Mutagenic Analysis of the Expression Platform of a Model Purine Riboswitch

Parker Juels

Department of Biochemistry

University of Colorado at Boulder

Defended April 6th, 2023

Thesis Advisor:

Dr. Robert T. Batey, Department of Biochemistry

Committee Members:

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

Representative)

Dr. J. Harrison Carpenter, Department of Ecology and Evolutionary Biology

Table of Contents

Disclaimer

Data for this thesis was collected in collaboration with Jesus Alicea and Lisa Hansen. Collaborating allowed for faster collection of data for the work intensive process of screening. It also allowed data collection to move forward when one member of the project was unable to conduct experiments or collect data. Additionally, the Gateway Library was also analyzed and discussed by Lisa Hansen for her honors thesis titled "Identifying Design Principles for the Expression Platform of the B. subtilis Adenine-Responsive pbuE Riboswitch Through a Genetic Screen", using some of the same data.

1 Abstract

Riboswitches are RNA based transcription and translation controllers. What makes them unique is their ability to recognize and up or down regulate transcription or translation based upon the presence or absence of a small molecule (ligand)[1]. Riboswitches have two key domains, an aptamer domain and an expression platform. The aptamer domain is responsible for binding and recognizing the ligand. The expression platform changes confirmation upon ligand binding to adapt a helix that could either halt or allow transcription and translation to occur[2]. The *PbuE* adenine-responsive riboswitch isolated from *Bacillus subtilis* has served as a model system for understanding and studying riboswitch structure and function. The structure and function of the aptamer domain of this switch has been well established[3]. However, the expression is less well established. The general sequence and structure is known but its function and sequence elements need to be studied in greater detail. To determine the role of strand exchange and the importance of key sequence elements to strand exchange, mutagenic analysis of a key sequence element involved in strange exchange was conducted. The results of this analysis suggest the importance of base pairing, especially towards the end of the helix formed during strand exchange. Additionally, a sequence element connecting the aptamer domain and expression platform was studied. However, this screening did not provide usable conclusions and further screening and analysis is necessary.

2 Introduction

2.1 Genetic Regulation

Bacterial cells are subject to constant fluctuations in their environment. Additionally, they have to respond to the changing needs of the cells as it transitions through cellular cycles, proliferation, and eventually apoptosis[4]. Genetic products such as proteins and ribozymes are responsible for executing many of the processes necessary for cellular survival[5]. As such, genetic regulation is necessary for responding to changing environmental and internal conditions of bacterial cells.

The major genetic pathway that needs to be regulated is the progression from DNA, to RNA, and finally to proteins[6]. DNA is how genetic material is stored in cells and needs to be transcribed to an mRNA transcript. The mRNA acts as a messenger that brings information stored in the DNA to ribosomes where it can be translated to proteins that can then preform a function necessary to cellular survival.

DNA → mRNA → Protein
Transcription   Translation

Figure 1: Schematic of Central Dogma. Transcription based controls regulate the blue arrow and either increase or decrease the rate of mRNA synthesis. Translation based controls regulate the red arrow and up regulate or down regulate the rate by which mRNA is transcribed into protein products.

2.2 RNA as a Regulator

While mRNA's major role is serving as a messenger between stored genetic information and protein synthesis, certain RNAs also serve as important genetic regulators. Four important categories of regulatory RNA are small nuclear RNA, MicroRNA, small interfering RNA, ribosomal RNA, and, most importantly for this project, riboswitches[7].

Small nuclear RNA

Small nuclear RNA join with proteins to form small nuclear ribonucleoproteins. These ribonucleoprotiens are involved in splicing introns out of pre-mRNA[8]. When DNA is originally transcribed to RNA, it includes regions of non-coding RNA (known as introns) that need to be removed by splicing to leave only coding regions (exons). Additionally, removing alternative sections of the genome, known as alternative splicing, leads to different mRNA transcripts which in turn create different proteins. snRNA are important in recognizing section of the pre-mRNA that need to be removed[8]. This is an important regulatory role. Differing snRNA in a cell can lead to recognition of different sequences and thereby creation of different proteins. Cells can utilize this to respond to their environment and needs by creating certain snRNA that will create splicesomes that splice at certain sites depending on what protein or enzyme is needed[8].

Micro-RNA

Micro-RNAs are non-coding RNA that bind to certain sequences of DNA or mRNA to up regulate or halt transcription[9]. They typically arise from introns that are spliced from pre-mRNA. They regulate both transcription and translation. They inhibit translation of mRNA by binding to

the 3' UTR and blocking translation initiation[9]. Micro-RNA can also bind to promoter regions in

DNA to up regulate translation of down stream genes by recruiting other transcription factors.

Micro-RNA has also been linked to certain disease in animals[9].

Small Interfering RNA

   Small interfering RNA are another transcriptional control. Small interfering RNA bind to

mRNA transcripts[10]. They are designed to have complimentary sequences to genes that need to

be down regulated. Small interfering RNA interact with proteins to recruit splicesomes to RNA

of interest. The si-RNA, in coordination with certain proteins, splice the mRNA at specific

sequences[10]. Cut up RNA can then not be translated and gene product can not be made. Si-RNA

can be regulated upstream by altering the rate of si-RNA transcription and catabolism[10].

2.3 Introduction to Riboswitches

   The RNA based genetic control of interest for this project was the riboswitsch.

Riboswitches are RNA based transcriptional and translational controls mostly found in bacteria

and archaea . They are found in the 5'UTR of mRNA. What makes them unique is their ability to

turn off and on in response to the presence of a certain molecule, known as a ligand[1]. A variety of

riboswitches with unique ligands have been identified.

   Purine riboswitches are biologically significant switches that regulate genetic expression

through binding purine nucleotide bases[11]. These riboswitches regulate the bio-availability of

purine nucleotides by controlling the rate of expression of genes involved in purine synthesis or

purine transport[12]. Purine riboswitches represent ideal models for studying the kinetics,

mechanism, and regulatory role of riboswitches due to their relatively small size, simple readily

available ligand, simple architecture, and well behaved folding pattern[13]. *Bacillus subtilis*

bacteria have an adenine responsive riboswitch that regulates the *ydhL* gene[14]. Isolation,

manipulation, and examination of this riboswitch has provided researches a deeper understanding

of riboswitch function. This switch also serves as a template for the switches studied in this

project.

2.4 General Riboswitch Structure
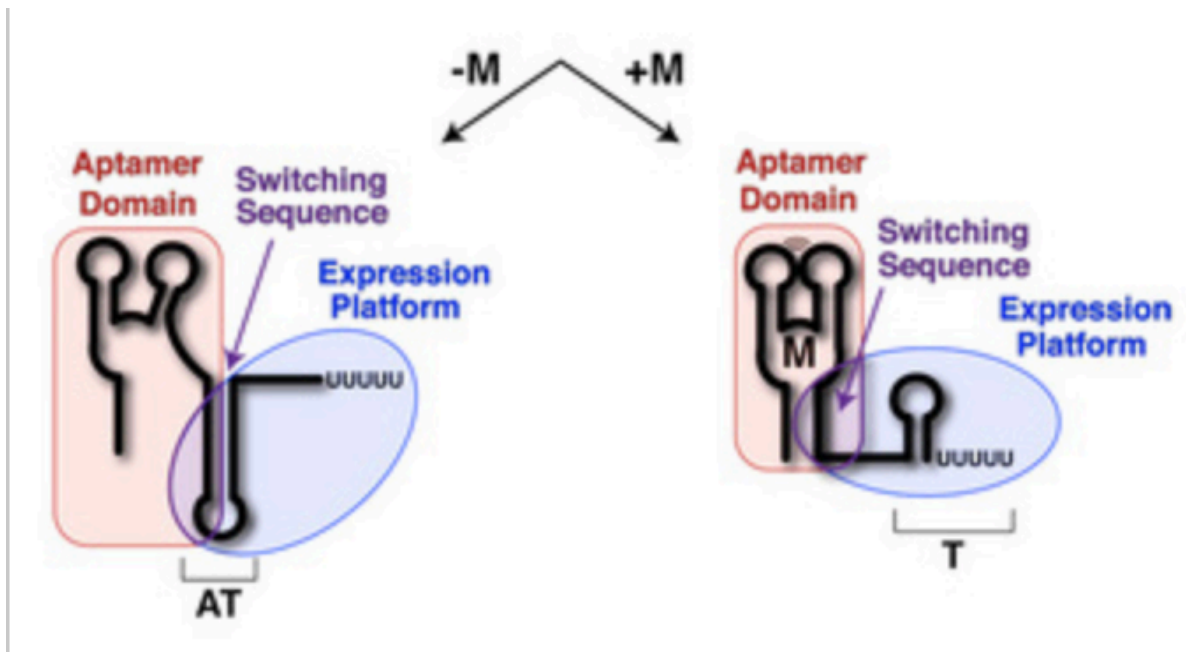
Figure 2: General structure of a riboswitch that responses to ligand "M". The left side demonstrates the riboswitch without ligand bound and the right side demonstrates the riboswitch with ligand bound. The aptamer domain is highlighted in red and the expression platform is highlighted in blue. Taken from Edwards and Batey (2010).

Riboswitches have two key domains that help it recognize ligand and change confirmation to initiate/halt transcription or translation. The first key domain is the aptamer domain. The aptamer domain consists of a loop with a unique sequence of base pairs that are ideal located to bind to a ligand of interest[11]. When bound to the ligand, a conformational change in the aptamer domain occurs which, in turn changes the expression domain and alters transcription or translation rate[16].

The interactions between the ligand and aptamer domain are optimized by a variety of specific base paring, interactions between helical loops of the riboswitch, and chemical interactions between the ligand and sequence[16]. The aptamer domain of many purine riboswitches have been well examined. A common motif within purine riboswitch aptamer domain is a triple helix structure where base pairs between the RNA strands stabilize the secondary structure[17,18]. This structure creates a binding pocket for ligand[17,18]. The ligand binding pocket gains specificity through specific base pairs in the pocket and the interaction this base pair has with ligand. A single pyrimidine base pair substitution in this pocket is responsible for the over 20,000 fold difference in recognition of guanine or adenine in specific riboswitches[12,19].
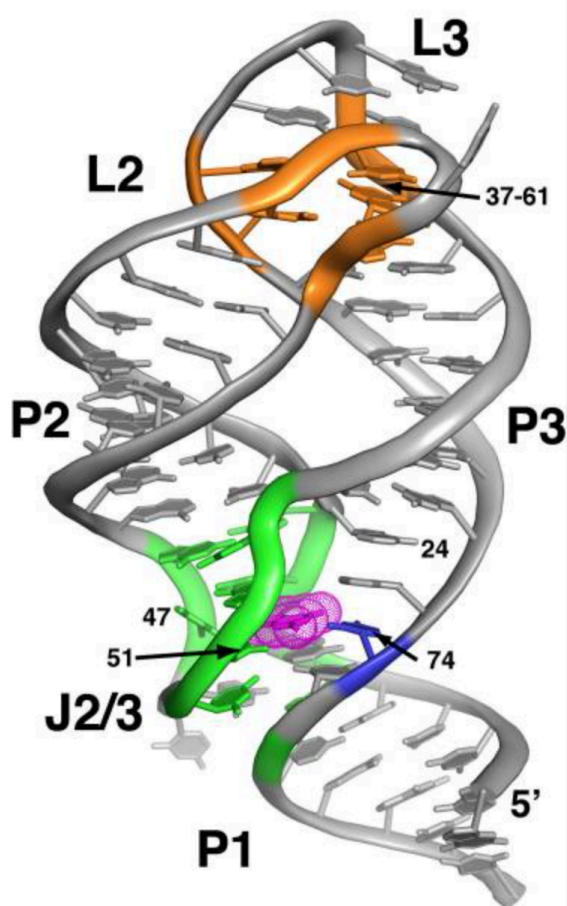
Figure 3: 3D structure of the aptamer domain in purine responsive riboswitches. mRNA is in gray, orange, blue, and green. The blue RNA base is responsible for distinguishing between guanine and adenine. The pink structure is the bound purine. Taken from Batey (2012).

