**Sequence constraints for efficient strand invasion for co-transcriptional gene regulation by an adenine sensing riboswitch**

**OR**

**Elucidating crucial elements of co-transcriptional strand invasion using an adenine sensing riboswitch**

**OR**

**Gleaning design principles of co-transcriptional strand invasion using an adenine sensing riboswitch**

**OR**

**Optimizing the functional core of an adenine sensing riboswitch informed by strand invasion**

**Lisa Hansen, Jesus Alicea, Ava Altenbern, Alexandra Brown, Parker Juels, Mia Kim, Aishah Hassan Scott Odierno, Matthew Payne, Peyton Roeder, and Robert T. Batey\***

Department of Biochemistry, University of Colorado, Boulder, CO, 80309-0596, USA

\* To whom correspondence should be addressed. Tel: 1(303) 492-2159; Fax: 1(303) 492-5894; Email: [Robert.Batey@colorado.edu](mailto:Robert.Batey@colorado.edu)

Keywords:

*Key points:*

**Abstract**

**Introduction: INCOMPLETE**

INTRO

Strand invasion involves at least three nucleic acid strands, two of which (substrate and incumbent strands) are in a duplex, and a third that is single stranded (invader strand) with similar or identical sequence composition to the incumbent (36535150, 34752075, 21258382). The process of invading or exchanging happens when the invader strand fully displaces the incumbent, creating a novel duplex (35659535, 24402831). While the process where both strands compete for pairing with the substrate is random, the formation of a toehold can favor displacement by the invader (36535150). In this case, a toehold is a set of base pairs between the substrate and invading, which is allowed when there is a single stranded overhang on the substrate strand (36535150, 21258382, MORE).

Within stand invasion, there are multiple stages that must be understood, including toehold formation, branch migration, and displacement (**Figure 1A**). Branch migration is the stochastic process in which the invading and incumbent strands compete for base pairing with the substrate strand (34752075, 10949296). This process can be biased towards the invading strand in the presence of a toehold because the invading strand will necessarily form a longer duplex (34752075, MORE). Based on previous studies, factors that influence the rate of invasion include the length of the toehold, the number of mismatches (non-Watson Crick or non-wobble pairing), the location of the toehold on the substrate strand, and the type of nucleic acid present ([25762335](https://pubmed.ncbi.nlm.nih.gov/25762335), 34752075, 21258382).

Riboswitches serve as an example of highly folded, functional RNAs that often employ strand invasion to regulate associated genes (15919195, MORE). These elements have two principal components: a receptor domain, and a regulatory domain. The receptor domain, also called the aptamer domain, is a highly structured and conserved region that binds a ligand with high specificity (21925376, 14523920, 15173824). The regulatory domain, or expression platform, is a dynamic region that creates a genetic output by folding co-transcriptionally based on the ligand input in the receptor domain (20943759).

Riboswitches that control at the level of transcription must fold co-transcriptionally to exert control on the RNA polymerase before it escapes the element. The adenine responsive pbuE riboswitch from B. subtilis is an ON switch that folds co-transcriptionally. The interconversion between the two mutually exclusive states proceeds through strand invasion of the 3’ end into the ligand binding pocket (15919195**).**

* **Hole in understanding?**
* **This study**

This study probes crucial elements of strand invasion within an RNA based system *in vivo* through a genetic screen. The RNA employed is a modified variant of the *pbuE B. subtilis* Adenine responsive, optimized to increased dynamic range and highlight key features of RNA folding (33259551). Canonically, it is a transcriptional switch that regulates expression of a purine efflux pump in *B. subtilis*, turning ON expression when excess adenine is present in the local environment (16931335, 25550163). Transcriptional riboswitches are under immense temporal constraints as they must survey the environment and either form the ON or OFF conformation before the RNA polymerase escapes the switch. The formation of the stable OFF conformation is only accomplished when the 3’ end of the riboswitch successfully invades into P1, displacing the initial duplex. This strand invasion is toehold-mediated by the duplex region P4, which stabilizes the 3’ end to invade into the intact P1 helix (). This RNA is therefore a key model for understanding strand displacement *in vivo* as it coordinates the co-transcriptional folding of RNA and ligand-gated strand displacement to make genetic decisions in a living system.

