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

**Introduction:**

Riboswitches are regulatory regions of RNA that exert genetic control by coupling metabolite sensing with folding. These elements are located in the 5’ untranslated region of many bacterial genes, though also found in Archaea and few Eukarya (31128223), and function in the absence of proteins to regulate gene output in a cis-fashion (15173824). 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). These genetic elements are of great interest for synthetic biology or clinical applications as potential tools because of their ability to function as sensors that can produce a genetic output. To leverage these natural elements as tools, investigation must be done to understand underlying principles that govern the sequence landscape of important regions within the riboswitch.

While much work has been done in the past two decades to characterize riboswitches, the expression platform remains elusive. The unifying feature of this region between riboswitches is its ability to fold and exert allosteric control in response to aptamer occupancy, but that can be achieved in many ways and control gene expression at various levels (15173824). Current knowledge about riboswitch design comes from studies surveying natural riboswitches (25015992, 36150954) or evolving novel aptamers through techniques like SELEX that often modify existing riboswitch scaffolds (36617976, 28092358). Limitations to these approaches are three-fold. First, the focus is placed on the aptamer domain, usually with the goal of evolving a novel sensor for a chosen effector ligand. Second, many of these analyses produce answers without providing insight into the principals governing these sensors. And third, many of these aptamers are developed *in vitro,* potentially limiting their portability into other systems. This has created a lack of understanding both about overarching strategies to design nucleic acid-based sensors, and more specifically how to design and manipulate a regulatory domain.

The *pbuE* adenine responsive riboswitch from *B. subtilis* is a simple and small riboswitch, making it an ideal model system for the study of riboswitches as it is well characterized and folds with high fidelity (25573585, 16201765, 25550163, 16931335, 24590258). 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). 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. Although they are referred to as riboswitches, the conformation with the terminator helix is far more thermodynamically stable, making the system work more as a co-transcriptional “fuse” than a switch (25573585,19595806, 15173824). The formation of the stable OFF conformation is only accomplished when the 5’ 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 5’ end to invade into the intact P1 helix ().

Strand invasion is a prolific phenomenon in nucleic acid biology, relevant in various contexts from riboswitches to homologous recombination to CRISPR/Cas9 gene editing (37095744). For a system with a single stranded region invading into a duplex each migration step, that is the formation of a basepair for one duplex at the expense of the other duplex, takes between 10 and 20 µs (592403, 6264399, 37095744). During the invasion, the two strands will step forward and backward in a random walk process until the invading duplex is fully formed, ending the exchange process (37095744). A bias exists for the terminator helix over the aptamer domain because P(T) is inherently longer than P1, and P3 due to the toehold (P4) and lack of ssRNA (J3/1). It has also been shown that the stability of P4, the toehold, will determine the rate at which invasion proceeds (37095744, https://doi.org/10.1023/A:1023928811651, 19894722, 24019238). This bias reinforces the switch being more of a fuse and emphasizes the importance of ligand binding and understanding how the sequence identity of each strand in this regulatory element either allows or disallows the ability of the ON and OFF switch to form.

While techniques exist to evolve novel aptamers (), the ability to optimize and understand principles behind the regulatory domain remains a bottleneck in the ability to utilize riboswitches are tools. To that end, Drogalis and Batey completed a mutagenic analysis of regions of the expression platform of the *pbuE* riboswitch to glean information about how modifying the expression platform would impact the signal amplitude between the ON- and OFF-states of the switch (33259551).



The resulting sequence, schematized in Figure X, 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 and an AA mismatch was introduced in P3, proximal to L3, to promote strand invasion into the aptamer domain and destabilize the tertiary structure of the binding pocket. The mutagenic analysis, while yielding a high preforming variant did not aim elucidate the design principles behind this improved switching ability.

In this study, we aim to understand the design principles governing a transcriptional riboswitch’s ability to fold co-transcriptionally and attenuate gene expression (19595806). 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 with limited manipulation to key areas of the switch based on the parameters of the system.