Mutational analysis of the PbuE adenine-responsive riboswitch isolated from *Bacillus subtilis* has been thoroughly conducted and has provided information on how the aptamer domain forms and what sequence elements are vital for ligand recognition. This mutational analysis demonstrated that high base pair conservation is needed between L1 and L2, at the three helix junction, and in the P1 helix (Figure 4)[11]. This demonstrated that not only is riboswitch sequence relevant to ligand recognition, but there are specific sequence elements that require conservation to ensure the switch can fold into and switch between confirmations [11].
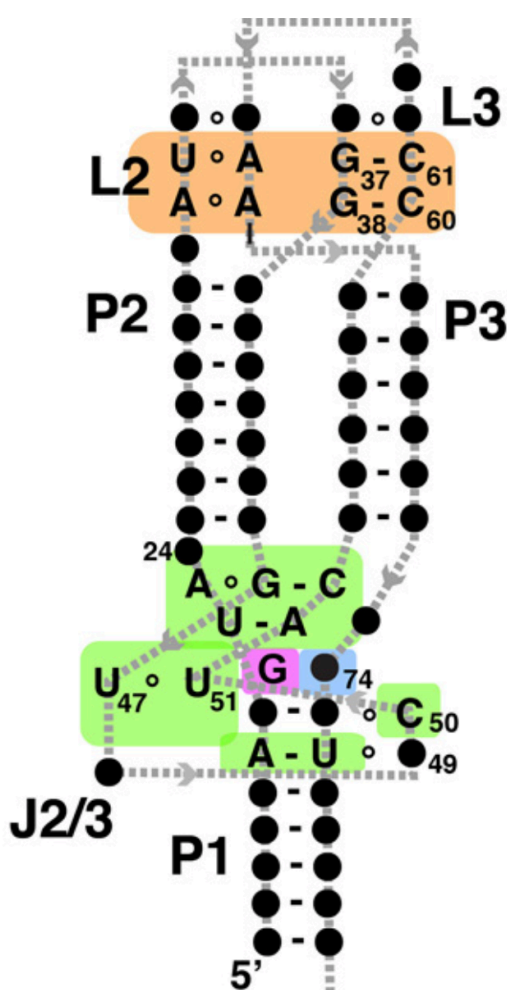
Figure 4: Two-dimensional structure of purine aptamer domain. Conserved regions, regions conserved in over 97% of active mutated switches, are colored and have nucleotide labeled. Watson-Crick base pairs are shown by dashes and non-cananocal base pairs are shown with open circles. Taken from Batey (2012).

The other major domain in a riboswitch is the expression platform which is typically located after the aptamer domain, further from the 5' end of the mRNA[1]. This section is typically in the general shape of a hairpin loop and is directly responsible for conducting the regulation. It typically has two confirmation an "OFF" confirmation that halts transcription or translation and an "ON" confirmation that allows translation or transcription to continue[20].

While the general structure of and mechanism of this domain are known (Figure 5), the exact role of sequence elements and the exact mechanism of switching confirmation within the expression platform is not well established[21]. Compared to the aptamer domain, where extensive research has been conducted on sequence elements[11], there has been little research on the importance and role of sequence elements in the expression platform[21]. As with the studies on the expression platform, PbuE adenine-responsive riboswitch isolated from *Bacillus subtilis* may provide an ideal system for studying motifs within the expression platform.

The three major areas of interest within the expression platform of the PbuE adenine-responsive riboswitch are the P1 helix, P5 helix, nucleator loop, and poly-uridine track (Figure 5)[22]. The P1 helix connects the aptamer domain to the expression platform. It disassociates when the riboswitch transitions to its "OFF "confirmation[22]. The nucleator loop (P4/L4 loop) is under the P1 helix and remains relatively unchanged between "OFF" and "ON" structure[22]. The P5 helix, also known as the terminator helix, is a large hairpin loop that forms only in the "OFF' confirmation and is responsible for causing transcription termination[22]. Finally, the poly-uridine track follows the rest of the expression platform and causes the RNA polymerase to pause. The

excess of uridine causes the proofreading mechanism of RNA polymerase to momentarily pause

transcription, providing time for the riboswitch to recognize ligand and adopt the correct

confirmation[23,24]. Also of key interest is the process by which the P1 helix is intruded upon

through creation of the P5 helix.
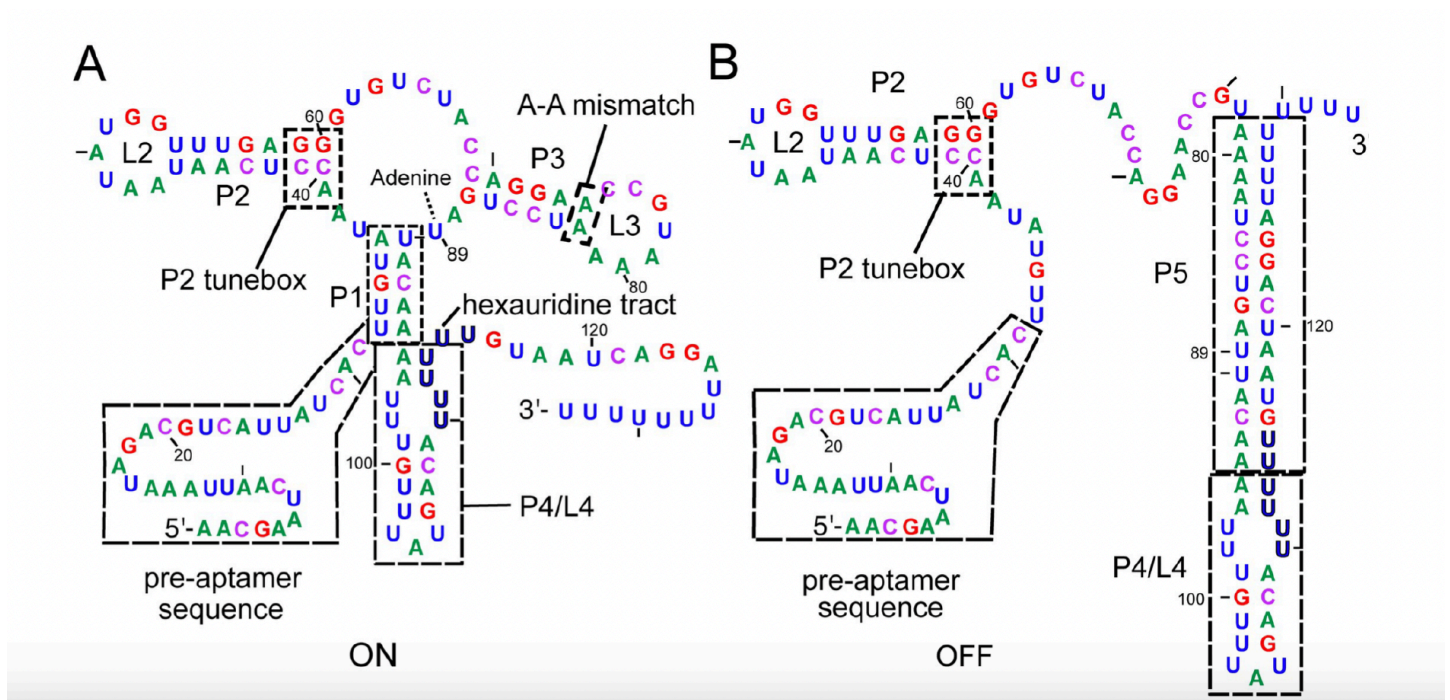
Figure 5: PbuE adenine-responsive riboswitch secondary structure. Figure A shows the

riboswitch with Adenine bound and B shows the structure of the riboswitch without adenine.

Labeled are the P1 helix, P5/terminator helix, P4/L4/nucleator loop, and hexauridine track of the

expression platform. Taken from Drogalis and Batey (2020).

2.5 General Mechanics of Riboswitches

Transcription controlling riboswitches act as rho-independent terminators of transcription[3]. As the pre-mRNA transcript is elongated the 5' UTR is the first section made[20]. When the riboswitch sequence is transcribed it immediately begins forming its secondary shape and binding a ligand[25]. There are two types of these riboswitches constitutively "ON" and constitutively "OFF" switches[3,26]. Constitutively "OFF" switches adopt the "OFF" confirmation immediately upon being transcribed and only switch to the "ON" confirmation upon ligand binding. Constitutively "ON" riboswitches are automatically in the "ON" confirmation until a ligand is bound and the switch changes confirmation into the "OFF" confirmation.

If a riboswitch is in its "ON" confirmation transcription continues as usual. If it is instead in the "OFF" confirmation it forces RNA polymerase off the DNA molecule and transcription is terminated. This transcription termination is a Rho independent termination of transcription. Traditionally, transcription in bacteria is terminated by a protein called rho recognizing a rho utilization site in the newly transcribed mRNA. This site is usually C rich and once the protein is bound it travels in the 5' to 3' direction and once it reaches the RNA polymerase it forces it to stall and fall off[27]. Transcription termination that is independent of proteins is, therefore, often referred to as Rho-independent termination.

This termination is achieved through the formation of a long hairpin loop followed by a poly-uridine tail[22]. The purpose of the poly-uridine tail is to slow down the RNA polymerase. Repetitive uridine bases in DNA causes RNA polymerase to stutter and slow down, giving time

for mRNA transcribed before to fold into structures and recognize ligand[23,24]. This slowed pace

allows the already transcribed riboswitch sense the environment to bind ligand or remain without

ligand and adopt the proper structure to allow transcription to halt or proceed[3].
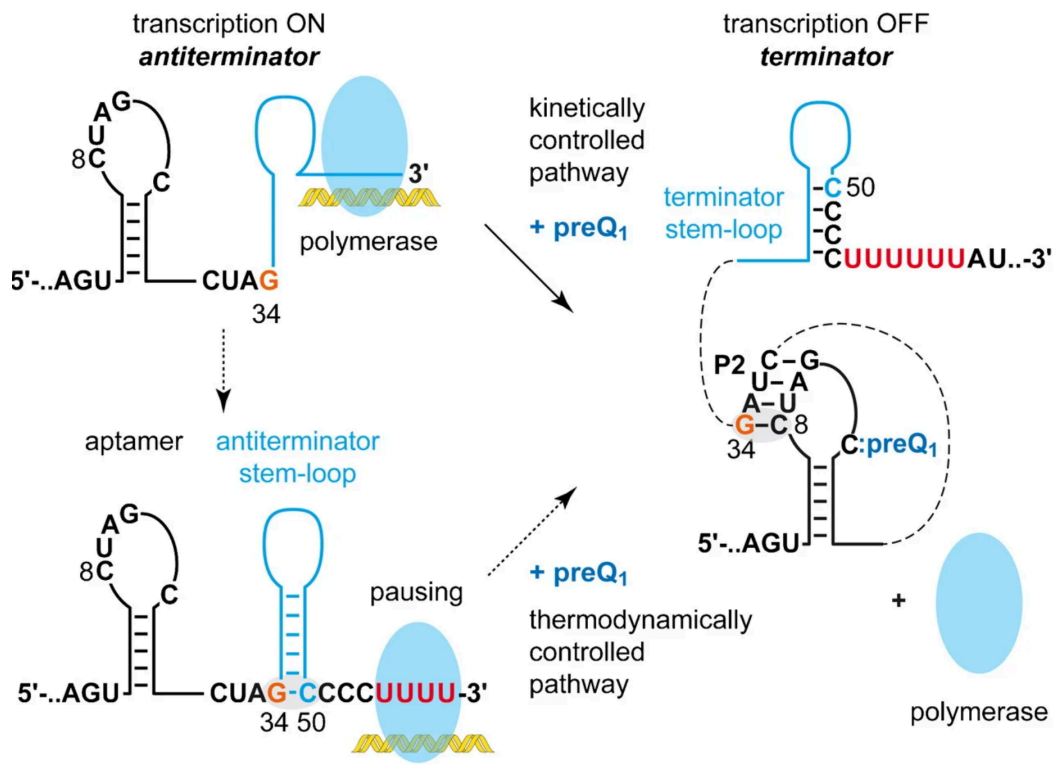
Figure 6: Schematic of transcriptional control by a preQ1 sensing riboswitch. This figure

demonstrates pausing upon reaching poly-uridine track, binding of ligand, and formation of

terminator loop. The formation of termination loop causes RNA polymerase and DNA template

disassociation. Taken from Reider, Kreutz, and Micura (2010).