The modified parental switch that the following genetic screen is based on greatly outperformed the 5.6-fold switching of *pbuE* with a near 100-fold induction in cells. Modifications were made under the assumption that Watson-Crick (WC) base pairing in both the receptor and regulatory domains would create the best switch (33259551). Therefore, the switch became highly symmetrical, save the aptamer domain, and regions thought to be key to nucleation of the terminator helix (such as P4) were enriched with GC pairs. In addition to changing the primary sequence, the pre-aptamer sequence was shortened by 27 nucleotides to reduce misfolding. The prior mutagenic analysis, while yielding a high preforming variant did not aim elucidate the design principles behind this improved switching ability.

This study, through multiple genetic screens, offers insights into primary and secondary structural requirements for the processes of strand-invasion and nucleation that allow for the dynamic properties of this unique regulatory element. The regulatory domain of the riboswitch was evaluated through multiple discrete libraries, each randomizing a different region of interest. Those libraries were analyzed through qualitative and quantitative methods to assess the riboswitch functionality, and functional variants were extracted and sequenced. Through the analysis of regions in the terminator helix, we find that strand invasion nucleation residues are the most essential and highly conserved. When looking at paired regions or regions in competition between the aptamer and terminator helix, balance between the two paired regions must be tuned to optimize functionality. From these data it seems that the development of novel biosensors that use this scaffold can be easily designed i*n vivo* with limited manipulation to key areas near the initial 3-way branch point of the switch based on the parameters of the system.

**Results and Discussion**

**Creation of artificial phylogeny through parallel genetic screens**

Before investigating novel riboswitches, previously tested riboswitches were interrogated through cell-based assay alongside the parental riboswitch for this genetic screen, *NH5*. This included the wildtype pbuE responsive purine riboswitch (WT) and the wildtype with a deletion of the 27 nucleotides in the pre-aptamer domain (del27\_WT), each of which were expected to underperform compared to *NH5*. Consistent with the work by Drogalis and Batey (), the del27\_WT variant had improved turn ON, slightly leakier turn OFF, and increased fold induction in comparison to the full WT variant (Figure XA). *NH5*, as suspected, displayed robust repression in absence of ligand, and expression in presence, with a fold induction 6x larger than the WT constructs.

Through screening of multiple regions of interest within the expression platform, 550 unique sequences were extracted and characterized with a cell-based activity assay. While *NH5* prevailed as the best solution for many of the libraries, a broad range of repression and fold induction was surveyed, ranging from those that outperformed *NH5*, to variants unable to function even as well as the WT riboswitch (Figure 2B). After confirming that the control riboswitches behaved as expected, libraries were constructed looking at 3 overarching themes. Theme one was toehold formation, so two libraries were created to characterize the toehold domain which nucleates the formation of the terminator element. These libraries included one focusing on the paired region that is the toehold itself, P4, and the terminal loop adjacent to that element L4. Three overlapping libraries were designed to understand the possible sequence space of the invading strand when the incumbent and substrate helix remains constant. The three libraries (P1 invader, P1-P3 invader, P3 invader) each include different obstacles including paired and unpaired regions of the aptamer domain. The second theme was characterizing the invading strand. Theme three was the exchange between the mutually exclusive ON and OFF states evaluating how the strength of one helix compares to the strength of the competitor helix through libraries that randomized regions in both helices. These includes at the site of initial invasion (P1-P4 exchange and P1-P4 exchange minimized), and a region further upstream where P1 is competing with the terminator helix (P1-P(T) exchange).

