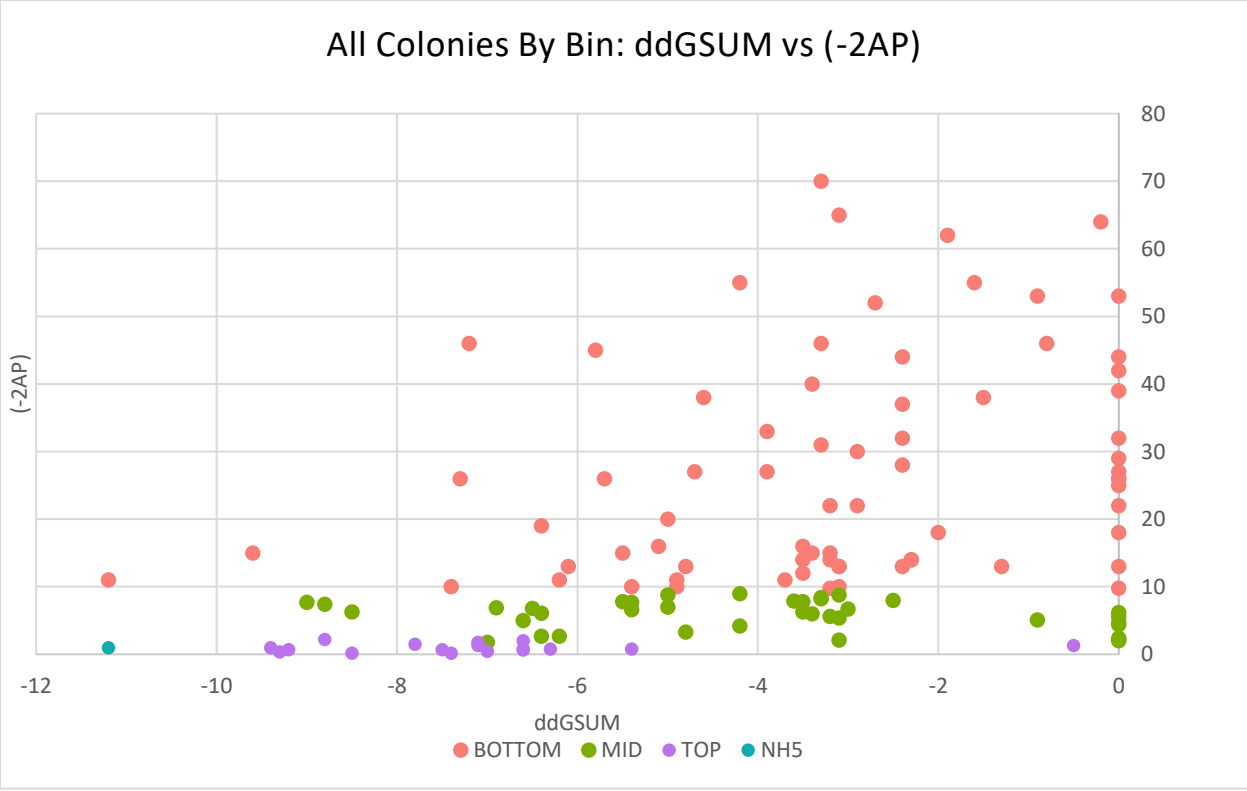


Figure 19. Graph investigating base pair stacking energies by the comparison of ΔΔG summation values vs repression in the absence of 2AP. The stacking energies between three base pairs of the randomized region and two additional conserved neighboring couples were totaled and assessed against the minimal expression values in a negative culture. A preference for stability within the efficient switches in the "Top" bin was suggested as a result.

Summations of ΔΔG were inclusive of not only the 3 original randomized base pairs, but also of the “nearest neighbor”. Therefore, the G-C pair proximal and the A-U pair distal to the polyuridine loop were also considered in the determination of helical stability. Divided by each bin, the decrease in -2AP repression values and increase in the -ΔΔG measurement (indicating an increase in bond strength) shows a correlation to riboswitch efficiency. However, only one colony was able to surpass the -ΔΔG value of NH5, and because of a high repression value, this sequence was sorted into the “Bottom” bin. Similar outliers within the “Top” divison that display low readings of stability may allude to the presence of alternative interactions between nucleotides within the RNA sequence, including asymmetrical internal bulging. Overall, the toehold of P4, although tolerant of various combinations of base pairing, highly prefers stable stacking interactions throughout the stem.