Similarly, translational control riboswitches have the "OFF" and "ON" confirmations. When in the "ON" confirmation the switch is folded in a way that allows the ribosome to bind normally and transcription to proceed. However, when the "OFF" confirmation is adopted, the switch forms a long hairpin structure that blocks the initiation sequence or prevents ribosome binding therefore blocking transaction[29]. A common mechanism is the formation of a hairpin loop that involves the Shine-Dalgarno sequence (Ribosome Binding Site) which blocks recognition of the mRNA by ribosome. This blocks ribosome binding and therefore prevents translation from occurring[29].
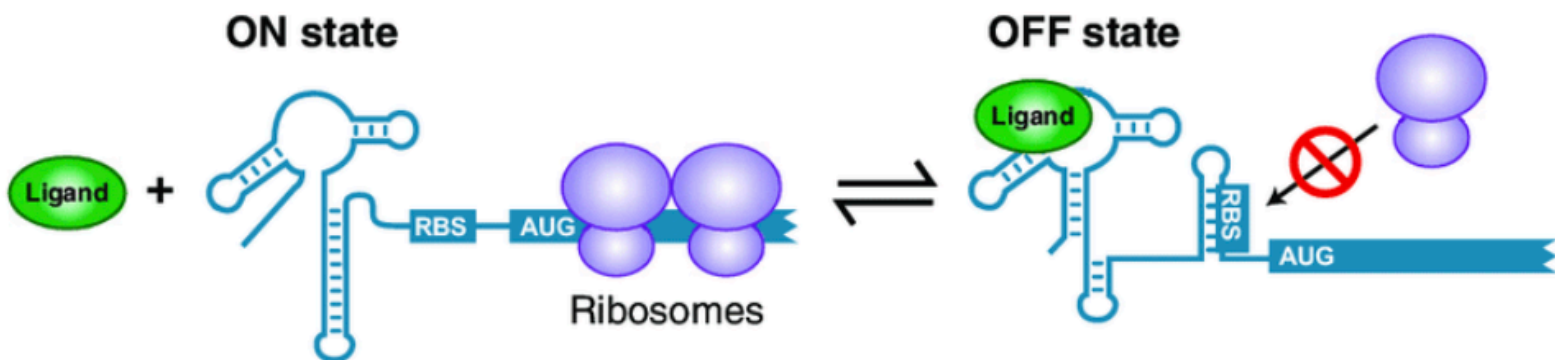
Figure 7: Diagram of translational control for constitutively "ON" riboswitch. The left side of the diagram demonstrates a riboswitch without ligand bound and the ribosome binding site is available for ribosome binding. The right side demonstrates the riboswitch with ligand bound, ribosome binding site unavailable because it is part of a hairpin loop, and ribosome unable to bind. Taken from Caron et al. (2012).

The key to both transcriptional and translational control with riboswitch is the ability to disrupt formed secondary structure in order to form an alternative structure[31]. The process by which this occurs is known as strand exchange. Strand exchange starts at a toe hold where the base pairs of a single stranded DNA region begin binding to complementary sequences on another strand[32]. This alternative strand, in the case of most riboswitches, is already bound to another complementary sequence. It is important that the base pairs at the start of strand exchange are strong in order to ensure the favorability of the process[32]. How strand exchange works within riboswtiches and what sequence elements are essential in ensuring proper exchange is an area of great interest for modern biochemists.
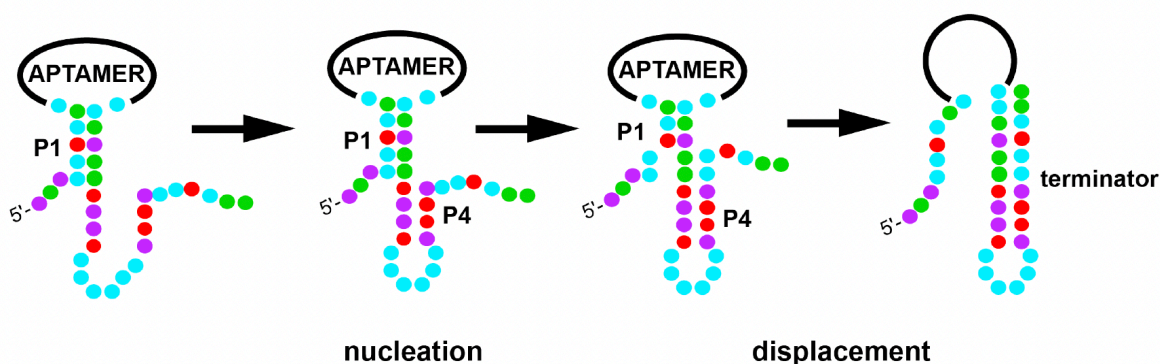
Figure 8: Diagram demonstrating strand exchange in a model purine riboswitch. Uridine nucleotides are shown in teal, adenine nucleotides are shown in green, guanine nucleotides are shown in red, and cystine nucleotides are shown in purple. The leftmost strand invades the P1 helix, displacing the rightmost strand and changing confirmations. Notice the perfect base pair matches between the invading strand and its complementary strand in the helix. Taken from Hansen (2022).

2.6 NH5

In order to understand the key elements and create an idealized artificial riboswitch,

Drogalis and Batey preformed mutagenic analysis of the PbuE adenine-responsive riboswitch

isolated from *Bacillus subtilis*[22]. The wild type PbuE adenine-responsive riboswitch is a

transcriptional regulator[14]. Without ligand bound, the switch folds into a terminator loop which

prevents transcription from continuing. When ligand is bound, the switch is able to disrupt the

terminator loop and create a secondary structure which allows for the continuation of

transcription[22].

Through this mutagenesis, Batey and Drogalis were able to develop NH5, an idealized

version of the wild type switch with an over 26x increase in fold induction[22]. This means the

engineered switch is significantly better at changing confirmation and exerting control then the

wild type. A few key differences between wild type and NH5 can be noted. This optimized

switch has a significantly shrunk nucleator helix, which now only consists of 4 G-C base pairs.

The pre-aptamer domain was also truncated in the optimized switch[22]. Most of these changes

occurred in the expression platform, the aptamer domain was largely preserved. The ideal ligand

for this switch is 2 amino purine (2AP), a derivative of adenine.
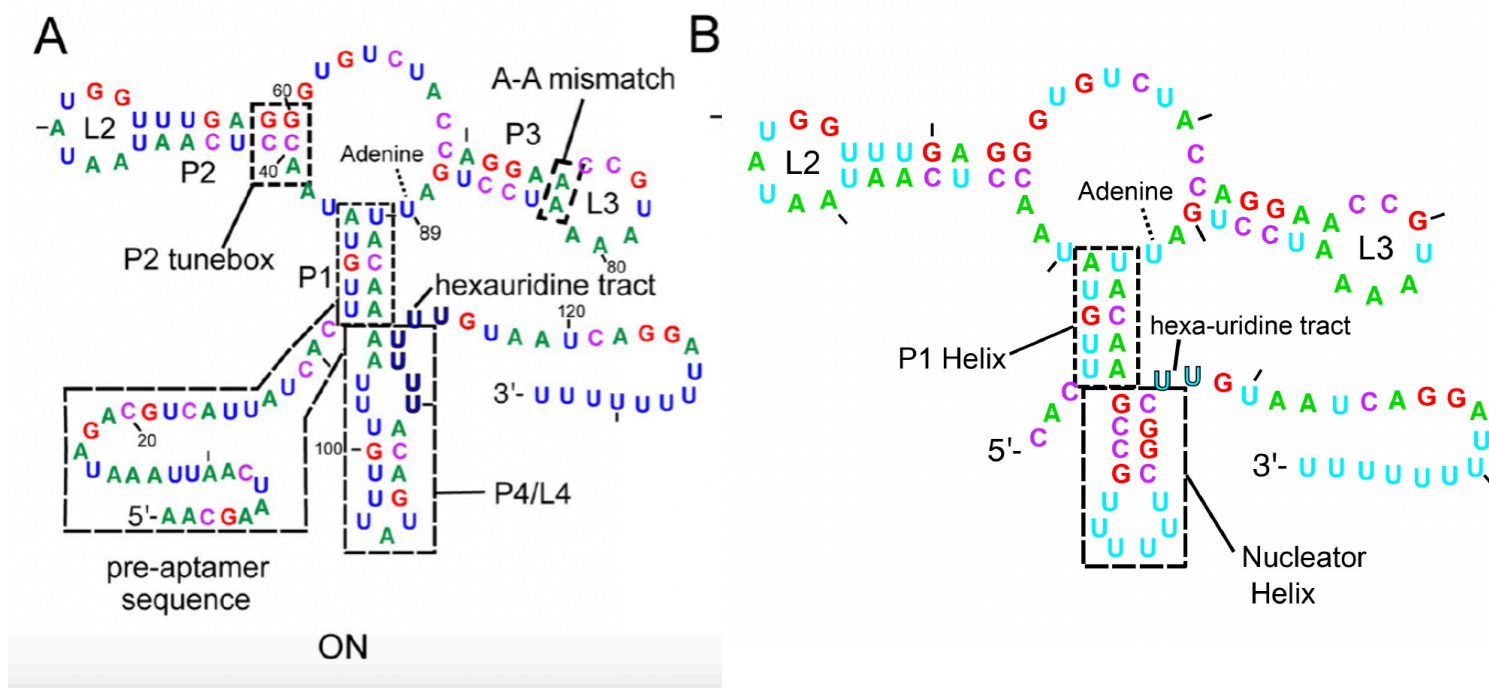
Figure 9: Side-by-Side structures of wild type (A) and optimized NHS (B) riboswitches. Of particular notes is the truncated pre-aptamer domain and the shrunken and optimized P1 helix. The nucleator helix is reduced to four G-C pairs and the uridine bump is removed. Adapted from Drogalis and Batey, 2020.