**Toehold formation**

Inclusion of a toehold in a strand invasion system is imperative to ensure high fidelity displacement (). To successfully terminate, the terminator helix must form at the expense of the binding pocket, and it must occur when the 3’ end of the helix is within the exit channel of the RNA polymerase (36631609), which accommodates 6-7 nucleotides of ssRNA (9843952). If those conditions are not met, the RNA polymerase will proceed regardless of ligand binding. P4 in the NH5 riboswitch serves as a toehold that includes 3 base pairs which are present in both conformations of the riboswitch (**Figure 1B**). Two libraries were designed to interrogate this region, the P4 library and the L4 library. L4, the terminal loop proximal to P4, was of interest because of the co-transcriptional nature of this riboswitch. To form a productive toehold, it was hypothesized that there would be sequence requirements that would either slow the processivity of the RNA polymerase or impact the nucleation of the toehold (Figure X). The P4 library includes the two base pairs in P4 distal to L4 and an additional adjacent base pair that is in competition with P1. In the parental *NH5*, L4 is a polyuridine tract and P4 is a GC rich stem. The L4 library was of interest because a the hexauridine tract in the NH5 parental mutant is hypothesized to be a “programmed pause” site for the RNA polymerase in *E. coli* and showed marked improvements for both expression and repression compared to mutants without this feature (33259551), and it has been shown that that motif pauses RNAP even in a heterologous system with a *B. subtilis* riboswitch expressed in *E. coli* (28541183). But, when tested against a variant with 5’UUCCUU rather than 5’UUUUUU, activity was very similar, so the sequence requirements of this region were unclear following mutagenesis (33259551).

To assess the sequence requirements of the two regions, two libraries were designed, each with 6 nucleotides randomized. Each library was screened with a different strategy based on performance. The L4 library produced numerous highly functional switches which rivaled *NH5*, so screening was halted due to the lack of insight and large burden of screening that region. Only ~200 sequences were observed of this library, which yielded 77 functional sequences, emphasizing the abundance of robust switchers within that library (table S#). Though only 4.8% of observing any potential sequence were screened in the L4 library, the data produced indicated that continued screening by this approach was unlikely to be informative. The requirements of the P4 library, which includes the three base pairs in P4 distal to L4, were much more stringent based on colonies screened, ~14,000, and functional sequences extracted, 127 (table S#). The P4 library was screened until ~97% of potential sequences had been observed (table S#).

Ultimately, the L4 library had a glaring lack of sequence preference. Every position was observed as any nucleotide, a large portion of observed riboswitches (~39%) maintained the ability to function, and many variants functioned comparably to *NH5*. No sequence variants were in the bottom bin, 8 were in the middle bin and 69 were in the top. While pausing can be important for regulation by riboswitches and the polyuridine tract has been seen to slow the movement of the RNAP in E. coli (28541183), these findings are consistent with work done by the Lucks group on the ZTP riboswitch who also saw a lack of functional impact from disrupting a predicted pause site (36864761). The region may not be important for pausing, but rather may serve to facilitate formation of the toehold so that P(T) can form efficiently within the exit channel of RNAP to stop transcription. Based on the activity of these variants, a 6-nucleotide loop appears to allow for efficient helix formation. LOOP LENGTH SERIES?

In the P4 library, all variants in the top bin had productive base pairing in the second base pair (n-1 of invasion for P1) and the majority (83%) had productive pairing in the other two pairs. This indicates pairing is important in this region, but the strength/identity of each pair may change depending on the sequence of the competing P1 strand. When looking at the three bins, it is most informative to look at secondary structure rather than primary sequence. When comparing percentage of productive base pairs across bins the top bin has a higher percentage of productive base pairs. The middle and bottom bin have similar proportions of pairing at the position distal to L4, but increased pairing in the other two base pairs.

Of note, one sequence which switched each of the 3 base pairs to AU (R9.2A4) rather than the parental GC expressed 9x more FP in the absence of ligand than *NH5*. But, in the presence of 2AP it expressed slightly more than *NH5*, resulting in a marked change in fold induction likely due to reduced stability of the toehold. Many colonies in the bottom bin also formed 3 productive base pairs, so the formation of any WC or wobble pair is not sufficient for robust switching. For example, one colony that also forms 3 AU pairs (R6N.2C3) and differs only in the directionality of base pairs 2 and 3. That sequence lacks the ability to express or repress near the parental or the other sequence with 3 AU pairs. One sequence, with a fold induction less than a tenth of *NH5*, still maintains 3 GC base pairs. The only difference is that the first base pair has been changed from GC to CG. The deleterious impact of this change may be related to stacking energies or interactions with the pre-aptamer sequence, but overall suggests that analyzing the strength of each base pair will not directly translate to the in vivo activity of the system.