IV. Discussion

Of the 127 unique sequences discovered throughout the examination of the Hairpin-Stem Library, three offer insights into favored organizational patterns (Figure 20). Two, taken from the "Top" bin of switchers, exhibit repression values lower and fold inductions higher than that of the NH5 variant. The third, however, holds one of the highest expression measurements in the absence of 2AP, and thus retains a fold induction of only 2.5. This assessment of both high performing and low performing colonies allows for a better understanding of the Hairpin-Stem region's flexibility and overall tolerance of certain mutations.

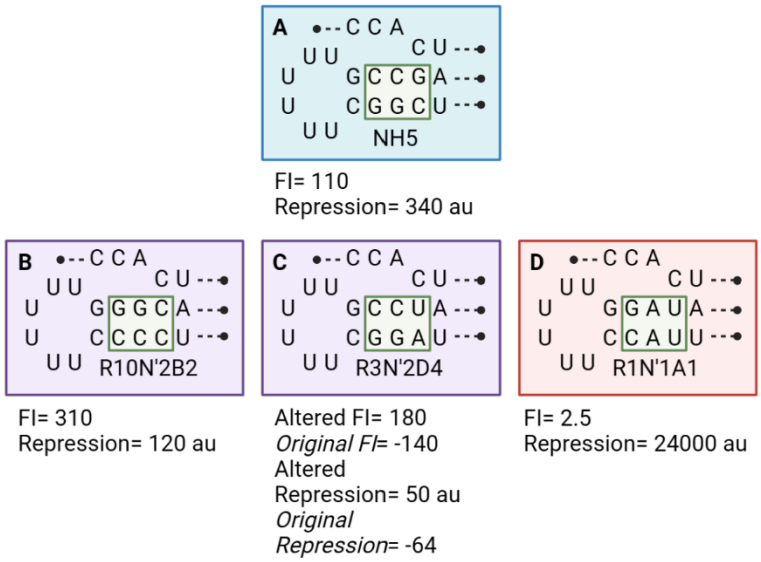


Figure 20. Selected functional switchers from the P4 Hairpin-Stem Library data set. The colors highlighting the entire structure of each selected sequence correspond to their particular bin, with NH5 highlighted in blue, the "Top" bin switchers (R10N'2B2 and R3N'2D4) in purple, and the "Bottom" bin colony (R1N'1A1) in red. **(A)** Original schematic of the NH5 variant, including elements of the pre-aptamer sequence and the P1 region. The fold induction and OD corrected repression values (without the subtraction of pBR322) averaged throughout the study are displayed below the diagram. **(B)** The secondary structure of the colony that exhibited the highest fold induction value, R10N'2B2. **(C)** Schematic for R3N'2D4, the sequence that exhibited the lowest original repression value without the subtraction of pBR322. Due to its original negative value, the repression measurement was altered to a standardized value of 50. **(D)** Structure for R1N'1A1, the sequence that exhibited the highest original repression value and one of the lowest fold inductions.

In considering sequences that promote repression without ligand binding, encourage the formation of the anti-terminator helix in the presence of the metabolite, and display an unusual base pairing preference, the colony labeled R10N'2B2 becomes a subject of interest. As the most efficient riboswitch pulled from the data set, the general allowance of a cytosine-cytosine mismatch in base pair position 1 alludes to an extreme acceptance of weak couplings in the distal region of P4. While able to be naturally present in differing RNA elements taking the structure of hairpin-stem, a CC mismatch commonly avoids the adoption of a perfect helix and forms an internal symmetrical loop[30]. However, the existence of a competing cytosine nucleotide in the P1 region that also forms pairing interactions with nucleotide 1 causes further obstacles. Nonetheless, the proclivity for strong GC coupling in Base Pair 2 and 3 is demonstrated.