The mutational analysis utilized to create NH5 has served as a guide for further research into what makes a riboswitch function best and why NH5 is such a high performer. Previous research has demonstrated that the uridines in the loop at the base of the nucleator helix play no specific role and high preforming switches can have any nucleotide[33]. Additionally, mutagenesis on the G-C stem of the nucleator helix has preliminarily demonstrated that strong base pairs are vital in formation of a high fidelity switch[33].

2.7 Understanding Expression Platform of NH5

Of interest for continued understanding of NH5 and other riboswitches it can serve as a model for, is research into the 3' tail of the riboswtich that invades the nucleator helix to form a termination helix. This is the focus of this project. In order to examine the invasion two libraries were set up. For the purpose of this paper, libraries are collections of mutated NH5 riboswitches that are identical except for random mutations in 5-8 specific nucleotides. The first of these libraries is the gateway library, a 6 nucleotide library in the 3' tail that is involved in invading the nucleator loop to form a terminator loop. The second library is a 5 nucleotide library on the 5' end of NH5 that is involved in formation of the P1 and P3 helixes of the aptamer domain in the "ON" confirmation. By understanding this library we would gain a further understanding of how strand exchange leads to confirmation change within the aptamer domain in an ideal riboswitch.

A   B
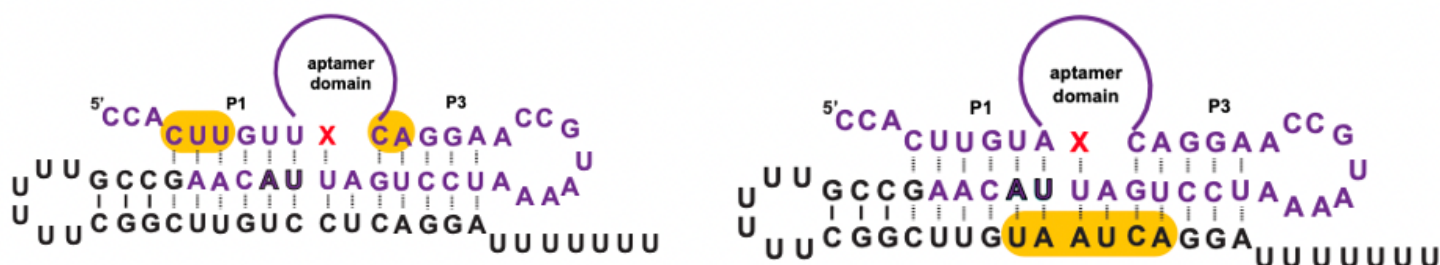
Figure 10: Schematic of libraries of interest. A represents the P1-P3 library with mutated nucleic acids highlighted in yellow. B represents the gateway library with mutated nucleic acids highlighted in yellow. Dashed and solid lines demonstrate base pairings that are made in one of the two confirmations.

3 Methods

In order to study the libraries of interest, mutational analysis must be conducted on the areas of interest. This mutational analysis consist of creation of riboswitches with targeted mutations followed my screening for riboswitches with preserved activity. A general schematic of the work flow of the mutagenesis and screening is demonstrated below.
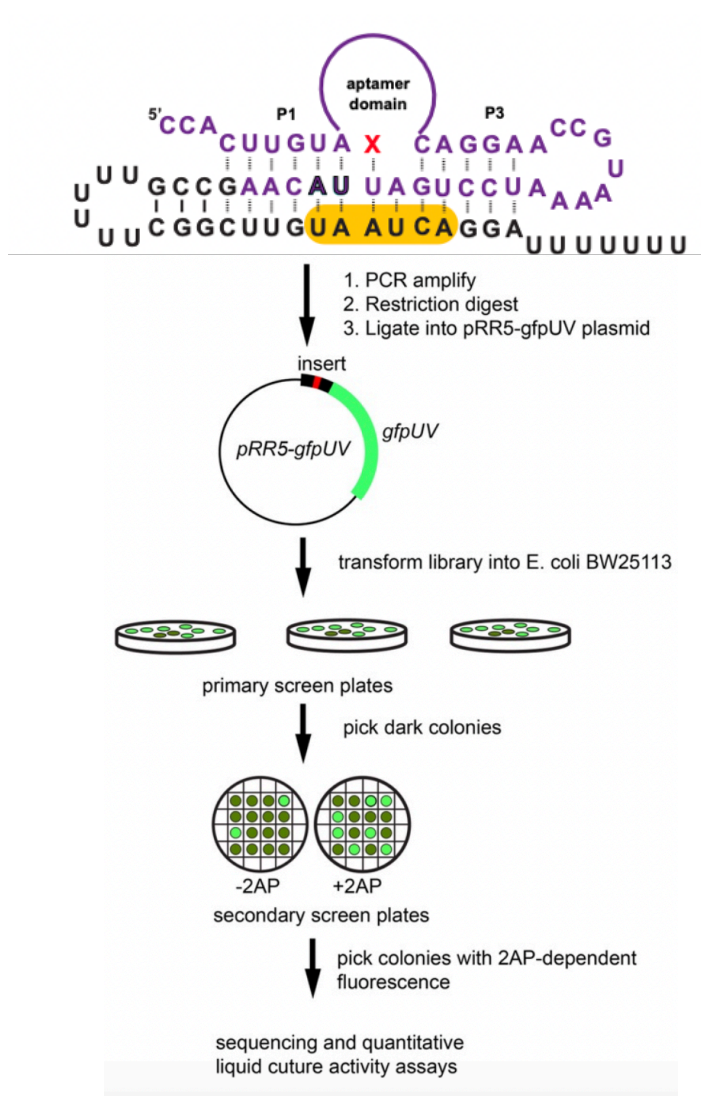
Figure 11: General figure of the workflow for mutagenic analysis and screening of libraries. This schematic demonstrates the process done for the Gateway library. Each step will be discussed in further detail in the following pages. Adapted from Hansen (2022).

3.1 Mutagenesis + Cloning + Ligation

PCR

To create the riboswtiches with targeted mutagenesis, a modified PCR reaction was

conducted. The inserts were made from two inner Ultramers from Integrated DNA Technologies,

4 primers (Shown in Table 1), and two outer gBlocksтм h (5'GENGblock NsiI and 3'GEN-

Gblock_HindIII) that were utilized to create a 288 nucleotide long oglionucleotide. The final

product was suspended at a concertaiotn of 100 μM in ddH2O. 1X Pfu buffer, 10 mM dNTPs, 10

nM of the inner oligonucleotides, 3'_SpeI_HindIII_pRRadaptor, and the library inserts, 1 μM of

each outer primer and a 1:25 ratio of Pfu polymerase were utilized to create the inserts[34].

Once the inserts were successfully made, PCR was utilized to amplify the

oligonucleotide. The PCR consisted of 12 rounds of a thermodynamic cycling between 95C,

60C, and 72C. The reaction was heated to 95C for 30 seconds to denature the double stranded

oligonucleotide. Next, the reaction was cooled to 72C for 45 seconds to allow the heat stabilized

polymerase to extends the newly synthesized strands. Finally, the reaction mixture is cooled back

to 60C for 30 seconds to allow the double stranded oligonucleotide to reform. At the end of the

12 rounds the mixture was left at 72C for an additional ten minutes to ensure double stranded

products are isolated.

After the PCR, the oligonucleotide products were subjected to tests to ensure success.

The reaction solution was run on 2% agarose gel. The products were compared to a DNA ladder

to ensure the oligonucleotides are the desired length. If the length is correct, the product was

isolated through the manufactures protocol for double stranded DNA less then 200 base pairs and

utilization of the Omega Biotek EZNA Cycle Pure Kit.

Name	Sequence (5' to 3')
Del27_NH5_lib_gateway	TTTACGGGCATGCATAAGGCTCGTATAATATATTC CANNNGTATAACCTCAATAATATGGTTTGAGGGT GTCTACCAGGAACCGTAAAATCCTGATTACAAGC CGTTTTTTCGGCTTGNNNNNNGGATTTTTTTTATT TACTAGTACATTTAAGTAAAGGAGTT
Del27_NH5_lib_P1P3	TTTACGGGCATGCATAAGGCTCGTATAATATATTC CACTTGTATAACCTCAATAATATGGTTTGAGGGTG TCTACNNGGAACCGTAAAATCCTGATTACAAGCC GTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTA CTAGTACATTTAAGTAAAGGAGTT
3'_SpeI_HindIII_pRRadaptor	GCATGCAAGCTTGGCGTAATCATGGTCATAACAA ACTCCTTTACTTAAATGTACTAGTA
5'GEN-Gblock_NsiI	TTTACGGGCATGCATAAGGCTCGTATA
3'GEN-Gblock_HindIII	AGGCATGCAAGCTTGGCGTAATCATGG
pRR_Forward	GCGCTAGCCACAGCTAACAC

Table 1: Sequences of relevant oligonucleotides for this project

Processing of Oligonucleotide and Ligation into Plasmid

The newly synthesized oligonucleotide must be treated for stability and ligated into a

plasmid so it can be expressed in *e. Coli* cells. The plasmid being utilized, pRR-*gfpUV*, contains

the ampicillin and carbenicillin resistance genes, sequences necessary for expression and

replication, and the gene for a fluorescence protein. The insert/riboswitch will be ligated in front

of the fluorescent protein genes so fluorescence can serve as a measure of riboswitch activity.

In order to ligate the plasmid and oligonucleotide together they were both digested with the same restriction enzymes. The enzymes used for this project was NsiI-HF and HindIII-HF from New England Biolabs. The insert was mixed in a 50 μL reaction with 1X CutSmart Buffer, and 20 units of both of the restriction digest enzymes. The reaction was let to run at 37C for an hour. The plasmid was digested in a 100 μL reaction with the same buffer, 40 units of the same enzymes, and 10 units of calf intestinal phosphatase to prevent re-annealing. Both products were treated the Omega BiotekEZNA Cycle Pure Kit and the insert was treated with ethanol. In order to maximize the ligation yield the insert was treated with T4 PNK. The insert was combined with 1 mM ATP, 1XT4 PNK buffer, and 10 units of T4 PNK for two hours at 65C.

Tranformation

Once the insert is successfully ligated into the plasmid the combined plasmid and insert must be transformed into cells. The cells used for this project were KEIO parental (E. coli K-12 BW25113). KEIO parental cells were chemically prepared using the rubidium chloride method. The cells were incubated with plasmid and 2XYT media for 30 minutes to an hour at 37C. After the incubation, they were transferred to CSB (defined media) plates with carbenicillin. For each library, it was tested through trail and error whether primary screening should be completed in the presence or absence of 2-AP. The Gateway library was plated on plates without 2-AP while the P1-P3 library was plated on plates with 2-AP. The plates were then left at 37C to grow overnight.