**Invasion Strand**

The next family of libraries seeks to understand the sequences that can successfully invade a non-randomized, WC duplex. The 3’ end of this riboswitch must invade into P1 and P3 to create P(T), a process facilitated by toehold formation in P4 and J3/1. The motivation for the design of *NH5* was to maximize the number of WC base pairs, both in the substrate-incumbent helix and the substrate-invader helix. Three libraries were deigned to determine if that is the optimal design by varying only the invader strand. The P1 invader (n=5) includes the 3’ end of P4 and the invasion of P1, P1-P3 invader (n=6) spans P1 to J3/1 to P3, and the P3 invader (n=5) spans P3.

Each of the three libraries were screened until there more than 99% probability than any one sequence would be observed (18567049). Of the two 5-nucleotide libraries, P1 invader yielded 17 unique and functional switches after observing ~5,100 colonies and the P3 invader yielded 15 unique and functional sequences after observing ~9,000 colonies (Table S#). For the 6-nucleotide library, P1-P3 invader, 141 unique and functional sequences were extracted after observing ~23,000 variants (Table S#).

Overwhelmingly, it seems that NH5 was the best solution for the invader strand across the three libraries. The P1-invader library, encompassing the transition into P1, had one variant with a higher fold-induction than *NH5* by changing the GC pairing formed at the third base pair to a GU wobble. But that variant failed to repress as well as *NH5*, meaning the decrease in stability likely limits the ability of the 3’ end to invading. The most conserved positions, likely the workhorses of the invasion process, were the first and second positions across all bins. These positions are part of the toehold (n-2 and n-1 positions of invasion into P1) and do not compete with P1, reinforcing the importance of the toehold for successful invasion. The n-1 position was invariant, even amongst low preforming switches and likely creates a stabilizing WC GC pair to initiate invasion into P1.

The P1-P3 invader failed to produce a variant with a fold-induction as high as NH5, with only 3 out of the 141 sequences falling in the top bin. However, one sequence was able to repress better than *NH5*, due to a single point mutation at position 5 which again changed a GC to a GU. Position 5 is the first that must compete with P3, and the weakening of that base pair is likely allowed due to the preceding toehold formed between the invader strand and J3/1. Across bins, there is high conservation in nucleotide identity at positions 4 and 5, corresponding to the n-1 and n positions of invasion into P3.

The final invasion library, P3 invader, had no sequences that preformed near *NH5* and had 0 sequences falling in the top bin. All sequences found in the middle bin were single point mutants at position 1, 2 or 5, but all of them had a dramatic impact on repression reflected in a far lower fold-induction. While WC pairing clearly leads to the best functional performance in this area, there was high nucleotide identity at positions 3 and 4 in both bins (more so in the middle bin) which indicates again that not all positions are created equal. This library aligns directly with P3 and has no ssDNA binding partners, but these two positions, n-2 and n-1 of the P(T) terminus, may be important for overall P(T) stability to actually effect change in RNAP activity. One variant in the mid bin with a point mutation at position 5 converting a UA pair to a UG pair had 7.4x more FP expression without ligand and 0.76x expression with ligand, again indicating the importance of a WC GC pair at the n-1 position. No point mutations were found for the n-2 position after extensive screening which may indicate that variation at that position does not allow for P(T) formation without a compensatory mutation. This is supported by the fact that all mutations of the n-2 position (position 4) are accompanied by at least 3 other mutations which may change the entire base pairing network in the region.

**Strand Exchange**

The final theme sought to combine insights from the toehold and invader to look at the more complicated interplay between the three strands. Three libraries were constructed including nucleotides in each of the strands predicted to actively be competing during the initiation or branch migration. The P1-P(T) exchange (n=6) is a slightly larger library looking at sequence space for two base pairs that are P1 for the ON state and P(T) for the OFF state. The other two libraries, P1-P4 exchange (n=5) and the minimized P1-P4 exchange (n=3), specifically zoom in on the gate between P4 and P1 to focus on the initiation of the strand exchange process.

The larger P1-P(T) exchange library was screened until there was a 96.4% of observing any one sequence through observation of ~14,000 colonies which gave 102 unique and functional sequences (Table S#). The P1-P4 exchange library gave 97 unique functional sequences after observing ~5,000 colonies, with a probability of observing any given sequence of 99.3% (Table S#). MINIMIZED LIBRARY HERE.