As an example of a high performance riboswitch that continuously exhibited repression values below the pBR322 background, R3N'2D4 conserved the identity of nucleotides 2 through 5 from the original NH5 variant. Further supporting the notion that a GC rich base of the hairpin-stem promotes the nucleation of the P4 region, increased stability closer to the L4 loop allows for the successful formation of the terminator helix. Base Pair 1, however, is once more altered to the weak Watson-Crick base pair conformation of U-A, presenting the common preference of "Top" bin switches for a pyrimidine, specifically uracil, in the N1 position.

Finally, the poor behavior of the "bottom" bin colony labeled R1N'1A1 portrays the importance of retaining canonical pairs in the central coupled position. Although holding a GC identity in the randomized site closest to the L4 loop, the AA and UU mismatches present in the respective BP1 and BP2 locations completely destabilize the invasion of the 3' RNA strand into the P3 region. By retaining a fold induction above the value of 1, the riboswitch proves its ability to, nonetheless, form an anti-terminator helix in the presence of ligand. However, the high

repression measurement alludes to the sequence's failure to nucleate a stable P4 region to support the general process of strand invasion.

Concurrent with previous examinations of the *pbuE* adenine-responsive riboswitch variant NH5, the ability to form a competent terminator helix in an environment without a sufficient concentration of ligand is highly dependent on Watson-Crick nucleotide pairing preferences at the base of the Hairpin-Loop. Therefore, a higher measure of stability is exhibited, promoting a stable toehold region to support the complete nucleation of P4 and the subsequent invasive terminator mechanism. However, through a more intensive screening of the N=6 nucleator region, a proclivity in efficient switches for weaker nucleotide pairs in the position furthest from the L4 sequence is revealed. Often exhibiting repression values lower that of NH5, the motivations behind why this unstable addition to a kinetically driven riboswitch are unknown. Ultimately, the results of this study not only provide further evidence that the formation of a toehold is crucial for the act of strand invasion, but also gave potential methods of adjustment to further improve the activity of the NH5 riboswitch. Without a proper combination of nucleotide interactions within the nucleator component, the terminator helix cannot successfully be formed in the absence of the ligand, and genetic expression is poorly regulated.

IV. Conclusion

Subjects of Future Investigations into the P4 Hairpin-Stem Library

As an introduction into the development of a highly efficient novel riboswitch, the examination of NH5's P4 region allows for a better understanding of the mechanism by which an RNA sequence alone can successfully modulate genetic expression at the transcriptional level. While only a single part in the effort to advance future design principles of synthetic riboswitches, a more thorough comprehension of the Hairpin-Stem component would foster insights into the process of strand invasion. Despite this study investigating the various elements contributing to effective P4 nucleation (including stacking energies, base pair preferences, and strong relationships between repression value and fold induction), questions about possible interactions between the P4 region and other elements of the NH5 riboswitch remain. The strong proclivity for weaker nucleobase interactions at the Base Pair 1 position is still not widely understood, providing reason for a further review of the Hairpin-Stem section.

With the potential to improve the performance of NH5, forthcoming studies of the P4 library may focus on how competition between nucleotide 6 and the cytosine in the P1 region encourages specific base pair partiality. Through a similar mutagenic process of cloning and screening, this prospective experiment would once again center on the randomization of the 6 nucleotides in the P4 Hairpin-Stem distal to the polyuridine loop. However, the mutation of the P1 cytosine to the purine derivative adenine, as seen in the altered NH5 structure in figure 21, would reveal how additional components of the expression platform contribute to the mechanism of strand invasion.