3.2 Screening

Primary Screen

After the plates are allowed to incubate overnight, a primary screen had to be conducted

to identify colonies with functioning riboswitch. Plates with 2-AP (P1P3 library) were examined

for bright colonies while plates without 2-AP were examined for dull colonies (Figure 12). Since

the NH5 riboswitch is a constitutively "OFF" riboswitch, functioning riboswitches would halt

transcription and expression of associated genes without ligand. However, with ligand,

transcription and expression of associated genes should be high. Through ligation, the riboswitch

became associated with *gfpUV*, the gene for a strong fluorescence protein. Therefore, when the

gene is expressed the colonies should glow under a UV wand. Using this visualization, colonies

were chosen from the primary plates and transitioned to grid plates (Figure 13).
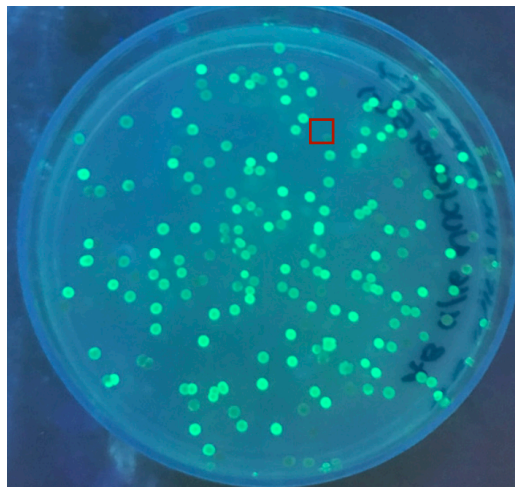
A **B**
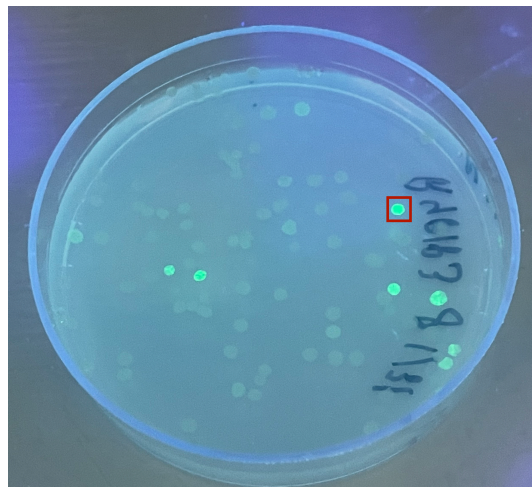
Figure 12: Example of primary plates with colony that would be picked for further examination surrounded by red square. Plate A represents plating on a negative 2AP plate where colonies with functioning riboswitches are dark. Plate B represents a positive 2AP plate where colonies with functioning riboswitches are bright. The Gateway library would be similar to Plate A and the P1P3 library would be similar to plate B.

Colony counting software was also utilized to keep track of the number of colonies observed and statistical likely-hood that a certain percent of the possible sequences within the library had been observed. Ideally, we would reach 100% coverage, but that is a statistical improbability. Instead, both the gateway and P1P3 library have reached 95%-99% coverage. The equation that was utilized to calculate this is shown below (Equation 1).

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

Equation 1: Equation to solve for the probability of observing every colony within a library given the number of colonies screened. *n* is the number of randomized nucleotides in the library (5 for both P1P3 and Gateway) and P is the probability of observing every possible colony in the library (34 Lisa).

Secondary Screen

Once winner are identified from the primary plates, they are subjected to a second round of screening to ensure they are actually functioning switches and not always stuck on or off. This process is conducted by creating two sets of CSB and carbenicillin plates with a grid drawn on. One set of the plates has 2AP included and is labeled with a "+" and one set does not have 2AP and is labelled "—" (Figure 13). The colonies are then transferred onto corresponding grid squares on a plus and minus plate. The plates are then placed in an incubator at 37C overnight. 14-16 hours later the plates are examined for colonies that are dull on the minus plates and bright on the positive plates under a UV lamps. These colonies have riboswitches that are correctly functioning and switching confirmation.
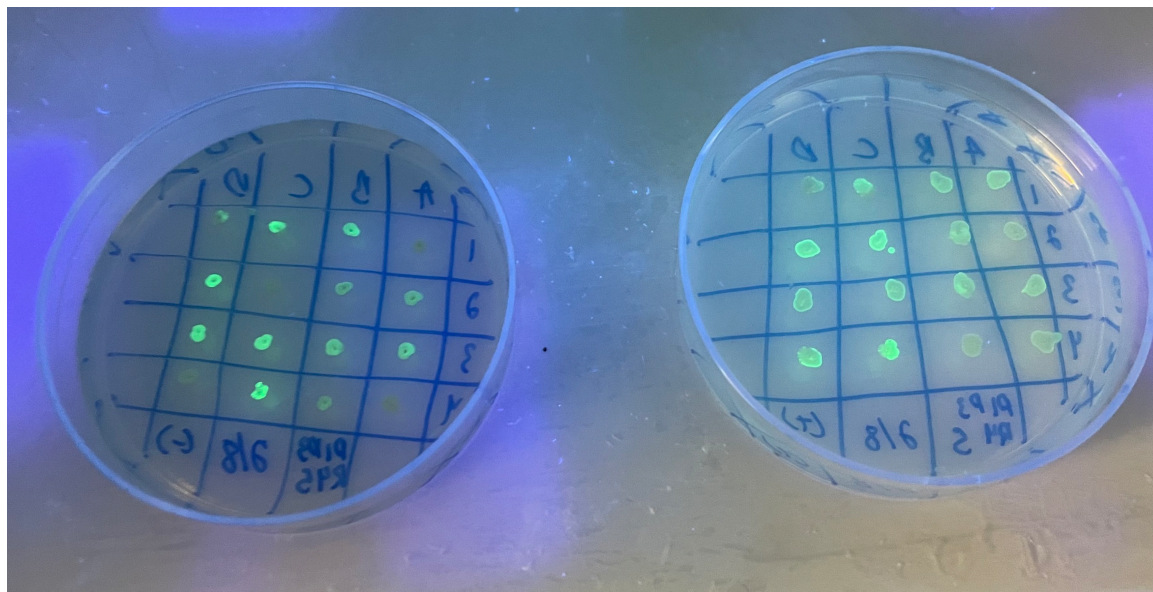
Figure 13: Grid plate example. The plate on the left does not have 2AP while the plate on the

right does. Switches that are dark on the left plate but bright on the right plate would be

considered colonies with functioning riboswitches. An examples of this would be colony 2C.

3.3 Sequencing

Colonies that are selected from the secondary screen must be further studied to

understand their gene sequence and activity.

First, the colonies are sequenced to isolate the exact nucleotides in their randomized

region. A sample of each colony is incubated at 37C overnight with 3mL of 2xYT media and 3μL

of ampicillin. 14-16 hours later, the plasmids within the colonies are isolated using a Omega

Biotek EZNA Plasmid DNA Mini Kit I. The plasmid is then suspended in 50μL of de-ionized

water. 10μL of the plasmid is then mixed with 5μL of PRR-F1 primers and sent to QuintaraBio for Sanger Sequencing. Sequences that come back with mutations outside of the area of interest or identical sequences to NH5 are not included in the final data set. All other sequences were placed in a FASTA file (appendix) and were subjected to an activity assay.
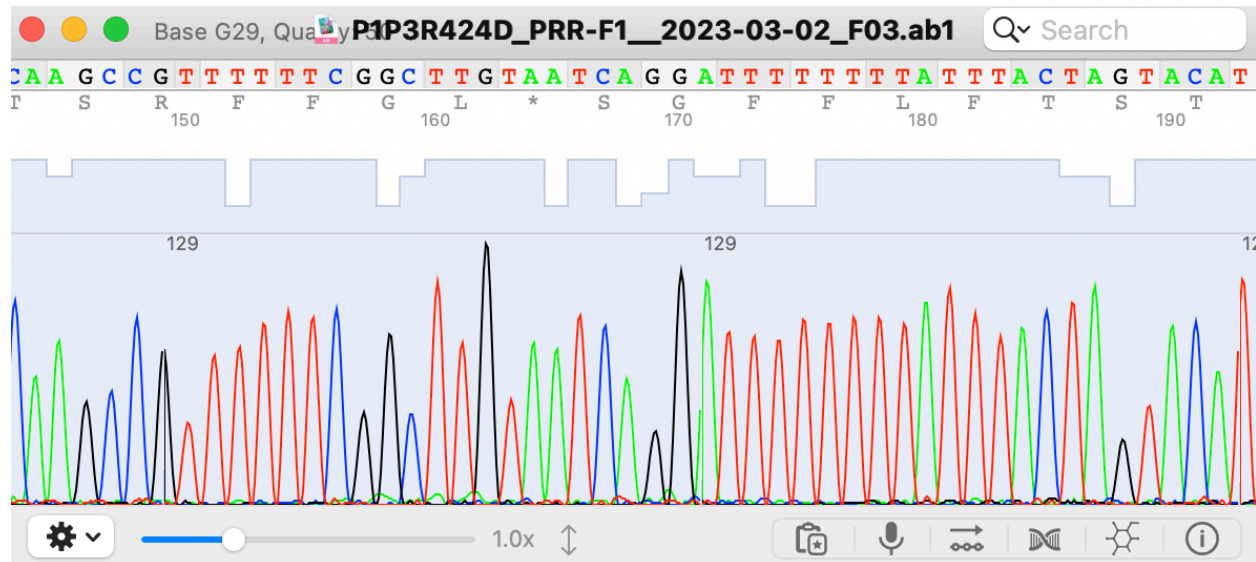
Figure 14: Example of data received from QuintaraBio. Each peak represents a nucleotide. The blue corresponds to a C, the black a G, the red a T, and the green an A. The shaded blue area represents the confidence interval.

3.4 Activity Assay

All of the screened and sequenced colonies are then subjected to an Activity Assay. A sample of each colony is incubated at 37C overnight with 3mL of 2xYT media and 3μL of ampicillin. 14-16 hours later, a sample of the colony is transferred into a tube labeled "+" and a

tube labeled "-". Each tube has 3μL of ampicillin and 3mL of CSB media. However, the positive tube also has 15μL of 100mM 2AP. The cells are then placed back into the 37C incubator.

Approximately 6 hours later, when the colonies reach a cell optical density of 0.4-0.6, they are removed from the incubator. The fluorescence of the contents of each cell tube was then measured. This was done by pipetting 200μL of each tube into a Corning Incorporated Costar 3603 96 well plate. In order to ensure consistency, each tube was pipetted into three wells each day for three consecutive days of measurement. Additionally, 200 μL of a control (5 μg/mL fluorescein) was added to a well to serve as a baseline for determine gain. The fluorescent measurement was conducted by the TECAN Infinite M200 Pro Plate reader at an excitation wavelength of 510 nm. The cell optical density of each well was also measured. This measurement was take at a wavelength of 600 nm.

In order to ensure accuracy of these measurements three controls were employed. First, NH5 was utilized as a positive control to test the fluorescence of a cell with a functioning riboswitch. Additionally, PBR-322 served as a negative control. PBR-322 is a plasmid without any fluorescence protein. By measuring the activity of cells without fluorescence protein, we gained a baseline for cell fluorescence that can be subtracted from the fluorescence of other colonies. Finally, gfpUV served as a positive control. These colonies had a plasmid with the gene for gfpUV, but no regulatory element. These colonies should then be constantly bright.

3.5 Calculations

In order to understand the data collected from the activity assay, three key calculations had to be conducted on the raw data. First the fluorescence must be corrected against the cell density to account for any differences in growth rate (Equation 2). Next, the fluorescence must be corrected for the background fluorescence of a normal cell (Equation 3). Finally, the fold induction must be calculated. This figure represents the degree to which fluorescence changes when ligand is present or absent (Equation 4). It therefore shows how effective the riboswitch is and the degree to witch it is able to suppress and encourage transcription.

