**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**

Much of the functional power of nucleic acids comes from their ability to strand exchange. This phenomenon is observed in contexts ranging from riboswitches to homologous recombination to CRISPR/Cas9 gene editing (37095744, 36535150). 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). The process of invading or exchanging happens when the invader strand fully displaces the incumbent, creating a novel duplex (35659535). 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).

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). 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). The existing literature has focused on studies examining the process of strand invasion in silico and in vitro, providing valuable insights to pivot into biological systems ([25762335](https://pubmed.ncbi.nlm.nih.gov/25762335),). Unfortunately, those studies lack the complexities of co-transcriptional folding and complex interactions between the RNA and RNAP that govern the process within cells.

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 and form the terminator helix (P(T)) (). The formation of P(T) triggers transcription termination and must occur in the exit channel of the RNA polymerase (RNAP) to cause global conformational changes and termination transcription (36631609). This indicates that strand invasion in vivo should be considered in the context of the RNAP (). 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**

**Toehold formation**

Inclusion of a toehold in a strand invasion system is imperative to ensure high fidelity displacement. 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 flexibility of the region (Figure X). The P4 library spanned 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). 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 via 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#).

**Conclusions**

**Material and Methods**

*Molecular Cloning of Riboswitch Libraries into Reporter Vector*

Riboswitches were amplified by PCR with consistent outer primers and a varied inner primer depending on library (Table #). PCR products were amplified for 12-20 rounds to minimize bias, product was confirmed on 2% agarose gel. Riboswitch libraries were then cloned into a reporter vector that was designed from the low-copy pBR327 plasmid (23654267, 25550163) using standard molecular cloning techniques (https://doi.org/10.1002/abio.370050118). The riboswitches were ligated upstream of *gfpUV* to control expression of fluorescent protein.

*Primary and Secondary Screening of Libraries*

Libraries were then transformed into *E. coli* Keio parental cells (BW25113) () with standard protocol. For those libraries that destabilized the expression platform (LIST THEM), transformants were plates onto CSB plates supplemented with 100 µg/mL carbenicillin. For the remaining libraries (LIST), which destabilized the aptamer domain were plated on CSB plates supplemented with 100 µg/mL carbenicillin and 0.5 mM 2AP. After overnight growth at 37 °C, plates were illuminated with a 3XX nm light to visualize *gfpUV* expression. These plates were photographed and the number of colonies on each plate was counted with the software Opencfu. Based on the number of colonies counted on each plate, likelihood of observing any one sequence was calculated using the second equation in *Reetz, MT et al* (18567049). When illuminated, those colonies that were dim on liganded plates or bright on unliganded plates were transferred onto a pair of gridded plates (one with ligand and one without) and again incubated overnight at 37 °C. Gridded plates were again illuminated with the same light after overnight growth and those colonies with were dim on the unliganded plates and bright on the liganded plates were selected to be screened further as they appeared to be functional riboswitches.

*Cell-based Activity Assay*



Figure 1:

Supplemental Table: Library completion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Library | nt varied | observed1 | functional2 | prob.3 |
| Anchor | 5 | 5,118 | 17 | 99.3 |
| Anchor\_exchange (P1\_gateway) | 5 | 5,069 | 97 | 99.3 |
| Gateway | 6 | 22,799 | 141 | 99.6 |
| L4 | 6 | 202 | 77 | 4.81 |
| P4 | 6 | 14,145 | 127 | 96.8 |
| P1\_exchange | 6 | 13,573 | 102 | 96.4 |
| P3 | 5 | 9,029 | 15 | >99.9 |