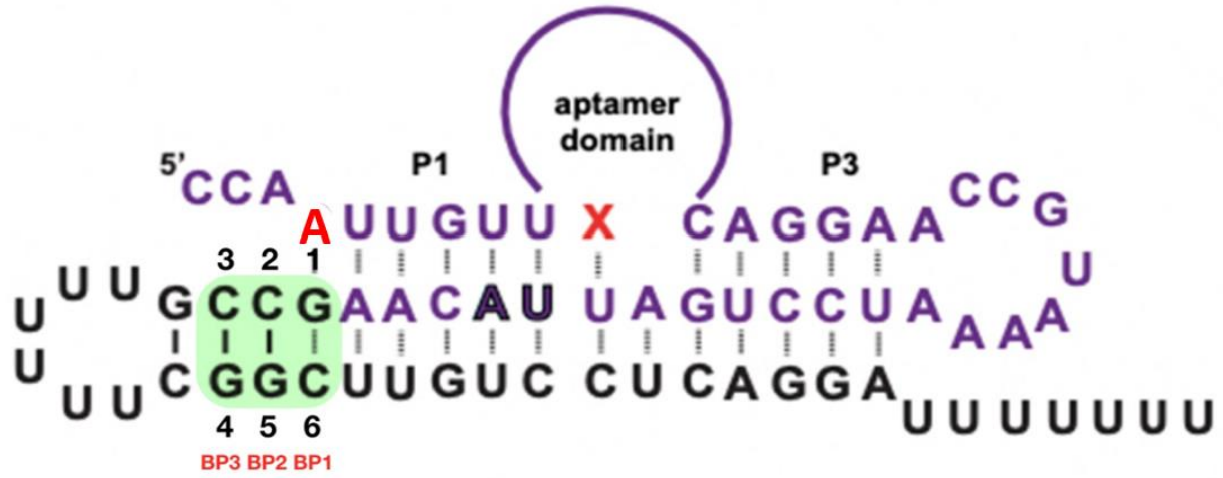


Figure 21. Structure of a possible NH5 P4 expression platform for the focus of a future study. To test whether or not unstable base pair preferences in the BP1 position is due to competing interactions with N1 in the P1 region, the cytosine nucleotide can be adjusted to an adenine. Altered from the original figure created by Robert T. Batey, PhD.

Applications of the Riboswitch

In the creation of a functional riboswitch that exhibits a proficient intrinsic transcriptional attenuator, the possibilities of both biological and synthetic applications expand. The ability to produce an artificial expression platform that holds the capacity to function both within cells and *in vitro* will allow for the enhancement of studies focusing on the aptamer domain. Used as a "parental" strain, the standardization of a high performance regulatory region may permit the development of aptamers that respond to specific metabolites, therefore creating novel biosensors. However, without extensive knowledge and control over the region that physically transforms between *ON* and *OFF* conformations, the ability to target specific molecules in order to cease or induce genetic expression would be futile.

A complete understanding of the strand invasion mechanism is therefore necessary in manipulating the expression platform element, and while this research has direct implications on the riboswitch, the modulation of other systems is also applicable. For example, antisense

oligonucleotides (ASOs) and the CRISPR-Cas9 genome editing technology both have flaws in terms of application that may be corrected with a better understanding of the process of displacement. When regarding CRISPR, the invasion of DNA is highly dependent on the perfect pairing of the target sequence. Strand invasion occurs to form the guide DNA duplex, and depending on the efficiency of this mechanism, off target effects may occur. Aptamers have also been exploited to improve CRISPR productivity by regulating the Single-guide RNA function, grafting control elements into the sgRNA sequence that are contingent on the binding of a ligand to create an *ON* secondary conformation[31].

The research of riboswitches, although a strong focus of recent biochemical study, remains a developing field. As an outstanding example of genetic regulation without the intervention of additional DNA or protein elements, the riboswitch is an exclusive structural component of mRNA that could also aid in the development of nucleic acid-based therapeutics. For instance, studies focusing on the deliverance of gene therapies by vectors developed from the adeno-associated virus (AAV) require a safe and efficient method of controlling expression during and after transport. Due to their small genomic footprints, flexible structures, and non-immunogenic nature, riboswitches appeared as the perfect candidate[31]. Although successful in modulating some biological processes in mammalian cells, like antibody expression, further progress in the isolation of a standardized expression platform must occur to permit clinical application.