Based on the P1-P(T) exchange library sequences, pairing in P1 seems to be more important than pairing in P(T). The best performer had 3 GC pairs and a GU pair in P4 proximal to L4. Productive pairing is important in this region but variation in pair strength between P1 and P4 may improve function. Of those 15 sequences in the top bin, 100% formed a WC or GU wobble pair for both P1 pairs while 60% formed the P4 pair proximal to L4 and 80% formed a productive pair in P4 distal to L4. That trend continues through the bins, and it seems that the P1 pair proximal to the aptamer domain is the most important, with conservation above 90% in all bins, and the P4 pair proximal to L4 is least important with less than half of the 42 variants in the middle bin and 0% of the bottom variants forming a WC or wobble pair at that position. In previous work investigating this riboswitch in vivo it was established that the minimum size of P1 while still maintaining the ability to switch in response to 2AP was 3 base pairs (25550163). The highly conserved base pair in this library is the third base pair from the aptamer domain, adjacent to the two pairs that have ligand dependent interactions, and while it is a conserved RY in phylogeny 4/15 of the top bin switches has YR at that pair (25550163).

Unlike the P1-P(T) exchange library, P1 pairing seems to be unnecessary in the P1-P4 exchange library. The highest functioning switch in terms of both repression and FI had 3 AU base pairs and none of the sequences that functioned better than *NH5* had a GC pair in P1. All of those sequences formed at least 1 production (WC or wobble) base pair in P4 and none of them had two GC pairs in P4 like *NH5* does. This indicates that NH5 could be improved by weakening the pairing in this region or removing the ability to form the terminal base pair in P1. In response to those findings, multiple directed mutants were created that were not pulled out of the screen. Since there was such a marked increase in function by substituting the GC pairs for AU pairs, two mutants were designed with one AU pair and one GC pair to determine if one position was more favorable and if it was due to the change at both pairs or instead if it was driven by a single pair. A mutant was also designed that maintained the *NH5* sequence at all positions except for the first base at the 5’ end of P1. That base was changed from a C to an A, henceforth called *NH5-A*, and served as an example of a switch with a longer P4 and shorter P1 due to elimination of that GC pair. These three mutants created a mutagenic cycle to elucidate the differences between the best functioning variant created in the screen and the parental *NH5.*

The mutagenic cycle revealed that no single designed mutant was able to outperform R3AE6C4 or the parental *NH5*. Changing either the L4 proximal or L4 distal base pair in this library from a GC pair to an AU pair resulted in 0.8x fold-induction of *NH5.* While the distal AU pair mutant was able to repress better than *NH5*, it was unable to express as much FP in the presence of 2AP resulting in a reduced fold-induction. The NH5-A mutant interestingly had the same activity as NH5 in the absence of 2AP, but only expressed half as much FP in the presence of 2AP, again impacting the overall fold induction and indicating that there are benefits to having the first base pair of P1. This mutagenic cycle indicated that these mutations in concert allow for such a robust repression of FP (1/5th that of *NH5*) while maintaining robust gene turn on in the presence of ligand.

MINIMIZED LIBRARY

**Plug and Play**

**Conclusions**

**Material and Methods**

Library construction

Each library was constructed via PCR with 3 outer primers and one library specific Ultramer (Integrated DNA Technologies) (sequences in Table S#). The libraries were created via PCR with Q5 DNA polymerase (New England Biolabs) and amplified with 12 or 20 rounds of amplification with an annealing temperature of 55 °C. Constructs were confirmed on 2% agarose gel and cleaned with E.Z.N.A® Cycle Pure Kit (Omega Bio-Tek). Libraries were then digested with the high-fidelity restriction digest enzymes NsiI and HindIII according to manufacturer’s protocol (New England Biolabs) and again cleaned with the E.Z.N.A® Cycle Pure Kit (Omega Bio-Tek). To improve ligation efficiency, digested libraries were then treated with PNK (New Englad Biolabs). The libraries were ligated into a low-copy plasmid containing an ampicillin resistance cassette and *gfpUV* as a reporter gene downstream of the site of HindIII cut site. The plasmid was ligated into the vector using T4 DNA ligase by the manufacturers protocol (New England Biolabs).