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

Equation 2: Equation for OD corrected fluorescence

OD Corrected Fluorescence = (OD corrected fluorescence) − (median background fluorescence)

Equation 3: Equation for OD Corrected Fluorescence. Median background fluorescence comes from the median of the fluorescence values of PBR-322 from the day of data collection.

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

Equation 4: Equation for Fold Induction.

4 Results

4.1 Gateway Library

Screening of the gateway library provided clear data on the importance of specific sequence elements within the region of interest. The three major takeaways from this library were the importance of the conservation of base pair at nucleotide four and five (base pairs are counted from 5' to 3'), the importance of a purine at nucleotide six, and that no mutated sequence preforms as well as NH5.

Number of Colonies Observed on Primary Screen	22,799
Number of Colonies Picked for Secondary Screen	1136
Number of Colonies Picked from Secondary Screen	321
Number of Colonies that had Fold Induction Over Two	142
Fraction of Variants Observed	99.6%

Table 2: Numerical Summary of Colonies Observed Through Gateway Screen

The gateway library is located in the sequence that invades the termination loop to instead form the P4 helix in the activated riboswitch. Typically with strand exchange, strong base pairing is vital to ensure the process is favorable enough to displace the non-complementary strand of the formed helix[32]. Knowing this, the sequence similarity of each isolated sequence to NH5 was determined. Since only one side of the P4 helix is undergoing mutagenesis, the more similar the sequence is to NH5 the more base pairs it will share with the strand it is invading. It was found that there is a connection between the number of conserved nucleotides and

effectiveness of the riboswitch (Graph 1). However, there are clear outliers to this trend with one

of the top preforms only having three conserved nucleotides. Additionally, looking at the

moderate switchers, the pattern of conserved sequences creating better switchers becomes

muddled.
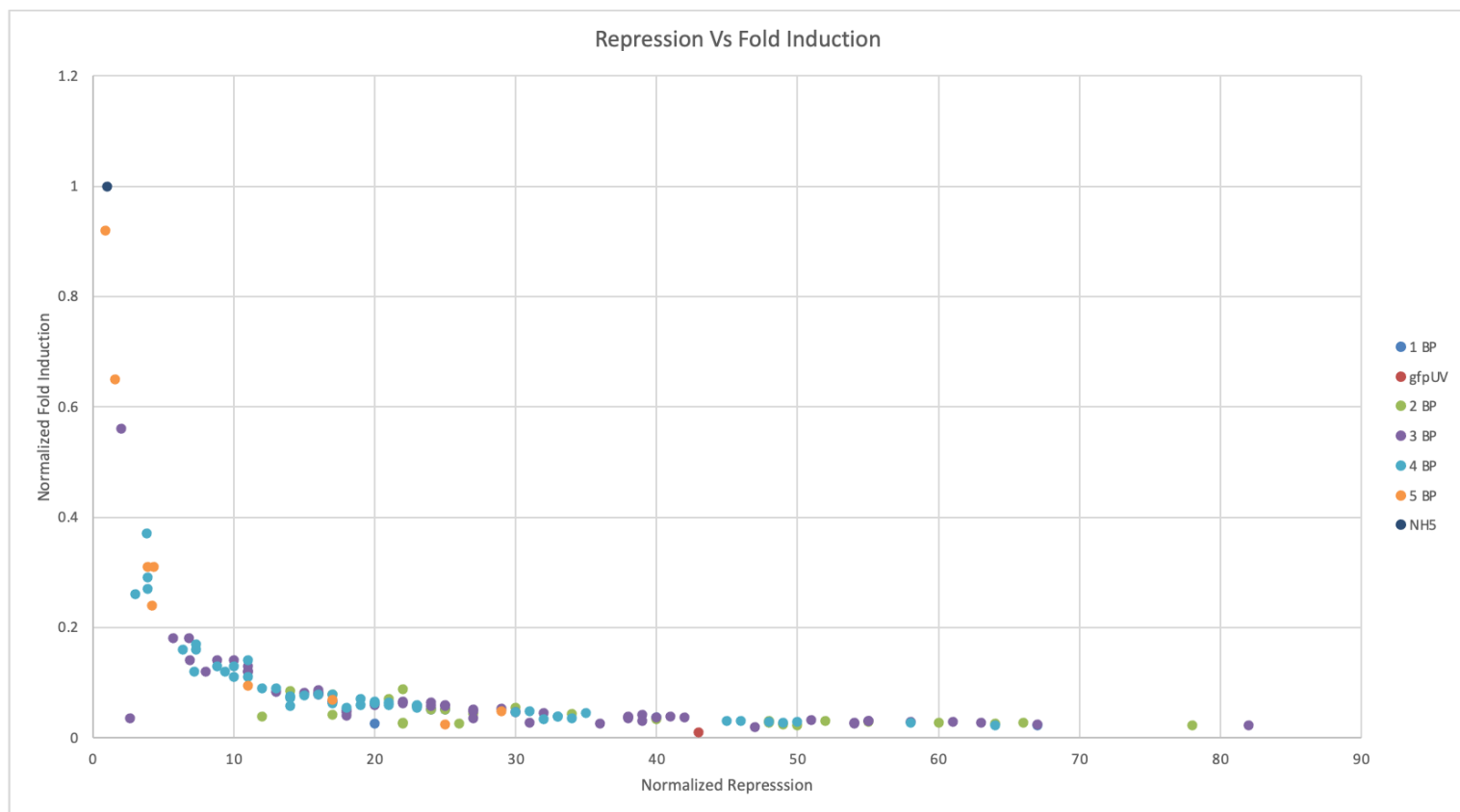
Graph 1: Graph of Fold Induction normalized to NH5, vs OD Corrected Fluorescence in the

absences of 2AP normalized to NH5. Each data point is color coded based upon the number of

nucleotides conserved/number of base pairs made with the P4 helix. There is a trend that the

more conserved nucleotides/base pairs the better the switch represses and induces, however there

are many notable outliers.

Base paring with the P4 helix clearly plays a role in the mechanism carried out by the sequence within the gateway library. However, there is more to it than base pairing across the region. This does not explain why a sequence with only three base pairs formed creates a successful sequence and why some sequences with five base pairs formed preform poorly.

Isolating each of the six randomized nucleotides and evaluating their conservation independently can provide greater insight. The top quarter of sequences, based upon fold induction, were studied. Specifically, what nucleotide was found at each base pair in each nucleotide for each sequence was recorded (Graph 2). While every nucleotide showed a preference for preserving NH5's sequence, there was less conservation at base pair 1-3, suggesting base pairing is less important at those locations. On the other hand, base pair 4 showed 89% of the high preforms conserved the U and base pair 5 showed 83% of the high performers conserving the C. Additionally, base pair 6 showed 97% of the high performers placing a purine (G or A) at the site (which was originally an A). As Guanine is able to form a very successful nonWatson-Crick base pair with U, this base also showed a necessity for preserving base paring (PMID: 11256617).
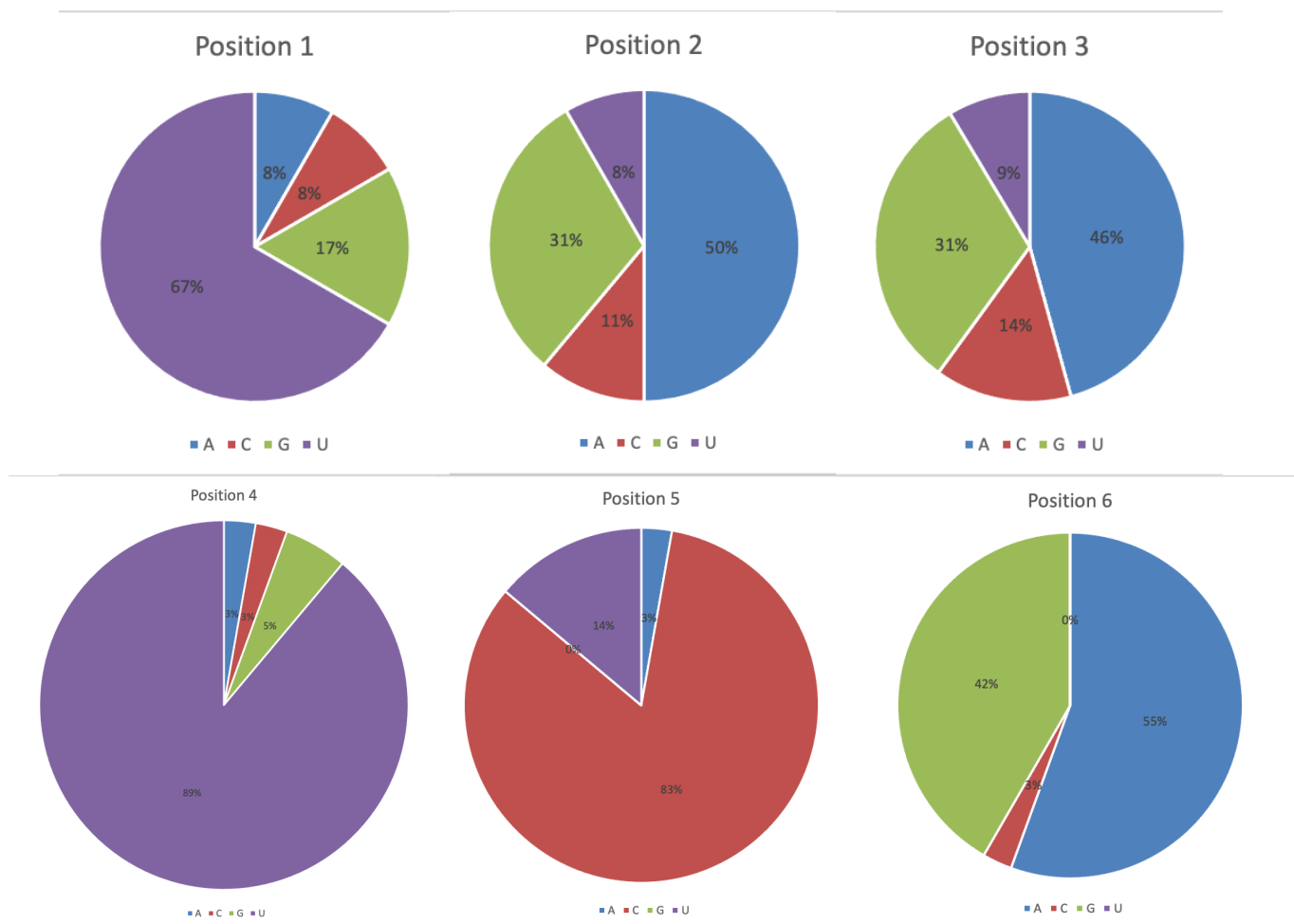
Graph 2: Nucleotide placed at each position in the gateway library for the top 1/4 sequences based upon fold induction. Position 1 was U in NH5, Position 2 was A in NH5, Position 3 was A in NH5, Position 4 was U in NH5, Position 5 was C in NH5, and Position 6 was A in NH5.
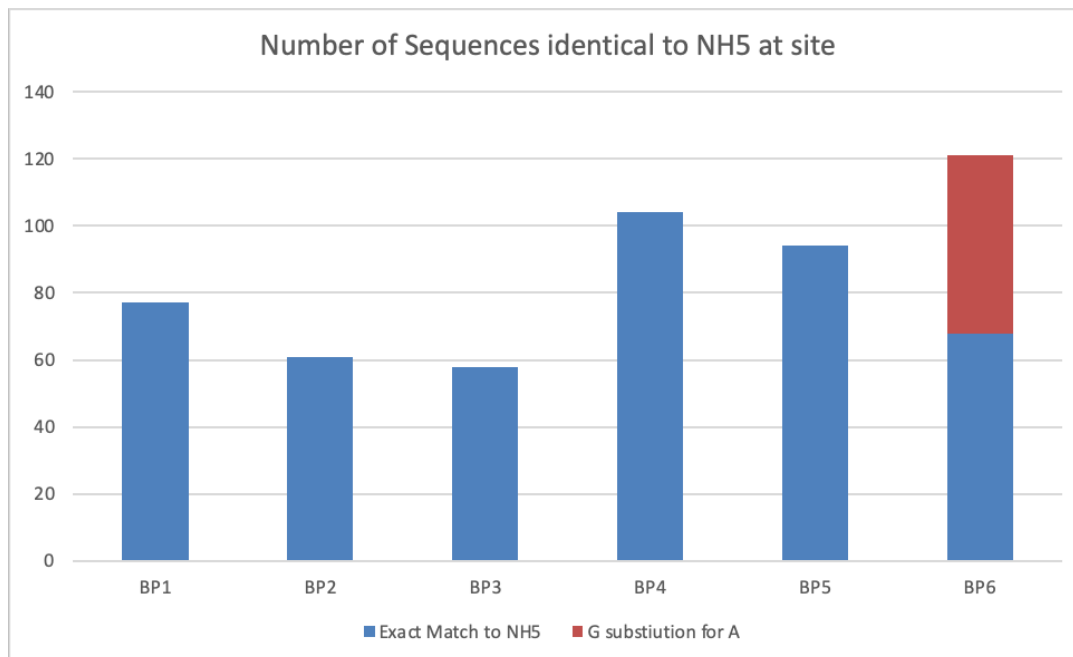
Graph 3: Number of sequences in all 142 studied that shared match for NH5 at each nucleotide position. BP6 has an extra red bar that demonstrates the number of A to G substations made. This demonstrates the importance of a purine at BP6.