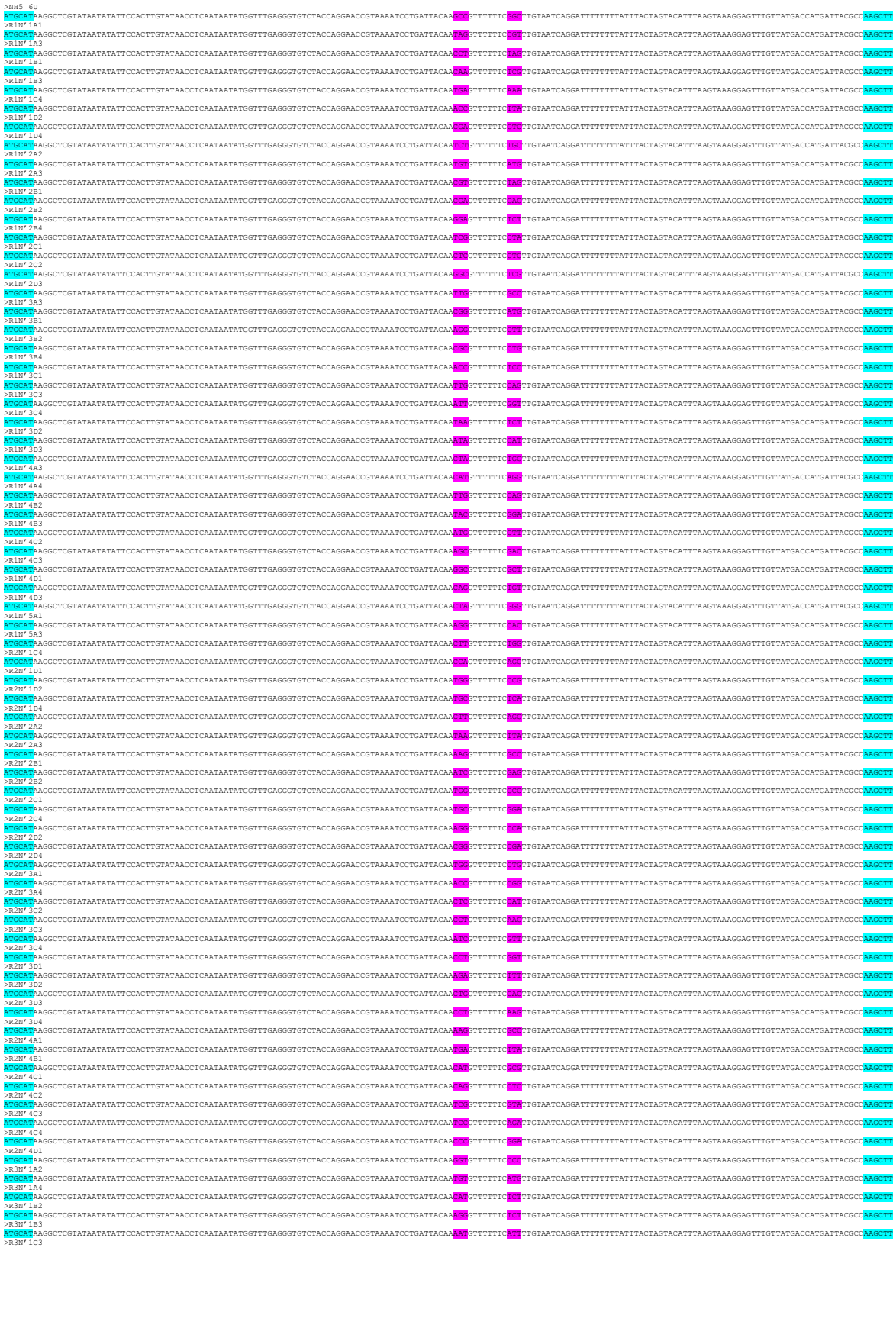
References

1. Su, Y., Liu, C., Fang, H. et al. Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microb Cell Fact* 19, 173 (2020). https://doi.org/10.1186/s12934-020-01436-8
2. Skarstad, K., & Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harbor perspectives in biology*, 5(4), a012922. https://doi.org/10.1101/cshperspect.a012922
3. CRICK F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, 138–163.
4. Johnson, G. E., Lalanne, J. B., Peters, M. L., & Li, G. W. (2020). Functionally uncoupled transcription-translation in Bacillus subtilis. *Nature*, *585*(7823), 124–128. https://doi.org/10.1038/s41 586-020-2638-5
5. Babitzke, P., & Yanofsky, C. (1993). Reconstitution of Bacillus subtilis trp attenuation in vitro with TRAP, the trp RNA-binding attenuation protein. *Proceedings of the National Academy of Sciences of the United States of America*, 90(1), 133–137. https://doi.org/10.1073/pnas.90.1.133
6. Henkin, T.M. and Yanofsky, C. (2002), Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays*, 24: 700-707. https://doi.org.colorado.idm.oclc.org/10.1002/bies.10125
7. Ciampi M. S. (2006). Rho-dependent terminators and transcription termination. *Microbiology* (Reading, England), 152(Pt 9), 2515–2528. https://doi.org/10.1099/mic.0.28982-0
8. Johnson, G. E., Lalanne, J. B., Peters, M. L., & Li, G. W. (2020). Functionally uncoupled transcription-translation in Bacillus subtilis. *Nature*, 585(7823), 124–128. https://doi.org/10.1038/s41586-020-2638-5
9. Farnham, P. J., & Platt, T. (1981). Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic acids research*, 9(3), 563–577. https://doi.org/10.1093/nar/9.3.563
10. McAdams, N. M., & Gollnick, P. (2015). Characterization of TRAP-Mediated Regulation of the B. subtilis trp Operon Using In Vitro Transcription and Transcriptional Reporter Fusions In Vivo. In M. Boudvillain (Ed.), RNA Remodeling Proteins: Methods and Protocols (pp. 333–347). Springer New York. https://doi.org/10.1007/978-1-4939-2214-7_20
11. Potter, K. D., Merlino, N. M., Jacobs, T., & Gollnick, P. (2011). TRAP binding to the Bacillus subtilis trp leader region RNA causes efficient transcription termination at a weak intrinsic terminator. *Nucleic acids research*, 39(6), 2092–2102. https://doi.org/10.1093/nar/gkq965
12. Breaker R. R. (2012). Riboswitches and the RNA world. Cold Spring Harbor perspectives in biology, 4(2), a003566. https://doi.org/10.1101/cshperspect.a003566