Directed mutant construction

Loop series preparation

Mix and match

Phenotypic screen of potential switches

Libraries were transformed into chemically competent *E. coli* K-12 BW25113 cells and plated onto chemically defined salt broth (CSB) plates with 1.2 % agar and ampicillin following ligation. The exchange libraries that destabilized P1 helices (P1) were grown on plates with 0.5 mM 2-Aminopurine and the invasion stand and toehold libraries that destabilized P4 or P(T) helices were grown on plates without ligand. Colonies were grown overnight at 37 °C and *gfpUV* expression was visualized with a 366 nm light. Plates were photographed and counting statistics were preformed using a software package from Opencfu (3.9.0). On liganded plates, bright colonies were selected and on unliganded plates dim colonies were selected. Colonies that were selected were transferred onto a set of gridded plates, one liganded and one unliganded, and those plates were grown overnight at 37 °C. Gridded plates were then visualized with the 366 nm light and variants that were dim on unliganded plates and bright on liganded plates were selected due to ligand inducible expression and moved forward to further characterization.

Sequence and quantification of ability to attenuate gene expression

Selected colonies were plasmid prepped with the E.Z.N.A® Plasmid DNA Mini Kit I, (Omega Bio-Tek) and sent for Sanger sequencing (QuintaraBio). Variants were also subjected to a cell-based activity assay to quantify ligand induced gene expression. Variants were inoculated overnight in 2xYT media supplemented with ampicillin spinning at 37 °C. Overnight growths were then diluted 1:1000 into 2 subcultures of defined CSB media supplemented with ampicillin. One of the two subcultures contained 0.5 mM 2-Aminopurine which the other did not contain ligand. Subcultures were grown spinning at 37 °C for about 6 hours until an OD600 range of 0.4-0.5 (mid-log phase growth) was achieved (). Subcultures were transferred into a Costar® 96-well half area microplate with 3 wells containing 200 µL for each subculture. Each plate included 200 µL of a 5.0 µg/mL fluorescein standard to set gain for fluorescent measurements. Plates were read with the Tecan Infinite M200® PRO plate reader and the OD 600 and fluorescence in arbitrary units were measured. The excitation and emission wavelengths for the measurements were 395 nm and 510 nm respectively. Each measurement was done in technical and biological triplicate. Each biological replicate of variants was accompanied by 3 controls, pBR327 to correct for the background fluorescence of cells, PRR5\_*gfpUV* which constitutively expresses the reporter gene, and the *NH5* parental riboswitch.

Analysis of activity assay and library screen

To determine the probability of observing any one sequence for each library given the colonies counted from initial plating and the size of the randomized library was calculated based on the second equation proposed by *M. Reetz et al*. relating these factors (18567049). Libraries were screened until a threshold of 95% was passed unless otherwise stated.

The fluorescence values measured were processed in excel by dividing each fluorescence value by its OD600 and then subtracting the median OD600-corrected pbr327 fluorescence value (Microsoft). Those OD600- and pBR327-corrected values were then compiled into a .csv file which was then analyzed by a library specific R markdown, the R markdowns are open source and freely available (). There were two colonies in the P4 library that had negative repression values due to their ability to repress as much or more than the pBR327 control. Those variants were given an artificial repression value of 50 for comparison with other colonies, a value lower than the parental.

Processing of data included removal of outliers for OD600-correctly fluorescent values for measurements with and without ligand. The liganded and unliganded measurements were separated and outliers were determined to be those data points found more than 1.5-times the interquartile range away from the outer quartiles and were removed, a decision made assuming a Gaussian distribution (https://doi.org/10.1007/978-981-10-7563-6\_53). Data from the two conditions were then combined and the standard error of the mean and ratio of fluorescence with 2AP to fluorescence without 2AP (fold induction) were calculated. FASTA files were created by combining the sequences for each colony with values from the activity assay to create an activity associated sequence alignment (library sequence alignments found in XXX). To facilitate discussion of large data sets, each library was also divided into 3 groups based on performance. Those with a fold induction >=0.5-times that of *NH5* and a repression value <1.5-times that of *NH5* were in the top bin. The middle bin was variants with <0.5 *NH5*-correctedfold induction with a repression value of <=9-times that of *NH5.* Bottom bin included all other variants.



Figure 1: STRAND INVASION AND NH5. A schematic of strand invasion with the NH5 adenine responsive riboswitch (walk through steps). B secondary structure of ON and OFF states of NH5.

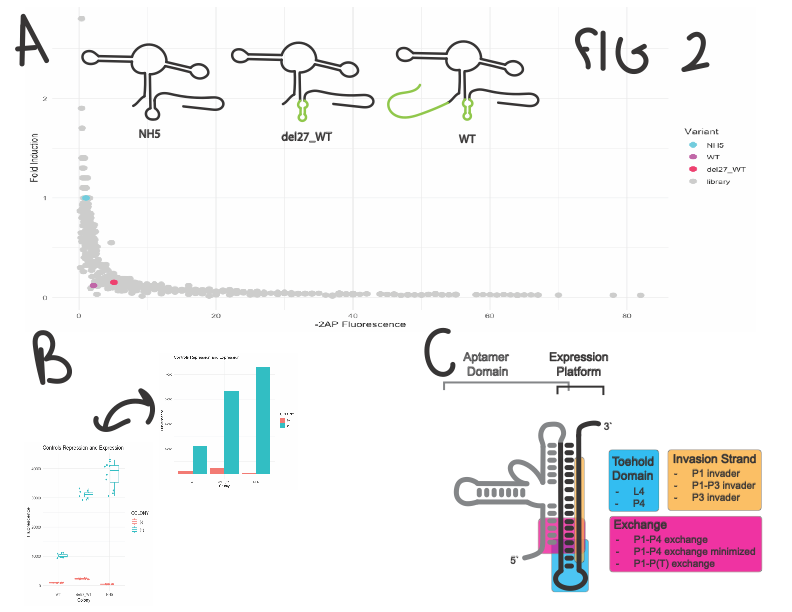


Figure 2: controls and libraries. A: controls highlighted in front of all libraries (P1-P4 exchange minimized not included). Cartoons of three controls shown in top corner, large structural changes in P4 and the pre-aptamer sequence are highlighted in green. B. activity of NH5, WT, and del27\_WT in presence and absence of 2AP. C. schematic of NH5 and regions surveyed.

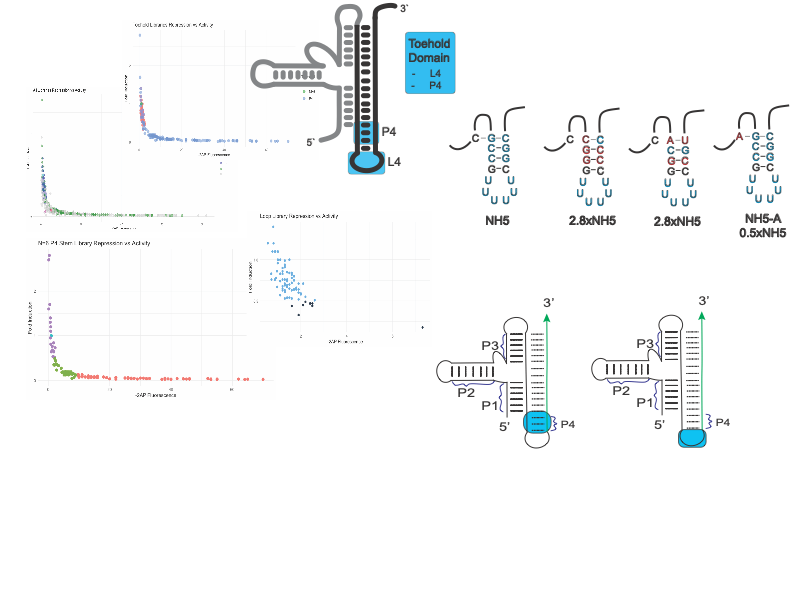


Figure 3: Toehold Graph, schematic, specific variants

Supplemental Table: Library completion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Library | nt varied | observed | functional | prob. |
| P1 invader  (Anchor) | 5 | 5,118 | 17 | 99.3 |
| P1-P4 exchange  (Anchor\_exchange P1\_gateway) | 5 | 5,069 | 97 | 99.3 |
| P3 invader  (Gateway) | 6 | 22,799 | 141 | 99.6 |
| L4  (loop) | 6 | 202 | 77 | 4.81 |
| P4  (nucleato/P4 stem) | 6 | 14,145 | 127 | 96.8 |
| P1-P(T) exchange  (P1\_exchange) | 6 | 13,573 | 102 | 96.4 |
| P3 invader  (P3) | 5 | 9,029 | 15 | >99.9 |