Analysis of individual nucleotide positions demonstrates the importance of conserved sequence element in the 3' half of the gateway library. Armed with the importance of a U at BP4, a C at BP5, and a purine at BP6; the effectiveness of switches was reevaluated (Graph 4). This demonstrated that all the switches with high fold induction and low repression numbers had all three of these important sequence elements. The high performer with only 3 conserved base pairs, was found to contain all three important elements. Additionally, there is a clearer trend amongst the moderate performers. The more important identities, the better the sequences seemed to perform.
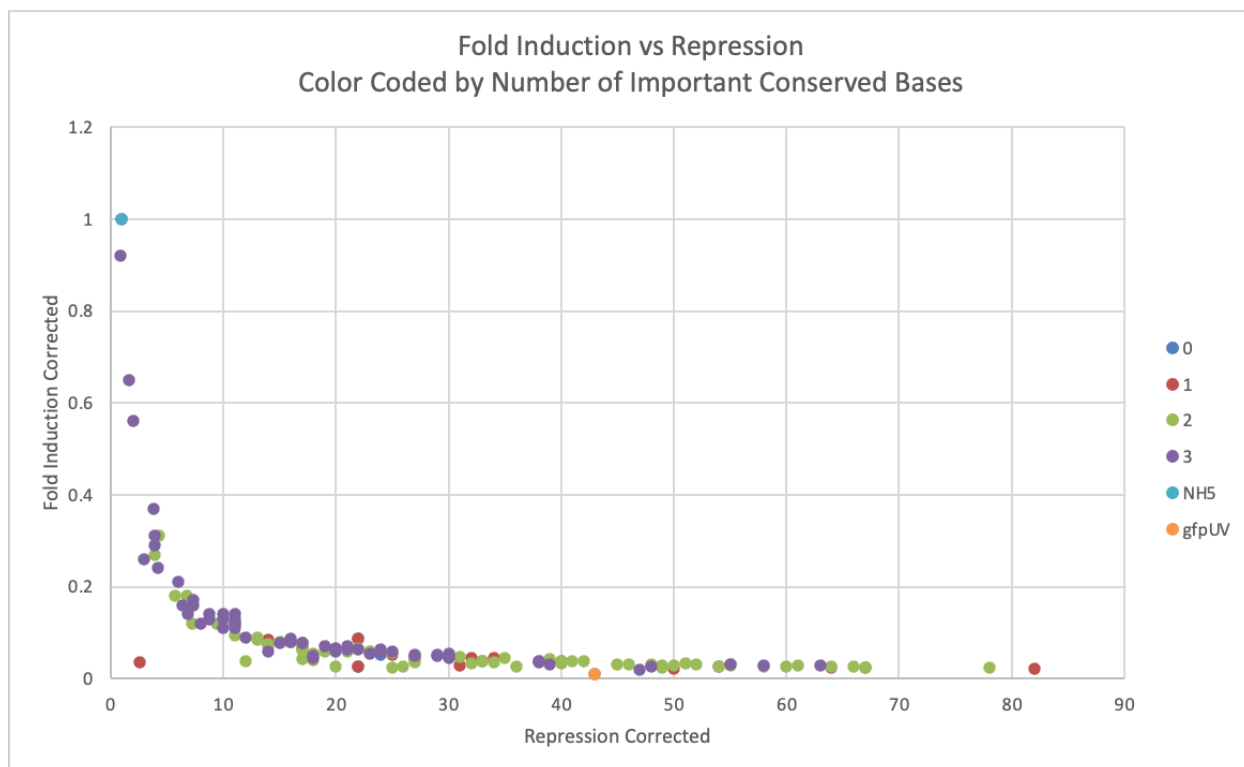
Graph 4: Graph of Fold Induction normalized to NH5, vs OD Corrected Fluorescence in the absences of 2AP normalized to NH5. Each data point is color coded based upon the number of important elements conserved from NH5. These elements were a U at BP4, a C at BP5, and a purine at BP6. There is a clear trend that the more important elements preserved, the better the switch represses and induces.

4.2 P1P3 Library

Screening of the P1P3 library appears to be unsuccessful. This library is involved in the P1 helix that connects the aptamer domain and expression platform. Previous mutagenic experiments focused on the aptamer domain of the *PbuE* adenine-responsive riboswitch have suggested that conservation of this domain would be important for riboswitch activity[3]. However, in wild type riboswitches, there is a single A-A mismatch in this region that was

resolved when creating NH5[22]. Therefore, at least three riboswitches with high fold induction should have been isolated with mutations in that site, a phenomenon that was not observed. Instead, with over 99% of the possible colonies observed, only five were selected and only one had a fold induction greater then two (Chart 3). Additionally, while screening NH5 was re-isolated three times and seven sequences that appeared to be effective had mutations outside of the area of interest (Chart 3). This suggests an error in the selection method or the creation of the library that prevented isolation of effective switchers and abundance of mutated riboswitches.

Number of Colonies Observed on Primary Screen	5,237
Number of Colonies Picked from Secondary Screen	15
Number of Colonies that had Fold Induction Over Two	1
Number of Times NH5 was Reisolated	3
Number of Mutated Sequences Chosen from Secondary Screen	7
Fraction of Variants Observed	99.4%

Chart 3: Numerical Summary of Colonies Observed Through P1P3 Screen

5 Discussion and Future Experiments

The first major takeaway that mutagenesis has revealed about the expression platform is the high efficiency of NH5. No mutated sequence within either library isolated a higher performer then NH5. In the gateway library, the closest sequence to NH5 had a single nucleotide substitution, an A for a G at nucleotide 6, and still had a fold induction only 92% that of NH5. In addition to the five nucleotide similarity to NH5, it also had a G in the sixth spot where any purine was highly tolerated. The next closest sequence had a fold induction 65% that of NH5. Slight deviations from the NH5 parent sequence have dramatic effects on the ability of the riboswitch to control transcription. This mutagenesis reiterated the success of Drogalis and Batey's creation of NH5 as an idealized riboswitch (Drogalis & Batey (2020).

The analysis of the Gateway Library demonstrated the importance of location of base pairs in initiating strand exchange. This library revealed that, while preservation of base pairing generally is vital to stand exchange, base pairing in nucleotides 4-6 are the most important in preserving riboswitch activity. In NH5 these three nucleotides form U-A, C-G, and A-U base pairs. Additionally, as Figure 15 shows, these nucleotides are the last to invade and the closest to the end of the newly formed helix. The preservation of base pairing in the end of the invading strand suggests the importance of a strong hold to maintain the newly formed helix and finish strand exchange.
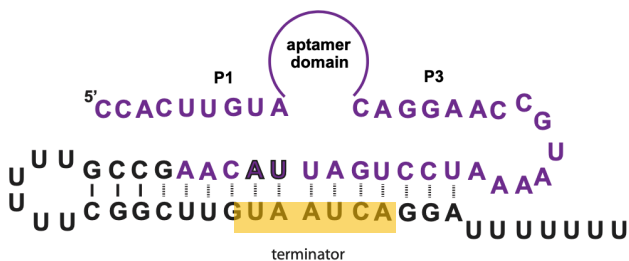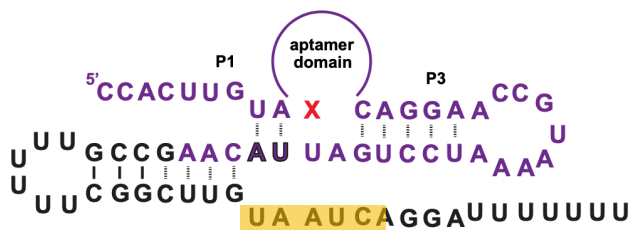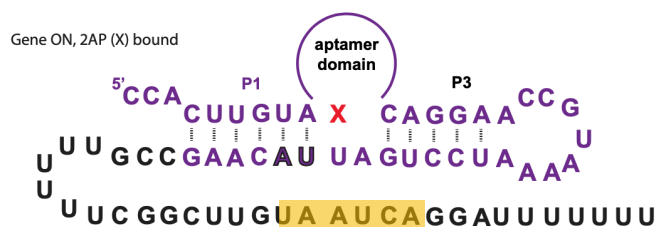
Figure 15: Schematic of strand invasion in NH5 with gateway library highlighted in yellow. Adapted from Lisa Hansen (2022).

The major takeaway from the P1P3 library is that continued screening needs to occur. The screening did not provide an adequate number of non-mutated functional riboswitches. The sequences that were isolated did not align with expected hypothesizes and understanding of the rest of the expression platform. To remedy this confusion, a new round of screening should be done and the P1P3 library should be reevaluated and reanalyzed based upon this new screening.

In order to understand the importance of nucleotide sequence and the mechanism of other sections of the expression platform, other members of the Batey Laboratory are conducting mutagenic analysis on alternative libraries. The end goal is to have an in-depth analysis of how

each section of the NH5 riboswitch's expression platform works to cause conformational change. A few of these libraries have also isolated sequences with better fold induction then NH5. Once the screening of the entire expression platform is completed, an idealized NH5 can be designed. In addition, the understanding gained on strand exchange, base pairing, and RNA folding can be applied to other model systems and utilize to advance fields of biology, biochemistry, and medicine.

References

(1) 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

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

(3) 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

(4) Gao, R., Mack, T. R., & Stock, A. M. (2007). Bacterial response regulators: versatile regulatory strategies from common domains. *Trends in biochemical sciences*, *32*(5), 225–234. https://doi.org/10.1016/j.tibs.2007.03.002

(5) Serganov, A., & Patel, D. J. (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature reviews. Genetics*, *8*(10), 776–790. https://doi.org/10.1038/nrg2172

(6) "Central Dogma." 2023. *National Human Genome Research Institute*. https://www.genome.gov/genetics-glossary/Central-Dogma (March 12, 2023).

(7) Nicole Giese Rura. 2019. "Regulatory RNA: The Sculptors of Gene Expression." *Whitehead Institute of MIT*. https://wi.mit.edu/news/regulatory-rna-sculptors-gene-expression (March 12, 2023).

(8) Valadkhan, S., & Gunawardane, L. S. (2013). Role of small nuclear RNAs in eukaryotic gene expression. *Essays in biochemistry*, *54*, 79–90. https://doi.org/10.1042/bse0540079

(9) Macfarlane, L. A., & Murphy, P. R. (2010). MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics*, *11*(7), 537–561. https://doi.org/10.2174/138920210793175895

(10) Dana, H., Chalbatani, G. M., Mahmoodzadeh, H., Karimloo, R., Rezaiean, O., Moradzadeh, A., Mehmandoost, N., Moazzen, F., Mazraeh, A., Marmari, V., Ebrahimi, M., Rashno, M. M., Abadi, S. J., & Gharagouzlo, E. (2017). Molecular Mechanisms and Biological Functions of siRNA. *International journal of biomedical science : IJBS*, *13*(2), 48–57.

(11) Gilbert, S. D., Love, C. E., Edwards, A. L., & Batey, R. T. (2007). Mutational analysis of the purine riboswitch aptamer domain. *Biochemistry*, *46*(46), 13297–13309. https://doi.org/10.1021/bi700410g

(12) Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., & Breaker, R. R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, *113*(5), 577–586. https://doi.org/10.1016/s0092-8674(03)00391-x

(13) 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

(14) Mandal, M., & Breaker, R. R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology*, *11*(1), 29–35. https://doi.org/10.1038/nsmb710

(15) Citation: Edwards, A. L. & Batey, R. T. (2010) Riboswitches: A Common RNA Regulatory Element. *Nature Education* 3(9):9

(16) Hu, G., Li, H., Xu, S., & Wang, J. (2020). Ligand Binding Mechanism and Its Relationship with Conformational Changes in Adenine Riboswitch. *International journal of molecular sciences*, *21*(6), 1926. https://doi.org/10.3390/ijms21061926

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

(18) Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R., & Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry & biology*, *11*(12), 1729–1741. https://doi.org/10.1016/j.chembiol.2004.11.018

(19) 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

(20) 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

(21) Porter, E. B., Marcano-Velázquez, J. G., & Batey, R. T. (2014). The purine riboswitch as a model system for exploring RNA biology and chemistry. *Biochimica et biophysica acta*, *1839*(10), 919–930. https://doi.org/10.1016/j.bbagrm.2014.02.014

(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) Widom, J. R., Nedialkov, Y. A., Rai, V., Hayes, R. L., Brooks, C. L., 3rd, Artsimovitch, I., & Walter, N. G. (2018). Ligand Modulates Cross-Coupling between Riboswitch Folding and Transcriptional Pausing. *Molecular cell*, *72*(3), 541–552.e6. https://doi.org/10.1016/j.molcel.2018.08.046

(24) Scull, C. E., Dandpat, S. S., Romero, R. A., & Walter, N. G. (2021). Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control. *Frontiers in molecular biosciences*, *7*, 607158. https://doi.org/10.3389/fmolb.2020.607158

(25) Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T., & Lucks, J. B. (2016). Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nature structural & molecular biology*, *23*(12), 1124–1131. https://doi.org/10.1038/nsmb.3316

(26) Boyapati, V. K., Huang, W., Spedale, J., & Aboul-Ela, F. (2012). Basis for ligand discrimination between ON and OFF state riboswitch conformations: the case of the SAM-I riboswitch. *RNA (New York, N.Y.)*, *18*(6), 1230–1243. https://doi.org/10.1261/rna.032177.111

(27) Jain, S., Gupta, R. & Sen, R. Rho-dependent transcription termination in bacteria recycles RNA polymerases stalled at DNA lesions. *Nat Commun* 10, 1207 (2019). https://doi.org/10.1038/s41467-019-09146-5

(28) Rieder, U., Kreutz, C., & Micura, R. (2010). Folding of a transcriptionally acting preQ1 riboswitch. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10804–10809. https://doi.org/10.1073/pnas.0914925107

(29) Breaker R. R. (2018). Riboswitches and Translation Control. *Cold Spring Harbor perspectives in biology*, *10*(11), a032797. https://doi.org/10.1101/cshperspect.a032797

(30) Caron, M. P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., & Lafontaine, D. A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(50), E3444–E3453. https://doi.org/10.1073/pnas.1214024109

(31) Cheng, L., White, E. N., Brandt, N. L., Yu, A. M., Chen, A. A., & Lucks, J. B. (2022). Cotranscriptional RNA strand exchange underlies the gene regulation mechanism in a purine-sensing transcriptional riboswitch. *Nucleic acids research*, *50*(21), 12001–12018. https://doi.org/10.1093/nar/gkac102

(32) Drury, M. D., & Kmiec, E. B. (2003). DNA pairing is an important step in the process of targeted nucleotide exchange. *Nucleic acids research*, *31*(3), 899–910. https://doi.org/10.1093/nar/gkg171

(33) Hansen, Lisa N. 2022. "Identifying Design Principles for the Expression Platform of the B. subtilis Adenine-Responsive pbuE Riboswitch Through a Genetic Screen."

(34) Molecular Cloning: A Laboratory Manual, 3rd ed., Vols 1,2 and 3 J.F. Sambrook and D.W. Russell, ed., Cold Spring Harbor Laboratory Press, 2001, 2100 pp., soft cover | Sigma-Aldrich http://www.sigmaaldrich.com/ (accessed 2022 -03 -20).

Appendix A: Gateway Library Fasfa

```
>NH5 6U_ 99
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G3B2_91
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G2C3_64
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R2G1B3_56
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'5G4C3_37
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'5G1D2_31
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G1B2_31
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R2G2A2_29
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'5G2C2_27
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'2G4D4_26
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G2A1_24
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G2D2_20
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G6A1_18
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCGGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R2G2C2_18
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACTTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G1A2_17
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'4G1B4_16
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTATTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G3D1_16
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTCTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G1B1_14
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'5G1B4_14
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTTTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G3C4_14
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G7A2_14
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAAATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'2G2C2_13
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G2A3_13
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10G1C4_13
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAAGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G5C1_12
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGACCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R2G1B2_12
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G2C1_12
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G2D3_12
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAGTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G1B2_11
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCGGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G2A4_11
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGCTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'4G3A1_11
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAGATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G1C4_9.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'5G5D4_8.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAAACAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'4G4A1_8.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCTTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G2A1_8.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGAGCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G6C4_8.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G5A2_8.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACGACGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G2A4_8.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCTATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G6C4_8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G2A4_8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G6B2_7.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAATCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G2B3_7.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTATGAGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G3A4_7.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGTGTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G1A3_7.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1B1_7.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGATCCGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G3C2_7.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCGGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G1D4_7.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAAGAGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1D3_7.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G1D4_7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGTATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G5B3_7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAAGAAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G1D1_6.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAGGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1A1_6.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCGGCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G2C4_6.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAGTCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R2G1A1_6.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAATTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G6B1_6.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G1A3_6.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCCATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G6A1_6.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G2B4_6.2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAATCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G1B3_6.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G4C1_5.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAGGCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G2A1_5.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'2G5C3_5.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGCATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G2C4_5.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCGGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G2C4_5.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGTTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G2A1_5.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAATTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G2A2_5.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAGATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G2A3_5.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G7D1_5.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGAATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'6G1D1_5.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAGCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G6A2_5.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGTGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G6B2_5.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCTTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G6D4_5.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGCGTACGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G1A1_5.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCTTTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G3B4_5.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTATGACGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G6B4_5.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAAGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G1C2_5
```

```
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10G1D1_5.0
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGCCTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G5D1_4.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTGCCGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G1C2_4.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCTTTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G5D3_4.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'3G2D4_4.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGATGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G8A2_4.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAATCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G6B3_4.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAGCTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G2C3_4.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCTTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1A3_4.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACGCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G3B1_4.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGAGAGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'2G3D2_4.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCAGAGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G1D4_4.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGCCCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G5A4_4.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCTATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G5C4_4.0
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGGGCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G6C2_3.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGTATCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G1B2_3.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G3C1_3.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGATCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G1A2_3.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCGATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'1G2B1_3.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAAATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G3B2_3.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGAGTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'6G2B3_3.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'2G6A1_3.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGATTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G1D4_3.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTGCAAGCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G3C4_3.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAAACGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1C3_3.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAGTTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G3B2_3.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCATGAGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'2G4A1_3.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCAGCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G3G2_3.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G2B2_3.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCGTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'6G2B2_3.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGATAACAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G4B2_3.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAAATACGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1B2_3.1
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTATTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'6G3B1_3.0
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCATCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G3B1_3.0
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAATTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R10G1C1_3.0
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGGTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G2A1_3.0
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTGTTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G2A4_2.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGGTCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G6B3_2.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCACTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G1D3_2.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGATGAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G3C3_2.9
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGCCTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G7A3_2.8
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTACCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R1G1B4_2.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGACCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R'6G3B2_2.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGGCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R6G6D3_2.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTATCAAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G3C1_2.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAGGATGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G1C3_2.7
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTGTCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R11G8A3_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTATCCGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G2D2_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTATCTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G6B4_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTTTTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G6D2_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCACTTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G1D2_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGGTTTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G1D2_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTGCCGGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G6B2_2.6
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACGTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G3C2_2.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCGGGTTAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R13G3D2_2.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTCGGCGGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G3A2_2.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGAAGCCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G4C2_2.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATAAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G4C2_2.5
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTGGCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G5B1_2.4
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGGTCAAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G1D3_2.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGTGTCGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G2B3_2.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTTATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R3G3B4_2.3
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTACGAGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G7B1_2.2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGGCATTGGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R8G3B1_2.2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAACCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R4G5C1_2.2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTGAGCTGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>R5G2D3_2
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGCTCTCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
```

Appendix B: P1P3 Library Fasta File

>NH5_99
ATGCATAAGGCTCGTATAATATATTCCACTTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>P1P3R431D_16.3
ATGCATAAGGCTCGTATAATATATTCCAACGGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>P1P3R424A_1.0
ATGCATAAGGCTCGTATAATATATTCCATTAGTATAACCTCAATAATATGGTTTGAGGGTGTCTACTGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>P1P3R421A_.95
ATGCATAAGGCTCGTATAATATATTCCAGACGTATAACCTCAATAATATGGTTTGAGGGTGTCTACTTGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>P1P3R424D_0.80
ATGCATAAGGCTCGTATAATATATTCCATCAGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCTGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT
>P1P3R272D1_.40
ATGCATAAGGCTCGTATAATATATTCCAACTGTATAACCTCAATAATATGGTTTGAGGGTGTCTACCAGGAACCGTAAAATCCTGATTACAAGCCGTTTTTTCGGCTTGTAATCAGGATTTTTTTTATTTACTAGTACATTTAAGTAAAGGAGTTTGTTATGACCATGATTACGCCAAGCTT