13. Wachter A. (2010). Riboswitch-mediated control of gene expression in eukaryotes. *RNA biology*, 7(1), 67–76. https://doi.org/10.4161/rna.7.1.10489

14. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., & Gelfand, M. S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression. *Trends in genetics*: TIG, 20(1), 44–50. https://doi.org/10.1016/j.tig.2003.11.008
15. Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harbor perspectives in biology*, 3(6), a003533. https://doi.org/10.1101/cshperspect.a003533
16. Wickiser, J. K., Winkler, W. C., Breaker, R. R., & Crothers, D. M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular cell*, 18(1), 49–60. https://doi.org/10.1016/j.molcel.2005.02.032
17. Hammann, C., & Westhof, E. (2007). Searching genomes for ribozymes and riboswitches. *Genome biology*, *8*(4), 210. https://doi.org/10.1186/gb-2007-8-4-210
18. 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. *Journal of molecular biology*, *359*(3), 754–768. https://doi.org/10.1016/j.jmb.2006.04.003
19. Edwards, A. L.; Batey, R. T. Riboswitches: A Common RNA Regulatory Element. Nat. Educ. 2010, 3 (9), 1:9.
20. Mandal, M., Breaker, R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**, 29–35 (2004). https://doi.org/10.1038/nsmb710
21. Batey R. T. (2012). Structure and mechanism of purine-binding riboswitches. *Quarterly reviews of biophysics*, *45*(3), 345–381. https://doi.org/10.1017/S0033583512000078
22. 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
23. Johansen, L. E., Nygaard, P., Lassen, C., Agersø, 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). *Journal of bacteriology*, *185*(17), 5200–5209. https://doi.org/10.1128/JB.185.17.5200-5209.2003
24. Marcano-Velázquez, 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. *The Journal of biological chemistry*, *290*(7), 4464–4475. https://doi.org/10.1074/jbc.M114.613497
25. Hansen, L. (2022) Identifying Design Principles for the Expression Platform of the B. subtilis Adenine-Responsive pbuE Riboswitch Through a Genetic Screen. https://scholar.colorado.edu/concern/undergraduate_honors_theses/6395w837t
26. Brown A. (2021). Genetic Study of the Bacillus subtilis pbuE Adenine-Responsive Riboswitch. https://scholar.colorado.edu/concern/parent/gb19f691f/file_sets/8623hz72x
27. Sambrook J, Russell DW et al. (2001*) Molecular cloning: a laboratory manual Third ed*. Cold Spring Harbor Laboratory Press.
28. Barker, K. *At the Bench: A Laboratory Navigator*, Updated ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2005.

29. Abraham Punnoose, J., Thomas, K. J., Chandrasekaran, A. R., Vilcapoma, J., Hayden, A., Kilpatrick, K., Vangaveti, S., Chen, A., Banco, T., & Halvorsen, K. (2023). High-throughput single-molecule quantification of individual base stacking energies in nucleic acids. *Nature communications*, *14*(1), 631. https://doi.org/10.1038/s41467-023-36373-8
30. Tavares, T. J., Beribisky, A. V., & Johnson, P. E. (2009). Structure of the cytosine-cytosine mismatch in the thymidylate synthase mRNA binding site and analysis of its interaction with the aminoglycoside paromomycin. *RNA (New York, N.Y.)*, *15*(5), 911–922. https://doi.org/10.1261/rna.1514909
31. Tickner, Z. J., & Farzan, M. (2021). Riboswitches for controlled expression of therapeutic transgenes delivered by adeno-associated viral vectors. *Pharmaceuticals*, *14*(06), 554.

Appendix: P4 Hairpin Library Stem FASTA File



ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCGGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'1C4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCCAATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'1D3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAGGGTTTTTCCCTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'1D4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAAGGTTTTTCCAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2A1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTGGTTTTTCCTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2A3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGGATTTTTTCGTATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2B3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACATGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2B4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTTGTTTTTCGAATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2C1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATAGTTTTTCCATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2C2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTCGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2C3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCTGTTTTTCGTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2D3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCTGTTTTTCGTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'2D4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTCGGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'3A1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGGGTTTTTCCCTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'3A2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAACAGTTTTTCGTTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'3A4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTGGTTTTTCCTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'3B1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'3B4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAACTGTTTTTCAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3N'3D3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGGTTTTTCGTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'1A2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTTGTTTTTCTAATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'1C2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTGGTTTTTCGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'1D2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAGGGTTTTTCCAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'1D4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAAGGTTTTTCCATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'2A4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGGGTTTTTCCTTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'2B2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAAGTTTTTCTTATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'2B4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCAGTTTTTCGTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4N'2D1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACAGGTTTTTCATGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5N'2C1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTTGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5N'2D2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATAGTTTTTCTACTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5N'2D4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCCTTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1A1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFGGGTTTTTCTTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1A2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFGGGTTTTTCCTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1A3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFCTGTTTTTCAGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1A4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCAAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1B3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFAGGTTTTTCCTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1C2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATGGTTTTTCCCTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1C4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFTGGTTTTTCTATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1D1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAGGGTTTTTCCTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'1D2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTGGTTTTTCTGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'2C2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFGATTTTTTCGCATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6N'2C3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATTGTTTTTCAATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1A1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGGGTTTTTCGTATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1A2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTGGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1B1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAACGGTTTTTCGTTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1B3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAAGGTTTTTCCTATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1C1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFTGGTTTTTCGGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1C2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFGATTTTTTCGACTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1C3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCTGTTTTTCAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'1D2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATGGTTTTTCTTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2A2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGCGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2A4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFTAGTTTTTCTGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2B3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAAGGGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2B4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGGGGTTTTTCCGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2C4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCGAATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2D3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGGGTTTTTCGGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R7N'2D4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATGGTTTTTCGGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'1A1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTTGTTTTTCGAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'1A3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTGGTTTTTCCGATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'1B1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFGTGTTTTTCGCATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'1C4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFCAGTTTTTCTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'1D2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGGGTTTTTCTTGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'2A2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACFAGGTTTTTCFGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'2B2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACFAGGTTTTTCFGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8N'2B3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGGGGTTTTTCCGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R9N'1B1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTTGTTTTTCAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R9N'1C3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAFGGGTTTTTCCGGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R9N'2A1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAATTGGTTTTTCCATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R9N'2A4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAATTGTTTTTCAATTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10N'1B1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCGGTTTTTCCGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10N'1C2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACCATGTTTTTCATGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10N'2B1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACTTGTTTTTCTAGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10N'2B2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAACGGGTTTTTCCCGTTGTAATCAGGATTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT

>R10N'2C3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAACGGTTTTTTCGGTTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10N'2C4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAACGGTTTTTTCGTTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT