UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Applying riboswitches for novel sensing and chemistry

Permalink
https://escholarship.org/uc/item/05n8k3nq

Author
Truong, Johnny

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Applying riboswitches for novel sensing and chemistry

By

Johnny Truong

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ming C. Hammond, Co-Chair
Professor Matthew B. Francis, Co-Chair
Professor David Savage
Professor Anne Baranger

Spring 2019

Abstract

Applying riboswitches for novel sensing and chemistry

by

Johnny Truong

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Ming C. Hammond, Chair

Riboswitches are cis-regulatory structured RNA elements capable of controlling expression of downstream genes by binding to small molecule ligands. These naturally evolved RNA elements possess remarkable affinity and selectivity for their small molecule ligands, high folding efficiencies, and thermostability for functioning in cellular environments. Due to these properties, a number of riboswitch-based technologies have emerged such as riboswitch reporters, aptazymes, and RNA-based fluorescent (RBF) biosensors which all have wide applications for detection, imaging, and regulatory circuits. While riboswitch reporters and aptazymes have been robustly studied to better understand how to improve their function, there are fewer studies that expand on RBF biosensor development. Here, novel approaches towards expanding the functional repertoire of RBF biosensors and systematically probing their properties are described.

First, we show that engineering circular permutations of the riboswitch aptamer domain yields functional biosensors for *S*-adenosyl-L-methionine (SAM), using the SAM-I riboswitch as our model. We reveal that this design can enhance fluorescence turn-on and ligand binding affinity compared to the non-permuted topology. Expanding upon these established design principles, novel biosensors for the ligand guanidine was developed. Two novel designs were added to our existing repertoire that generated functional RBF biosensors using the architecture of the guanidine-I riboswitch. A new base-pair mutation strategy was applied on these guanidine biosensors which, resulting in modest changes to biosensor activation speeds just from single base-pair mutations. Lastly, riboswitches were explored as potent scaffolds to generate a self-labeling ribozyme. Various natural or engineered riboswitches for the electrophilic ligand, SAM, were screened for reactivity with an analog, Hey-SAM, as a proxy to measure ribozyme activity. In collaboration with Agilent Labs, a high-throughput method was developed for probing and screening latent ribozyme activity using a microarray platform. The efforts and strategies put forth here use riboswitches outside their native context for applications in detection and catalysis further showcasing the broad utility of riboswitch-based tools.

TABLE OF CONTENTS

Acknowledgements

These past 5 years have been an intense personal and professional journey. I would not be where I am today without the support of so many others, with only just a fraction of people highlighted here.

First, I would like to thank Dr. Ming Hammond, for guidance and support as my research advisor. I have always appreciated your rigor and passion with regards to scientific research. It has helped me grow immensely as a scientist and a person. This journey would not have been possible without your mentorship and patience for my own development. I am so grateful for you to have given me this opportunity.

And also of course I could not do this without other members in the Hammond Lab. I'll always be happy to know we are not only colleagues but friends too. To the senior graduates: Cindy and Zach, you both taught me how to be "high-brow" in regards to research and my personal life and I will always remember you both fondly in those regards. Yichi, you are one of the most brilliant, technical scientists I have ever met, had the pleasure of knowing, and was mentored by. Among that, I will always appreciate our bubble tea adventures where we talked less about science and more about other interesting facets of life. To the fellow graduates: Andrew, Todd, and Rebekah; our time in the lab has been such a crazy rollercoaster with so many ups and downs but I'm grateful that we could all support each other throughout this experience. I don't think any other set of labmates would have made adjusting to Utah as easy as it has been. Andrew, while you're definitely the quietest of us all, you've always been vocal whenever I've needed any experimental help or feedback. No one else in the lab comes close to sharing my views on how transformative a music festival can be. Todd, I have always admired how you balance your easy-going personality with such a level-head throughout so many difficult situations. Also, I can't remember how many countless HOTS games we've played together. Rebekah, when we're not discussing science, we are definitely talking about all types of gossip that exist. I don't know anyone else I've commiserated and shared with more than you in this lab and I will always remember our graduate experience fondly. To current lab members: Wyatt, you have been a huge life-saver in our transition to Utah and I'm excited to hear about the great things you're going to do in medical school. Zhesen, while you're a huge grump in the lab, you're the first post-doc to join us here in Utah and it has been great learning from the most "outdoorsy" guy I know. Kylee and Scotty, you both are the future of the lab and while it's a huge task, I can't imagine two other capable individuals to depend on. Additionally, I'd like to thank all other lab alumni I have had the pleasure of knowing and working with: Debojit, Jongchan, Xavier, James, Lucy, Jian, Wanda, Cindy L., Bao. I wish you all the best in your current endeavors.

Beyond my research advisor, there have been so many scientific professionals that have played a huge role in my pursuit of a PhD. Thanks to Bo, Paige, and Joel at Agilent Technologies for being such wonderful collaborators. You three have been so supportive of the research and receptive to valuable discussions for furthering the project. Thanks to Dr. Anne Baranger, Dr. Matt Francis, Dr. David Savage, and Dr. Don Rio for serving on my qualifying exam committee and giving me valuable insights with my research project. Thanks to Dr. Carolyn Bertozzi, for teaching one of the best courses I have taken in my time as a student and being a huge inspiration to me as a scientific and personal role model.

Throughout this PhD, I have made so many amazing friends that made this graduate school experience unforgettable. To some of my closest cohort friends, Rachel, Helen, Jenna, Alec, John, Pratima, and Gloria: You all have seen me at some of worst and best points these past 5 years and I'm so grateful for creating so many memories with you all and others. To my roommates Steven and Andre: I had such a great time living together with you guys and could not imagine anyone else who puts up with my lifestyle. My days outside of Berkeley would be so dull and uninteresting without my SF family, Derek, Brian, Danny, Tammy, (and sometimes Tim). While I was initially an atypical friend for you all, you all stuck by me and gave me an opportunity to learn and grow from all of you. I don't know any other individuals who have accepted me as willingly and easily as you all did and have been as flexible with my personal circumstances. I sincerely hope we remain in one another's lives as much as we can and I can't thank you enough for all the support you've given me during graduate school.

Additionally, I could not get to where I am today without the support of family. I have always been appreciative of my humble upbringing as it continues to mold me into the man I am today. To my uncles, aunts, and cousins, you've all been a huge help with understanding my career choices so far and supporting me in them. To my parents: Dad, while we have a lot of different views and aren't as close, I will always appreciate the times you've surprised with me with how much you genuinely care. Mom, thank you for your continuous support throughout my career decisions although you personally don't understand why I make them. You've done so much for this family and one day I hope to be able to return that in kind. To my brother, Tommy, thank you for always trying your best. I know we haven't always gotten along but I am so happy to see you as a wonderful husband and father. To my sister, Melina, thanks for always checking-in and keeping me updated about what goes on in Georgia. I'm so proud of our relationship now as adults and I hope that it continues no matter where we both may be.

Finally, thank you to RJ, my first and maybe last boyfriend. It is crazy to think we met less than a year ago and under very bizarre circumstances. However, I have never been more excited than when I'm talking to you. You've taught me so much about loving someone unconditionally without sacrificing who I am as a person. I can't wait to start this next chapter of my life together with you.

CHAPTER 1

Introduction to riboswitch-based technologies

To date, the centrality of RNA has far exceeded its classical roles of coding messenger RNAs, processing tRNAs, and catalytic ribosomal RNAs. The rapid and ongoing discovery of new types of non-coding regulatory RNAs further showcases the functional diversity for an essential biomolecule in nature. With the wealth of biology uncovered with non-coding RNAs, there has also been vast development of engineered RNA tools in conjunction. The context for noncoding RNAs in biology can vary such as those that utilize ribonucleoprotein complexes including ribosomal RNAs, microRNAs, and clustered regularly interspaced short palindromic repeats (CRISPR) RNAs. Ribosomal RNAs are catalytic in nature and essential for protein translation while microRNAs and CRISPR RNAs can regulate gene expression through targeting RNA transcripts or dsDNA respectively (Cech and Steitz, 2014). Independently-functioning non-coding RNAs can include riboswitches and ribozymes which often exist in untranslated regions of mRNA transcripts. With new types of RNAs being uncovered at a rapid rate, there also has been a surge in expanding the roles of classic RNAs in parallel. Chemical modifications on classical RNA motifs have been shown to possess much biological significance. For instance, methyl-N6-adenosine (m^6A) on messenger RNAs has been discovered to be involved in gene regulatory processes crucial to development and disease in eukaryotes (Schaefer et al., 2017). Additionally, the structure determination of the human ribosomal 80s subunit has illuminated ribosomal RNA modifications that could have potential implications towards health and development as well (Natchiar et al., 2017). Akin to the discovery of histone and DNA modifications that yielded the era of epigenetics, these biologically relevant RNA modifications have given birth to the exciting field of epitranscriptomics.

Beyond these diverse functions found in RNA, their structural capabilities are also astounding when compared to their protein counterpart. While proteins can utilize up to a combination of 20+ naturally available amino acids with a variety of functional groups, there exists much less chemical diversity for nucleic acids. RNAs utilize four different nucleobases with the identical phosphodiester linkages and ribose-sugars. Despite the limited subset of monomers, RNAs can still adopt a number of complex three-dimensional structures through electrostatics, hydrogen bonding, and base-stacking. Interestingly, nucleic acids like RNA and DNA can both be utilized in similar, complex folds but the former possesses an additional hydroxyl group on 2'-C of the ribose sugar enabling additional interactions. These structural capabilities have been exploited towards the discovery and development of "RNA aptamers", nucleic acid sequences that can bind small molecules.

1.1 Nucleic Acid Aptamers for Detection

Aptamers are considered the nucleic acid equivalent to antibodies. Aptamers and antibodies both utilize complex biological scaffolds that confer their specificity for a particular ligand. Their respective global architectures are able to encapsulate their target through various intermolecular interactions. While antibodies are known for their characteristic "Y-shape" and ligand-binding Fab (fragment, antigen-binding) regions, RNA aptamers can possess diverse global structures . Due to this, aptamers can differ in their size amongst one another and tend to be a magnitude smaller in weight than antibodies. On average, aptamers possess a molecular weight of 5 – 15 kDa while antibodies can

typically range anywhere from 15 – 100 kDa (Groff et al., 2015). These considerations play a huge role when we think about developing aptamers to be as robust and effective as antibodies for detection purposes.

For decades, antibodies have been engineered and developed into research tools, diagnostics, and even therapeutics. Antibody-based assays such as immunoblotting and enzyme-linked immunosorbent Assay (ELISA) have become standard biochemical techniques. Despite the wide-spread adoption of antibodies as a detection platform, there still exists limitations. First, developing antibodies for a novel molecular target is a lengthy and resource-intensive procedure. Animal subjects are injected with the ligand of interest to elicit a natural immune response in which antibodies for the ligand are harvested. This process typically requires months for antibodies to be developed and the use of different animal subject creates batch-to-batch variation, limits production scale, and drives up costs (Lakhin et al., 2013; Baker, 2015). Second, antibodies need to be handled carefully as the require appropriate temperature and pH conditions during transport and storage. Third, development of antibodies for small molecular ligands (under 100 Daltons) can be challenging as their detection sensitivity and specificity can suffer (Groff et al., 2015). While antibodies are widely employed for diagnostic purposes, these reasons have been have propelled the field to seek better alternatives.

The advances in aptamer discovery positions them nicely to answer the call for antibody-alternatives. The generation of aptamers is highly cost-effective and can be done as quickly as several weeks through an *in vitro* technique known as “Systematic Evolution of Ligands for Exponential Enrichment (SELEX)” (Ellington and Szostak, 1990; Tuerk and Gold, 1990). With SELEX, a library of randomized oligonucleotides undergoes a selection pressure that enriches sequences capable of binding and their identities are uncovered by downstream DNA sequencing. Recent technological innovations in nucleotide synthesis and sequencing contributes to lower costs and the continued appeal of aptamer development (Stoltenburg et al., 2007). Additionally SELEX can be used to develop aptamers for ligands that are toxic to cells or that do not generate an immune response. Aptamers can also be stored and transported in solid form after lyophilization or precipitation which allows for more flexible handling. Aptamers can also exhibit higher selectivity for their cognate ligand over analogs compared to their antibody counterpart which is the case for the theophylline aptamer which discriminates its ligand over the close analog caffeine over 10^4 times (Zimmermann et al., 2000). Lastly, aptamers can also easily accommodate chemical modifications to the phosphate backbone or 5’ or 3’ termini which minimally affect affinity while increasing desired properties like resistance to degradation (Thiviyanathan and Gorenstein, 2012).

While aptamers generated *in vitro* have numerous potential advantages over antibody production, there are still challenges to address. Aptamers and antibodies are selected for their ability to bind the ligand of interest, but they rarely elicit strong conformational changes upon doing so. Large conformational changes have been useful in constructing probes or sensors as they can communicate the presence of a ligand through structural modulation. This property has been vital in FRET-based detection methods which is dependent on the distance of the fluorophore donor and acceptor (Marx, 2017). Additionally, *in vitro* selected aptamers are optimized to bind their ligand in controlled environments as opposed to the cellular setting where thermostability and

folding capabilities are an issue. To address these problems, we can turn to naturally-evolved RNA aptamers called “riboswitches” as a potential solution.

1.2 Riboswitches: Naturally-evolved RNA Aptamers

Overview

Riboswitches were first termed and described in 2002, when various researchers concurrently found evidence of allosteric gene regulation of mRNAs in the presence of ligands like thiamine pyrophosphate, vitamin B12, and flavin mononucleotide (FMN) (Breaker et al., 2002; Mironov et al., 2002; Winkler, 2002). Essentially, these RNAs could “switch” gene regulation “ON-or-OFF” based off a ligand-induced conformation change. Riboswitches are commonly located in the 5’ untranslated regions (5’-UTR) of messenger RNAs primarily in bacteria and are comprised of two critical domains: an aptamer domain and expression platform. In the presence of ligand, the aptamer domain exhibits a conformational switching that also affects the expression platform in a way to induce gene regulation. Expression platforms vary widely in sequence identity and regulatory mechanisms that control gene expression. The most prevalent mechanisms of expression platforms are transcriptional termination and translational initiation. Transcriptional termination occurs when the ligand creates a conformational change that destabilizes an anti-terminator structure and favors the formation of an intrinsic terminator hairpin which can stall the RNA polymerase from elongating the mRNA. This mechanism is common to riboswitches controlling genes for metabolites such as S-adenosyl-L-methionine (SAM) as its accumulation causes an OFF-state similar to a negative feedback loop (Wang and Breaker, 2008). Conversely, translation initiation occurs when the ligand causes a conformation change revealing a ribosome-binding site (Shine-Delgarno sequence in bacteria) and allows for the ribosome to bind and initiate protein translation triggering an ON-state. This is commonly observed with riboswitches pertaining to signaling molecules like cyclic di-GMP (Sudarsan et al., 2008).

Riboswitch Mining and Characterization

To date, there are over 40 distinct riboswitch classes that have been experimentally validated to bind a specific ligand and induce gene regulation. Riboswitch discovery has been greatly accelerated with the availability of genomic data and computational power of modern CPUs. One modern workflow for riboswitch mining (Stav et al., 2019) starts with taking whole bacterial genomes and isolating regions between identified/putative genes. These intergenic regions are then analyzed for open reading frames (ORFs) which are cross-referenced for homologous non-coding RNA sequences. Algorithms for structure prediction are then used to create a consensus model which displays the sequence context where these riboswitches can be found. At this point, these riboswitch candidates may have a predicted function if they are found upstream of a particular gene since in most cases, they interact with a ligand related to genes they control. For instance, the riboswitch for *S*-adenosylhomocysteine (SAH) was easily identified due to the identified genes downstream of the RNA motif being related to SAH metabolism (Weinberg et al., 2007). The most recent riboswitch discoveries have added nucleoside diphosphates and guanosine tetraphosphate

(Sherlock et al., 2018a, 2019) to the repertoire of riboswitch ligands and demonstrates the greater complexity of riboswitch discovery. While these RNA motifs were known for some time, they were so similar in secondary structure that were considered to be a subtype of the guanidine-binding riboswitch, ykkC-I (Battaglia and Ke, 2018). Upon further analysis, it was determined that this subtype encompassed several classes of riboswitches that were shown to bind to PRPP, ppGpp, and nucleoside diphosphates (Sherlock et al., 2018b, 2018c, 2019).

Once riboswitch candidates are identified, there is extensive characterization to identify their ligands. Most validation of riboswitches utilizes biochemical techniques such as in-line probing or reporter assays (Sudarsan et al., 2008). In-line probing can elucidate structural and affinity information as it relies on structure-dependent cleavage of the RNA in the presence of ligand. Regions of reduced/increase RNA cleavage is usually observed with and without the ligand present and identify regions that undergo a conformational change when the ligand is bound. Additionally, the riboswitch candidate is typically examined in a cellular context where the riboswitch is put upstream of a reporter gene such as β-galactosidase, GFP, or luciferase and is transformed into mutant cell types which exhibit different endogenous levels of the ligand. In this way, the activity of the reporter gene is dependent on the amount of the ligand present in the cell. Aside from validation, there are a number of characterization techniques that are used to elucidate global structural information for riboswitches such as x-ray crystallography, nuclear magnetic resonance (NMR), and small angle x-ray scattering (SAXS). Currently there are over 39 riboswitch structures that have been solved using these structure determination techniques with that number will only continue to grow with the discovery of new riboswitches. To date, there are over 24 RNA motifs that are presumed to be riboswitches but whose ligands have not yet been established which represents a wealth of new biology and structures to be discovered (Greenlee et al., 2018).

1.3 Riboswitch-based Sensors

The discovery of riboswitches has led to the utilization of their aptamer domains to create various molecular sensors. The modularity of RNA allows for a number of complex RNA moieties to be fused in such a way that their original function can be conducive to a detection method. This “plug-and-play’ strategy has been robustly demonstrated with the generation of tools including riboswitch reporters, aptazymes and fluorescent biosensors (Fig. 1.2).

Riboswitch Reporters

Riboswitches can be inserted upstream of a reporter gene such that the reporter’s activity/expression is linked to the amount of ligand that can interact with the riboswitch. These riboswitch-based reporters have been developed using various genes and outputs such as luciferases for luminescence, GFP for fluorescence , and LacZ for colorimetric measurements. Most riboswitch discoveries are validated using a LacZ reporter for their discernable signal changes that rely on enzymatic activity. A recent example was used to uncover the guanidine riboswitch where a LacZ reporter was used that to correlate higher activity upon increased cellular levels of guanidine

(Nelson et al., 2017). GFP riboswitch reporters for cyclic-di-GMP have also been used to screen and validate cyclic-di-GMP degrading enzymes or diguanylate cyclases among various bacterial candidates (Zhou et al., 2013). Additionally, luciferase reporters were instrumental in understanding rare riboswitches found in eukaryotes where the TPP riboswitch was found to regulate alternative splicing of introns (Li and Breaker, 2013).

Aptazymes

Riboswitch aptamer domains can also be fused with known ribozymes to generate a ligand-dependent ribozyme which has been conventionally termed "aptazyme". An engineered aptazyme was developed to detect cyclic-di-GMP through fusing the Vc2 cyclic-di-GMP riboswitch aptamer with the hammerhead ribozyme sequence through an optimized transducer stem. (Furukawa et al., 2014). Additionally, aptazymes can be useful for regulating gene expression upon a ligand input demonstrated by engineered purine aptazymes that were functional when transfected into HeLA cells.(Zhong et al., 2016). Recently, they have been used to create ligand-dependent Cas9-mediated genome editing by masking guide RNAs into an aptazyme until the ligand is present inducing cleavage and releasing a functional guide RNA (Tang et al., 2017). While this type of riboswitch-based sensor has mostly been synthetically engineered, there also exists a natural aptazyme in the case of glucosamine-6-phosphate (GlcN6P) activated ribozyme, glmS. The glmS ribozyme utilizes GlcN6P in a direct interaction which catalyzes the cleavage of the RNA.

RNA-based Fluorescent Biosensors (RBFs)

The conformational switching of riboswitch aptamer domains in the presence of a target ligand has also been exploited in the development of fluorescent biosensors. These biosensors are composed of a riboswitch aptamer that is fused to a fluorogenic RNA aptamer. In the presence of the target ligand, these sensors will induce a conformational change that activates the fluorogenic aptamer and giving off fluorescent signal (Fig. 1.3). These allosteric sensors have been shown to be modular and highly adaptable to sense a number of ligands (**Table 1.1**, **Fig. 1.4**). The first demonstrated sensor of this type used the aptamer domains for theophylline, adenosine triphosphate (ATP) and flavin mononucleotide (FMN) and fused them to the fluorophore-binding aptamer for malachite green (Stojanovic and Kolpashchikov, 2004). However, their use was limited to *in vitro* applications as malachite green is a DNA intercalator and cytotoxic to cells. More recent developments in fluorogenic aptamers for *in vivo* applications have led to the discovery of the Spinach aptamer which is able to bind the pro-fluorescent dye 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), an analog of the GFP chromophore (Paige et al., 2011). Using Spinach and its improved variants Spinach2 and Broccoli, numerous riboswitch-based biosensors have been developed for ligands such as ADP, ATP, SAM, cyclic di-AMP, cyclic di-GMP, bacterial cGAMP, and S-adenosyl-L-homocysteine ((Paige et al., 2012a; Kellenberger et al., 2013a, 2013b, 2015a, 2015c; Bose et al., 2016; Su et al., 2016a; Litke and Jaffrey, 2019).

1.4 Broadening the Utility of Riboswitch-based Tools

While many tools have been constructed using naturally evolved RNA aptamers, the subset of 40 riboswitch classes and ligands limits the chemical diversity for molecular detection and application of riboswitch-based tools. In order to circumvent these issues, there has been many efforts put forth towards developing synthetic riboswitches either completely *de novo* or by rational design and synthesizing modified riboswitch ligands to uncover new biology or applications.

Engineered / Artificial Riboswitches

Riboswitches also display some flexibility in their binding pocket and have been shown to accommodate structural analogs of their natural ligands when key nucleotides are mutated. For instance, the natural riboswitches for the ligands guanine, adenine, and 2'-deoxyguanosine all maintain a core secondary structure but a single C-to-U mutation can change the specificity of a guanine riboswitch to bind adenine and this same mutation in 2'-deoxyguanosine riboswitches shifts affinity towards 2′-deoxy-2,6-diaminopurine nucleoside (Mandal and Breaker, 2004; Kim et al., 2007). In another case, point mutations of RBF biosensors from GEMM-I cyclic di-GMP riboswitches shifted selectivity towards bacterial cGAMP which later led to the discovery of cGAMP riboswitches (Kellenberger et al., 2013). Our lab has been demonstrated the utility of this mutagenesis approach by creating new RBF biosensors. GEMM-II cyclic di-GMP riboswitches were altered with a G-to-A mutation, shifting its binding selectivity from cyclic di-GMP towards the mammalian signal 2',3'-cGAMP (Bose et al., 2016). Additionally, adenine riboswitches have been systematically mutated to selectively bind orthogonal ligands ammeline and azacytosine over their cognate ligand after 3 base-pair mutations (Dixon et al., 2010).

Concurrently, there have been established efforts to push *in vitro* SELEX to not only select for aptamers but for synthetic riboswitches that induce a conformational change when bound to a ligand. It has been shown that already evolved aptamers like theophylline and tetracycline can be successfully adapted to regulate gene expression in live cells similar to riboswitches despite not initially selecting for that activity (Suess et al., 2004; Hanson et al., 2005). However, more recent efforts have seen a variety of modified SELEX methods that incorporate riboswitch scaffolds to select for switching upon ligand binding. One method involves using conserved motifs from guanine and cyclic di-GMP riboswitches to scaffold SELEX in order to retain robust folding properties and switching when bound to the ligand (Porter et al., 2017). Additional selection methods have enabled *de novo* discovery of functional riboswitches such as those created for target ligands such as explosive compound, trinitrotoluene and the antibiotic, ciproflaxin (Groher et al., 2018; Harbaugh et al., 2018).

Unnatural Riboswitch Ligands

The other side of improving riboswitch-based tools does not always have to involve the RNA aptamer but can involve modifying riboswitch ligands or finding non-natural ligands that interact with them. Attachment of nitrobenzyl photocaging moieties

on ligands like theophylline and guanine demonstrated riboswitches could be used for optogenetic control (Walsh et al., 2014; Dhamodharan et al., 2018). Additionally, a panel of cobalamin-dye analogs with varying fluorescent properties have been synthesized which allow for multicolor imaging when appending a cobalamin riboswitch onto a RNA of interest (Braselmann et al., 2018). These strategies modify the cognate ligands in areas that are solvent-accessible and do not interact with the aptamer allowing for the native RNA riboswitch to still recognize these analogs In addition to modifying riboswitch ligands, recent efforts have explored screening small molecule libraries to see if new classes of molecules can interact with a native riboswitch. In one study, the preQ1 riboswitch was found to bind small molecules with a dibenzofuran core in a binding mode different than the native ligand but could still modulate an expression platform (Connelly et al., 2019).

1.5 Approaches for RNA Tagging

Another application for aptamers is for tagging RNA but this area has been relatively underexplored for riboswitch aptamers. Generally, there are two approaches to RNA-based tagging: 1) Non-covalent labeling and 2) Covalent labeling. Depending on the tagging method, the labeled RNA sequence can be studied *in vitro* with post-transcriptional labeling to profile RNA properties or *in vivo* with live cell imaging and dynamics.

Non-covalent approaches

One of the earliest methods utilizes the MS2 coat protein which binds the MS2 RNA hairpin motif and fusion of the MS2 coat protein with GFP allows for tracking of a messenger RNA (Bertrand et al., 1998). However, the MS2 system has several limitations such as the number of stem loops that can be added as it can affect routine cloning and balancing signal with protein fusion levels because non-specific aggregation of MS2-GFP fusion proteins can occur which obscures analyses like microscopy (Weil et al., 2010). Other non-covalent alternatives that have arisen involve appending fluorogenic RNA aptamers onto the target RNA sequence. Improvements from the Spinach system mentioned earlier has brought about aptamers such as Broccoli and Corn which are optimized for *in vivo* imaging of RNAs and the latter being photostable enough to quantitate RNA Pol III transcription (Filonov et al., 2014; Song et al., 2017). Other fluorogenic aptamers systems such as the Mango (Dolgosheina et al., 2014) and the DIR (Tan et al., 2017) aptamers have also been developed which bind different pro-fluorescent dyes which are not analogs of fluorescent protein chromophores. Additionally, the diverse spectral properties of these aptamer systems have been showcased through the development of an aptamer-based Forster resonance energy transfer (FRET) system utilizing the Spinach and Mango aptamers (Jepsen et al., 2018). Recently, riboswitch aptamers for cobalamin were used as a RNA tag as cobalamin acts as a quencher when in proximity to a fluorescent dye but when interacting with the riboswitch, this quenching is reduced and one observes turn-on fluorescence of the conjugated dye (Braselmann et al., 2018).

Covalent approaches

Most covalent approaches utilize biorthogonal click reactions in combination with synthetic, enzymatic or chemo-enzymatic methods (George and Srivatsan, 2017). The azide-alkyne cycloaddition and the alkene-tetrazine inverse electron demand Diels-Alder (IEDDA) are two examples that have been robustly demonstrated for RNA tagging (Paredes et al., 2011). Once these moieties are installed on the RNA sequence, they are able to react with their cognate partner efficiently and rapidly. By conventional solid-phase phosphoramidite chemistries for oligonucleotide synthesis, it is easy to site-specifically install click moieties on a small RNA sequence to react with its functional partner. Despite the hydrolytic instability of RNAs, they have been shown to react efficiently in copper azide-alkyne cycloadditions (CuAAC) with better stabilizing ligands that reduce spontaneous RNA cleavage (Hong et al., 2009). Another method to prevent copper-mediated degradation of RNA is to utilized strain-promoted azide-alkyne cycloadditions (SPAAC) using faster-reacting cyclooctynes (Singh et al., 2012). Recent efforts have also shown that DNA-templated synthesis better promotes the alkene-tetrazine IEDDA between complementary RNA fragments (Šečkutė et al., 2013).

For enzymatic and chemo-enzymatic methods, the incorporation of these reactive functional groups onto RNAs utilizes 1) an RNA-interacting enzyme and 2) analogs containing the desired click moieties. Enzymatic methods are distinguishable as they mainly pertain to tagging the RNA during transcription by an RNA polymerase and a NTP analog. T7 RNA polymerases have robustly demonstrated to be promiscuous enough to incorporate clickable UTP analogs. Both 5-ethynyluridine and 7-ethynyl-8-aza-7-deazaadenosine were shown to be non-specifically incorporated into nascent RNA transcripts and visualized by fluorescence microscopy in live cells (Jao and Salic, 2008; Zheng and Beal, 2016). Chemo-enzymatic approaches rely on natural or engineered RNA-processing enzymes. Various RNA methyltransferases naturally exist and have been demonstrated to label RNAs with click moieties by incorporating analogs of the methyl cofactor, SAM (Motorin et al., 2011; Tomkuvienė et al., 2012; Plotnikova et al., 2014). Another enzyme class, tRNA agmatidine synthetase (Tias) has also been shown to site-specifically label RNAs with azide or alkyne functional groups for subsequent imaging in mammalian cells (Li et al., 2015).

While there are a number of covalent approaches to label RNAs, only a few labeling approaches exist that use RNA machinery such as ribozymes for bond formation and cleavage. One method has been to do SELEX for a ribozyme that reacts with itself onto a nucleic acid sequence of interest. The earliest this has been demonstrated with is a self-labeling ribozyme that can attach with an iodoacetamide moiety conjugated with biotin (Wilson and Szostak, 1995). More recently, another RNA selection for self-labeling ribozymes yielded one construct which could react with fluorescein-iodoacetamide to fluorescently tag itself (Sharma et al., 2014). While no instances are known in nature, one self-labeling ribozyme was found to react with epoxide probes through a selection using a genomic RNA pool from organisms with a high percentage of non-coding RNAs (McDonald et al., 2014). In addition to ribozymes that act upon themselves, a polymerase ribozyme capable of elongating short, oligonucleotide primers was found to incorporate NTP analogs to generate fluorescently-labeled RNAs (Samanta et al., 2018). While using NTP analogs to tag

RNAs has largely been accomplished through polymerase enzymes, this is one of the first instances that polymerase ribozymes can utilize this strategy as well.

1.6 Outlook

In summary, RNA aptamers are powerful alternatives to established detection gold-standards such as antibodies and naturally-evolved aptamers like riboswitches are have been rapidly adapted into robust sensors. While many detection platforms exist, RBF biosensors are highly attractive due to the their modularity with synthetic or natural riboswitch aptamers and can be more accessible as they operate using fluorescence outputs which most laboratory instrumentation can detect. Uncovering insights towards the development of RNA-based sensors is highly desirable.

In this thesis, I first focus on expanding the utility and scope of RNA-based fluorescent biosensors by establishing new design strategies. In Chapter 2, a general design strategy was pioneered in which circular permutations of riboswitch aptamers were applied to make better performing biosensors for *S*-adenosyl-L-methionine (SAM). We demonstrated a new biosensor topology which could be generalizable to adapting other riboswitch scaffolds. In Chapter 3, we apply the engineering insights gained from Chapter 2 toward the development of functional guanidine biosensors using the ykkC riboswitch scaffold. Then, a novel mutational strategy using base-pair shuffling was applied to these sensors to understand how RBF biosensor kinetics are tuned, resulting in changes to response speed.

Additionally, there is little precedence of RNA-based tagging methods which utilize riboswitches. One study uses cobalamin riboswitches as a non-covalent RNA tag for multicolor imaging as the riboswitch can tightly bind to a panel of cobalamin-dye analogs but this is the only instance to our knowledge (Braselmann et al., 2018). With collaborators at Agilent Technologies, we investigated a novel covalent RNA tagging strategy using SAM riboswitches in Chapter 4.

1.7 Figures

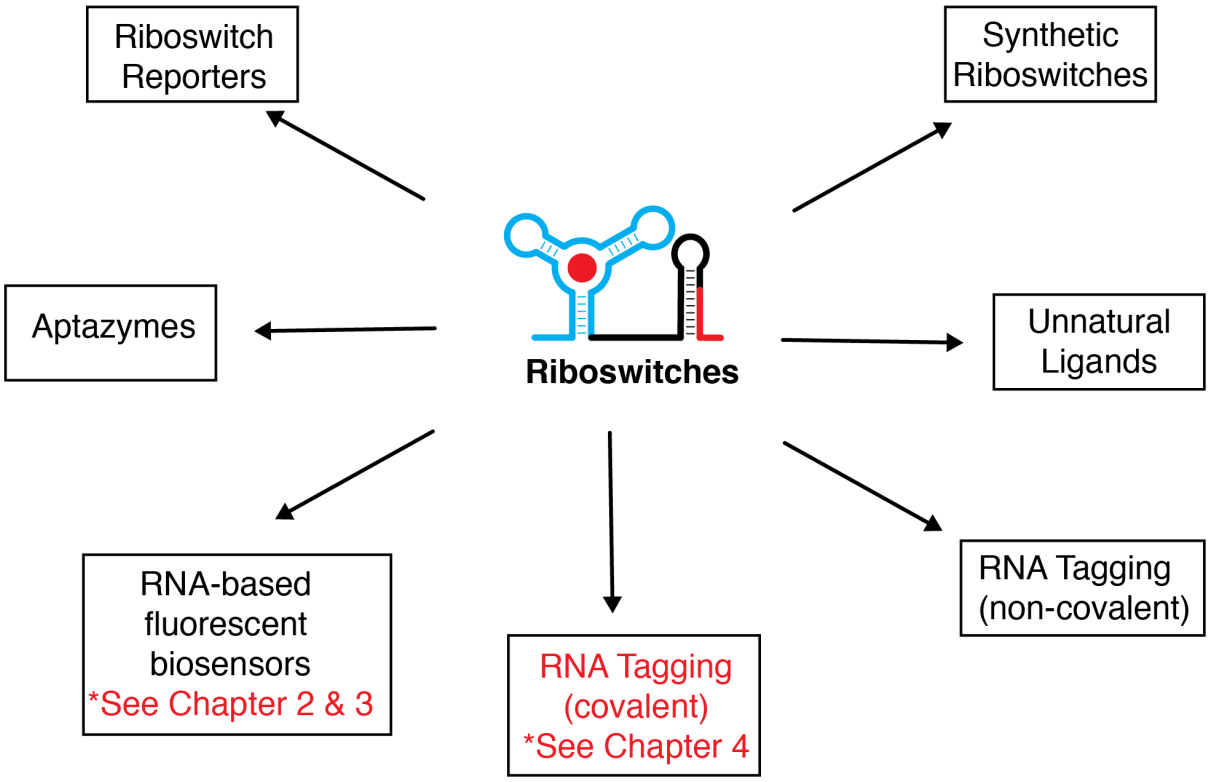


Figure 1.1. Demonstrated applications of riboswitches outside their native context A variety of riboswitch applications used for either tool development, detection and imaging. Red text highlights areas that have not been established or investigated until this thesis.

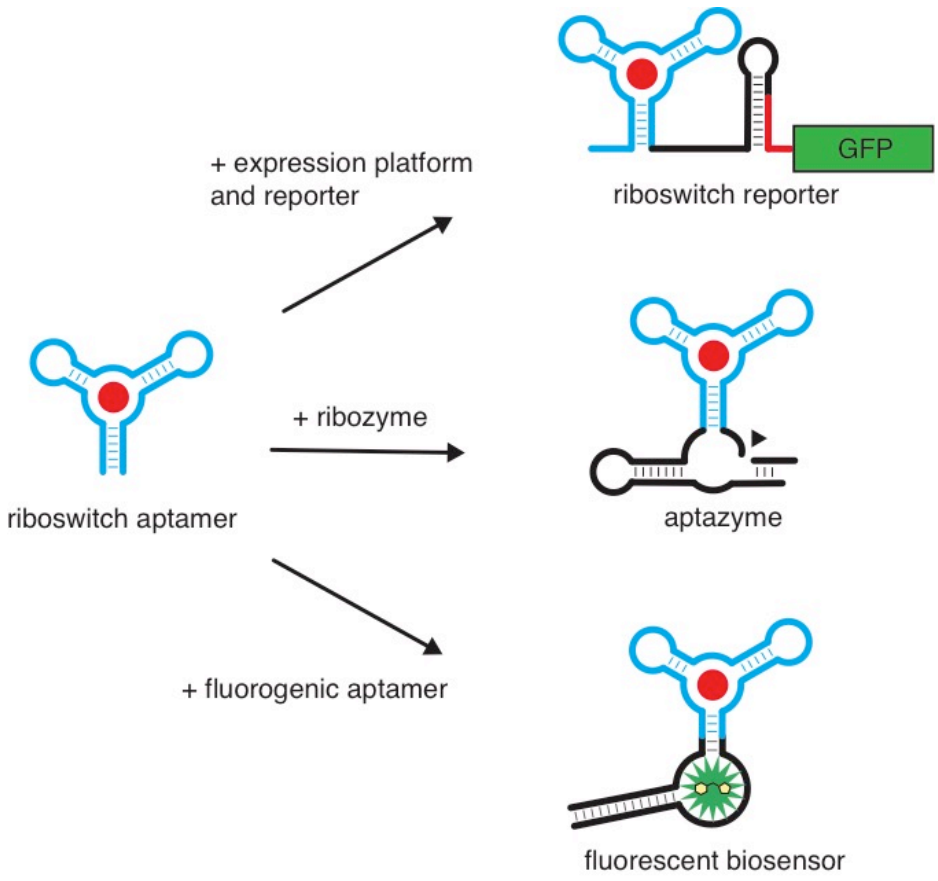


Figure 1.2. Modular approach to construct riboswitch-based sensors
Riboswitch aptamers can be fused to other functional RNA domains like an expression platform, ribozyme, or fluorogenic dye-binding aptamer to generate riboswitch reporters, aptazymes, and fluorescent biosensors, respectively.

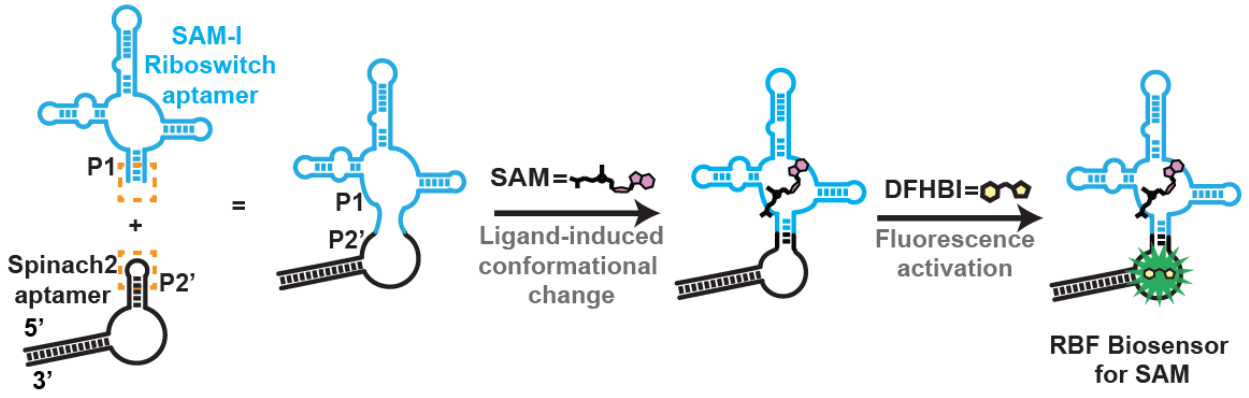


Figure 1.3. Construction of RNA-based fluorescent biosensors using fluorogenic aptamers

Fusion of riboswitch aptamer domain (blue) to a dye-binding fluorogenic aptamer (black), creates a RNA-based fluorescent biosensor. The ligand S-adenosyl-L-methionine (pink) binds to its native riboswitch aptamer domain which allows the fluorogenic aptamer to bind the pro-fluorescent dye (yellow) and activate its fluorescence (green) in a RNA-dye complex.

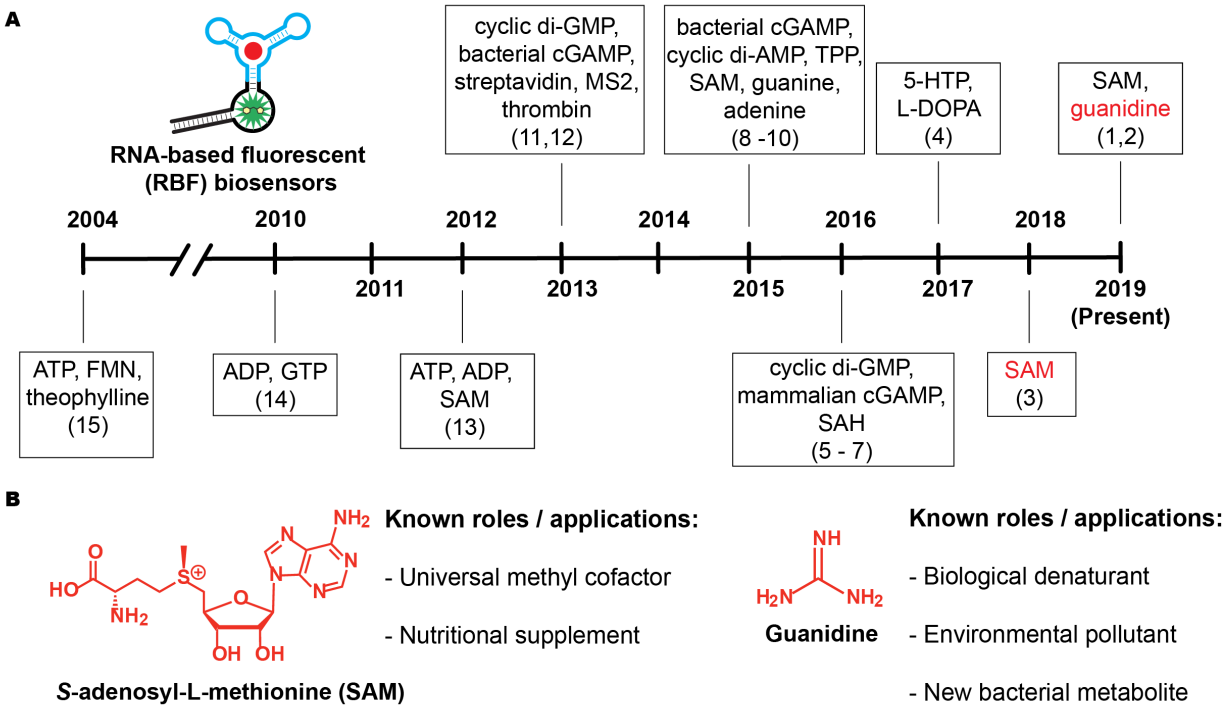


Figure 1.4. History and development of RBF biosensors for a variety of ligands (a) Timeline of RNA-based fluorescent biosensor development. Numbers correspond to the organization of Table 1.1 with full references. Red text highlights ligands that RBF biosensors were developed in this thesis. (b) Background information and structures for the ligands S-adenosyl-L-methionine and guanidine.

Table 1.1. **Literature precedent of RNA-based fluorescent biosensors and the metabolites they detect**. Publications are sorted by most recent first. Red text indicates content that will appear or originates from this thesis.

#	Biosensor Type	Fluorogenic Aptamer	Ligand(s) Detected	Reference / Year
1	Allosteric	Spinach2	Guanidine	See Chapter 3
2	Allosteric (Circular)	Broccoli	S-adenosyl-L-methionine (SAM)	(Litke and Jaffrey, 2019)
3	Allosteric	Spinach2	SAM	(Truong et al., 2018) See Chapter 2
4	Allosteric	Broccoli	5-HTP, L-DOPA	(Porter et al., 2017)
5	Allosteric	Spinach2	Cyclic di-GMP	(Wang et al., 2016)
6	Allosteric	Spinach2	Mammalian cGAMP	(Bose et al., 2016)
7	Allosteric	cpSpinach2	S-adenosyl-L-homocysteine (SAH)	(Su et al., 2016)
8	Spinach riboswitch	Spinach	Thiamine 5'- pyrophosphate SAM, guanine, adenine	(You et al., 2015)
9	Allosteric	Spinach	Cyclic di-AMP	(Kellenberger et al., 2015b)
10	Allosteric	Spinach	Bacterial cGAMP	(Kellenberger et al., 2015a)
11	Allosteric	Spinach	Streptavidin, thrombin, MS2 coat protein	(Song et al., 2013)
12	Allosteric	Spinach	Cyclic di-GMP and bacterial cGAMP	(Kellenberger et al., 2013)
13	Allosteric	Spinach	ATP, ADP, SAM	(Paige et al., 2012)
14	Allosteric	BFR	ADP, GTP	(Furutani et al., 2010)
15	Allosteric	Malachite Green	ATP, theophylline, FMN	(Stojanovic and Kolpashchikov, 2004)

1.1 References

Baker, M. Reproducibility Crisis: Blame It on the Antibodies. *Nature* **2015**, *521* (7552), 274–276.

Battaglia, R. A.; Ke, A. Guanidine-Sensing Riboswitches: How Do They Work and What Do They Regulate? *Wiley Interdiscip. Rev. RNA* **2018**, *9* (5), e1482.

Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M. Localization of ASH1 MRNA Particles in Living Yeast. *Mol. Cell* **1998**, *2* (4), 437–445.

Bose, D.; Su, Y.; Marcus, A.; Raulet, D. H.; Hammond, M. C. An RNA-Based Fluorescent Biosensor for High-Throughput Analysis of the CGAS-CGAMP-STING Pathway. *Cell Chem. Biol.* **2016**, *23* (12), 1539–1549.

Boussebayle, A.; Torka, D.; Ollivaud, S.; Braun, J.; Bofill-Bosch, C.; Dombrowski, M.; Groher, F.; Hamacher, K.; Suess, B. Next-Level Riboswitch Development—Implementation of Capture-SELEX Facilitates Identification of a New Synthetic Riboswitch. *Nucleic Acids Res.* **2019**.

Braselmann, E.; Wierzba, A. J.; Polaski, J. T.; Chromiński, M.; Holmes, Z. E.; Hung, S.-T.; Batan, D.; Wheeler, J. R.; Parker, R.; Jimenez, R.; et al. A Multicolor Riboswitch-Based Platform for Imaging of RNA in Live Mammalian Cells. *Nat. Chem. Biol.* **2018**, *14* (10), 964–971.

Breaker, R. R.; Winkler, W.; Nahvi, A. Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression. *Nature* **2002**, *419* (6910), 952–956.

Cech, T. R.; Steitz, J. A. The Noncoding RNA Revolution-Trashing Old Rules to Forge New Ones. *Cell* **2014**, *157* (1), 77–94.

Connelly, C. M.; Numata, T.; Boer, R. E.; Moon, M. H.; Sinniah, R. S.; Barchi, J. J.; Ferré-D'Amaré, A. R.; Schneekloth, J. S. Synthetic Ligands for PreQ1 Riboswitches Provide Structural and Mechanistic Insights into Targeting RNA Tertiary Structure. *Nat. Commun.* **2019**, *10* (1), 1501.

Dhamodharan, V.; Nomura, Y.; Dwidar, M.; Yokobayashi, Y. Optochemical Control of Gene Expression by Photocaged Guanine and Riboswitches. *Chem. Commun.* **2018**, *54* (48), 6181–6183.

Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chem. Biol.* **2014**, *9* (10), 2412–2420.

Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R. Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *J. Am. Chem. Soc.* **2014**, *136* (46), 16299–16308.

Furukawa, K.; Gu, H.; Breaker, R. R. In Vitro Selection of Allosteric Ribozymes That Sense the Bacterial Second Messenger C-Di-GMP; Humana Press, Totowa, NJ, 2014; pp 209–220.

Furutani, C.; Shinomiya, K.; Aoyama, Y.; Yamada, K.; Sando, S. Modular Blue Fluorescent RNA Sensors for Label-Free Detection of Target Molecules. *Mol Biosyst* **2010**, *6* (9), 1569–1571.George, J. T.; Srivatsan, S. G. Posttranscriptional Chemical Labeling of RNA by Using Bioorthogonal Chemistry. *Methods* **2017**, *120*, 28–38.

George, J. T.; Srivatsan, S. G. Posttranscriptional Chemical Labeling of RNA by Using Bioorthogonal Chemistry. *Methods* **2017**, *120*, 28–38.

Goodson, M. S.; Bennett, A. C.; Jennewine, B. R.; Briskin, E.; Harbaugh, S. V.; Kelley-Loughnane, N. Amplifying Riboswitch Signal Output Using Cellular Wiring. *ACS Synth. Biol.* **2017**, *6* (8), 1440–1444.

Greenlee, E. B.; Stav, S.; Atilho, R. M.; Brewer, K. I.; Harris, K. A.; Malkowski, S. N.; Mirihana Arachchilage, G.; Perkins, K. R.; Sherlock, M. E.; Breaker, R. R. Challenges of Ligand Identification for the Second Wave of Orphan Riboswitch Candidates. *RNA Biol.* **2018**, *15* (3), 377–390.

Groff, K.; Brown, J.; Clippinger, A. J. Modern Affinity Reagents: Recombinant Antibodies and Aptamers. *Biotechnol. Adv.* **2015**, *33* (8), 1787–1798.

Groher, F.; Bofill-Bosch, C.; Schneider, C.; Braun, J.; Jager, S.; Geißler, K.; Hamacher, K.; Suess, B. Riboswitching with Ciprofloxacin-Development and Characterization of a Novel RNA Regulator. *Nucleic Acids Res.* **2018**, *46* (4), 2121–2132.

Hanson, S.; Bauer, G.; Fink, B.; Suess, B. Molecular Analysis of a Synthetic Tetracycline-Binding Riboswitch. *RNA* **2005**, *11* (4), 503–511.

Harbaugh, S. V; Goodson, M. S.; Dillon, K.; Zabarnick, S.; Kelley-Loughnane, N. Riboswitch-Based Reversible Dual Color Sensor. **2017**, 1–24.

Harbaugh, S. V.; Martin, J. A.; Weinstein, J.; Ingram, G.; Kelley-Loughnane, N. Screening and Selection of Artificial Riboswitches. *Methods* **2018**, *143*, 77–89.

Jang, S.; Jang, S.; Noh, M. H.; Lim, H. G.; Jung, G. Y. Novel Hybrid Input Part Using Riboswitch and Transcriptional Repressor for Signal Inverting Amplifier. *ACS Synth. Biol.* **2018**, *7* (9), 2199–2204.

Jepsen, M. D. E.; Sparvath, S. M.; Nielsen, T. B.; Langvad, A. H.; Grossi, G.; Gothelf, K. V; Andersen, E. S. Development of a Genetically Encodable FRET System Using Fluorescent RNA Aptamers. *Nat. Commun.* **2018**, *9* (1), 18.

Kellenberger, C. A.; Wilson, S. C.; Hickey, S. F.; Gonzalez, T. L.; Su, Y.; Hallberg, Z. F.; Brewer, T. F.; Iavarone, A. T.; Carlson, H. K.; Hsieh, Y.-F.; et al. GEMM-I Riboswitches from Geobacter Sense the Bacterial Second Messenger Cyclic AMP-GMP. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (17), 5383–5388.

Kellenberger, C. A.; Chen, C.; Whiteley, A. T.; Portnoy, D. A.; Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic Di-AMP. *J. Am. Chem. Soc.* **2015**, *137* (20), 6432–6435.

Kellenberger, C. A.; Chen, C.; Whiteley, A. T.; Portnoy, D. A.; Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic Di-AMP. *J. Am. Chem. Soc.* **2015**, *137* (20), 6432–6435.

Kellenberger, C. A.; Wilson, S. C.; Sales-Lee, J.; Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic Di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc.* **2013**, *135* (13), 4906–4909.

Kim, J. N.; Roth, A.; Breaker, R. R. Guanine Riboswitch Variants from Mesoplasma Florum Selectively Recognize 2'-Deoxyguanosine. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (41), 16092–16097.

Lakhin, A. V.; Tarantul, V. Z.; Gening, L. V. Aptamers: Problems, Solutions and Prospects. *Acta Naturae* **2013**, *5* (4), 34.

Li, S.; Breaker, R. R. Eukaryotic TPP Riboswitch Regulation of Alternative Splicing Involving Long-Distance Base Pairing. *Nucleic Acids Res.* **2013**, *41* (5), 3022–3031.

Litke, J. L.; Jaffrey, S. R. Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts. *Nat. Biotechnol.* **2019**.

Lotz, T. S.; Suess, B. Small-Molecule-Binding Riboswitches. *Microbiol. Spectr.* **2018**, *6* (4).Mandal, M.; Breaker, R. R. Adenine Riboswitches and Gene Activation by Disruption of a Transcription Terminator. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 29–35.

Mandal, M.; Breaker, R. R. Adenine Riboswitches and Gene Activation by Disruption of a Transcription Terminator. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 29–35.

Marx, V. Probes: FRET Sensor Design and Optimization. *Nat. Methods* **2017**, *14* (10), 949–953.

McDonald, R. I.; Guilinger, J. P.; Mukherji, S.; Curtis, E. a; Lee, W. I.; Liu, D. R. Electrophilic Activity-Based RNA Probes Reveal a Self-Alkylating RNA for RNA Labeling. *Nat. Chem. Biol.* **2014**, *10*, 1049–1054.

Mironov, A. S.; Gusarov, I.; Rafikov, R.; Lopez, L. E.; Shatalin, K.; Kreneva, R. A.; Perumov, D. A.; Nudler, E. Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **2002**, *111* (5), 747–756.

Morse, D. P.; Nevins, C. E.; Aggrey-Fynn, J.; Bravo, R. J.; Pfaeffle, H. O. I.; Laney, J. E. Sensitive and Specific Detection of Ligands Using Engineered Riboswitches. *J. Biotechnol.* **2018**, *272–273*, 22–32.

Paige, J. S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S. R. Fluorescence Imaging of Cellular Metabolites with RNA. *Science (80-.).* **2012**, *335* (6073), 1194–1194.

Porter, E. B.; Polaski, J. T.; Morck, M. M.; Batey, R. T. Recurrent RNA Motifs as Scaffolds for Genetically Encodable Small-Molecule Biosensors. *Nat. Chem. Biol.* **2017**, *13* (3), 295–301.

Samanta, B.; Horning, D. P.; Joyce, G. F. 3′-End Labeling of Nucleic Acids by a Polymerase Ribozyme. *Nucleic Acids Res.* **2018**, 1–6.

Samanta, B.; Horning, D. P.; Joyce, G. F. 3'-End Labeling of Nucleic Acids by a Polymerase Ribozyme. *Nucleic Acids Res.* **2018**, *46* (17), e103.Schaefer, M.; Kapoor, U.; Jantsch, M. F. Understanding RNA Modifications: The Promises and Technological Bottlenecks of the "Epitranscriptome". *Open Biol.* **2017**, *7* (5).

Sharma, A. K.; Plant, J. J.; Rangel, A. E.; Meek, K. N.; Anamisis, A. J.; Hollien, J.; Heemstra, J. M. Fluorescent RNA Labeling Using Self-Alkylating Ribozymes. *ACS Chem. Biol.* **2014**, 1680–1684.

Sherlock, M. E.; Sadeeshkumar, H.; Breaker, R. R. Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. *Biochemistry* **2019**, *58* (5), 401–410.

Sherlock, M. E.; Sudarsan, N.; Breaker, R. R. Riboswitches for the Alarmone PpGpp Expand the Collection of RNA-Based Signaling Systems. *Proc. Natl. Acad. Sci.* **2018**, *115* (23), 6052–6057.

Song, W.; Filonov, G. S.; Kim, H.; Hirsch, M.; Li, X.; Moon, J. D.; Jaffrey, S. R. Imaging RNA Polymerase III Transcription Using a Photostable RNA–Fluorophore Complex. *Nat. Chem. Biol.* **2017**, *13* (11), 1187–1194.

Song, W.; Strack, R. L.; Jaffrey, S. R. Imaging Bacterial Protein Expression Using Genetically Encoded RNA Sensors. *Nat. Methods* **2013**, *10* (9), 873–875.

Stav, S.; Atilho, R. M.; Mirihana Arachchilage, G.; Nguyen, G.; Higgs, G.; Breaker, R. R. Genome-Wide Discovery of Structured Noncoding RNAs in Bacteria. *BMC Microbiol.* **2019**, *19* (1), 66.

Stojanovic, M. N.; Kolpashchikov, D. M. Modular Aptameric Sensors. *J. Am. Chem. Soc.* **2004**, *126* (30), 9266–9270.

Su, Y.; Hickey, S. F.; Keyser, S. G. L.; Hammond, M. C. *In Vitro* and *In Vivo* Enzyme Activity Screening via RNA-Based Fluorescent Biosensors for *S* -Adenosyl- l -Homocysteine (SAH). *J. Am. Chem. Soc.* **2016**, *138* (22), 7040–7047.

Sudarsan, N.; Lee, E. R.; Weinberg, Z.; Moy, R. H.; Kim, J. N.; Link, K. H.; Breaker, R. R. Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science* **2008**, *321* (5887), 411–413.

Suess, B.; Fink, B.; Berens, C.; Stentz, R.; Hillen, W. A Theophylline Responsive Riboswitch Based on Helix Slipping Controls Gene Expression in Vivo. *Nucleic Acids Res.* **2004**, *32* (4), 1610–1614.

Tan, X.; Constantin, T. P.; Sloane, K. L.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A. Fluoromodules Consisting of a Promiscuous RNA Aptamer and Red or Blue Fluorogenic Cyanine Dyes: Selection, Characterization, and Bioimaging. *J. Am. Chem. Soc.* **2017**, *139* (26), 9001–9009.

Tang, W.; Hu, J. H.; Liu, D. R. Aptazyme-Embedded Guide RNAs Enable Ligand-Responsive Genome Editing and Transcriptional Activation. *Nat. Commun.* **2017**, *8*, 15939.

Thiviyanathan, V.; Gorenstein, D. G. Aptamers and the Next Generation of Diagnostic Reagents. *Proteomics. Clin. Appl.* **2012**, *6* (0), 563.

Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249* (4968), 505–510.Walsh, S.; Gardner, L.; Deiters, A.; Williams, G. J. Intracellular Light-Activation of Riboswitch Activity. *Chembiochem* **2014**, *15* (9), 1346–1351.

Weil, T. T.; Parton, R. M.; Davis, I. Making the Message Clear: Visualizing MRNA Localization. *Trends Cell Biol.* **2010**, *20* (7), 380–390.

Weinberg, Z.; Barrick, J. E.; Yao, Z.; Roth, A.; Kim, J. N.; Gore, J.; Wang, J. X.; Lee, E. R.; Block, K. F.; Sudarsan, N.; et al. Identification of 22 Candidate Structured RNAs in Bacteria Using the CMfinder Comparative Genomics Pipeline. *Nucleic Acids Res.* **2007**, *35* (14), 4809–4819.

Wilson, C.; Szostak, J. W. In Vitro Evolution of a Self-Alkylatlng Ribozyme. *Nature* **1995**, *374* (6525), 777–782.

Winkler, W. C. An MRNA Structure That Controls Gene Expression by Binding FMN. *Proc. Natl. Acad. Sci.* **2002**, *99* (9), 15908–15913.

Xiu, Y.; Jang, S.; Jones, J. A.; Zill, N. A.; Linhardt, R. J.; Yuan, Q.; Jung, G. Y.; Koffas, M. A. G. Naringenin-Responsive Riboswitch-Based Fluorescent Biosensor Module for Escherichia Coli Co-Cultures. *Biotechnol. Bioeng.* **2017**, *114* (10), 2235–2244.

You, M.; Litke, J. L.; Jaffrey, S. R. Imaging Metabolite Dynamics in Living Cells Using a Spinach-Based Riboswitch. *Proc. Natl. Acad. Sci.* **2015**, *112* (21), E2756–E2765.

Zhong, G.; Wang, H.; Bailey, C. C.; Gao, G.; Farzan, M. Rational Design of Aptazyme Riboswitches for Efficient Control of Gene Expression in Mammalian Cells. *Elife* **2016**, *5*.

Zhou, H.; Zheng, C.; Su, J.; Chen, B.; Fu, Y.; Xie, Y.; Tang, Q.; Chou, S.-H.; He, J. Characterization of a Natural Triple-Tandem c-Di-GMP Riboswitch and Application of the Riboswitch-Based Dual-Fluorescence Reporter. *Sci. Rep.* **2016**, *6* (1), 20871.

Zimmermann, G. R.; Wick, C. L.; Shields, T. P.; Jenison, R. D.; Pardi, A. Molecular Interactions and Metal Binding in the Theophylline-Binding Core of an RNA Aptamer. *RNA* **2000**, *6* (5), 659–667.

CHAPTER 2

Accessing RNA-based fluorescent biosensors for S-adenosyl-L-methionine through novel topologies

Portions of this work were published in the following scientific journal:

Truong, J., Hsieh, Y. F., Truong, L., Jia, G. & Hammond, M. C. Designing fluorescent biosensors using circular permutations of riboswitches. *Methods* **143,** 102–109 (2018).

2.1 Abstract

RNA-based fluorescent (RBF) biosensors have been applied to detect a variety of metabolites in vitro and in live cells. They are designed by combining the ligand sensing domain of natural riboswitches with in vitro selected fluorogenic aptamers. Different biosensor topologies have been developed to accommodate the diversity of riboswitch structures. Here we show that circular permutation of the riboswitch ligand sensing domain also gives functional biosensors, using the SAM-I riboswitch as our model. We reveal that this design can enhance fluorescence turn-on and ligand binding affinity compared to the non-permuted topology.

2.2 Introduction

Riboswitches have been applied for classic synthetic biology applications including inducible gene regulation and metabolic engineering (Chappell et al., 2015; Villa et al., 2017). More recently, detailed structural information of riboswitch secondary and tertiary structure has enabled their combination with in vitro selected fluorogenic aptamers to generate RNA-based fluorescent (RBF) biosensors for metabolite imaging and detection (Hallberg et al., 2017) RBF biosensors are designed so that binding of a target ligand induces the riboswitch conformational change that allows a dye to bind and turn on fluorescence. In this way, the biosensor generates a fluorescent signal only in the presence of a target metabolite. To date, RBF biosensors have been applied for high-throughput screening of enzyme activity in vitro and in vivo, imaging metabolites and signals in Gram positive and Gram negative bacteria, and anaerobic imaging of bacterial signals. (Hallberg et al., 2017)

Different biosensor topologies have been explored to broaden the rational design strategy to accommodate the diversity of riboswitch and other aptamer folds. The first riboswitch-based biosensors utilized “transducer modules” consisting of randomized stem sequences to connect to the dye-binding aptamer called Spinach, which binds the profluorescent dye 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) ((Paige et al., 2012b) One improvement our group made to biosensor design was to use the natural P1 stem from native riboswitches, which simplified the transducer module design and permitted libraries of riboswitch aptamers to be screened for finding the optimal biosensor (Kellenberger et al., 2013a, 2015c; Wang et al., 2016)). The conventional topology (RS-Sp, **Fig. 2.1A**) thus uses a P1-P2’ stem architecture for the transducer module, where P1 refers to the first pairing stem of the ligand-sensing riboswitch and P2’ stem refers to the second pairing stem of the dye-binding aptamer, which could be either Spinach or an improved fluorogenic aptamer variant, Spinach2 (Strack et al., 2013)

The conventional riboswitch-Spinach fusion design is effective for many riboswitch aptamers, but is not suitable when the terminal ends of the ligand-sensing domain do not form a P1 stem. For example, the S-adenosyl-L-homocysteine (SAH) riboswitch contains a terminal pseudoknot that places the 5’ and 3’ ends far apart. To address this challenge, our group constructed a biosensor by making a circular permutation of the Spinach2 aptamer and fusing it to the P2 stem of the SAH riboswitch (RS-cpSp, **Fig. 2.1B**) (Su et al., 2016b). An alternate strategy is to incorporate Spinach

to mimic the expression platforms found in native riboswitches. The ligand-free riboswitch aptamer interferes with the Spinach aptamer folding to block DFBHI binding, but target ligand binding restores the dye-binding aptamer (**Fig. 2.1C**) (You et al., 2015b).

Here we showcase another strategy to access new biosensor designs that uses circular permutations of a riboswitch aptamer (cpRS-Sp, **Fig 2.1D**). To our knowledge, there are no natural examples of circularly permutated riboswitches. However, circularly-permuted hammerhead ribozyme variants are found in diverse genomes (Hammann et al., 2012) and some pre-tRNAs form circularized intermediates before maturation (Soma et al., 2007). To investigate these new biosensor topologies, the SAM-I riboswitch was chosen as a model 4-way junction RNA fold to explore novel permuted architectures. This riboswitch class has been extensively characterized with structural, biochemical, and biophysical studies (Winkler et al., 2003; Montange and Batey, 2006; Heppell et al., 2011), which suggested that a circularly-permuted SAM-I riboswitch could be fused with a fluorogenic aptamer via different pairing stems. These constructs were generated, screened, and characterized as new biosensors utilizing circularly-permuted riboswitch topologies, then were compared to traditional biosensor architectures.

2.3 Results

Screen of different biosensor topologies

The SAM-I riboswitch has a 4-way junction architecture. Biochemical and x-ray crystallographic information have shown that the pairing stem element P2 is involved in formation of a pseudoknot that is important for SAM binding (Montange and Batey, 2006). Thus, modifications at the P2 stem were excluded from our biosensor design, but the other pairing stems P1, P3, and P4 became candidates for the transducer stem. Biosensors that fuse Spinach2 to the riboswitch P1 stem are the conventional RS-Sp design. In contrast, biosensors that fuse Spinach2 to the riboswitch P3 or P4 stems would represent two new cpRS-Sp designs, which require the P1 stem to be closed by a loop sequence (GCAA, **Fig. 2A**). In all cases, the riboswitch or circularly permuted riboswitch was fused to the P2' stem of Spinach2, thus the nomenclature for the different topologies is P1-P2', P3-P2' or P4-P2'. A small library of 32 biosensor candidates were designed based on these criteria: 1) Four SAM-I riboswitches that have been previously characterized were chosen, (Winkler et al., 2003; Montange and Batey, 2006; Sudarsan et al., 2006) as it has been shown that sampling riboswitches from diverse phylogeny can generate highly fluorescent and well-folded RNA biosensors (Wang et al., 2016) 2) For each riboswitch sequence, the three biosensor topologies described above were designed; 3) Based on empirically derived rules for length of the transducer stem, (Paige et al., 2012) each P1-P2' construct had two possible stem lengths and each P3-P2' and P4-P2' had three possible stem lengths. The 32 biosensors were ordered as commercial DNA oligonucleotides, amplified by PCR, synthesized by in vitro transcription, and tested for fluorescence response to ligand, all in a 96-well format. In this initial high-throughput screen, 12 candidate biosensors showed a response to SAM with greater than 1.5x fluorescence activation (**Fig. 2.2B**). Interestingly, 8 hits were cpRS-Sp P4-P2' designs and the remaining hits were the conventional RS-Sp P1-P2' design. None of the P3-P2'

constructs were functional as biosensors, even though we separately verified that both P3 and P4 permutants of the Bs riboswitch aptamer retain good binding affinity to SAM (**Fig. 2.3**).

Further in vitro analysis of the 12 hits was performed at physiologically relevant conditions (37 °C and 3 mM $MgCl_2$) to predict their performance in bacteria (**Fig. 2.4A**). Three initial hits of the CpRS-Sp P4-P2' design with varying stem lengths derived from Polaribacter irgensii SAM-I did not give fluorescence turn-on in response to SAM in this secondary screen until higher ligand concentrations (200 μM), which is consistent with their observed poorer affinity in the initial screen (**Fig. 2.4B**). The remaining biosensors still retained activity with similar fold activations.

Selectivity and sensitivity of cpRS-Sp biosensors for SAM

To compare cpRS-Sp and RS-Sp biosensor designs, constructs Bc 4-5 and Bc 1-5 were chosen as representative models. Both biosensors were derived from the same parent SAM-I riboswitch from Bacillus clausii and use the same base pair length in the transducer (**Fig 2.5A**). We tested their selectivity for SAM versus S-adenosyl-L-homocysteine (SAH), which differs in one methyl group from the natural ligand. Previous studies have shown that SAM-I riboswitches are 550-fold selective for SAM over SAH (Montange et al., 2010). The two riboswitch-based biosensors also appear fully selective for SAM over SAH at equimolar concentrations. Bc 4-5 and Bc 1-5 were measured to have dissociation constants (K_d) for SAM of ~0.7 and ~1.7 μM, respectively. Furthermore, Bc 4-5 had ~2-fold higher maximal fluorescence then Bc 1-5. Thus, Bc 4-5, a cpRS-Sp design, has better performance than Bc 1-5, a conventional RS-Sp design.

In vivo detection of SAM with a cpRS-Sp biosensor using flow cytometry

While the secondary screen at physiologically relevant conditions suggests these biosensors would function in a cellular context, we wanted to verify their activity in bacterial cells. Efforts to image metabolites such as SAM are important as it is the required cofactor for enzymatic methylation and is a key metabolite in cellular sulfur metabolism ((Burgess, 2013) Additionally, E. coli have been shown to contain sub-millimolar levels of SAM (Bennett et al., 2009) and this high abundance would allow us to measure maximal fluorescence turn-on response in cells expressing biosensor constructs. To validate that cellular fluorescence was due to ligand-dependent turn on of the biosensors, mutant biosensors were generated that carry mutations that ablate ligand binding in the SAM-I riboswitch (Winkler et al., 2003). Additionally, Spinach2 and a Spinach2 mutant (Strack et al., 2013)that cannot bind DFHBI were included to compare the maximal fluorescence of our biosensors relative to the parent dye-binding aptamer Spinach2 and the background fluorescence of biosensor mutants to the inactive Spinach2. Each of the constructs were inserted into a tRNA scaffold on a pET31b plasmid (Kellenberger et al., 2015b). E. coli cells transformed with pET31b plasmids containing each construct were analyzed by flow cytometry. In all cases, robust fluorescence was observed for functional biosensors while their corresponding mutants showed close to background fluorescence (**Fig. 2.6**). Similar to the in vitro results, Bc 4-5 was brighter in vivo than Bc 1-5. Furthermore, two cpRS-Sp designs, Bs 4-4 and Bc 4-5, possess mean fluorescence intensity values higher than the parent Spinach2 aptamer. While their signal is saturated in this E. coli strain, these biosensors may be useful in auxotrophic strains to

study biosynthetic enzyme mutants or SAM transporters. As we have seen previously (Wang et al., 2016), by surveying diverse riboswitch sequences and a limited set of transducer stem lengths, we obtained biosensors with a range of affinities.

An unexpected effect of SAM on biosensor and Spinach2 fluorescence

While re-investigating biosensor function at higher SAM concentrations, we observed that fluorescence turn-on was decreased at 500 μM and higher ligand concentrations (**Fig. 2.7**). These results were reminiscent of a prior observation we made, that 3 mM and higher ATP reduced Spinach2 fluorescence (Su et al., 2016), likely due to competitive displacement of DFHBI from the dye-binding pocket. Similarly, we found that Spinach2 fluorescence was significantly decreased at 500 μM and higher SAM. These surprising results lead us to suggest that the Spinach2 aptamer may have general affinity for adenosine-containing metabolites. Although the K_d for SAM or ATP is much poorer (500-fold or higher) than for DFHBI, the intracellular concentrations of these metabolites are in the range that they can compete for binding. This explains at least in part why observed fluorescence turn-on is generally lower in vivo than in vitro. In practice, we have found that using 50 μM or higher DFHBI gives sufficient signal for experiments in E. coli. Also, we always perform in vivo experiments with Spinach2 as a control to compare to our Spinach2-based biosensor. Biosensor function is validated for sensing changes in metabolite levels only if the same conditions result in no fluorescence change in the Spinach2 control.

2.4 Discussion

This study shows that circularly-permuted riboswitches can provide an additional scaffold for designing functional RNA-based fluorescent biosensors. In fact, our results suggest it may be possible to generate more sensitive and higher fluorescent biosensors from a parent riboswitch by surveying different transducer stem topologies. There are a number of natural riboswitches or synthetic aptamers besides the SAM-I class that possess multi-stem junctions, such as the PreQ1 or FMN riboswitches ((Serganov et al., 2009; McCown et al., 2014) for which this strategy could be adapted to generate biosensors for their respective ligands. Also, the ability to exploit different connections to riboswitch folds may be beneficial to make biosensors with other dye-binding aptamers besides the Spinach scaffold, such as the Mango and SRB-2 aptamers for example (Sunbul and Jäschke, 2013; Dolgosheina et al., 2014) or to make other RNA devices (Chappell et al., 2016). Although naturally circularly-permuted riboswitches have not yet been found, our results show that permuted aptamer domains can retain or even improve on ligand binding and conformation switching, which suggests that this method may be promising for making synthetic riboswitches for gene regulation as well.

While P4-P2' cpRS-Sp constructs were functional, biosensors utilizing a P3-P2' topology were not. We verified at least in one example that the circularly permuted riboswitch aptamers still bind SAM with good affinities. There are two other potential explanations for this result. One is that the SAM-I riboswitch aptamer does not undergo a significant conformational change in the P3 stem upon ligand binding. A detailed study of the Bs yitJ SAM-I riboswitch discovered key differences in conformational changes with

and without the expression platform (Lu et al., 2010). The riboswitch aptamer with the expression platform displayed modest structural changes in the P3 stem upon ligand addition, but the aptamer alone showed little to no modulation in the presence of SAM. Our biosensor design incorporates only the SAM-I riboswitch aptamer, so likely does not modulate at the P3 stem efficiently to alter fluorescence output. Alternatively, even though the aptamer alone is capable of binding SAM, it is possible that the positioning of the Spinach2 aptamer at the P3 stem sterically hinders the riboswitch folding, or vice versa.

In vivo analysis by flow cytometry shows that cpRS-Sp sensors with P4-P2' topologies, Bs 4-4 and Bc 4-5, have higher mean fluorescence intensities than the Spinach2 aptamer alone (Fig. 2.5A). This directly contrasts with RS-cpSp sensors for SAH, which exhibit lower fluorescence than the fluorogenic aptamer alone (Su et al., 2016b). This suggests that the placement of circular permutation on the fluorogenic aptamer hinders maximal fluorescence more than creating a circular permutation on the riboswitch.

Additionally, there have been two previously developed biosensors for SAM that use the RS-Sp and Spinach-based riboswitch topologies (**Fig. 2.1A**, **C**). The first SAM biosensor fused the original Spinach aptamer to the Enterococcus faecalis SAM-III riboswitch, which is from a different riboswitch class that utilizes a 3-way junction (Lu et al., 2008; Paige et al., 2012b). The second biosensor was a Spinach-based riboswitch generated by fusing Spinach to the B. subtillis yitJ SAM-I riboswitch 5, which is one of the native riboswitches used in this study. Interestingly, the only functional biosensor construct derived from this riboswitch from our designs was Bs 4-4. The Bs 4-4 biosensor representing the cpRS-Sp topology has a dissociation constant of 1.0 μM, which is comparable to the reported dissociation constant of 1.2 μM for the Spinach-based riboswitch (You et al., 2015a)

2.5 Conclusions/Future Directions

Overall, this work validates our concept that circular permutations of riboswitch aptamers can add to existing design strategies for creating RNA-based fluorescent biosensors. To our knowledge, there has not been any previous biosensor designs which utilize circular permutations on the riboswitch aptamer. Furthermore, our success with modulating biosensor fluorescence via ligand-induced structural changes in the P4 stem suggests that SAM riboswitches with this alternate topology have the potential to regulate gene expression as well.

2.6 Figures

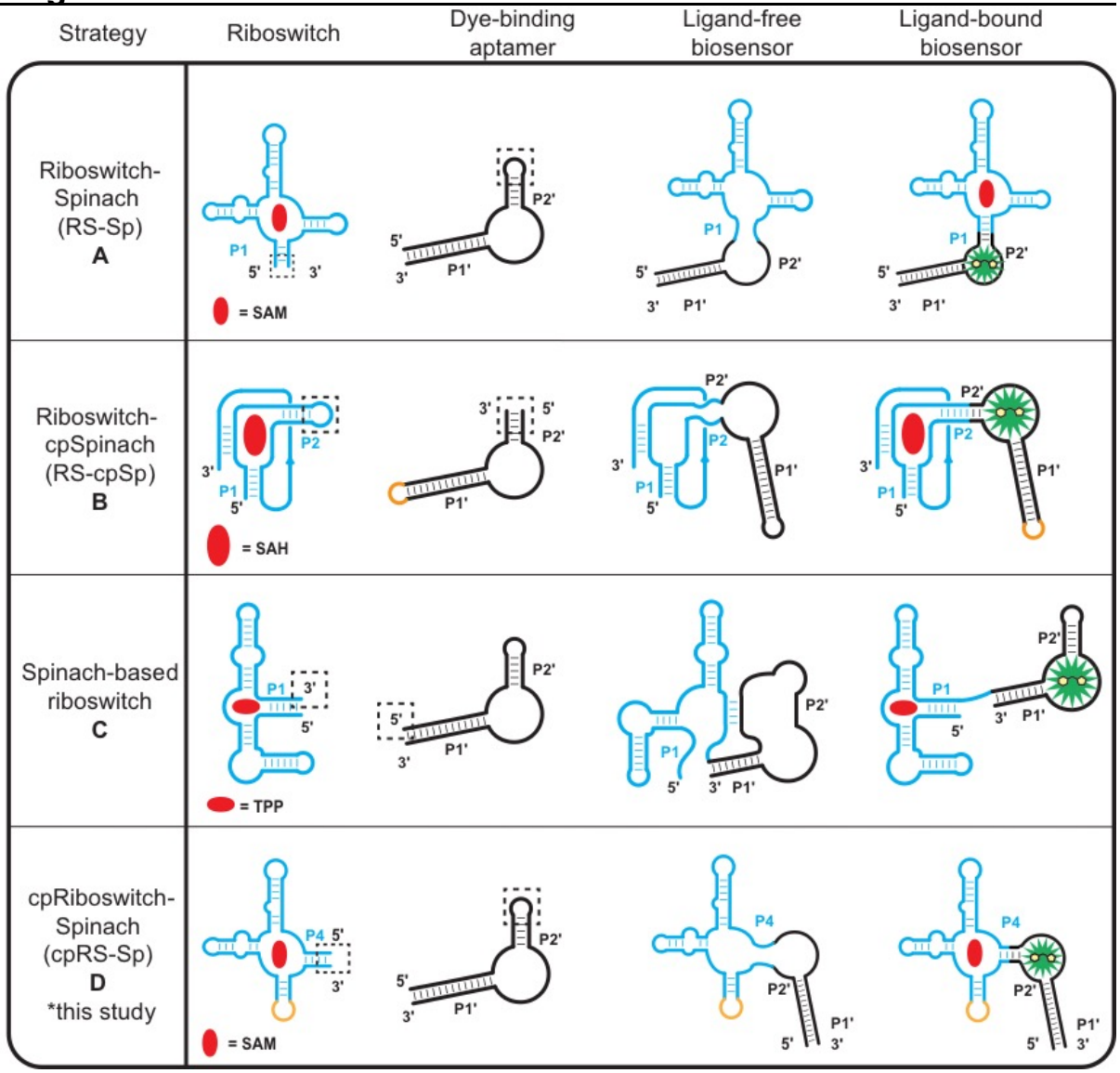


Figure 2.1. Design strategies for riboswitch-based fluorescent biosensors.
(a) The riboswitch–Spinach (RS-Sp) fusion strategy for biosensor development involves grafting the P1 stem of a riboswitch sensor domain to the P2' stem of Spinach. 4-7 (b) For riboswitches with non-pairing 5' and 3' ends, an internal stem of the riboswitch can be fused to a circular permutation of Spinach2 to generate a Riboswitch-circularly permuted Spinach2 design (RS-cpSp). (Su et al., 2016) (c) Spinach riboswitches can be created by replacing the gene regulatory expression platform of a riboswitch with the Spinach aptamer. (You et al., 2015) (d) A new biosensor strategy for any riboswitch with multiple pairing stems is to circularly permute the riboswitch and fuse it to the P2' stem of Spinach2 to generate a circularly-permuted Riboswitch-Spinach2 (cpRS-Sp) design. Riboswitch domains (blue) and their corresponding metabolites (red) are depicted with either the Spinach/Spinach2 aptamer (black) and 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI, green). "Sp" represents either Spinach/Spinach2 was used.

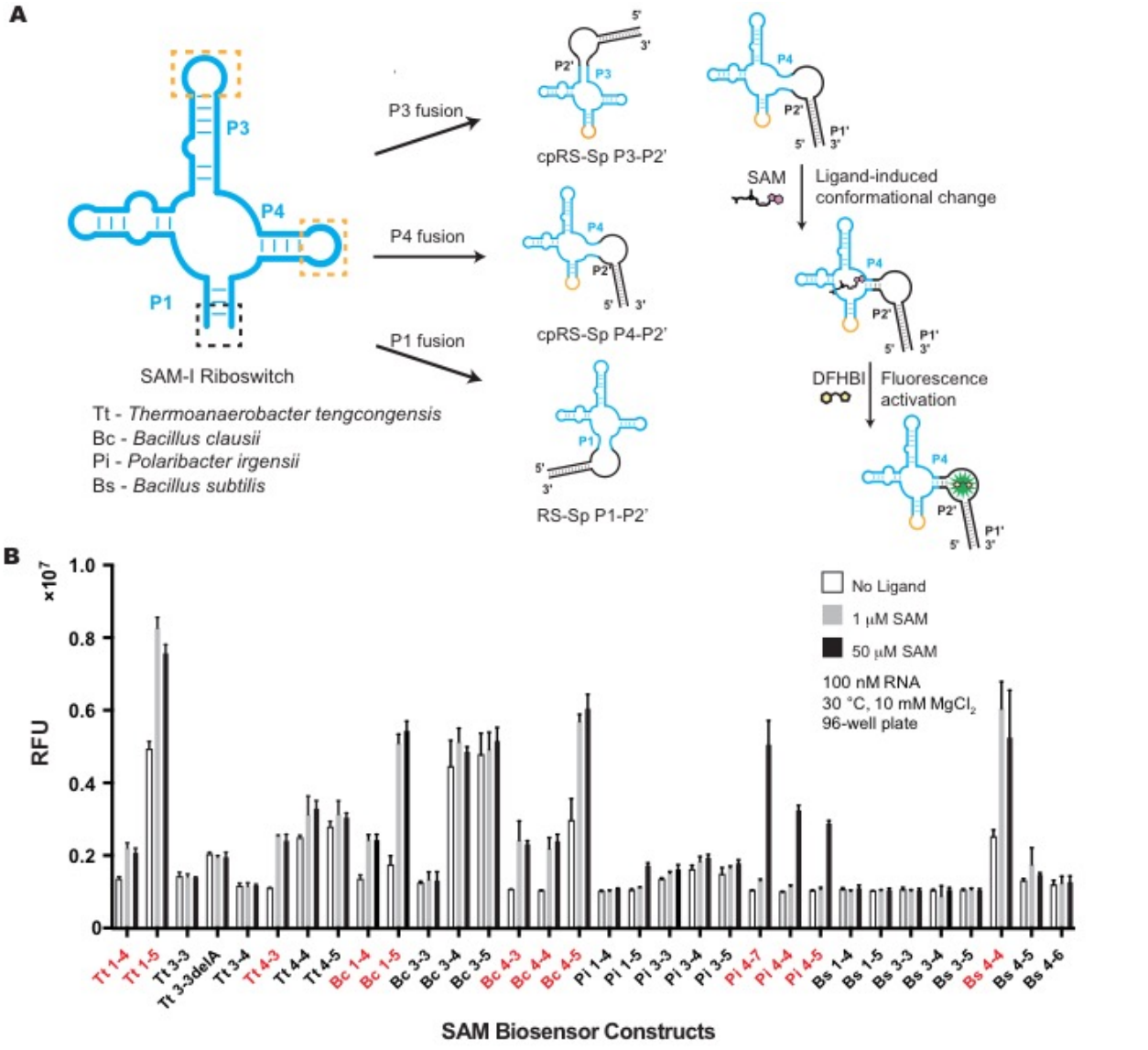


Figure 2.2. *In vitro* screen of cpRS-Sp biosensors derived from the SAM-I riboswitch class.
(a) Design of different SAM biosensor architectures and mechanism for fluorescence turn-on is depicted for the P4-P2' biosensor. (b) *In vitro* fluorescence response of SAM-I biosensors to SAM ligand. The nomenclature is based on the species origin of the sequence, the pairing stem of the SAM-I riboswitch that is fused to Spinach, and number of base pairs retained from the riboswitch stem, e.g. Bs 4-4 stands for SAM-I riboswitch from *Bacillus subtilis*, with Spinach fused to the P4 stem of the riboswitch that retains 4 base pairs. SAM biosensors chosen for further characterization (with greater than 1.5x fluorescence increase at 50 μM SAM) are indicated in red. Data shown are mean values ± standard deviation taken from two independent replicates.

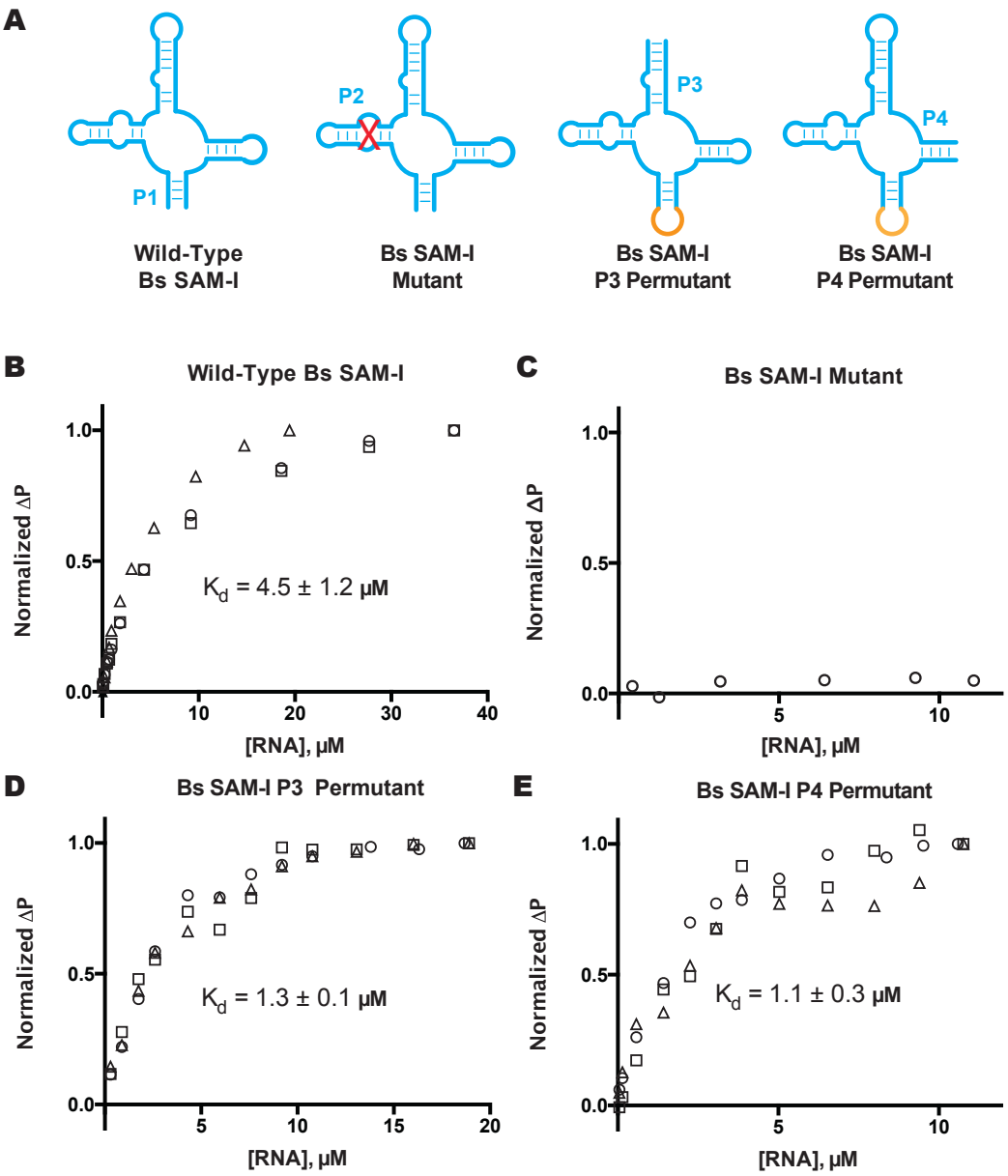


Figure 2.3. Ligand binding analysis of SAM-I riboswitch aptamers.
(a) Schematic of SAM-I riboswitch aptamer, mutant, and permutants. Fluorescence polarization data is shown for Cy5-labeled SAM analog with (a) wild-type Bs SAM-I (P1-7 stem), (b) Bs SAM-I mutant (P2 disruptive mutation), (c) Bs SAM-I P3 permutant (P3-5 stem), and (d) Bs SAM-I P4 permutant (P4-6 stem). Different symbols indicate independent replicates. Change in polarization (ΔP) was normalized to the maximal difference observed at the highest RNA concentration. The reported K_d value is the average of those calculated from best-fit curves with margin of error reported as SD. Data from (a) and (b) were reproduced with permission (Hickey and Hammond, 2014).

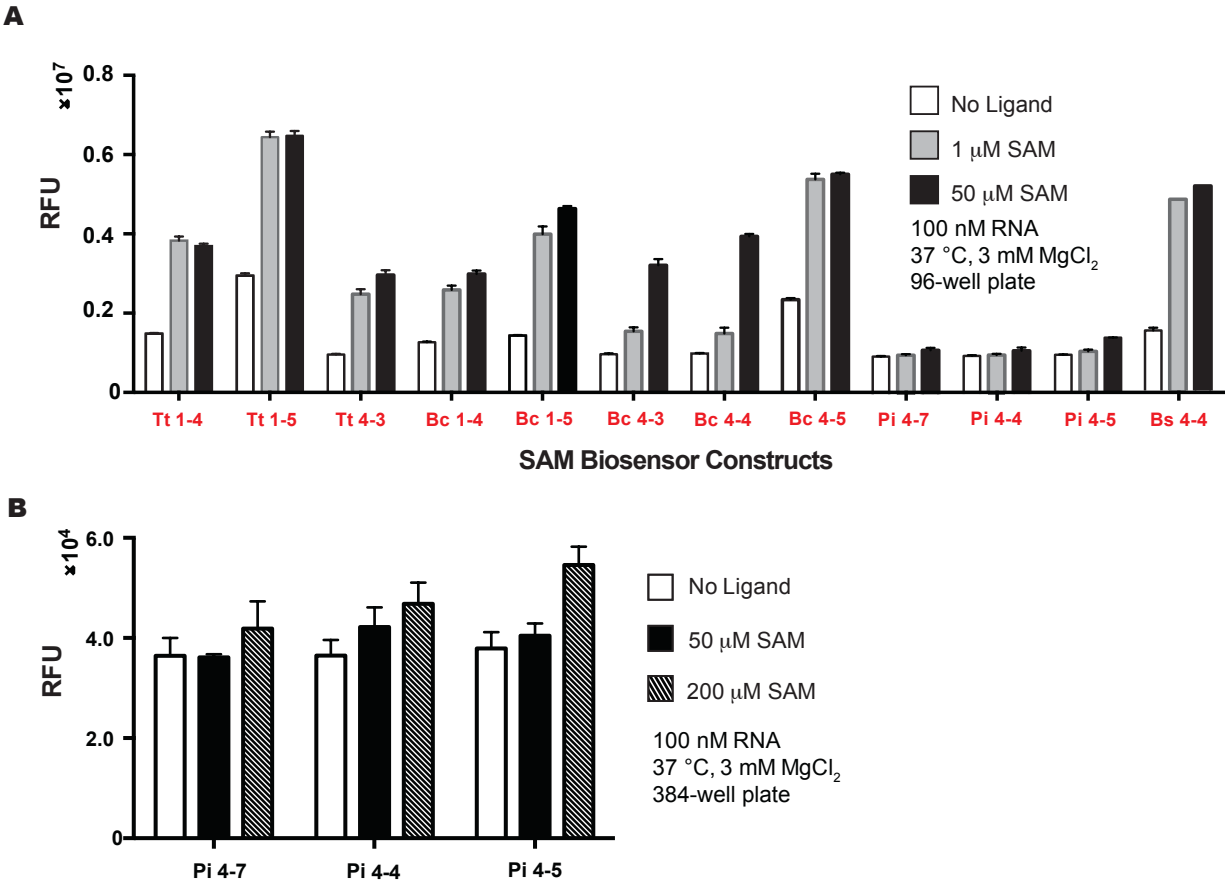


Figure 2.4. Re-screen of cpRS-Sp biosensors at physiologically relevant conditions.
(a) *In vitro* fluorescence response of SAM-I biosensors to SAM ligand at 37 °C and 3 mM $MgCl_2$. (b) Pi biosensors were re-screened at higher SAM concentrations. Data shown are mean values ± standard deviation taken from two independent replicates.

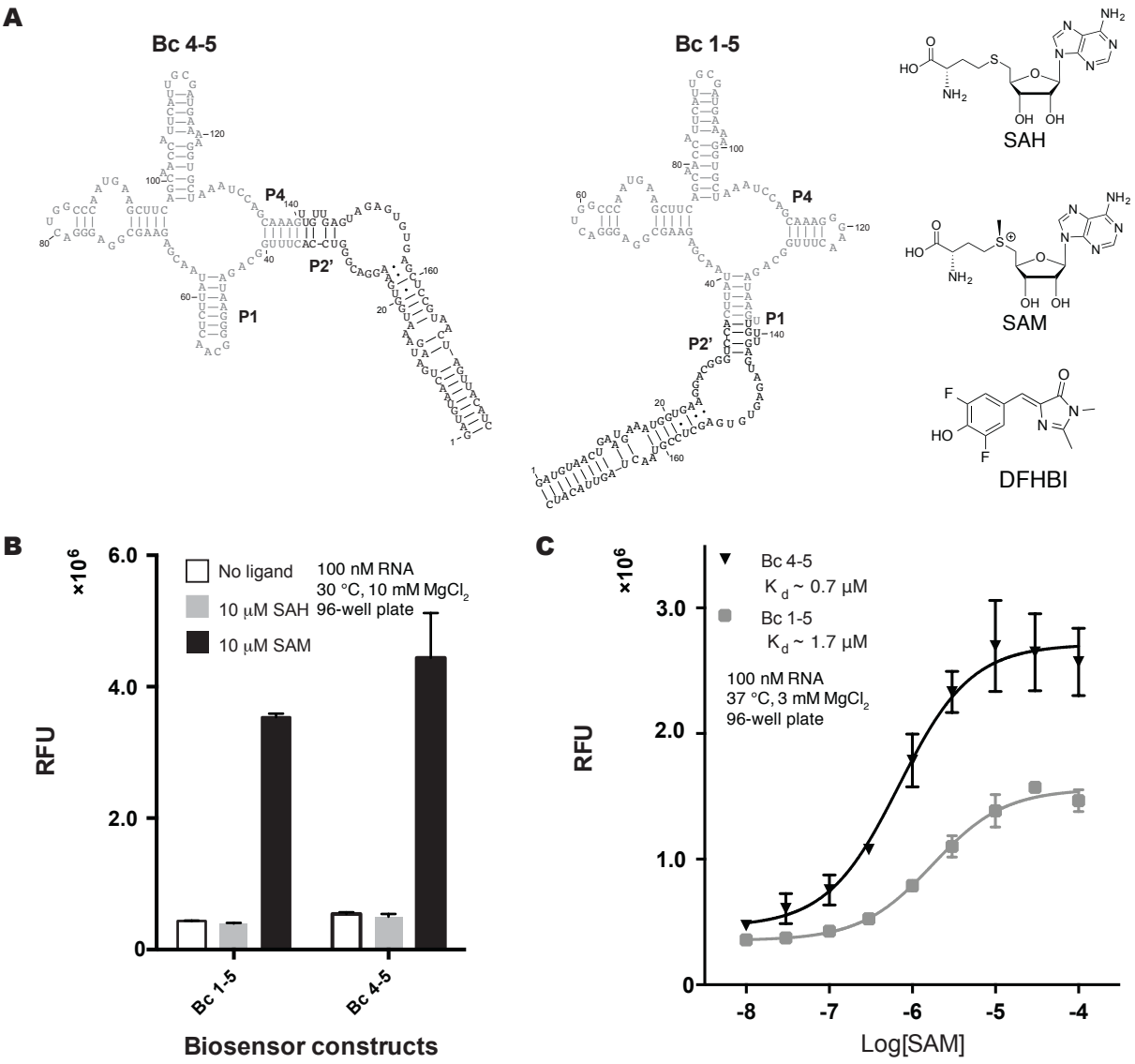


Figure 2.5. Selectivity and sensitivity of cpRS-Sp biosensors.
(a) Secondary structure models for Bc 1-5 and Bc 4-5 biosensors and chemical structures of the biosensor ligand, *S*-adenosyl-L-methionine (SAM), the structural analog *S*-adenosyl-L-homocysteine (SAH), and dye DFHBI. (b) Selectivity of biosensors for SAM over SAH. (c) *In vitro* analysis of biosensor binding affinities for SAM. Data shown are mean values ± standard deviation taken from at least two independent replicates.

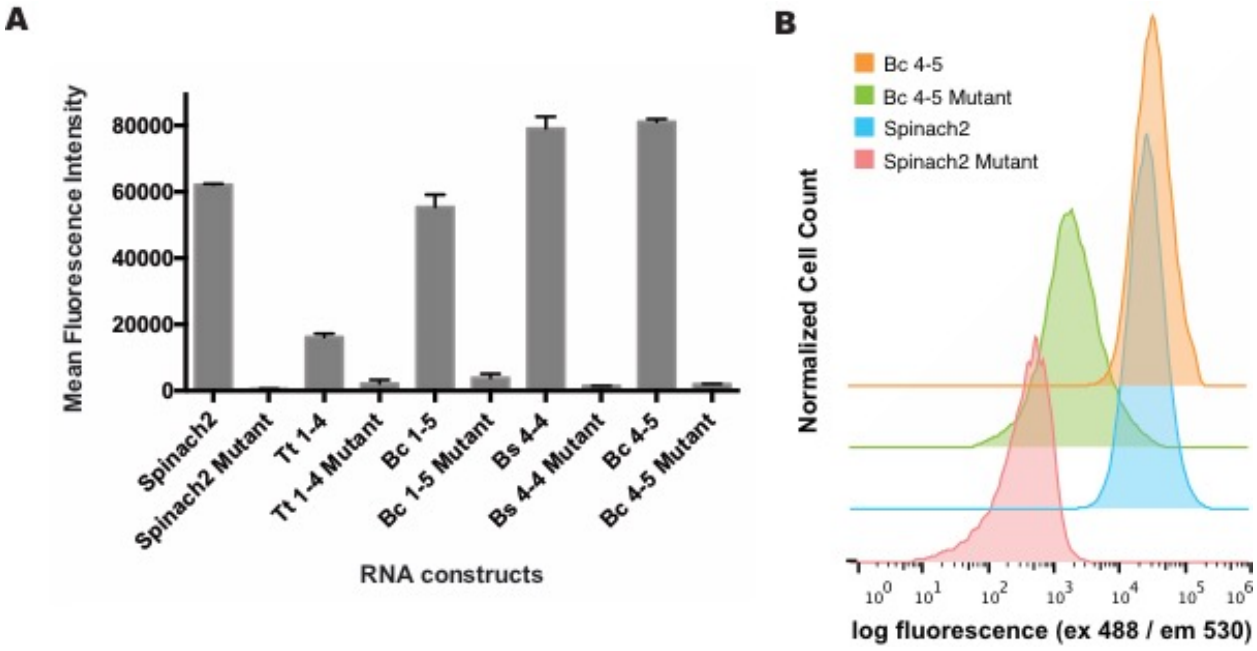


Figure 2.6. *In vivo* performance of SAM cpRS-Sp biosensors.
(a) Live cell fluorescence measured by flow cytometry for *E. coli* BL21* cells expressing plasmid encoding biosensors and incubated in media containing DFHBI. Mean fluorescence intensity was determined by analyzing 30,000 cells per replicate. All error bars represent standard deviation between technical replicates. (b) Representative flow histograms for cells expressing biosensors or controls.

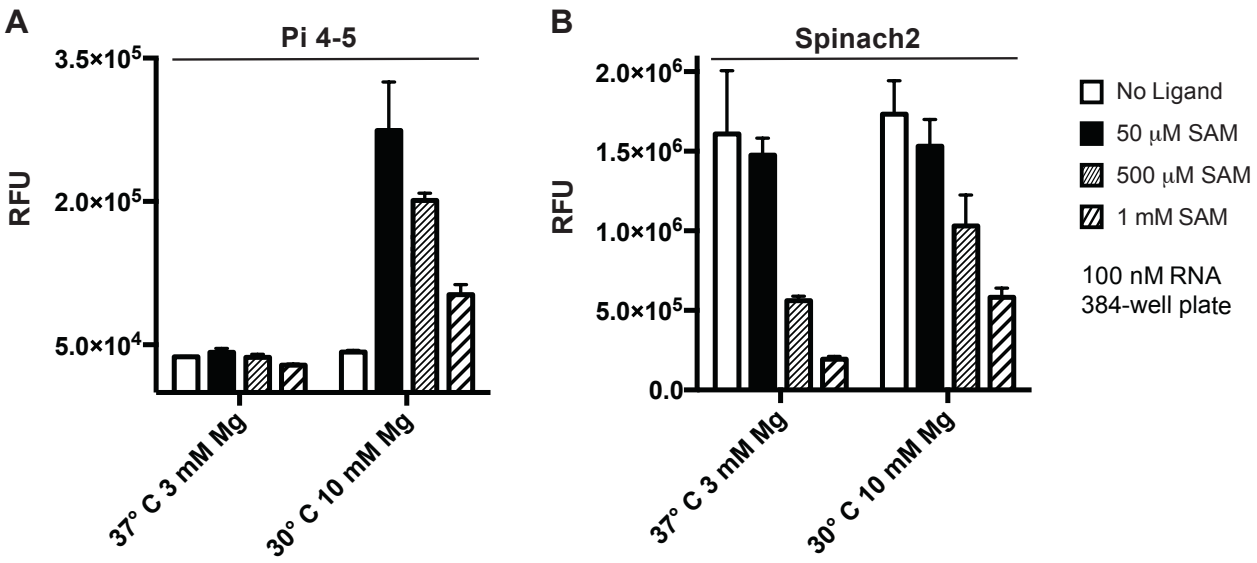


Figure 2.7. Effects of temperature, Mg^{2+}, and SAM on biosensor and Spinach2 fluorescence.
(a) *In vitro* fluorescence activation of (a) biosensor Pi-4-5 and (b) Spinach2 aptamer at 30 °C, 10 mM $MgCl_2$ and 37 °C, 3 mM $MgCl_2$ in the presence of 0, 50, 500, and 1000 μM SAM. Data shown are mean values ± standard deviation taken from two independent replicates.

Table 2.1. **Candidate biosensor sequences**. **Bold** sequences indicate the Spinach2 sequence, which flanks the riboswitch sequence. *Italic and underlined* indicate the part of the transducer stem derived from the native riboswitch. Blue indicates the tetraloop used for circular permutation of the P1 stem in the riboswitch.

Name	Sequence (5' to 3')
Tt 1-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TATC*AAGAGAGGTGGAGGGACTGGCCCGA TGAAACCCGGCAACCAGCCTTAGGGCATGGTGCCAATTCCTGCAGCGGTTTCGCTGAAA*GATG***TT GTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 1-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTATC*AAGAGAGGTGGAGGGACTGGCCCG ATGAAACCCGGCAACCAGCCTTAGGGCATGGTGCCAATTCCTGCAGCGGTTTCGCTGAAA*GATGA* **TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 3-3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***GGC*ATGGTGCCAATTCCTGCAGCGGTTTC GCTGAAAGATGAGAGGCAACTCTTATCAAGAGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAA CCA*GCC***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 3-3 delA	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***GGC*TGGTGCCAATTCCTGCAGCGGTTTCG CTGAAAGATGAGAGGCAACTCTTATCAAGAGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAAC CA*GCC***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 3-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CTGG*TGCCAATTCCTGCAGCGGTTTCGCT GAAAGATGAGAGGCAACTCTTATCAAGAGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAA*CCA G***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 4-3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CTG*AAAGATGAGAGGCAACTCTTATCAAGA GAGGTGGAGGGACTGGCCCGATGAAACCCGGCAACCAGCCTTAGGGCATGGTGCCAATTCCTG*C AG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 4-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***GCTG*AAAGATGAGAGGCAACTCTTATCAAG AGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAACCAGCCTTAGGGCATGGTGCCAATTCCTG *CAGC***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Tt 4-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CGCTG*AAAGATGAGAGGCAACTCTTATCAA GAGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAACCAGCCTTAGGGCATGGTGCCAATTCCTG *CAGCG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 1-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTAT*AACGAGAAGCGGAGGGACTGGCCCA ATGAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAATCCAGCAAAGGGAACTTTGGCAG *ATAA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 1-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CTTAT*AACGAGAAGCGGAGGGACTGGCCC AATGAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAATCCAGCAAAGGGAACTTTGGCA G*ATAAG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 3-3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***GAA*AAGGTGCTAAATCCAGCAAAGGGAACT TTGGCAGATAAGGGGCAACTCTTATAACGAGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAA CCA*TTC***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 3-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TGAA*AAGGTGCTAAATCCAGCAAAGGGAAC TTTGGCAGATAAGGGGCAACTCTTATAACGAGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAA CCA*TTCA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 3-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***ATGAA*AAGGTGCTAAATCCAGCAAAGGGAA CTTTGGCAGATAAGGGGCAACTCTTATAACGAGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGC AACCA*TTCAT***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 4-3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTG*GCAGATAAGGGGCAACTCTTATAACGA GAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAATCC AG*CAA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 4-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTTG*GCAGATAAGGGGCAACTCTTATAACG AGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAATC CAG*CAAA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bc 4-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CTTTG*GCAGATAAGGGGCAACTCTTATAAC GAGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAAT CCAG*CAAAG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 1-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TATC*AAGAAAGGCGGAGGGATTAGACCCA TTGAAGCCTTAGCAACCCTTTAGTAATAAAGAAGGTGCTAAATTCTACTCAATTATTCGTAATTGGAT A*GATA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 1-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTAT*CAAGAAAGGCGGAGGGATTAGACCC ATTGAAGCCTTAGCAACCCTTTAGTAATAAAGAAGGTGCTAAATTCTACTCAATTATTCGTAATTGGA TAG*ATAA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**

Pi 3-3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***AAG*AAGGTGCTAAATTCTACTCAATTATTCG TAATTGGATAGATAACAGCAATGTTATCAAGAAAGGCGGAGGGATTAGACCCATTGAAGCCTTAGCA ACC*CTT***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 3-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***AAAG*AAGGTGCTAAATTCTACTCAATTATTC GTAATTGGATAGATAACAGCAATGTTATCAAGAAAGGCGGAGGGATTAGACCCATTGAAGCCTTAG CAACC*CTTT***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 3-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TAAAG*AAGGTGCTAAATTCTACTCAATTATT CGTAATTGGATAGATAACAGCAATGTTATCAAGAAAGGCGGAGGGATTAGACCCATTGAAGCCTTA GCAACC*CTTTA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 4-7	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TAATTGG*ATAGATAACAGCAATGTTATCAA GAAAGGCGGAGGGATTAGACCCATTGAAGCCTTAGCAACCCTTTAGTAATAAAGAAGGTGCTAAAT TCTAC*TCAATTA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 4-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTGG*ATAGATAACAGCAATGTTATCAAGAA AGGCGGAGGGATTAGACCCATTGAAGCCTTAGCAACCCTTTAGTAATAAAGAAGGTGCTAAATTCTA C*TCAA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pi 4-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***ATTGG*ATAGATAACAGCAATGTTATCAAGA AAGGCGGAGGGATTAGACCCATTGAAGCCTTAGCAACCCTTTAGTAATAAAGAAGGTGCTAAATTCT AC*TCAAT***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 1-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TATC*AAGAGAAGCAGAGGGACTGGCCCGA CGAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACCAAGGTGCTAAATCCAGCAAGCTCGA ACAGCTTGGAA*GATA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 1-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***TTATC*AAGAGAAGCAGAGGGACTGGCCCG ACGAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACCAAGGTGCTAAATCCAGCAAGCTCG AACAGCTTGGAA*GATAA***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 3-3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CAT*GACCAAGGTGCTAAATCCAGCAAGCTC GAACAGCTTGGAAGATAAGAAGCAATTCTTATCAAGAGAAGCAGAGGGACTGGCCCGACGAAGCTT CAGCAACCGGTGTA*ATG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 3-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CCAT*GACCAAGGTGCTAAATCCAGCAAGCT CGAACAGCTTGGAAGATAAGAAGCAATTCTTATCAAGAGAAGCAGAGGGACTGGCCCGACGAAGCT TCAGCAACCGGTGTA*ATGG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 3-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***GCCAT*GACCAAGGTGCTAAATCCAGCAAG CTCGAACAGCTTGGAAGATAAGAAGCAATTCTTATCAAGAGAAGCAGAGGGACTGGCCCGACGAAG CTTCAGCAACCGGTGTA*ATGGC***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 4-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***CTTG*GAAGATAAGAAGCAATTCTTATCAAG AGAAGCAGAGGGACTGGCCCGACGAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACCAA GGTGCTAAATCCAG*CAAG***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 4-5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***GCTTG*GAAGATAAGAAGCAATTCTTATCAA GAGAAGCAGAGGGACTGGCCCGACGAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACCA AGGTGCTAAATCCAG*CAAGC***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bs 4-6	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA***AGCTTG*GAAGATAAGAAGCAATTCTTATCA AGAGAAGCAGAGGGACTGGCCCGACGAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACC AAGGTGCTAAATCCAG*CAAGCT***TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**

Table 2.2. **Sequences used in *in vivo* flow cytometry analysis**. *Italic* sequences indicates the tRNA scaffold that is appended onto the terminal ends. **Bold** sequences indicate the Spinach2 sequence, which flanks the riboswitch sequence. *Italic and underlined* indicate the part of the transducer stem derived from the native riboswitch. Blue indicates the tetraloop used for circular permutation of the P1 stem in the riboswitch. Red indicates the site of mutation(s) that abolish ligand binding to the aptamer.

Name	Sequence (5' to 3')
Spinach2	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**GTAGGCTGCTTCGGCAGCCTAC**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAG TTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA*
Spinach2 Mutant	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGccTCCA**GTAGGCTGCTTCGGCAGCCTAC**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGT TACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA*
Tt 1-4	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***TATC***AAGAGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAACCAGCCTTAGG GCATGGTGCCAATTCCTGCAGCGGTTTCGCTGAAA***GATG*****TTGTTGAGTAGAGTGTGAGCTCC GTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA*
Tt 1-4 Mutant	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***TATC***AgaAGAGGTGGAGGGACTGGCCCGATGAAACCCGGCAACCAGCCTTAGG GCATGGTGCCAATTCCTGCAGCGGTTTCGCTGAAA***GATG*****TTGTTGAGTAGAGTGTGAGCTCC GTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA*
Bc 1-5	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***CTTAT***AACGAGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAACCATTCATT GCGATGAAAAGGTGCTAAATCCAGCAAAGGGAACTTTGGCAG***ATAAG*****TTGTTGAGTAGAGTG TGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGC CA*
Bc 1-5 Mutant	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***CTTAT***AAgcAGAAGCGGAGGGACTGGCCCAATGAAGCTTCAGCAACCATTCATT GCGATGAAAAGGTGCTAAATCCAGCAAAGGGAACTTTGGCAG***ATAAG*****TTGTTGAGTAGAGTG TGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGC CA*
Bs 4-4	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***GCTTG***GAAGATAAGAAGCAATTCTTATCAAGAGAAGCAGAGGGACTGGCCCGA CGAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACCAAGGTGCTAAATCCAG***CAAGC*** **TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAG TCCCTGTTCGGGCGCCA*
Bs 4-4 Mutant	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***GCTTG***GAAGATAAGAAGCAATTCTTATCAgaAGAAGCAGAGGGACTGGCCCGAC GAAGCTTCAGCAACCGGTGTAATGGCGATCAGCCATGACCAAGGTGCTAAATCCAG***CAAGC*****T TGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGT CCCTGTTCGGGCGCCA*
Bc 4-5	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***CTTTG***GCAGATAAGGGGCAACTCTTATAACGAGAAGCGGAGGGACTGGCCCAA TGAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAATCCAG***CAAAG*****TTGTTGAGTAG AGTGTGAGCTCCGTAACTAGTTACATC**
Bc 4-5 Mutant	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA***CTTTG***GCAGATAAGGGGCAACTCTTATAAgcAGAAGCGGAGGGACTGGCCCAAT GAAGCTTCAGCAACCATTCATTGCGATGAAAAGGTGCTAAATCCAG***CAAAG*****TTGTTGAGTAGA GTGTGAGCTCCGTAACTAGTTACATC**

2.7 Materials and Methods

Reagents and oligonucleotides

DNA oligonucleotides for biosensor constructs were purchased as Ultramers from Integrated DNA Technologies (Coralville, IA) and other DNA oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA). S-Adenosyl-L-methionine (SAM), S-adenosyl-L homocysteine (SAH) were purchased from Sigma-Aldrich (St Louis, MO). DFHBI was synthesized following previously described protocols (Paige et al., 2011) and was stored as a 10 mM stock in DMSO at -20 °C. Chemically competent BL21 (DE3) Star cells were purchased from Life Technologies (Carlsbad, CA).

In vitro transcription

DNA templates for in vitro transcription were prepared by PCR amplification using Phusion DNA polymerase (NEB) from sequence-confirmed plasmids or Ultramer oligonucleotides (for screening experiment only) using primers that added the T7 polymerase promoter sequence at the 5' end. PCR products were purified either by a 96-well format ZR-96 DNA Clean-up kit (Zymo Research) for screening experiments or by QIAquick PCR purification kit (Qiagen) for characterization and application experiments. RNA was transcribed from DNA templates using T7 RNA polymerase in 40 mM Tris- HCl, pH 8.0, 6 mM MgClR2R, 2 mM spermidine, and 10 mM DTT. RNAs were either purified by a 96- well format ZR-96 Clean & Concentrator (Zymo Research) or by denaturing (7.5 M urea) 6% PAGE. RNAs purified by PAGE were visualized by UV shadowing and extracted from gel pieces using Crush Soak buffer (10 mM Tris-HCl, pH 7.5, 200 mM NaCl and 1 mM EDTA, pH 8.0). Purified RNAs were precipitated with ethanol, dried, and then resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Accurate RNA concentrations were determined by measuring the absorbance at 260 nm after performing a hydrolysis assay to eliminate the hypochromic effect due to RNA secondary structure (Wilson et al., 2014).

General procedure for in vitro fluorescence assays

In vitro fluorescence assays were carried out in binding buffer containing 100 nM RNA, 10 μM DFHBI, 40 mM HEPES, pH 7.5, 125 mM KCl, and 3 or 10 mM $MgCl_2$, as indicated in the figures. Other conditions, including temperature and concentration of ligand, were varied in different experiments as indicated. The RNA was renatured by heating to 72 °C for 3 min in the binding buffer then cooled to ambient temperature for 5 min prior to addition to the reaction solution. DFHBI was added to the solution containing buffer and RNA, and then ligand (or water for no ligand control) was added before fluorescence measurement. Binding reactions were performed in 100 μL volumes and were incubated at the indicated temperature in a Corning Costar 3915 96-well black plate or a Greiner Bio-One 384-well black plate in a Molecular Devices SpectraMax Paradigm Multi-Mode detection platform plate reader (Sunnyvale, CA). The fluorescence emission was measured during 30 to 60 min total with the following instrument parameters: 448 nm excitation, 506 nm emission.

Fluorescence polarization (FP) binding assays

Fluorescence polarization readings were carried out using a QuantaMaster spectrofluorometer (Photon Technology International) at excitation 646 nm, emission 662 nm. Samples were prepared in 50 ml of TBM buffer (90 mM Tris base, 89 mM boric acid, and 10 mM $MgCl_2$, pH 7.0) containing 1 mM of a Cy5-labeled SAM analog (C_8-Cy5) (Hickey and Hammond, 2014) and saturation binding experiments were performed with RNA concentrations ranging from 0 to 40 μM. RNA was added successively to the sample cuvette, and concentration values were corrected for added volume. Samples were equilibrated at 30 °C for 2 min prior to each FP measurement using tubing connecting the cuvette holder to a water bath.

Binding affinity analysis of SAM biosensors

To measure the binding affinities of SAM biosensors, fluorescence assays were performed with the following conditions: binding buffer with 10 mM $MgCl_2$, 37 ° C, 100 nM RNA, 10 μM DFHBI, and ligand SAM concentrations from 10 nM to 100 μM. The fluorescence of the sample with DFHBI but no RNA was subtracted as background to determine relative fluorescence units.

In vivo fluorescence assays by flow cytometry

Preparation of cell samples for flow cytometry was carried out by inoculating 3 mL of LB/carb media with 150 µL of an overnight culture of BL21 (DE3) Star cells containing either the pET31b-RS-Spor pET31b-cpRS-Sp constructs. Cells were grown aerobically to an OD_{600} ~ 0.5 - 0.6, then induced with 1 mM IPTG at 37 °C for 2 h. Cell density was measured by OD_{600}, and assuming that there are $1x10^9$ cells/mL for for an OD_{600} of 1, $4x10^8$ cells were sampled and pelleted at rt for 4 min at 3,700 rcf, washed once with PBS media at pH 7.0, then resuspended in PBS media containing 100 µM DFHBI. Cellular fluorescence was measured for 30,000 cells using an Attune NxT Acoustic Focusing Cytometer (Life Technologies).

2.8 References

Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli. *Nat Chem Biol* **2009**, *5* (8), 593–599.

Burgess, D. J. Epigenetics: Bacterial DNA Methylation Gets SMRT. *Nat. Rev. Genet.* **2013**, *14* (1), 4.

Chappell, J.; Watters, K. E.; Takahashi, M. K.; Lucks, J. B. A Renaissance in RNA Synthetic Biology : New Mechanisms , Applications and Tools for the Future. *Curr. Opin. Chem. Biol.* **2016**, *28* (2015), 47–56.

Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J. RNA Mango Aptamer-

Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chem. Biol.* **2014**, *9* (10), 2412–2420.

Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem.* **2017**, *86* (1), 515–539.

Hammann, C.; Luptak, A.; Perreault, J.; de la Pena, M. The Ubiquitous Hammerhead Ribozyme. *Rna* **2012**, *18* (5), 871–885.

Heppell, B.; Blouin, S.; Dussault, A.-M.; Mulhbacher, J.; Ennifar, E.; Penedo, J. C.; Lafontaine, D. a. Molecular Insights into the Ligand-Controlled Organization of the SAM-I Riboswitch. *Nat. Chem. Biol.* **2011**, *7* (6), 384–392.

Hickey, S. F.; Hammond, M. C. Structure-Guided Design of Fluorescent S-Adenosylmethionine Analogs for a High-Throughput Screen to Target SAM-I Riboswitch RNAs. *Chem. Biol.* **2014**, *21* (3), 345–356.

Kellenberger, C. a.; Wilson, S. C.; Sales-Lee, J.; Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic Di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc.* **2013**, *135* (13), 4906–4909.

Kellenberger, C. A.; Hallberg, Z. F.; Hammond, M. C. Live Cell Imaging Using Riboswitch-Spinach TRNA Fusions as Metabolite-Sensing Fluorescent Biosensors. *Methods Mol Biol* **2015a**, *1316*, 7–103.

Kellenberger, C. A.; Chen, C.; Whiteley, A. T.; Portnoy, D. A.; Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic Di-AMP. *J. Am. Chem. Soc.* **2015b**, *137* (20), 6432–6435.

Lu, C.; Smith, A. M.; Fuchs, R. T.; Ding, F.; Rajashankar, K.; Henkin, T. M.; Ke, A. Crystal Structures of the SAM-III/S(MK) Riboswitch Reveal the SAM-Dependent Translation Inhibition Mechanism. *Nat. Struct. Mol. Biol.* **2008**, *15* (10), 1076–1083.

Lu, C.; Ding, F.; Chowdhury, A.; Pradhan, V.; Tomsic, J.; Holmes, W. M.; Henkin, T. M.; Ke, A. SAM Recognition and Conformational Switching Mechanism in the Bacillus Subtilis YitJ S Box/SAM-I Riboswitch. *J. Mol. Biol.* **2010**, *404* (5), 803–818.

McCown, P.; Liang, J.; Weinberg, Z.; Breaker, R. Structural, Functional, and Taxonomic Diversity of Three PreQ1 Riboswitch Classes. *Chem. Biol.* **2014**, *21* (7), 880–889.

Montange, R. K.; Batey, R. T. Structure of the S-Adenosylmethionine Riboswitch Regulatory MRNA Element. *Nature* **2006**, *441* (7097), 1172–1175.

Montange, R. K.; Mondragón, E.; Tyne, D. Van; Garst, A. D.; Ceres, P.; Batey, R. T. Discrimination between Closely Related Cellular Metabolites by the SAM-I Riboswitch. *J. Mol. Biol.* **2010**, *396* (3), 761–772.

Paige, J. S.; Wu, K. Y.; Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science (80-.).* **2011**, *333* (6042).

Paige, J. S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S. R. Fluorescence Imaging of Cellular Metabolites with RNA. *Science (80-.).* **2012**, *335* (6073), 1194–1194.

Serganov, A.; Huang, L.; Patel, D. J. Coenzyme Recognition and Gene Regulation by a Flavin Mononucleotide Riboswitch. *Nature* **2009**, *458* (7235), 233–237.

Soma, A.; Onodera, A.; Sugahara, J.; Kanai, A.; Yachie, N.; Tomita, M.; Kawamura, F.; Sekine, Y. Permuted TRNA Genes Expressed via a Circular RNA Intermediate in Cyanidioschyzon Merolae. *Science (80-.).* **2007**, *318* (5849), 450–453.

Strack, R. L.; Disney, M. D.; Jaffrey, S. R. A Superfolding Spinach2 Reveals the Dynamic Nature of Trinucleotide Repeat-Containing RNA. *Nat. Methods* **2013**, *10* (12), 1219–1224.

Su, Y.; Hickey, S. F.; Keyser, S. G. L.; Hammond, M. C. *In Vitro* and *In Vivo* Enzyme Activity Screening via RNA-Based Fluorescent Biosensors for *S* -Adenosyl- l - Homocysteine (SAH). *J. Am. Chem. Soc.* **2016**, *138* (22), 7040–7047.

Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R. Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. *Science (80-.).* **2006**, *314* (5797).

Sunbul, M.; Jäschke, A. Contact-Mediated Quenching for RNA Imaging in Bacteria with a Fluorophore-Binding Aptamer. *Angew. Chemie - Int. Ed.* **2013**, *52* (50), 13401–13404.
Villa, J. K.; Su, Y.; Contreras, L. M.; Hammond, M. C. Synthetic Biology of Regulatory RNAs" Regulating with RNA in Bacteria and Archaea. **2017**.

Wang, X. C.; Wilson, S. C.; Hammond, M. C.; D.J., R.; M.J., D.; C.M., W.; V.T., L.; A.T., I.; H.K., C.; Y.-F., H. Next-Generation RNA-Based Fluorescent Biosensors Enable Anaerobic Detection of Cyclic Di-GMP. *Nucleic Acids Res.* **2016**, *44* (17), e139–e139.

Wilson, S. C.; Cohen, D. T.; Wang, X. C.; Hammond, M. C. A Neutral PH Thermal Hydrolysis Method for Quantification of Structured RNAs. *RNA* **2014**, *20* (7), 1153–1160.

Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R. An MRNA Structure That Controls Gene Expression by Binding S-Adenosylmethionine. *Nat. Struct. Biol.* **2003**, *10* (9), 701–707.

You, M.; Litke, J. L.; Jaffrey, S. R. Imaging Metabolite Dynamics in Living Cells Using a Spinach-Based Riboswitch. *Proc. Natl. Acad. Sci.* **2015**, *112* (21), E2756–E2765.

CHAPTER 3

Breaking riboswitch speed limits through engineering and rational design of a guanidine biosensor

Portions of this work will published in a scientific journal:

Truong, J., & Hammond, M. C. Approaches to jailbreak riboswitch speed: a case study with a novel guanidine biosensor. *In preparation*

3.1 Abstract

Riboswitch-based technologies like reporters, aptazymes, and RNA-based fluorescent (RBF) biosensors have wide applications for detection, imaging, and regulatory circuits. While riboswitch reporters and aptazymes have been robustly studied to enhance their function, there are few studies targeted towards RBF biosensors. Here, RBF biosensors for the ligand guanidine are developed as a case-study for tuning biosensor properties. Using the ykkC-I riboswitch aptamer, two novel biosensor designs were explored that generated functional biosensors. A subset of these biosensors were then systematically probed using a base-pair mutation strategy which has never been applied for investigating RBF biosensor performance. From this approach, we generated various biosensor mutants which exhibit modest improvements in their kinetics and fold-activation. This finding establishes our mutational strategy as an additional optimization method in the development of RBF biosensors.

3.2 Introduction

Riboswitches are cis-regulatory elements that been naturally evolved to possess high affinity aptamer domains. Upon binding their cognate ligand, they induce conformational changes that communicate with an expression platform to regulate genes. Their robust mechanisms have been exploited in the development and engineering of modular sensor technologies such as riboswitch reporters, aptazymes and RNA-based fluorescent (RBF) biosensors (Hallberg et al.). These tools have seen increasing adoption for applications in ligand detection since they possess a rare duality of being genetically encodable and functioning outside a cellular context (*in vitro* or cell-free environments). With the rapid growth of analytes for these riboswitch-based sensors, there is growing interest in systematically probing properties of these sensors.

Traditionally, riboswitch reporters and aptazymes are extensively used *in vivo* for experimental riboswitch validation and conditional gene regulation circuits, respectively, and as such have inspired the development of optimization methods. Some screening approaches completely randomize the identity of the expression platform reporters or communication modules in aptazymes which test for function and provide empirical guidelines for rational design (Zhong et al., 2016; Harbaugh et al., 2018). The development of cell-free transcription-translation systems also has enabled computational and *in vitro* studies to better understand components involved in function. One study investigating the thiamin pyrophopsphate (TPP) aptazyme compares how a single cytosine deletion can slow the kinetics of cleavage and how the addition of stabilizing base-pairs in the communication module raises activation 4-fold (Ichihashi et al., 2012). Another study uses an automated computational design to increase activation ratios of synthetic riboswitch reporters by constraining the aptamer domain and modeling free energies of a library of switching sequences (Espah Borujeni et al., 2016). This approach was reported to improve the activation of riboswitches like theophylline 383-fold *in vivo* and tetramethylrosamine 16-fold in a cell-free system. These detailed approaches have been used to characterize riboswitch reporter and aptazyme systems but have not been used directly to study RBF biosensors.

RBF biosensors are composed of a riboswitch aptamer that is fused to a fluorogenic RNA aptamer through a transducer module. In the presence of the target ligand, the transducer module communicates with the dye-binding aptamer allowing for a pro-fluorescent dye to bind and activate fluorescence. In this way, the biosensor generates a fluorescent signal which can be measured *in vitro* or live cells through standard laboratory instrumentation such as a plate reader or a flow cytometer respectively.

Previous work has demonstrated how the identity of the communication module plays a significant role in biosensor activation. While complete randomization of transducer modules generated functional sensor for SAM, TPP, and ATP (Paige et al., 2012b), our group discovered that using natural pairing elements of a riboswitch as the transducer module afforded biosensors that were brighter and/or faster than using the former approach. (Kellenberger et al., 2013a, 2015c; Su et al., 2016b; Wang et al., 2016; Truong et al., 2018). Additionally, we have demonstrated that sampling the phylogenetic diversity of riboswitches in bacteria is generalizable and can be robustly employed to tune performance. Using phylogenetic variants of the GEMM-I class, RBF biosensors for cyclic di-GMP were made to be 4-fold brighter and 30-fold faster than an initial biosensor designed from the structurally characterized *Vibrio cholera* riboswitch (Wang et al., 2016).

Our phylogenetic approach demonstrates that sequence changes to the riboswitch aptamer can greatly affect biosensor performance, including turn-on kinetics. However, for widespread riboswitch classes, it is impractical to test every possible phylogenetic sequence. As this initial library is pared down for screening, better-performing phylogenetic variants may exist but are not selected and tested. A generalizable strategy to tune kinetics of validated biosensors would serve as an attractive alternative than re-screening phylogeny. Additionally, enhancements within the riboswitch aptamer of RBF biosensors may also be transferrable to other riboswitch-based tools like reporters and aptazymes.

Optimization of activation speeds for biosensors is important to allow for dynamic monitoring of sensitive changes which can be overlooked in conventional experimental assays. For instance, if a biological process occurs on the order of minutes but a sensor fully activates within hours, it is not an accurate depiction of the dynamics. A riboswitch reporter for 2,4-dinitrotoluene tested in cells achieved a maximal fluorescent response in 5-14 h (Davidson et al., 2012; Harbaugh et al., 2017) but was shown that it could respond as fast 30 min in a cell-free transcription-translation system after optimization (Espah Borujeni et al., 2016). Phylogenetic RBF biosensors for cyclic di-GMP elicit a maximal response in 10 min *in vitro* with the fastest sensor possessing a half-maximal response times ($t_{1/2}$) of ~1 min. Previously, we have shown that these fast sensors were necessary to monitor changes (on the order of minutes) in cyclic di-GMP levels upon zinc depletion in *E. coli* (Yeo et al., 2018).

In this work, we explore a novel strategy to further tune RBF biosensor performance by systematically mutating conserved base-pairs in the riboswitch. To test this proof-of-principle, we developed a novel biosensor for the ligand guanidine using the recently discovered guanidine-I riboswitch (Nelson et al., 2017) as our model. Two novel designs were developed for this biosensor using insights from our lab's previous approaches. These functional designs were then utilized to screen a phylogenetic

biosensor library. Two phylogenetic biosensors were selected as scaffolds to generate a library of biosensor mutants to observe their effects on biosensor kinetics. These guanidine biosensors were then tested for their performance in live cells to determine if *in vitro* activation speeds could correlate to *in vivo* response times.

3.3 Results

Screen and optimization of topologies for functional guanidine biosensors

Guanidine riboswitches serve as attractive model candidates to develop into a biosensor for understanding kinetics *in vitro* or in live cells, as it has been shown that cells can uptake guanidine when it is exogenously added to the media. A biosensor for this target is highly desirable as it has just been established as a newly discovered bacterial metabolite and pathways to its accumulation are still largely unknown (Nelson et al., 2017). Beyond a biological role, guanidine and close derivatives such as nitroguanidine are known to be environmental pollutants and their detection in contaminated samples is of large interest (Panahi et al., 2016).

In the context of adapting a riboswitch into an RBF biosensor, adjacent 5' and 3' termini and stem loops that do not interact with the ligand are critical as these areas can be used to create transducer modules which communicate with the fluorogenic aptamer. The first class of riboswitches for this ligand, guanidine-I, possessed several advantages for biosensor adaptation over related classes such as the guanidine-II and guanidine-III riboswitches (Sherlock et al., 2016; Sherlock and Breaker, 2017). The overall structural folds for the guanidine-I riboswitch (Reiss et al., 2017) use only one out of three helical stem loops to form a ligand binding pocket compared to the guanidine-II type (Huang et al., 2017b) which utilizes both pairing stems loops to bind guanidine. Comparably, the guanidine-III possesses a complex pseudoknot involving its terminal ends and possess a stem loop which is involved with a RNA triplex to bind the ligand (Huang et al., 2017a).

While the guanidine-I class was the best candidate relative to the other riboswitch classes, two key challenges in biosensor design were pinpointed with further analysis of the *S. acidophilus* guanidine-I crystal structure (PDB:58T3). First, the 5' and 3' terminal ends of this riboswitch were quite distant, spanning almost 26 Å from one another. Previously, RBF biosensors from the cyclic di-GMP (PDB:3IRW) and SAM-I (PDB:4KQY) riboswitches possessed termini distances of 10 – 12 Å when grafting a fluorogenic aptamer or stem loop. Second, the structure outlined key tertiary contacts between the P1 and P3 helices mediated by guanidine that needed to be maintained but noted that the P1 stem loops were not directly involved in this interaction. These design challenges required modification of our pre-existing repertoire of design principles. Thus two new topologies were designed termed "Junction (J)" and "Linker (L)" to create a functional guanidine biosensor (**Fig. 3.1A**).

The junction topology uses an artificial transducer stem that creates an architecture reminiscent of a 3-way junction. This architecture is observed commonly in natural riboswitches such as the cyclic-di-GMP or guanine classes (Batey et al., 2004; Smith et al., 2009). We surmised that transducer stems from other RBF biosensors our lab has previously developed would allow them to retain their native switching

properties and form the basis of an artificial P0 stem which would communicate and be fused to the fluorogenic aptamer Spinach2 (**Fig. 3.1B**)

The linker topology instead utilizes an approach demonstrated in Chapter 2 (Truong et al., 2018), where circularly permuting riboswitches by connecting the 5' and 3' termini allow for the P1 to be fused to Spinach2. We hypothesized that guanidine could better enable tertiary interactions between P1 and P3 helices that would stabilize the transducer module and allow for fluorescent activation (**Fig. 3.1C**). While past biosensors use a stem loop to connect the terminal ends, here a flexible linker is necessary as the terminal ends of this riboswitch are much farther apart.

For these topologies, we envisioned a general design starting from an experimentally validated riboswitch sequence from the organism *Klebsiella pneumoniae* (Kpn). We hypothesized that the issue of the distant termini could be resolved by adding adenosine spacers to each end of the P0 stem in the junction topology and a poly-Adenosine (poly-A) tether in the linker design (**Fig. 3.2A,B**). This approach was inspired by previous work of Jewett and co-workers where ribosomal subunits were tethered using a poly-A linkers to span a 30-40 Å gap (Orelle et al., 2015). With this in mind, an initial set of 18 junction biosensors (9 different lengths spanning 1-3 adenosine residues x 2 different P0 stems) and 4 linker biosensors with 4 different poly-A tether lengths were screened for fluorescence activation in response to guanidine (**Fig. 3.2C**). Indeed, we observed 5 junction biosensors with over 1.2-fold activation in the presence of 1 mM guanidine. For the linker biosensors, we observed a constitutive, fluorescent signal greater than the parent aptamer Spinach2. This suggested that these constructs possessed an overly stabilized transducer module that was not dependent on guanidine binding and did not follow the mechanism we desired (**Fig. 3.2C**).

An optimization approach for both topologies was carried out to investigate other functional P0 stems for the junction design and to determine if destabilizing the transducer modules of the linker design could yield functional sensors (**Fig. 3.3A**). Sampling other artificial transducer stems could yield other functional junction biosensors with improved fold-turn on. Previous junction sensors were tested with the "A" and "B" stem identities and now three additional stems (C-E) were tested but only with spacer lengths that appeared functional (higher than 2 adenosine spacers) totaling 12 additional junction biosensors. Three out of the 12 constructs possessed a fold-turn greater than 1.4x and added the "C" and "D" stem identities to the functional repertoire (**Fig. 3.3B**). For the linker optimization, two constructs, L-4 and L-5 were chosen to create a series of 7 stem truncations which generated 14 linker biosensors to test. The truncations were achieved by deleting residues and base-pairs in the Spinach2 aptamer as to not disturb tertiary interactions of the P1 helix necessary to contact the P3 helix and bind guanidine (**Fig. 3.3A**). Screening of the truncated linker biosensors yielded 4 out of 14 constructs with greater than 1.2 fold-activation and identified 3 stem truncations that enabled ligand-responsive biosensors.

Sensitivity and selectivity of guanidine biosensors

To compare junction and linker biosensor designs, constructs Kpn J-,2,2-A and L-5 Tr 4 were chosen as representative models and experimental binding affinities were determined for both constructs (**Fig. 3.4A**). J-2,2-A and L-5 Tr 4 were measured to have dissociation constants (K_d) of 60 μM and 6.8 mM, respectively. In comparison, the

riboswitch aptamer alone was reported to have a dissociation constant of 20 μM based on in-line probing experiments (Nelson et al., 2017a). This result shows that linker designs have almost 100-fold poorer affinity than the junction topology. To understand the selectivity of these biosensors, Kpn J-2,2-A was tested against various structural analogs of guanidine. At 10 mM concentrations of each analog, there is slight fold-activation for aminoguanidine and agmatine but this response was much higher with the cognate ligand guanidine (**Fig. 3.4B**). Assuming the dissociation constant for aminoguanidine is at least 10 mM, (as it exhibits half-maximal fluorescence response of the biosensor at this concentration) these results suggest that the selectivity for guanidine over close analogs is greater than ~166-fold (10 mM / 60 μM). To probe if fluorescence activation required the functional riboswitch, we generated a mutation in the biosensor (J-2,2-M) that was previously shown to disrupt regulation of the riboswitch reporter (Nelson et al., 2017a). The G-to-C mutation displayed no fluorescence activation in the presence of guanidine (**Fig. 3.4B**).

Phylogenetic screen using functional topologies

Using guanidine-I riboswitches from 5 different organisms, a phylogenetic library for both topologies was designed incorporating previously derived empirical rules. We have shown that sampling riboswitches from diverse phylogeny can generate highly fluorescent and well-folded RNA biosensors (Wang et al., 2016). The phylogenetic junction library consists of three functional transducer stems (A,C,D) with four various spacer lengths giving a total of 60 new biosensors. As for the phylogenetic linker library, two stem truncations Tr 4 and Tr 5 were arbitrarily chosen to move forward using two poly-A lengths (L-4 and L-5), which provide 20 additional constructs. All 80 phylogenetic biosensors were screened for a response to 10 μM and 1 mM guanidine (**Fig. 3.7**) with a total of 14 hits which gave a fold-activation greater than 1.4 (**Fig. 3.6B**). An initial time-course for fluorescence activation of representatives Kpn J-2,2-A and Dru J-2,2-D are shown to demonstrate the phylogenetic differences between biosensors (**Fig. 3.7C**).

Mutational base-pair shuffling to tune *in vitro* kinetics

A mutational swapping strategy was then devised to mutate base pairs within pairing elements P1, P1b, and P2 at locations where the nucleotide identity is not conserved (**Fig. 3.7A**). This analysis was done separately for both Kpn J-2,2-A and Dru J-2,2-D where the identified base-pair positions were subjected to transposing the base identities or completely swapping them for an alternative purine and pyrimidine (**Fig. 3.7B**). In order to measure these subtle kinetic changes, our biosensor assay was modified to incorporate automated injection of the ligand and subsequent fluorescence detection at 0.5 s. A summary of the effects on the activation times for Dru J,2-2-D mutant biosensors is depicted and compiled (**Fig. 3.7C,D Table 3.1**). Overall, each biosensor had very different effects from the transpose and purine-pyrimidine (Pur-Pyr) swap libraries. The Dru J-2,2-D library withstood both type of mutations reasonably well with a higher percentage of Pur-Pyr swaps being functional mutants. The biggest observed improvement in turn-on kinetics was for the C6U-G44A mutant, which had a faster $t_{1/2}$ than WT by ~30 sec (**Fig. 3.7F**). For Kpn J-2,2-A, all Pur-Pyr swapping mutants resulted in non-functional biosensors (data not shown). In most cases, the Kpn

mutant biosensors appear maximally fluorescent independent of ligand addition or is below a reliable threshold for the biosensor (less than 1.2x fold activation).

In vivo testing and dynamics of biosensors with guanidine and analogs

The enhancement of biosensor kinetics *in vitro* could also be transferrable within a live cell setting. To verify that our biosensors work *in vivo*, Kpn J-2,2-A, its corresponding mutant Kpn J-2,2-M, and Spinach2 were appended with a tRNA scaffold and inserted into a pET31b plasmid (Kellenberger et al., 2015b). These plasmid constructs were then transformed into *E. coli* cells which were grown in auto-induction media overnight to express the encoded RNA construct. Each sample was pre-incubated in spent media containing DFHBI and later guanidine was added and samples were analyzed by flow cytometry. We expected the addition of guanidine to Kpn J-2,2-A to increase overall fluorescence, but we were surprised to observe that same change with Spinach2 and Kpn J-2,2-M, as they should not be responsive to the ligand (**Fig. 3.8A**). Representative histograms for each sample with guanidine added did not display any abnormal effects towards cell viability or distribution (**Fig. 3.8B**).

Due to variability between biological replicates, mean fluorescent intensity (MFI) values does not fully capture the total change displayed upon ligand addition. To better show this effect, we report the data as ΔMFI/MFI, which calculates the difference between the two MFI values at each time point ($MFI_{+guanidine} - MFI_{no\ ligand}$) and dividing by the $MFI_{+guanidine}$. With this, we tested the dynamic response of the Kpn J-2,2-A biosensor and Spinach2 aptamer when guanidine or related structural analogs were added to the media (**Fig. 3.8C**). Interestingly, urea was the only ligand that did not cause a change in Spinach2 or Kpn J-2,2-A fluorescence when added to the media while agmatine shows the biggest signal change. Comparing the structures of the tested ligands, it is apparent that the guanidine moiety and/or cationic character is responsible for this non-specific increase of fluorescence.

Additional testing of guanidine biosensors in phosphate-buffer-saline (PBS) media for an extended time-course of 2 hours displayed a greater overall signal change than using spent media (**Fig. 3.9A, 3.9B**). The use of PBS media lowers the overall background fluorescence and makes the fold-change with the addition of the ligand more noticeable. A small palette of phylogenetic biosensors were also constructed for *in vivo* flow cytometry assays. (**Fig. 3.9C**). While Spinach2 exhibits a ~2x non-specific fluorescence increase in the presence of guanidine, the functional biosensor constructs possess a >2-fold activation ratio which is likely attributed to a ligand-specific interaction with the biosensor (**Fig. 3.9D**).

3.4 Discussion

This study demonstrates the first instance of an RNA-based biosensor that can detect guanidine through two biosensor topologies developed for accommodating the guanidine-I riboswitch. As a chaotropic agent, guanidine is widely detected *in vitro* by HPLC and UV-Vis methods but our work displays the first biological sensor to do so.

While initial design junction and linker designs were inspired by natural RNA motifs and structural considerations, they both required intensive empirical screening to obtain functional sensors with varying brightness and affinities. Additionally, optimization based off a singular riboswitch from *Klebsiella pneumoniae* were found to be transferrable to sequences from other organisms and more applicable to the phylogenetic junction library (13/60, 26.6% hit rate) compared to the linker design (1/20, 5% hit rate). This is likely attributed to junction designs preserving the natural riboswitch sequence while the linker design truncations may have been more detrimental to the riboswitch folding than we initially proposed.

This development of a guanidine biosensor also showcases a new mutational swapping strategy to tune biosensor speeds through profiling base-pair interactions. Transposing base-pair residues (e.g. A-U to U-A) is a method commonly used to probe if pairing elements interact with the ligand and/or are necessary to mediate gene regulation by the riboswitch. A LacZ reporter using the *B.* subtilis guanidine-I riboswitch was still able to regulate gene expression after transposing two U-A base-pairs in the P1 stem (Nelson et al., 2017a). While these mutations compensate to still allow for riboswitch regulation, it has never been tested if they can affect the rate of riboswitch folding. To our knowledge, swapping the purine and pyrimidine identity of a base-pair (e.g. A-U to G-C) while maintaining its original position in structural pairing elements has never been employed in the study of riboswitches and RNA folding. It is interesting to note that the P1 and P2 pairing elements are predicted to be well-folded in the absence of ligand (Reiss et al., 2017) and aside from long-range tertiary interactions, they do not directly interact with guanidine. However, a large proportion of mutations render the biosensors highly fluorescent independent of guanidine which suggests that these mutations on the P1 and P2 have secondary effects on P3 stem or the P1-P3 interface which directly contact the ligand. Overall, these mutations yielded faster variants with a $t_{1/2}$ values of 2.68 – 3.28 min and slower variants with a $t_{1/2}$ value greater than 4.33 min.

The *in vivo* dynamics of our guanidine biosensors are hindered by the non-specific fluorescence of the Spinach2 aptamer when guanidine is added. This important control demonstrates that there is a secondary effect of guanidine in our flow cytometry assay that is independent of biosensor function. One likely hypothesis is that the cationic nature of the ligand could drive the cell permeability of the pro-fluorescent dye DFHBI. In media around a pH of 7, DFBHI can exist in a deprotonated hydroxy anion which could be interacting electrostatically with the positively-charged guanidine. This would explain why ligands with more cationic character such as agmatine exhibits a higher non-specific increase than related structural analogs (**Fig 3.7C**). It has been demonstrated that incorporating the guanidino moiety on a number of negatively charged bio- or small molecules can increase their overall cellular uptake (Stanzl et al., 2013). While a definitive mechanism has not been established, it is hypothesized that guanidine group can interact with negatively charged moieties, neutralizing their anionic character and making them more cell permeable. This finding further corroborates our hypothesis and poses a huge limitation for screening mutational enhancements which alter riboswitch folding speeds *in vivo*. While it is likely there is some fraction of fluorescence increase that is specific to interaction of the ligand with the biosensor, this can be difficult to deconvolute when investigating cellular activation times. However,

this could potentially be overcome if mutant variants were constructed into a riboswitch reporter rather than an RBF biosensor.

3.5 Conclusions/Future Directions

Overall, the development of a novel RBF biosensor for guanidine utilizes established biosensor approaches combined with important structural insights to create two novel biosensor designs. To our knowledge, there are no instances where RNA-based fluorescent biosensors have been investigated systematically using chimeric transducer stems from other natural riboswitches. Our success with adapting the guanidine-I riboswitch into a biosensor gives rise to the ability to tailor one's approach in generating RBF biosensors for any target ligand. Additionally, these biosensors were used as the first case-study to probe riboswitch kinetics through mutational swapping of conserved pairing elements and further demonstrates an alternative approach to improve biosensor speed and performance.

3.6 Figures

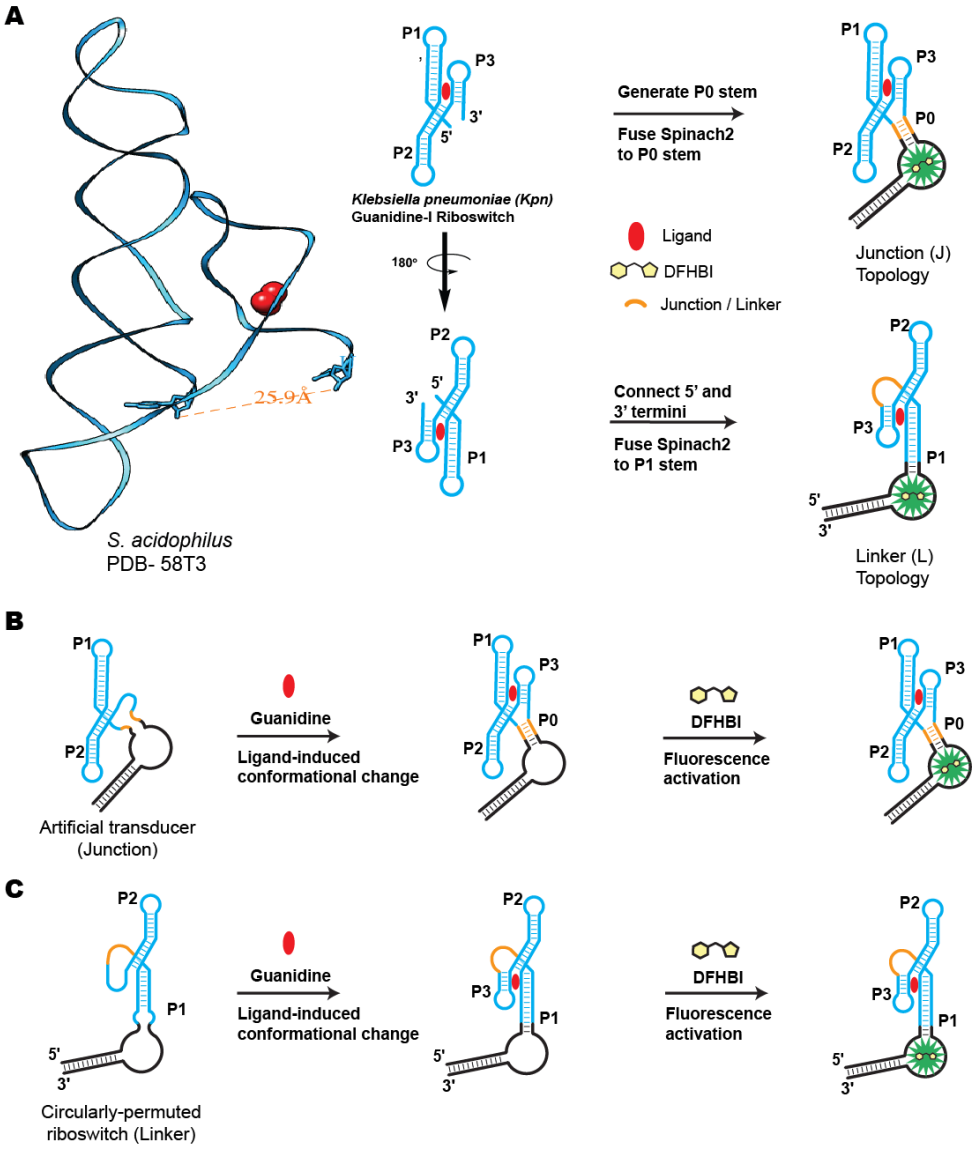


Figure 3.1. Initial design and proposed mechanisms of guanidine-I biosensors. (a) Design of two different topologies for a guanidine biosensor utilizing a 3-way junction (J) or a circularly permuted riboswitch with a poly-A linker (L) from the *Klebsiella pneumoniae* (Kpn) guanidine-I riboswitch. (b) Proposed mechanism of junction biosensors which utilizes an artificial transducer stem (orange) that is able to form upon the presence of ligand (red). (c) Proposed mechanism of linker biosensors which utilize a circularly-permuted riboswitch with a flexible linker region (orange) that allows for the ligand (red) to mediate tertiary interactions between the P1 and P3 helices.

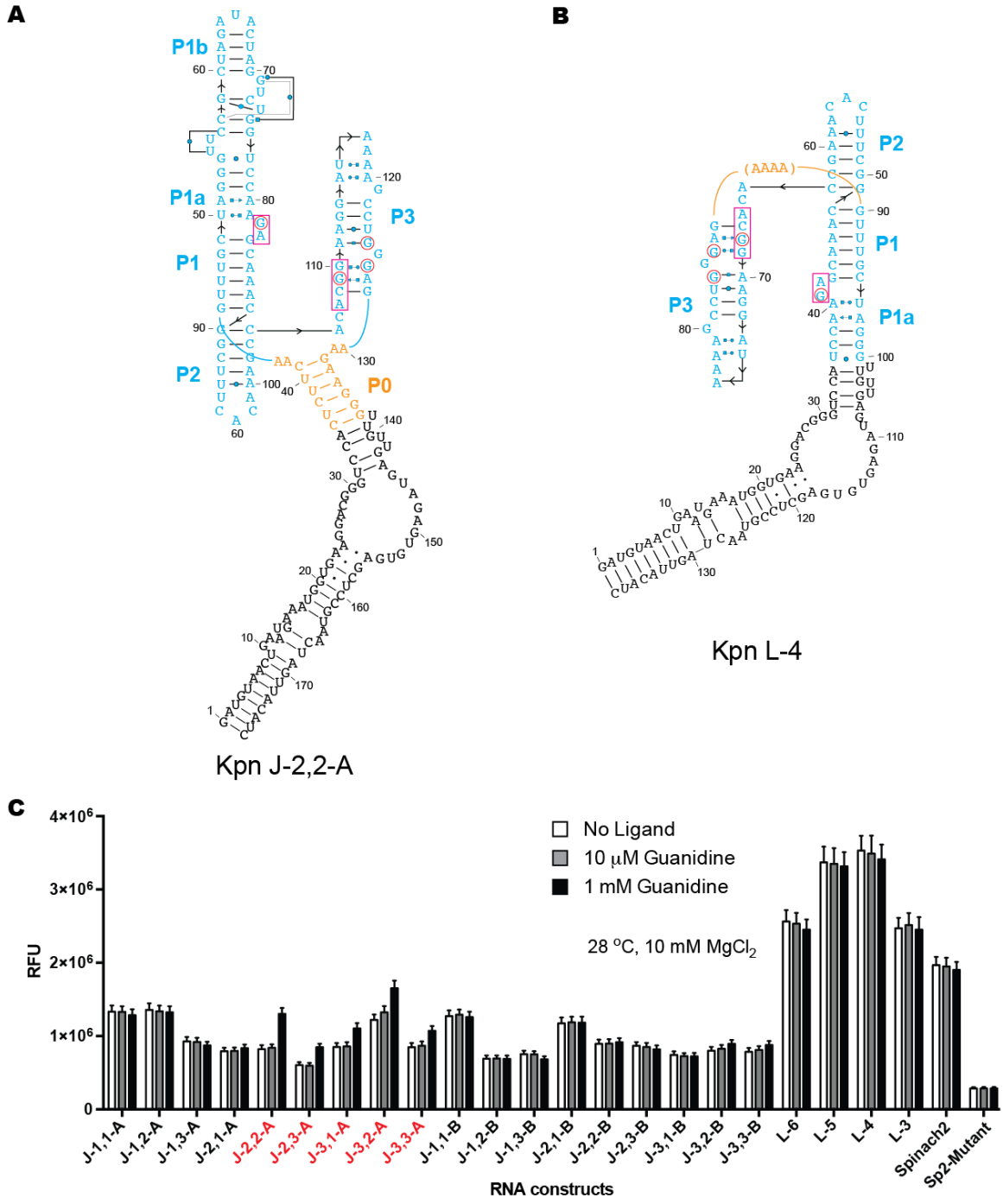


Figure 3.2. Secondary structures models and screening of guanidine-I biosensors. (a) Secondary structure models of a representative junction biosensor with 2 adenosine spacers on each end of the P0 stem. (b) Secondary structure models of a representative linker biosensors with a poly-A linker composed of 4 adenosine residues. Residues in red circles directly contact the ligand guanidine. (c) *In vitro* fluorescence response of biosensors to the ligand, guanidine. The nomenclature for junction (J) biosensors denotes the number of adenosines in the sequence, and the identity of the P0 stem, e.g. J-2,3-A stands for a junction biosensor that uses 2 adenosine spacers closer to the 5'-end and 3 adenosine spacers closer to the 3'-end, and uses stem sequence "A". The nomenclature

for linker (L) biosensors denotes the number of adenosines in the linker region. Guanidine biosensors with greater than 1.2x fluorescence increase at 1 mM guanidine) are indicated in red. Data shown are mean values ± standard error of the mean taken from two independent replicates.

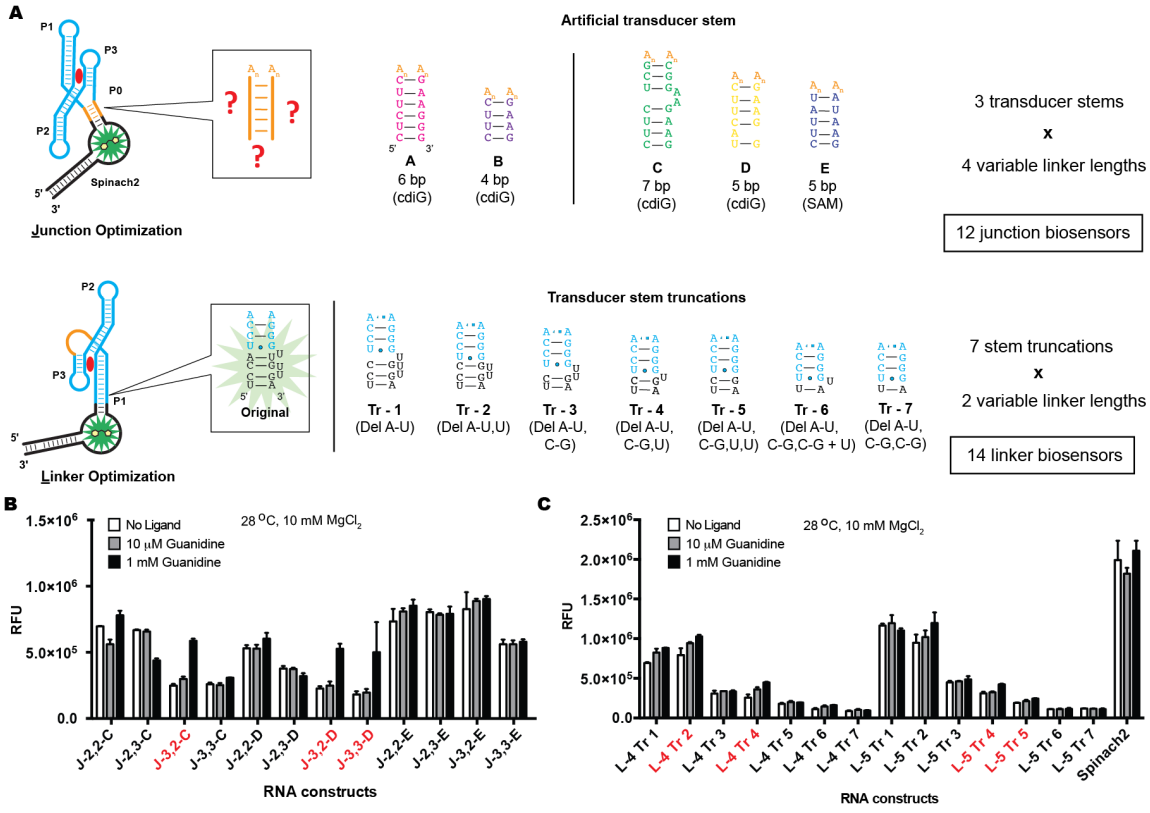


Figure 3.3. Sub-optimization of junction and linker biosensors for guanidine. (a) Optimization of junction biosensors screening for other functional artificial stem sequences. Previous linker constructs possessed overly stable transducer stems so subsequent truncations were tested to create functional sensors (b) *In vitro* fluorescence response of optimized junction biosensors to the ligand, guanidine. The nomenclature for junction (J) biosensors denotes the number of adenosines in the sequence, and the identity of the P0 stem. Guanidine biosensors with greater than 1.4x fluorescence increase at 1 mM guanidine) are indicated in red. (c) *In vitro* fluorescence response of optimized linker biosensors to the ligand, guanidine. The nomenclature for linker (L) biosensors denotes the number of adenosines in the linker region and the identity of the stem truncation. Guanidine biosensors with greater than 1.2x fluorescence increase at 1 mM guanidine) are indicated in red. Data shown are mean values ± standard error of the mean taken from two independent replicates.

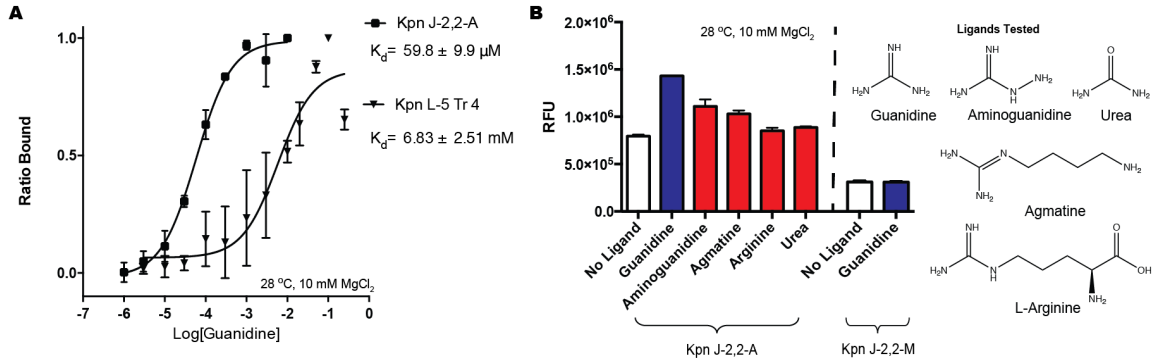


Figure 3.4. Sensitivity and selectivity of guanidine biosensors.
(a) Binding affinity determination of representative junction and linker biosensors. (b) *In vitro* fluorescence response of Kpn J-2,2-A and Kpn J-2,2-A Mutant (J-2,2-M) to guanidine (blue) and structural analogs (red) at 10 mM concentrations of ligand. Data shown are mean values ± standard error of the mean taken from at least two independent replicates.

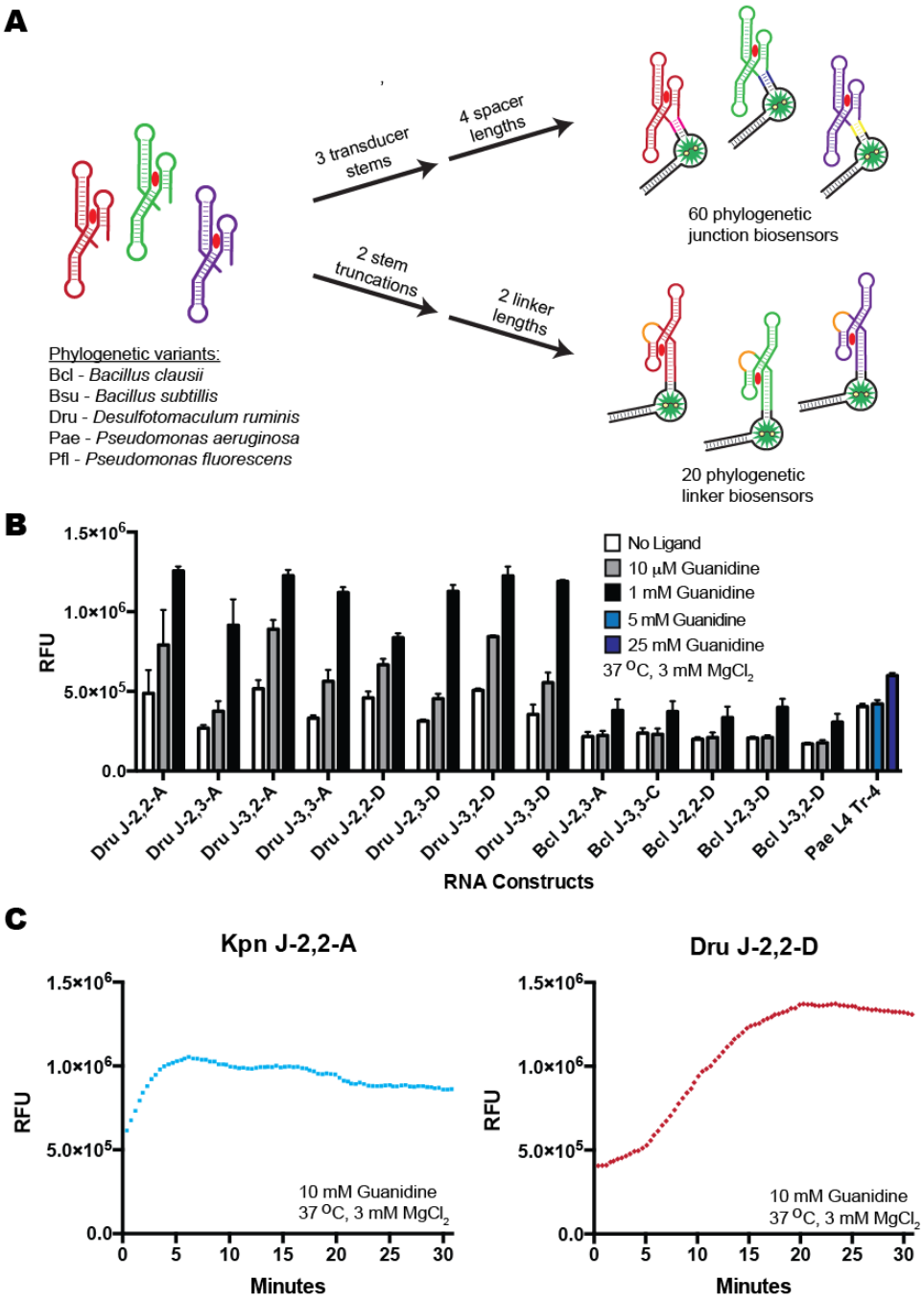


Figure 3.5. Phylogenetic library screening to obtain functional biosensors with differing properties.

(a) Phylogenetic junction library (60 total biosensors) and a linker library (20 total biosensors) from 5 different riboswitch sequences were constructed using empirical evidence from previously tested biosensors. (b) Selected phylogenetic hits with over 1.4x fold activation at concentrations of 1 mM guanidine. Data shown are mean values ± standard error taken from three independent replicates. (c) Comparison of kinetic turn-on for Kpn J-2,2-A and Dru J-2,2-D biosensors. Data shown are mean values from three independent replicate with error bars omitted for clarity.

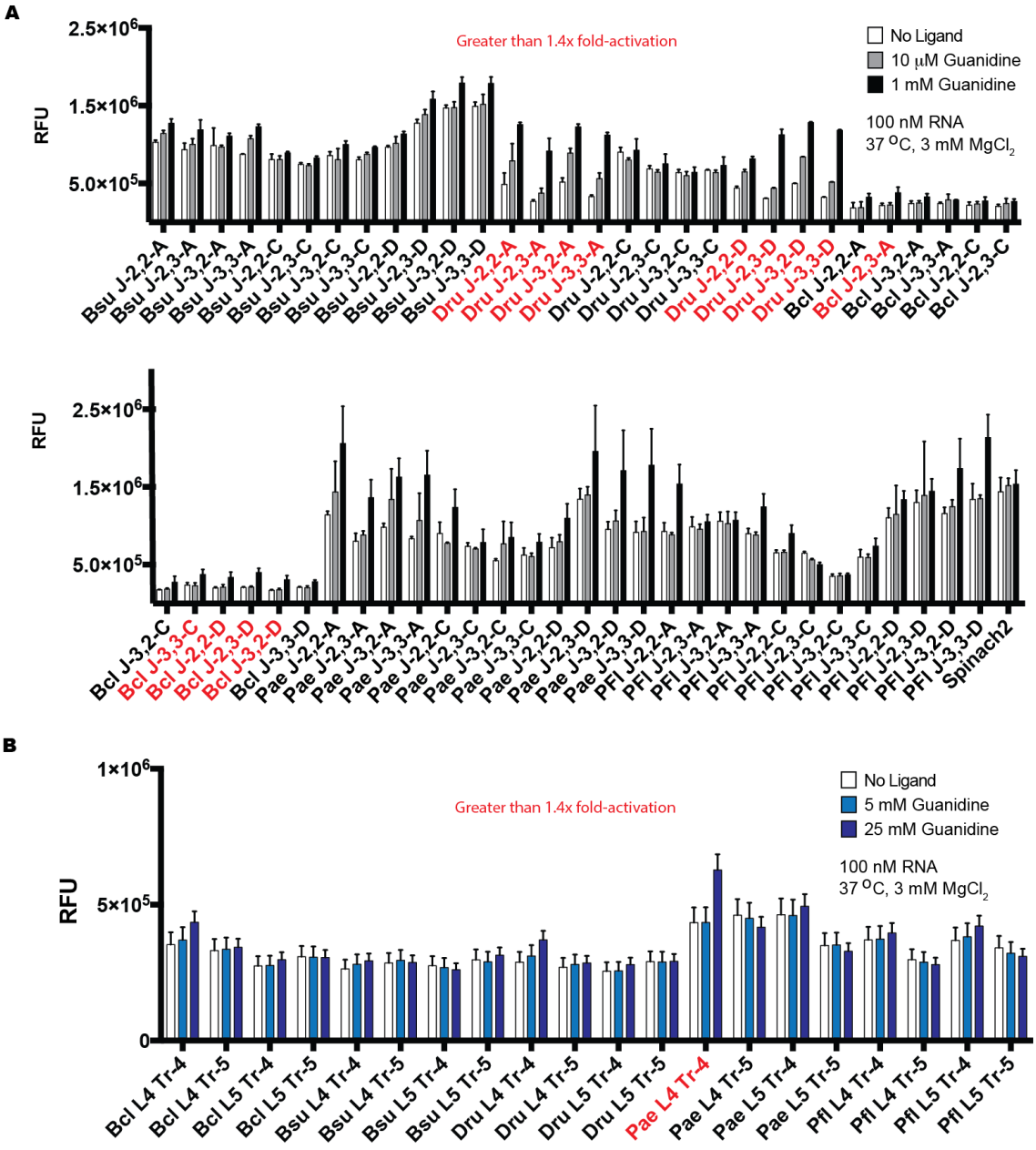


Figure 3.6. Screen of phylogenetic junction and linker biosensor libraries. (a) Phylogenetic junction library consisting of 60 total biosensors from 5 riboswitch sequences with 13 constructs with greater than 1.4x fold activation (red) . (b) Phylogenetic linker library consisting of 20 total biosensors from 5 riboswitch sequences with 1 construct with greater than 1.4x fold activation (red). Data shown are mean values ± standard error taken from three independent replicates

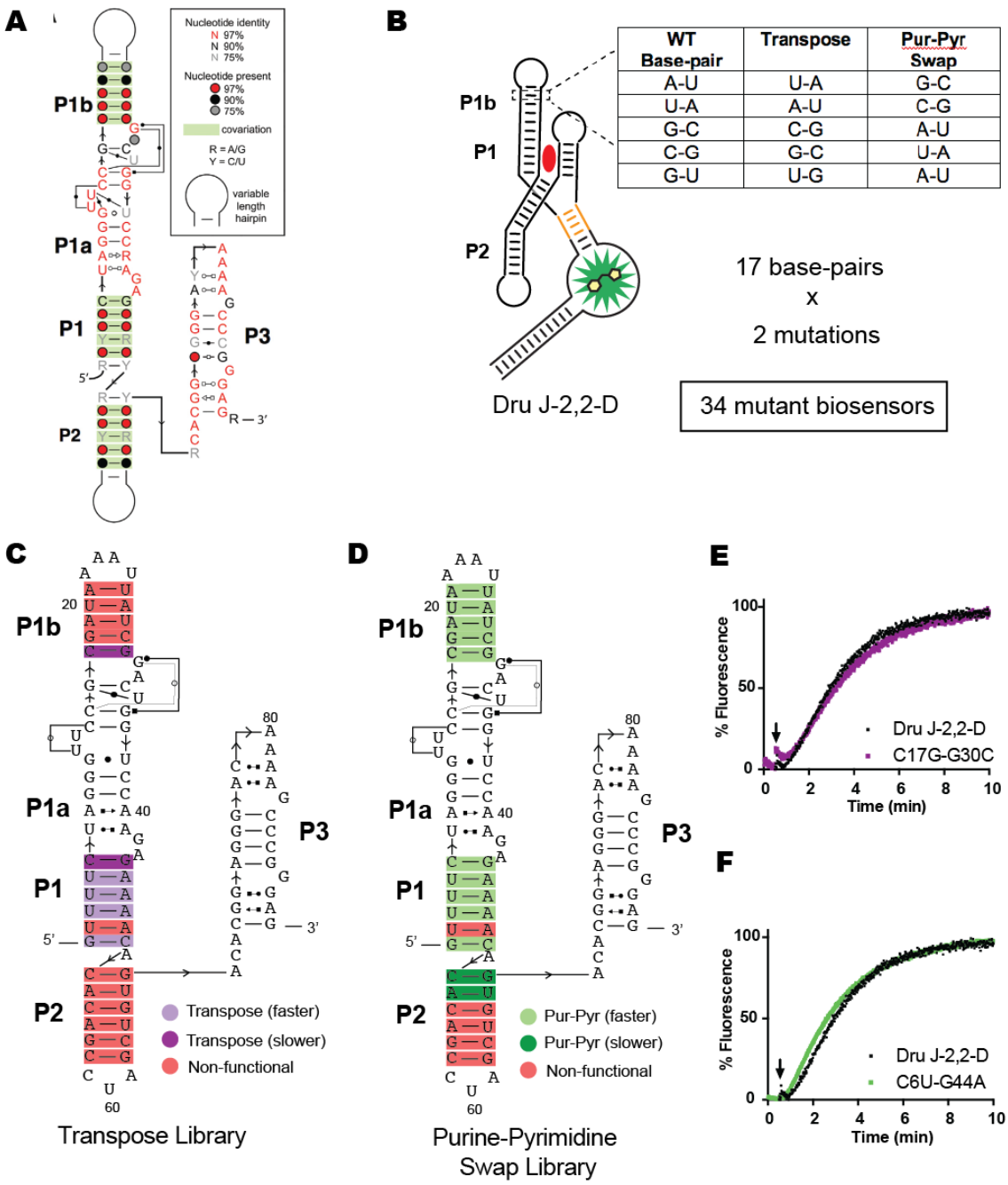


Figure 3.7. Mutational swapping strategy to tune biosensor kinetics.
(a) Consensus sequence of the guanidine-I riboswitch class detailing areas with high sequence conservation Reproduced with permission (Reiss et al., 2017). (b) Design of mutational swaps made in the P1, P1b, and P2 pairing elements for the Dru J-2,2-D biosensor. (c) Transpose library effects on the Dru J-2,2-D biosensor outlined by designated color coding. (d) Purine-pyrimidine swap library effects on the Dru J-2,2-D biosensor outlined by designated color coding. (e) Representative transpose mutant C17G – G30C with a slightly lower $t_{1/2}$ value. (f) Representative purine-pyrimidine swap mutant C6U – G44A with a slightly higher $t_{1/2}$ value. Data shown are mean values taken from two independent replicates with error bars omitted for clarity.

Table 3.1 . Summary of effects of mutational swapping on Dru J-,2-2-D biosensors. ">" means the apparent $t_{1/2}$ is estimated to be greater than the value because in the time-course used, the biosensor did not reach maximum activation.

	Transpose			Purine-Pyrimidine Swap		
WT Base-pair	**Mutation**	**Fold-Activation**	**$t_{1/2}$ (min)**	**Mutation**	**Fold-Activation**	**$t_{1/2}$ (min)**
Natural	None	2.7	3.15	None	2.7	3.15
G1-C49	C1-G49	3.9	>3.07	A1-U49	2.6	>3.11
U2-A48	A2-U48	1.5	3.04	C2-G48	1.4	3.13
U3-A47	A3-U47	2.5	2.81	C3-G47	2.4	3.00
U4-A46	A4-U46	2.4	2.88	C4-G46	2.9	2.91
U5-A45	A5-U45	1.9	3.12	C5-G45	2.9	3.00
C6-G44	G6-C44	3.1	>3.46	U6-A44	5.1	2.68
C17-G30	G17-C30	2.0	3.28	U17-A30	2.9	2.82
G18-C29	C18-G29	1.1	N/A	A18-U29	2.1	2.74
A19-U28	U19-A28	1.1	N/A	G19-C28	2.1	2.87
U20-A27	A20-U27	1.1	N/A	C20-G27	2.3	2.95
A21-U26	U21-A26	1.0	N/A	G21-C26	2.3	2.94
C53-G67	G53-C67	1.1	N/A	U53-A67	4.4	>3.64
A54-U66	U54-A66	1.0	N/A	G54-C66	2.3	>3.92
C55-G65	G55-C65	1.0	N/A	U55-A65	0.9	N/A
A56-U64	U56-A64	0.8	N/A	G56-C64	0.8	N/A
G57-C63	C57-G63	0.8	N/A	A57-U63	3.0	>4.33
C58-G62	G58-C62	0.8	N/A	U58-A62	0.9	N/A

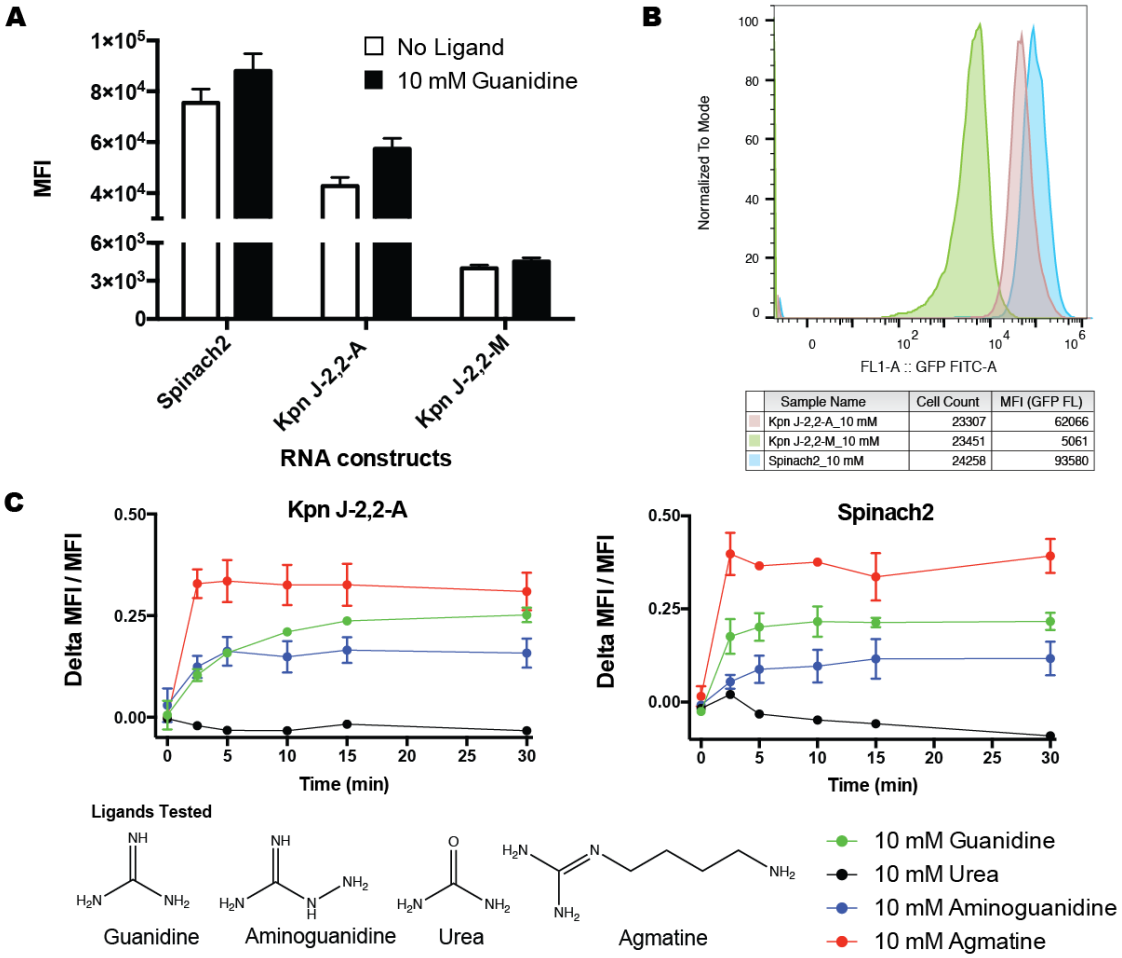


Figure 3.8. *In vivo* testing of guanidine biosensors with guanidine and structural analogs.
(a) Live cell fluorescence measured by flow cytometry for *E. coli* BL21* cells expressing plasmid encoding RNA constructs and preincubated in spent media containing DFHBI before adding guanidine at the indicated concentration. Mean fluorescence intensity was determined by analyzing 30,000 cells per replicate after 30 min of incubation with ligand. All error bars represent standard error between 3 biological replicates. (b) Representative flow histograms for cells expressing biosensors or controls with 10 mM ligand added to the media. (c) Dynamics of Kpn J-2,2-A and Spinach 2 with guanidine and analogs using flow cytometry. Delta MFI / MFI values were determined after analyzing 30,000 cells per each timepoint with samples with or without ligand. All error bars represent standard error between 3 biological replicates

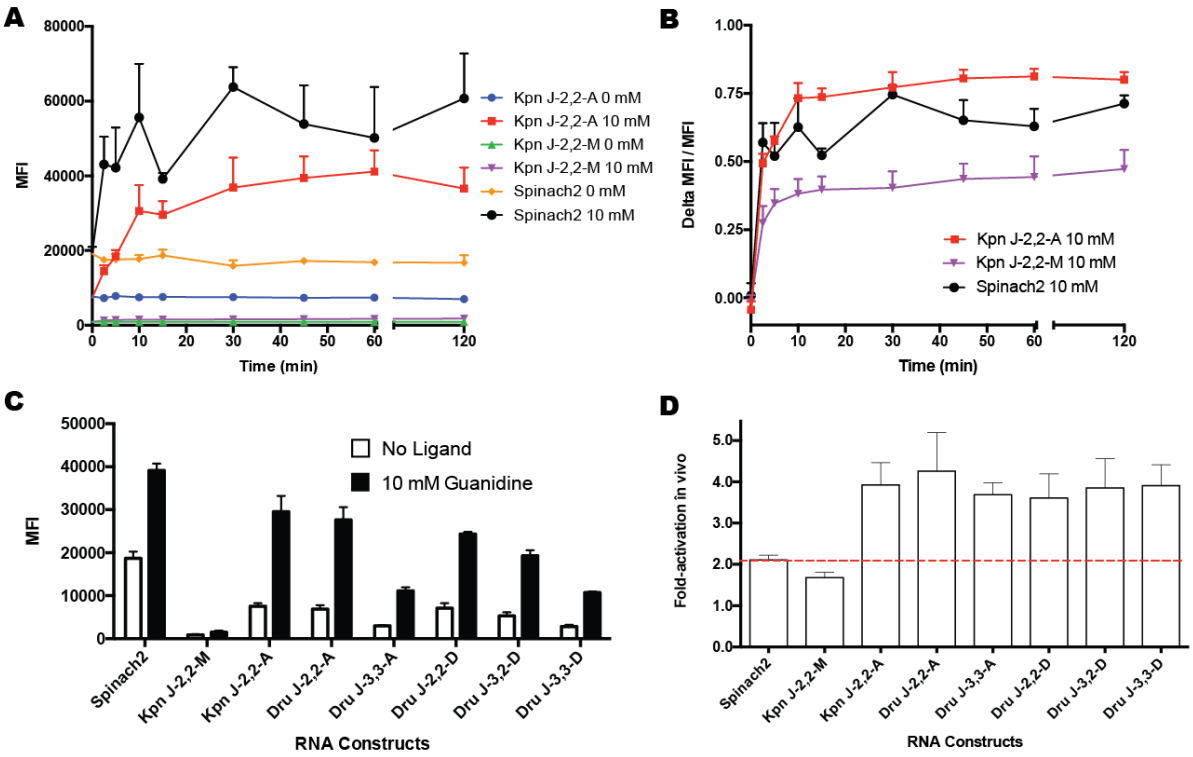


Figure 3.9. Extended *in vivo* dynamics and testing of guanidine biosensors in PBS media.

(a) Dynamics of Kpn J-2,2-A , Kpn J-2,2-M and Spinach 2 with guanidine over a 2-hour time course in PBS media with raw MFI values. (b) Identical to part (a) but plotting delta MFI / MFI values. (c) Live cell fluorescence measured by flow cytometry for *E. coli* BL21* cells expressing plasmid encoding RNA constructs and preincubated in PBS media at pH 7.0 containing DFHBI and taken after incubating with ligand for 15 min. (d) Quantitation of fold-activations using MFI values with and without ligand. Red dashed line indicates the level of non-specific increase observed by the Spinach2 aptamer. Values were determined after analyzing 30,000 cells per each timepoint with samples with or without ligand. All error bars represent standard error between 3 biological replicates.

Table 3.2. **Representative junction biosensor sequences used in this study**. **Bold** sequences indicate the Spinach2 sequence, which flanks the guanidine-I riboswitch sequence used. Orange indicates the artificial transducer stem or adenosine spacers for junction biosensors. *Underlined* indicates the adenosine spacers which are not part of the stem.

Name	Sequence (5' to 3')
Kpn J-2,2-A	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaGCTGGCTAGGGTTCCGGTTCACCG CGGTGAACGTCTGGTCCGAGAGCTGGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGG GAGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Kpn J-2,2-M (G to C)	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaGCTGGCTAGcGTTCCGGTTCACCGC GGTGAACGTCTGGTCCGAGAGCTGGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGGG AGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Kpn J-2,2-B	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttcaaGCTGGCTAGGGTTCCGGTTCACCGC GGTGAACGTCTGGTCCGAGAGCTGGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGGG AGaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Kpn J-2,2-C	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttctcgaaGCTGGCTAGGGTTCCGGTTCACC GCGGTGAACGTCTGGTCCGAGAGCTGGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCG GGAGaacgaaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Kpn J-2,2-D	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tacttcaaGCTGGCTAGGGTTCCGGTTCACCG CGGTGAACGTCTGGTCCGAGAGCTGGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGG GAGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Kpn J-2,2-E	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttataaGCTGGCTAGGGTTCCGGTTCACCGC GGTGAACGTCTGGTCCGAGAGCTGGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGGG AGaaataag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bcl J-2,2-A	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaTGCGCCTAGGGTTCCGCTTCATTTG TAAGGGCTGGTCCGAGAGGTGCACACGGCGTCTGCCGTGACACGGAGGGATAAAAGCCCGGGAG aagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bcl J-2,2-C	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttctcgaaTGCGCCTAGGGTTCCGCTTCATTT GTAAGGGCTGGTCCGAGAGGTGCACACGGCGTCTGCCGTGACACGGAGGGATAAAAGCCCGGGA Gaacgaaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bcl J-2,2-D	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tacttcaaTGCGCCTAGGGTTCCGCTTCATTTG TAAGGGCTGGTCCGAGAGGTGCACACGGCGTCTGCCGTGACACGGAGGGATAAAAGCCCGGGAG aagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bsu J-2,2-A	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaCTCTTCAAGTTTTCTAGGGTTCCGC ATGTCAATTGACATGGACTGGTCCGAGAGAAAACACATACGCGTAAATAGAAGCGCGTATGCACAC GGAGGGAAAAAAGCCCGGGAGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Bsu J-2,2-C	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttctcgaaCTCTTCAAGTTTTCTAGGGTTCCG CATGTCAATTGACATGGACTGGTCCGAGAGAAAACACATACGCGTAAATAGAAGCGCGTATGCACA CGGAGGGAAAAAAGCCCGGGAGaacgaaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTA CATC**
Bsu J-2,2-D	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tacttcaaCTCTTCAAGTTTTCTAGGGTTCCGC ATGTCAATTGACATGGACTGGTCCGAGAGAAAACACATACGCGTAAATAGAAGCGCGTATGCACAC GGAGGGAAAAAAGCCCGGGAGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Dru J-2,2-A	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaGTTTTCTAGGGTTCCGCGATAAAAT TATCGGACTGGTCCAAGAGAAAACACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAG aagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Dru J-2,2-C	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttctcgaaGTTTTCTAGGGTTCCGCGATAAAA TTATCGGACTGGTCCAAGAGAAAACACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGA Gaacgaaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**

Dru J-2,2-D	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tacttcaaGTTTTCTAGGGTTCCGCGATAAAAT TATCGGACTGGTCCAAGAGAAAACACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAG aagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pae J-2,2-A	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaGCCGACTAGGGTTCCGACTCGCTC GCGAGTGGCTGGTCCGAGAGTTGGCGACCTCCAGTGAGGTTACACGGCGGGATAAAAGCCCGGG AGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pae J-2,2-C	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttctcgaaGCCGACTAGGGTTCCGACTCGCTC GCGAGTGGCTGGTCCGAGAGTTGGCGACCTCCAGTGAGGTTACACGGCGGGATAAAAGCCCGGG AGaacgaaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pae J-2,2-D	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tacttcaaGCCGACTAGGGTTCCGACTCGCTC GCGAGTGGCTGGTCCGAGAGTTGGCGACCTCCAGTGAGGTTACACGGCGGGATAAAAGCCCGGG AGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pfl J-2,2-A	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**ctcttcaaGCTGACTAGGGTTCCGGCTCGCTA AGGCGAGTGGCTGGTCCGAGAGTCGGCGACCTCCAGTTGAGGTTACACGGCGGGATAAAAGCCC GGGAGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pfl J-2,2-C	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**cttctcgaaGCTGACTAGGGTTCCGGCTCGCTA AGGCGAGTGGCTGGTCCGAGAGTCGGCGACCTCCAGTTGAGGTTACACGGCGGGATAAAAGCCC GGGAGaacgaaagaag**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pfl J-2,2-D	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tacttcaaGCTGACTAGGGTTCCGGCTCGCTA AGGCGAGTGGCTGGTCCGAGAGTCGGCGACCTCCAGTTGAGGTTACACGGCGGGATAAAAGCCC GGGAGaagaagg**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**

Table 3.3. **Representative linker biosensor sequences used in this study**. **Bold** sequences indicate the Spinach2 sequence, which flanks the guanidine-I riboswitch sequence used. Orange indicates the poly-adenosine linker length for linker biosensors.

Name	Sequence (5' to 3')
Kpn L-4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 1	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCC**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 2	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCC**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 3	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 6	**GATGTAACTGAATGAAATGGTGAAGGACGGGT**tccgagagctggcgacctcggcgaggttacacggcgggataaaagcccgggagAAAAgctggctagggTAGAGTGTGAGCTCCGTAACTAGTTACATC
Kpn L-4 Tr 7	**GATGTAACTGAATGAAATGGTGAAGGACGGGT**tccgagagctggcgacctcggcgaggttacacggcgggataaaagccccgggagAAAAgctggctagggAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Bcl L-4 Tr 4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tccgagaggtgcacacggcgtctgccgtgacacggagggataaaagcccgggagAAAAtgcgcctagggTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Bcl L-4 Tr 5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA**tccgagaggtgcacacggcgtctgccgtgacacggagggataaaagcccgggagAAAAtgcgcctagggTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Bsu L-4 Tr 4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagaaaacacatacgcgtaaatagaagcgcgtatgcacacggagggaaaaaagcccgggagAAAAgttttcTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Bsu L-4 Tr 5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagaaaacacatacgcgtaaatagaagcgcgtatgcacacggagggaaaaaagcccgggagAAAAgttttcGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Dru L-4 Tr 4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccaagagaaaacacacagcctagctgtgacacggagggacaaaagcccgggagAAAAgttttctagggTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Dru L-4 Tr 5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccaagagaaaacacacagcctagctgtgacacggagggacaaaagcccgggagAAAAgttttctagggGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Pae L-4 Tr 4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagttggcgacctccagtgaggttacacggcgggataaaagcccgggagAAAAgccgactagggTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC

Pae L-4 Tr 5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagttggcgacctccagtgaggttacacggcgggataaaa gcccgggagAAAAgccgactaggg**GAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pfl L-4 Tr 4	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagtcggcgacctccagttgaggttacacggcgggataaa agcccgggagAAAAgctgactaggg**TGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**
Pfl L-4 Tr 5	**GATGTAACTGAATGAAATGGTGAAGGACGGGTC**tccgagagtcggcgacctccagttgaggttacacggcgggataaa agcccgggagAAAAgctgactaggg**GAGTAGAGTGTGAGCTCCGTAACTAGTTACATC**

Table 3.4. **Sequences used in *in vivo* flow cytometry analysis.** *Italic* sequences indicates the tRNA scaffold that is appended onto the terminal ends. **Bold** sequences indicate the Spinach2 sequence, which flanks the riboswitch sequence. Orange indicates the artificial transducer stem or adenosine spacers for junction biosensors. *Underlined* indicates the adenosine spacers which are not part of the stem. Red indicates the site of mutation(s) that abolish binding to the aptamer.

Name	Sequence (5' to 3')
Spinach2	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**GTAGGCTGCTTCGGCAGCCTAC**TTGTTGAGTAGAGTGTGAGCTCCGTAACTAG TTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA*
Kpn J-2,2-A	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**ctcttcaaGCTGGCTAGGGTTCCGGTTCACCGCGGTGAACGTCTGGTCCGAGAGCT GGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGGGAGaagaaggTTGTTGAGTAGA GTGTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGG CGCCA*
Kpn J-2,2-M	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**ctcttcaaGCTGGCTAGcGTTCCGGTTCACCGCGGTGAACGTCTGGTCCGAGAGCT GGCGACCTCGGCGAGGTTACACGGCGGGATAAAAGCCCGGGAGaagaaggTTGTTGAGTAGA GTGTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGG CGCCA*
Dru J-2,2-A	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**ctcttcaaGTTTTCTAGGGTTCCGCGATAAAATTATCGGACTGGTCCAAGAGAAAAC ACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAGaagaaggTTGTTGAGTAGAGT GTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCG CCA*
Dru J-3,3-A	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**ctcttcaaaGTTTTCTAGGGTTCCGCGATAAAATTATCGGACTGGTCCAAGAGAAAAC ACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAGaaagaaggTTGTTGAGTAGAGT GTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCG CCA*
Dru J-2,2-D	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**tacttcaaGTTTTCTAGGGTTCCGCGATAAAATTATCGGACTGGTCCAAGAGAAAAC ACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAGaagaaggTTGTTGAGTAGAGT GTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCG CCA*
Dru J-3,2-D	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**tacttcaaaGTTTTCTAGGGTTCCGCGATAAAATTATCGGACTGGTCCAAGAGAAAA CACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAGaagaaggTGTTGAGTAGAGT GTGAGCTCCGTAACTAGTTACATC***CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCG CCA*
Dru J-3,3-D	*GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCG***GATGTAACTGAATGAAATGGTGAAGGA CGGGTCCA**tacttcaaaGTTTTCTAGGGTTCCGCGATAAAATTATCGGACTGGTCCAAGAGAAAA CACACAGCCTAGCTGTGACACGGAGGGACAAAAGCCCGGGAGaaagaaggTTGTTGAGTAGAG TGTGAGCTCCGTAACTAGTTACATC

3.7 Materials and Methods

Reagents and oligonucleotides

DNA oligonucleotides for biosensor constructs were purchased as Ultramers from Integrated DNA Technologies (Coralville, IA) and other DNA oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA). Guanidine hydrochloride and all guanidine analogs were purchased from Sigma-Aldrich (St Louis, MO). DFHBI was synthesized following previously described protocols (Paige et al., 2011) and was stored as a 10 mM stock in DMSO at -20 °C. Chemically competent BL21 (DE3) Star cells were purchased from Life Technologies (Carlsbad, CA).

***In vitro* transcription**

DNA templates for in vitro transcription were prepared by PCR amplification using Phusion DNA polymerase (NEB) from sequence-confirmed plasmids or Ultramer oligonucleotides (for library screening) using primers that added the T7 polymerase promoter sequence at the 5' end. PCR products were purified either by a 96-well format ZR-96 DNA Clean-up kit (Zymo Research) for screening experiments or by QIAquick PCR purification kit (Qiagen) for characterization and application experiments. RNA was transcribed from DNA templates using T7 RNA polymerase in 40 mM Tris- HCl, pH 8.0, 6 mM $MgCl_2$, 2 mM spermidine, and 10 mM DTT. RNAs were either purified by a 96- well format ZR-96 Clean & Concentrator (Zymo Research) or by denaturing (7.5 M urea) 6% PAGE. RNAs purified by PAGE were visualized by UV shadowing and extracted from gel pieces using Crush Soak buffer (10 mM Tris-HCl, pH 7.5, 200 mM NaCl and 1 mM EDTA, pH 8.0). Purified RNAs were precipitated with ethanol, dried, and then resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Accurate RNA concentrations were determined by measuring the absorbance at 260 nm after performing a hydrolysis assay to eliminate the hypochromic effect due to RNA secondary structure (Wilson et al., 2014).

General procedure for in vitro fluorescence assays

In vitro fluorescence assays were carried out in binding buffer containing 100 nM RNA, 10 μM DFHBI, 40 mM HEPES, pH 7.5, 125 mM KCl, 200 mM NaCl, and 3 or 10 mM $MgCl_2$, as indicated in the figures. Other conditions, including temperature and concentration of ligand, were varied in different experiments and are indicated in the figures. The RNA was renatured by heating to 72 °C for 3 min in the binding buffer then cooled to ambient temperature for 5 min prior to addition to the reaction solution. DFHBI was added to the solution containing buffer and RNA, and then ligand (or water for no ligand control) was added before fluorescence measurement. Binding reactions were performed in 30 μL volumes and were incubated at the indicated temperature in a Greiner Bio-One 384-well black plate in a Molecular Devices SpectraMax Paradigm Multi-Mode detection platform plate reader (Sunnyvale, CA). The fluorescence emission was measured during 30 to 60 min total with the following instrument parameters: 448 nm excitation, 506 nm emission.

Binding affinity analysis of guanidine biosensors

To measure the binding affinities of guanidine biosensors, fluorescence assays were performed with the following conditions: binding buffer with 10 mM $MgCl_2$, 37 ° C, 100 nM RNA, 10 μM DFHBI, and ligand guanidine concentrations from 100 nM to 30 mM. The fluorescence of the sample with DFHBI but no RNA was subtracted as background to determine relative fluorescence units.

In vitro fluorescence turn-on kinetics

For kinetics experiments, RNA was refolded in binding buffer and incubated in binding buffer containing 10 μM DFHBI and incubated at the desired temperature for the experiment for 15 minutes. Then, addition of 10 mM guanidine of ligand was done manually or by automated injector module on a SpectraMax i3x plate reader (Molecular Devices). Fluorescence measurements were taken every 0.5 seconds before ligand addition (30 s) and then after addition (10 min). For experiments using manual ligand addition, there was an approximate dead time of 1 minute for the first fluorescence reading. Fluorescence values were then normalized against maximum fluorescence value exhibited for each biosensor.

Molecular cloning

For *in vivo* biosensor assays, biosensor sequences were appended with a tRNA scaffold through PCR overhangs and resulting products were subcloned into a pET31b plasmid using a double restriction digest and ligation with BglII and XhoI restriction sites.

In vivo fluorescence assays by flow cytometry analysis (Guanidine-responsive time course)

2.0 mL *E. coli* BL21 (DE3) Star cultures expressing biosensor from their respective pET31b-T7-tRNA constructs were grown in ZYP-5052 autoinduction media while shaking at 37 ˚C for 20 hours. A small aliquot of each culture was then saved, with the rest centrifuged at 13,000 rpm for 1 min and the supernatant extracted as spent media. DFHBI-1T was then added to both aliquots of spent media to a final concentration of 50 μM. Samples were incubated at room temperature for 10 minutes for first flow cytometry reading (zero time-point); following that, guanidine (to 10 mM final concentration in the sample) was added or water (for no ligand samples) and were shaken at 25 ˚C between each reading. Fluorescence measurements were taken at each time interval analyzing 30,000 events on an Attune NxT flow cytometer (Life Technologies) equipped with a 488 nm laser for excitation and 515/15 filter for emission. Data was analyzed with FlowJo (version 10.0.7).

3.8 References

Batey, R. T.; Gilbert, S. D.; Montange, R. K. Structure of a Natural Guanine-Responsive Riboswitch Complexed with the Metabolite Hypoxanthine. *Nature* **2004**, *432* (7015), 411–415.

Battaglia, R. A.; Ke, A. Guanidine-Sensing Riboswitches: How Do They Work and What Do They Regulate? *Wiley Interdiscip. Rev. RNA* **2018**, *9* (5), e1482.

Davidson, M. E.; Harbaugh, S. V; Chushak, Y. G.; Stone, M. O.; Kelley-Loughnane, N. Development of a 2, 4-Dinitrotoluene-Responsive Synthetic Riboswitch in E. Coli Cells. *ACS Chem. Biol.* **2012**, *8* (1), 234–241.

Espah Borujeni, A.; Mishler, D. M.; Wang, J.; Huso, W.; Salis, H. M. Automated Physics-Based Design of Synthetic Riboswitches from Diverse RNA Aptamers. *Nucleic Acids Res.* **2016**, *44* (1), 1–13.

Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem.* **2017**, *86* (1), 515–539.

Harbaugh, S. V; Goodson, M. S.; Dillon, K.; Zabarnick, S.; Kelley-Loughnane, N. Riboswitch-Based Reversible Dual Color Sensor. **2017**, 1–24.

Harbaugh, S. V.; Martin, J. A.; Weinstein, J.; Ingram, G.; Kelley-Loughnane, N. Screening and Selection of Artificial Riboswitches. *Methods* **2018**, *143*, 77–89.

Huang, L.; Wang, J.; Lilley, D. M. J. The Structure of the Guanidine-II Riboswitch. *Cell Chem. Biol.* **2017**, *24* (6), 695-702.e2.

Huang, L.; Wang, J.; Wilson, T. J.; Lilley, D. M. J. Structure of the Guanidine III Riboswitch. *Cell Chem. Biol.* **2017**, *24* (11), 1407-1415.e2.

Knappenberger, A. J.; Reiss, C. W.; Strobel, S. A. Structures of Two Aptamers with Differing Ligand Specificity Reveal Ruggedness in the Functional Landscape of RNA. *Elife* **2018**, 1–23.

Nelson, J. W.; Atilho, R. M.; Sherlock, M. E.; Randy, B. Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. **2017**, 220–230.

Orelle, C.; Carlson, E. D.; Szal, T.; Florin, T.; Jewett, M. C.; Mankin, A. S. Protein Synthesis by Ribosomes with Tethered Subunits. *Nature* **2015**, *524* (7563), 119–124.

Panahi, T.; Weaver, D. J.; Lamb, J. D.; Harrison, R. G. A New Approach for Trace Analysis of Guanidine Compounds in Surface Water with Resorcinarene-Based Ion Chromatography Columns. *Analyst* **2016**, *141* (3), 939–946.

Peselis, A.; Serganov, A. YkkC Riboswitches Employ an Add-on Helix to Adjust Specificity for Polyanionic Ligands. *Nat. Chem. Biol.* **2018**, *14* (September).

Reiss, C. W.; Xiong, Y.; Strobel, S. A. Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. *Structure* **2017**, *25* (1), 195–202.

Roh, J. H.; Kilburn, D.; Behrouzi, R.; Sung, W.; Briber, R. M.; Woodson, S. A. Effects of Preferential Counterion Interactions on the Specificity of RNA Folding. *J. Phys. Chem. Lett.* **2018**, *9* (19), 5726–5732.

Sherlock, M. E.; Breaker, R. R. Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. *Biochemistry* **2017**, *56* (2), 359–363.

Sherlock, M. E.; Malkowski, S. N.; Breaker, R. R. Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry* **2016**, acs.biochem.6b01270.

Smith, K. D.; Lipchock, S. V; Ames, T. D.; Wang, J.; Breaker, R. R.; Strobel, S. A. Structural Basis of Ligand Binding by a C-Di-GMP Riboswitch. *Nat. Struct. Mol. Biol.* **2009**, *16* (12), 1218–1223.

Stanzl, E. G.; Trantow, B. M.; Vargas, J. R.; Wender, P. A. Fifteen Years of Cell-Penetrating, Guanidinium-Rich Molecular Transporters: Basic Science, Research Tools, and Clinical Applications. *Acc. Chem. Res.* **2013**, *46* (12), 2944–2954.

Sudarsan, N.; Lee, E. R.; Weinberg, Z.; Moy, R. H.; Kim, J. N.; Link, K. H.; Breaker, R. R. Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science* **2008**, *321* (5887), 411–413.

Truong, J.; Hsieh, Y. F.; Truong, L.; Jia, G.; Hammond, M. C. Designing Fluorescent Biosensors Using Circular Permutations of Riboswitches. *Methods* **2018**, *143*, 102–109.

Yeo, J.; Dippel, A. B.; Wang, X. C.; Hammond, M. C. In Vivo Biochemistry: Single-Cell Dynamics of Cyclic Di-GMP in Escherichia Coli in Response to Zinc Overload. *Biochemistry* **2018**, *57* (1), 108–116.

Zhong, G.; Wang, H.; Bailey, C. C.; Gao, G.; Farzan, M. Rational Design of Aptazyme Riboswitches for Efficient Control of Gene Expression in Mammalian Cells. *Elife* **2016**, *5*.

CHAPTER 4

Reprogramming riboswitches from sensing to reacting

4.1 Abstract

The RNA World hypothesis illuminates the diverse capabilities of RNAs and how they could be responsible for the origin of life. Riboswitches which are hailed as modern descendants from this time could have been responsible for other functions before evolutionarily converging to bind ligands and induce gene regulation. One function we propose modern riboswitches diverged from was that they could catalytically react with the ligands they bind as well. If this latent activity already exists, then it's possible it has been overlooked during the discovery of riboswitches. Alternatively, riboswitches could be a potent scaffold to be engineered into a self-labeling ribozyme as they already tightly bind to their ligand. Here, we explore the functional versatility of natural or engineered riboswitches sensing and reacting with analogs of the electrophilic ligand *S*-adenosyl-L-methionine (SAM). To this end, in collaboration with Agilent Labs, a high-throughput method was developed for probing and screening latent ribozyme activity using a microarray platform.

4.2 Introduction

Examples of RNA machinery that are capable of self-replication, catalysis, and regulation all support the notion of a RNA world in which primordial organisms utilized RNA as the essential molecule of life (Pressman et al., 2015). Non-coding and cis-regulatory elements called "riboswitches" have largely been classified as descendants from these early organisms (Breaker, 2012). Conventionally, riboswitches are known for their ability to modulate gene expression upon binding their cognate ligand. However, it could be possible that early riboswitches were responsible for other functions before their evolution converged as the exquisite molecular receptors they are now. One intermediate function that riboswitches could have possessed is catalysis with the ligands they bind. This idea is even more tractable when considering two types of catalytic RNAs such as natural self-cleaving ribozymes and artificial self-alkylating ribozymes.

Self-cleaving ribozymes like the hammerhead and twister classes which are non-coding catalytic RNAs remain elusive in terms of biological roles despite being widespread among bacteria (Forster and Symons, 1987; Roth et al., 2014) . Additionally, the glmS ribozyme was discovered to function similarly to a riboswitch in that upon addition of the ligand, glucosamine-6-phosphate, it can catalyze its own self-cleavage (Winkler et al., 2004). As riboswitches demonstrate tight affinity for ligands they bind and ribozymes demonstrate catalysis, it could be envisioned that there are RNAs which combine these two functions. Alternatively, the selection of artificial self-alkylating ribozymes shows that RNAs are capable of more than phosphodiester cleavage mechanisms but also nucleophilic substitution with electrophilic ligands. This has been demonstrated with self-labeling ribozymes that can react with an iodoacetamide moiety to conjugate biotin or fluorescein onto itself (Wilson and Szostak, 1995; Sharma et al., 2014). Another self-labeling ribozyme selected from a genomic RNA pool was found to selectively react with epoxides probes (McDonald et al., 2014). These examples show that RNAs can catalyze their own reaction with unnatural

electrophiles and leads us to believe there are natural RNAs that can react with biological electrophiles as well.

Thus far, no known riboswitches have been reported to react catalytically with the ligands they bind. Among the different riboswitch ligands, we identified that the ligand *S*-adenosyl-L-methionine (SAM) could be a potent electrophile that riboswitches could react with. SAM is an essential cofactor for methylation as it donates the methyl group on its positively charged sulfonium center allowing for methylation of DNA, RNA, and proteins (Bennett et al., 2017). While the majority of these reactions utilize enzymes known as methyltransferases to catalyze these transformations, there exists the unexplored possibility that a ribozyme could catalyze methylation on itself or even other targets.

These insights have become the basis for the work outlined here. We hypothesized that SAM riboswitches could catalyze auto-methylation in addition to its sensing capabilities (**Fig 4.1A**). Looking at structural insights from *Bacillus subtilis* SAM-I riboswitch as a model (Lu et al., 2010), it was reasoned that the binding pocket of the riboswitch could accommodate a SAM analog which could be used to probe ribozyme activity (**Fig 4.1B**). In the last decade, a number of SAM analogs have been developed for profiling histone and DNA methylation ((Dalhoff et al., 2006; Gražvydas Lukinavičius et al., 2007; Islam et al., 2011, 2013; Wang et al., 2011; Kim et al., 2012; Wang and Luo, 2013) by incorporation of alkyne or azide chemical handles for subsequent functionalization by click chemistry. These precedents further supported our hypothesis to use a SAM analog to probe for latent ribozyme activity of native riboswitches. While we believed this function could exist in natural riboswitches, it may also be possible to engineer ribozyme activity through structure-guided mutagenesis approach where changing residues at the ligand binding pocket could increase reactivity. To accomplish this, we explored the development of a screening assay for ribozyme activity through a gel-based format and a high-throughput microarray platform.

4.3 Results

Synthesis of Hey-SAM, a proxy for latent ribozyme activity

Using a previously devised synthetic route (Hickey and Hammond, 2014) SAH (**9**) was synthesized a crucial precursor in our SAM analog synthesis (**Fig. 4.2**). This route relies on two critical intermediates: 5-chloro-5'-deoxyadenosine (**3**) and N-Boc-L-homocysteine (**7**). We accessed **3** in two steps by treating commercially-available adenosine (**1**) with thionyl chloride to produce a sulfinyl intermediate (**2**) and which was hydrolyzed to the vicinal diol using with NH_4OH. We obtained **7** in good yields after three steps. L-Homocysteine thiolactone (**4**) was dynamically resolved from the racemic D,L-homocysteine thiolactone mixture by generation of the mandelic acid salt. Our desired enantiomer, **5** readily precipitated while its D-enantiomer remained in the mother liquor and reacted with salicylaldehyde to form an imine that promotes racemization and is reversibly hydrolyzed. The free amine of **5** was then protected with di-tert-butyl dicarbonate to yield N-Boc-L-homocysteine thiolactone (**6**) and subsequent hydrolysis of the lactone under basic conditions generated **7**. Coupling partners **3** and **7** were reacted in strongly basic conditions at high concentrations to form N-Boc-*S*-

adenosyl-homocysteine (**8**). The Boc group was then removed under strongly acidic conditions to yield **9**, which was purified through HPLC.

With **9** in hand, we sought to generate (E)-hex-2-en-5-ynyl SAM, (Hey-SAM, **11**) by coupling to the respective alkyl halide (**Fig. 4.2**). A previous route to **11** was repeated (Wang et al., 2013) which uses (E)-1-bromo-hex-2-en-5-yne (**10**) to react with **9**. In this procedure, **10** is used crude with impurities as it has been reported to not be isolable. The resulting residue from the reaction of trans-1,4-dibromo-2-butene with ethynylmagnesium bromide produces **10** but also afforded a di-addition product from the Grignard reagent and unreacted starting material that was still present when monitoring by GC-MS. Although usage of crude **10** can produce **11** in milligram quantities, there is a large proportion of starting material **9**, that is unreacted, which required optimized HPLC purification to separate the desired product Hey-SAM. Hey-SAM was purified using a water and acetonitrile with 0.05% TFA (Solvent A and B) with a gradient over 30 min (0 to 10% solvent B over 25 min and then 10 to 70% solvent B over 5 min).

Demonstration of Hey-SAM binding to SAM-I riboswitches using a biosensor assay

Previously our group has developed a palette of RNA-based biosensors which specifically bind the ligand SAM and induce fluorescence activation (Truong et al., 2018). These sensors were applied to assess if the SAM-I riboswitch can accommodate the Hey-SAM analog into its binding pocket. One of the biosensors, Bs 4-4 is also derived from the organism *Bacillus subtillis* with sequence overlap to *B. subtillis* SAM-I sequence used for x-ray crystal structure determination (Lu et al., 2010). To our delight, we found that Bs 4-4 is responsive to SAM and Hey-SAM with apparent binding affinities of 0.7 μM and 2.4 μM, respectively but does not respond to SAH (**Fig.4.3**). It is known that the positively charged sulfonium center is highly recognized through electrostatic interactions within the binding pocket so the neutral sulfonium center on SAH renders it unable to bind and activate the biosensor (Lu et al., 2010). Hey-SAM still retains these electrostatics and it appears the added sterics of the alkyne handle does not hinder affinity greatly. This finding supports our initial hypothesis that Hey-SAM would serve as a suitable proxy for measuring reactivity with SAM riboswitches.

Ribozyme activity assay using a gel-based format and fluorescence analysis

After obtaining Hey-SAM and affirming its ability to bind to the SAM-I riboswitch through our biosensor assay, a robust method was designed to detect potential ribozyme activity of SAM riboswitches (**Fig 4.4**). Since riboswitches have been highly evolved to bind their ligand, their latent ribozyme activity of SAM riboswitches may be very inefficient. In this way, a highly sensitive assay would be necessary to detect labeling activity. It has been previously demonstrated that Hey-SAM can be synthesized *in situ* within live cells and used to profile chromatin modifications of engineered protein methyltransferases (Wang et al., 2011, 2013). These studies utilized copper click chemistry to conjugate biotin to modified chromatin form cell lysates and enriched for labeled chromatin by binding of streptavidin beads. We envisioned using click chemistry to enable a fluorescent output for ribozyme activity. Assuming the ribozyme self-reacts with Hey-SAM, the alkyne handle will be transferred to the RNA

and could be subsequently tagged with a fluorophore through a Cu-I azide-alkyne cycloaddition (CuAAC) reaction. These clicked products containing the fluorophore could be separated on a denaturing polyacrylamide gel and visualized using a fluorescence scanner with appropriate excitation and emission light.

Utilizing this assay, the reactivity of the *Bacillus subtilis* SAM-I riboswitch was tested with varying amounts of Hey-SAM (**Fig 4.5A**). Increasing the stoichiometric amount of the ligand Hey-SAM directly correlates with increasing the overall fluorescent signal observed, demonstrating that its reactivity is ligand-dependent. Additionally, a fluorescently-labeled *Bs* SAM-I riboswitch was used as a total loading control to quantify labeling activity (AlexaFluor 488 scan) and normalized to the amount of RNA loaded on the gel which was determined through subsequent gel staining with ethidium bromide (EtBr scan). At 1:1 stoichiometry for *Bs* SAM-I RNA and Hey-SAM, the % labeling was estimated to be below 0.1%. With this assay, we also sought to evaluate whether sequences from other riboswitch sequences could have differing reactivities with the ligand (**Fig 4.5B**,). Representatives from four other SAM riboswitch classes (SAM-II, -III, -IV, SAM-SAH) were evaluated alongside the tightly-binding *Bacillus clausii* (Bc) SAM-I riboswitch to compare reactivity to the Bs SAM-I RNA (**Table 4.1**). We observed that the relative labeling activity (0.42%) of the *Bc* SAM-I RNA was the greatest out of the tested constructs which could be attributed to its tight affinity (700 pM, Sudarsan et al., 2006).

Ribozyme library screening using microarray platform

While the labeling activity of native riboswitches could be detected, it was evident that this reactivity was not its primary function. However, we proposed that it could be possible to engineer this function into riboswitches through a structure-guided mutagenesis approach. In collaboration with Agilent Technologies, we sought to measure this ribozyme activity using their proprietary microarray technology as a screening platform (**Fig 4.6A**). Approximately 4000 SAM-I riboswitches sequences taken from the Rfam database (Accession: RF00162) and 30 different mutations were designed along the binding pocket that interacts with the ligand (Note: the computational design of library sequences were performed by Dr. Yichi Su, a former labmate and collaborator). These mutants were randomized from their native sequence at distinct nucleotide positions that make direct contacts with the ligand or were in less than a 10 Å proximity. This mutant DNA library, designed with a total of 119,878 sequences, was appended with unique barcodes and universal priming regions for amplification by PCR and introduction of a T7 promoter for *in vitro* transcription of the ribozyme library (**Fig 4.6B**). An oligonucleotide library was synthesized using Agilent's proprietary SurePrint technology (OLS 1.0) corresponding to a printed DNA microarray that possessed 122,000 customized capture oligos in duplication (a total of 244K unique features). Each unique barcode is complementary to a capture oligo that is printed on two locations on the microarray, permitting spatial resolution of signal for each riboswitch sequence.

The OLS 1.0 was amplified by PCR using universal primers that introduced the T7 promoter for downstream transcription. Accordingly, 10 µg of OLS 1.0 DNA was used for *in vitro* transcription with T7 RNA Polymerase and purified by spin column. The transcribed ribozyme library then was reacted with Hey-SAM and subsequent click

chemistry with a Cy5 dye generates a fluorescent output linked to ribozyme activity. In addition, the ribozyme library was non-specifically labeled with a Cy3 dye through Agilent's Universal Labeling System (ULS) to normalize for total RNA levels. Hybridization of the dual-labeled RNAs onto the microarray and subsequent wash-out left areas with overlapping Cy5 and Cy3 values which were quantitated to measure overall labeling efficiency of the ribozyme. Capture oligos which form stable hairpins were used as negative controls to normalize Cy5 and Cy3 labeling as these signals represent background noise.

Before samples were analyzed for microarray hybridization, the OLS 1.0 RNA library after transcription and OLS 1.0 RNA library after Cy5-labeling was subjected for nucleotide size analysis by Bioanalyzer (data not shown). These results identified that after *in vitro* transcription, the OLS 1.0 RNA was ~95% intact with an average size distribution of 160 nts (which correlates well with the expected distribution of ~165 nts). Similarly, after Cy5-labeling, the size distribution of the OLS 1.0 RNA library is similar to 160 nts but only ~70% of the OLS 1.0 Cy5-RNA was fully intact. The RNA degradation of the library is attributed to the CuAAC reaction as it has been reported copper facilitates production of hydroxy radicals which spontaneously cause RNA cleavage (Paredes and Das, 2011).

Within the microarray results, a statistical analysis was performed to determine among the 30 different mutation categories which ones have any significant labeling activity above or below background (**Table 4.2**). This analysis highlights the probability a mutation category yielded variants with labeling activity that is statistically significant. The probability that a mutation category yields ribozymes with activity ABOVE background is denoted by "mLogP EnrichedUp" values while the probability that a mutation category yields ribozymes with activity BELOW background is denoted by "mLogP EnrichedDown". It is important to note these values are not a measure of the total labeling efficiencies observed from a sequence. From these results, 3 nucleotide base-pair positions (N7-N494, N9-N492, and N6-N495) were identified that likely possess increased activity upon mutation. The mutation category G7-U494 was observed to possess a mLogP EnrichedUp of 12.21 indicating the likelihood this mutation yielded ribozymes with above background activity was very high. An in-depth visualization of these 3 base-pair positions from the SAM-I crystal structure shows that all 3 come in close proximity to the methyl group on the sulfonium center of the ligand SAM (**Fig. 4.7**). The U7-base from the natural base-pair U7-A494 comes into very close proximity to the donating methyl group of SAM which when mutated to a G7 residue, confers enhanced labeling activity likely due to better position and closer proximity.

Second-generation combinatorial mutation library screen

Using the insights from the OLS 1.0 microarray results, we sought to create a combinatorial mutation library termed OLS 2.0 (**Fig. 4.7**) which utilized the most enriched mutations at the 3 nucleotide base-pair positions: 1. N7-N494 (Position A) , 2. N9-N492 (Position B), and 3. N6-N495 (Position C). Using these three positions, families of single, double, and triple mutants were created with a total of 183 mutation categories with priority to combinations utilizing Position A (due to their probability values observed from the OLS 1.0 microarray results). Starting from approximately

1300 riboswitch sequences, we generated a total of 236,548 unique ribozyme sequences to screen using identical methodology for preparation and microarray hybridization. The second generation library was again synthesized by Agilent through an Oligonucleotide Library Synthesis platform (OLS 2.0) along with a DNA microarray that possessed up to 244,000 customized capture oligos as single replicates.

The OLS 2.0 was amplified by PCR using the same universal primers which introduced the T7 promoter for downstream transcription. Accordingly, 10 µg of OLS 2.0 DNA was used for *in vitro* transcription with T7 RNA Polymerase and purified by spin column. The transcribed ribozyme library then was reacted with Hey-SAM and subsequent click chemistry with a Cy5 dye and yielded spectral signature for ribozyme activity. In addition, the ribozyme library was non-specifically labeled with a Cy3 dye through Agilent's Universal Labeling System (ULS) to normalize for total RNA levels. Hybridization of the dual-labeled RNAs onto the microarray and subsequent wash-out left areas with overlapping Cy5 and Cy3 values which were quantitated to measure overall labeling efficiency of the ribozyme. Capture oligos which form stable hairpins were used as negative controls to normalize Cy5 and Cy3 labeling as these signals represent background noise.

While a probability analysis for enrichment was also performed for the OLS 2.0, the results did not yield significantly enriched mutation categories above or below background (data not shown). We believe this is attributed to the increased number of mutation categories (30 to 183) with 4-fold less sequences per category (4000 to 1300) which failed to show any clear trends among the data. Thus, we manually sorted through the data to obtain sequences with the highest labeling efficiency on the array. We identified the top 10 mutated sequences and compiled them with their respective wild-type sequences on the array (**Table 4.3**).

Biochemical validation of ribozyme activity of selected OLS 2.0 variants

To verify the labeling efficiencies observed on the microarray, we obtained variants for 3 ribozymes with the highest activity and their corresponding wild-type sequences for validation through a gel-based activity assay similar to previous methods (**Fig. 4.4**). These sequences were constructed to be almost identical to the sequences used on the microarray except with the universal reverse priming sequence being omitted. Once these sequences were reacted with Hey-SAM and subsequently clicked with a Cy5-azide, the products were purified and separated on a 6% denaturing PAGE-gel for analysis on a gel imager (**Fig. 4.9**). Scanning on appropriate Cy5 and EtBr channels allowed for visualization and quantification of relative labeling activity from all 6 sequences. To compare labeling activity from the microarray and on the gel, a relative labeling ratio was calculated for both based on the sequences used (**Table 4.4**). This relative labeling ratio is a proxy for understanding how close the enhancement of labeling activity from the microarray is mirrored on gel. OLS 2.0 Seq 5 is the only sequence that experiences a relative fold-enhancement in both cases that is very promising. From the OLS 2.0 microarray, the % labeling for Seq 5 is estimated to be almost 4.4%, an 8.3-fold improvement from its corresponding wild-type sequence.

4.4 Discussion

This microarray approach was able to screen more than 300,000 mutagenized riboswitch sequences for labeling activity. Despite the apparent false positive rate from the microarray, the platform was still able to identify an engineered ribozyme with OLS 2.0 Seq 5 that could be validated biochemically. The corresponding wild-type riboswitch exists in the organism *Staphylococcus aureus*, a member of the Firmicutes and a commonly known pathogen in skin infections. While it is unlikely in this case there is a relation of its ribozyme activity to its phylogenetic origin, self-alkylating ribozymes that react with epoxide electrophiles were isolated from organism genomes that possessed a high percentage of non-coding RNAs which can include catalytic RNAs or riboswitches (McDonald et al., 2014).

Comparing the results of OLS 1.0 and OLS 2.0 microarray iterations could also be insightful about potential residues involved with catalysis. OLS 2.0 Seq 5 is composed of a double mutation U7-G494 and U9-A492 corresponding to the nucleotide positions from *Bacillus subtilis* SAM-I crystal structure (**Fig. 4.7**). While no mutation category was found to be significantly enriched in the OLS 2.0, the U9-A492 mutant was the 2nd most enriched mutation in the OLS 1.0 and the U7-G494 mutation is a transposition of 1st most enriched G7-U494 mutation from the OLS 1.0.

4.5 Conclusions/Future Directions

In summary, we have developed a microarray screening platform to test activity for self-labeling ribozymes. While initial investigations looked to native riboswitches for this latent reactivity, we have found with robust screening and structure-guided mutagenesis that enhanced catalytic variants can exist such as OLS 2.0 Seq 5. Further characterization of OLS 2.0 Seq 5 could determine the site of labeling and grant insight about sequence requirements for this labeling activity. Other future work applying the mutations found in this catalytic variant to other phylogenetic sequences may explain its functional basis as well. This investigation and validation of reprogramming natural riboswitches into desired ribozymes is currently ongoing.

4.6 Figures

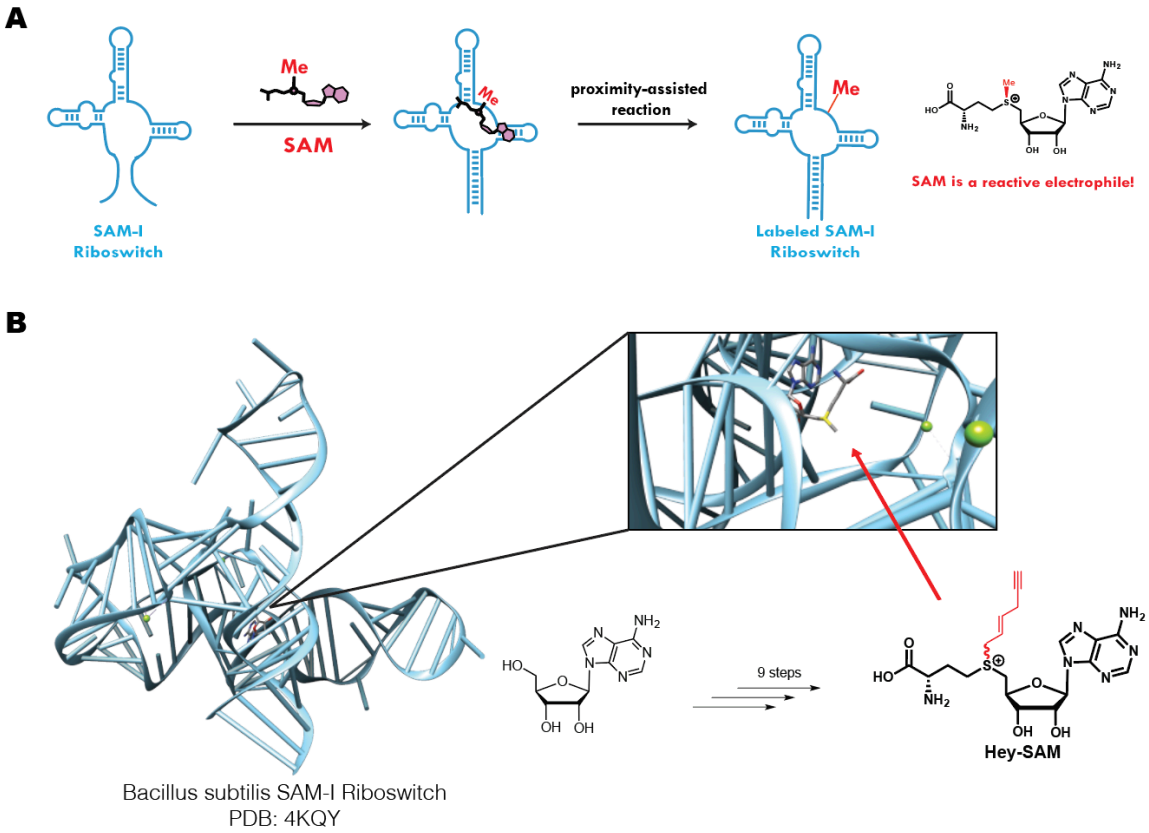


Figure 4.1. SAM-I riboswitches as potential ribozymes
a) SAM-I riboswitches could have activity acting as self-alkylating ribozymes that reacts with the electrophile SAM. b) The crystal structure of the SAM-I riboswitch shows the methyl group of SAM is solvent exposed with potential to accommodate SAM analogs.

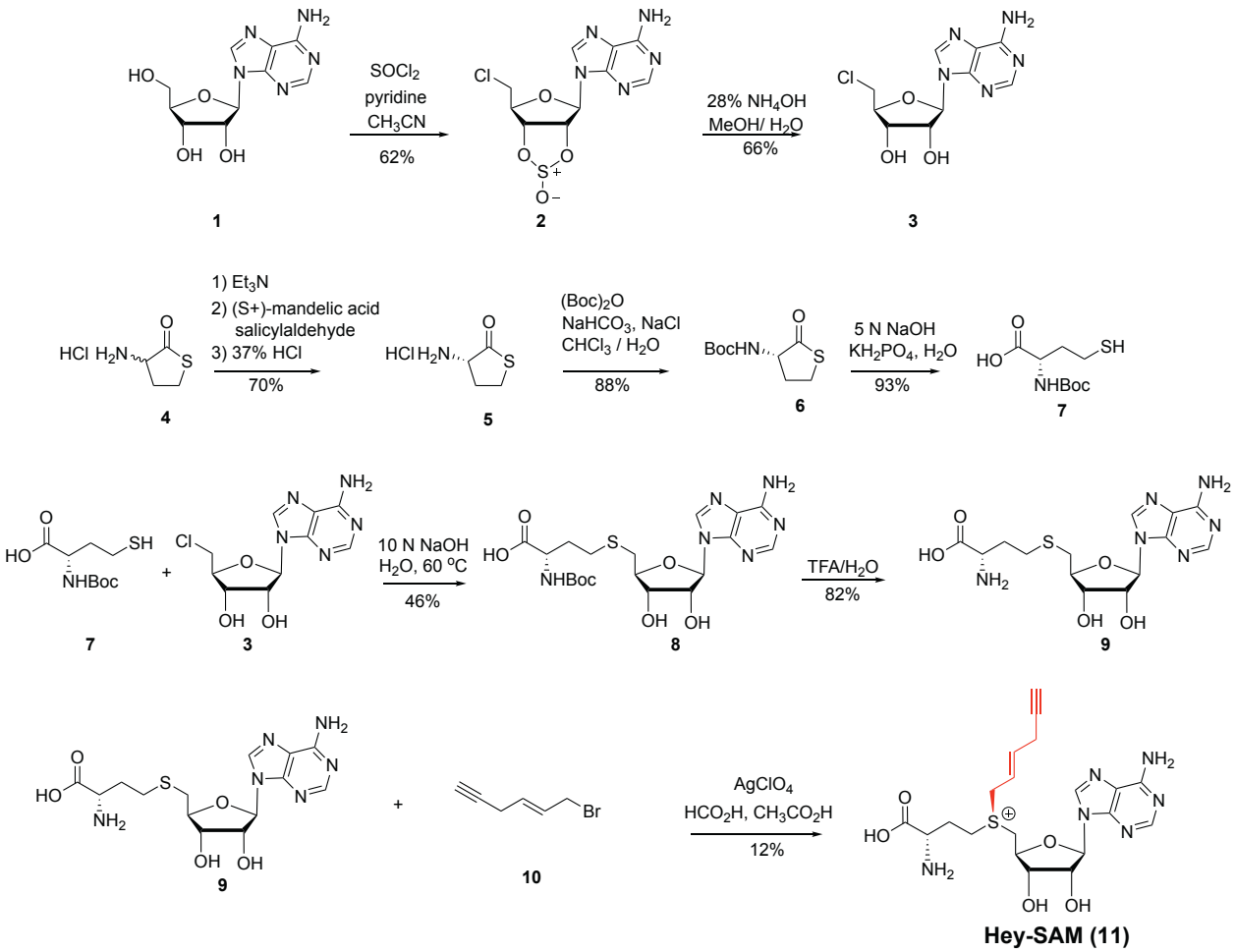


Figure 4.2. Synthesis of (E)-hex-2-en-5-ynyl SAM (Hey-SAM)
A scheme for the synthesis of Hey-SAM starting with cost-efficient and commercially available starting materials.

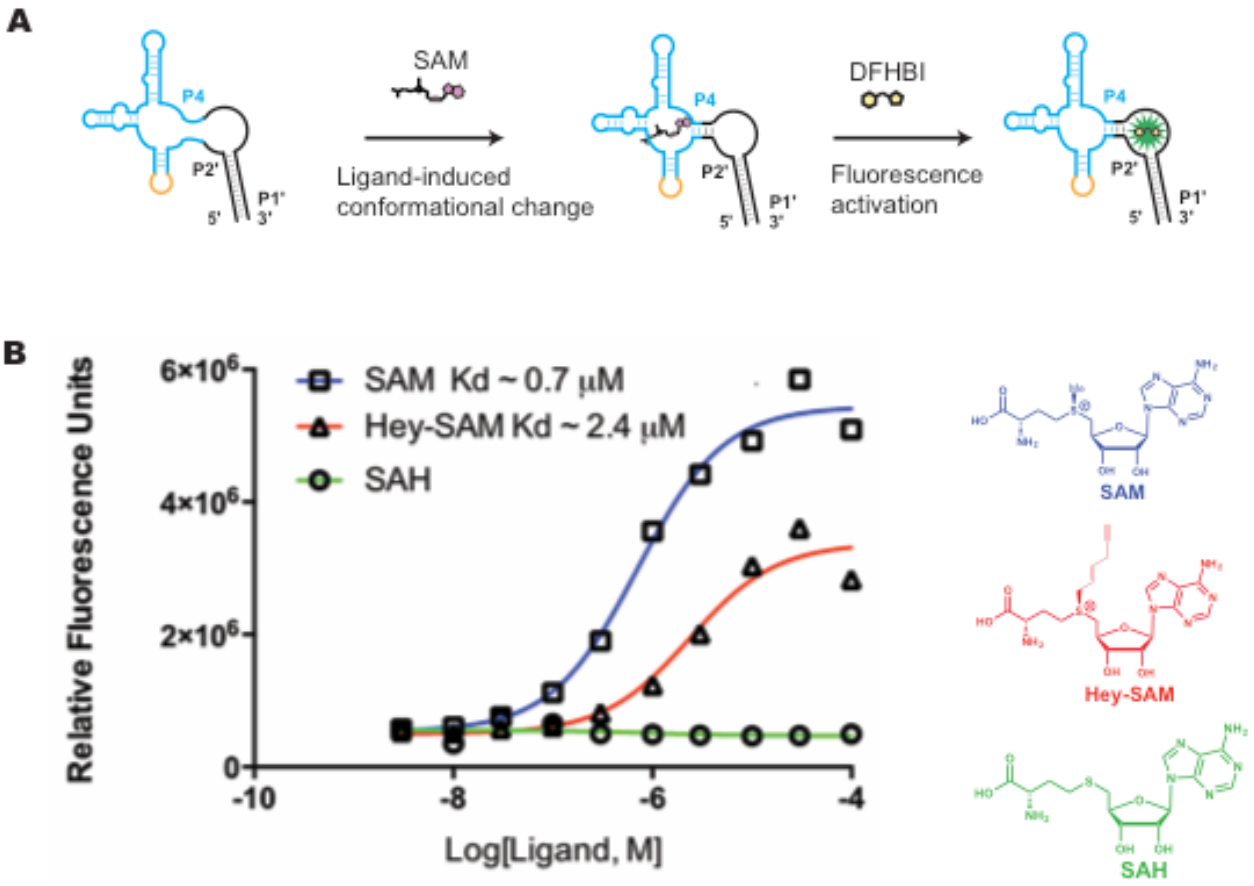


Figure 4.3. Detection of Hey-SAM by Bs 4-4, a known biosensor for SAM
(a) Mechanism for fluorescence turn-on is depicted for the biosensor designed from the *Bacillus subtilis* SAM-I riboswitch. (b) Binding affinity determination of biosensor when using SAM and analogs such as Hey-SAM and SAH. Data shown are values taken from one independent replicate.

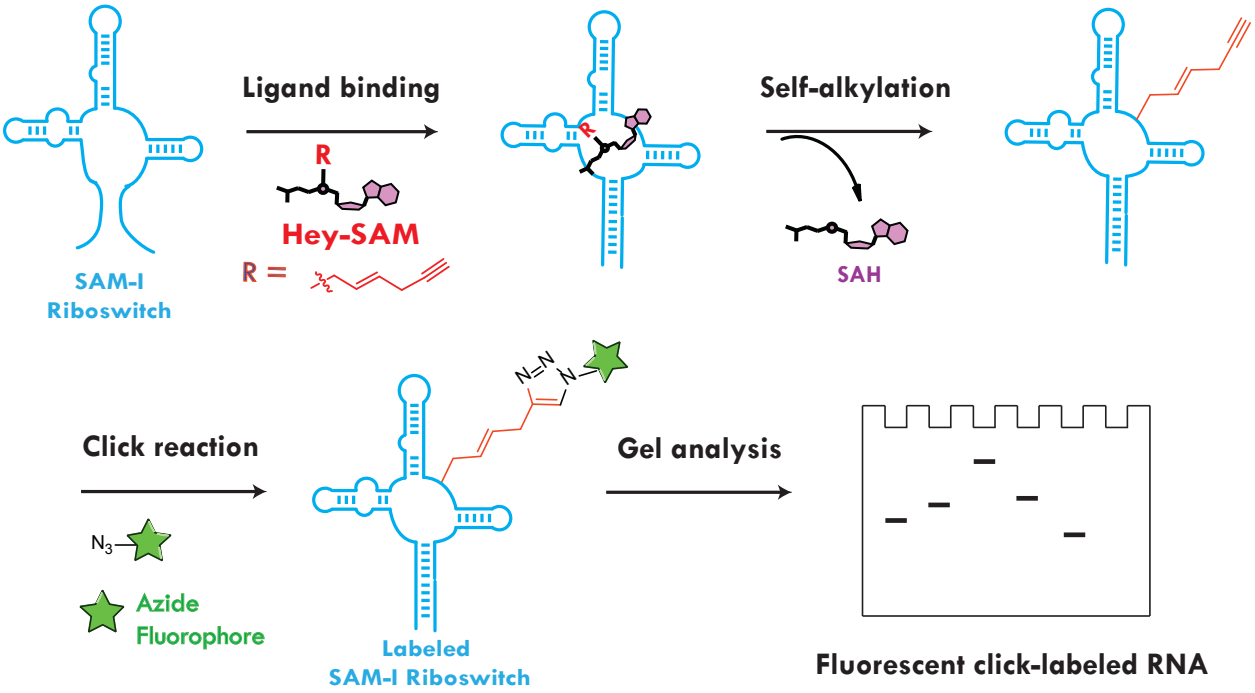


Figure 4.4. Ribozyme activity assay using fluorescent gel analysis
Ribozymes that react with Hey-SAM will be self-labeled with an alkyne handle that can be subsequently fluorescently-labeled by a CuAAC reaction with an azide-dye and fluorescent products can be separated on a denaturing urea-PAGE gel and analyzed.

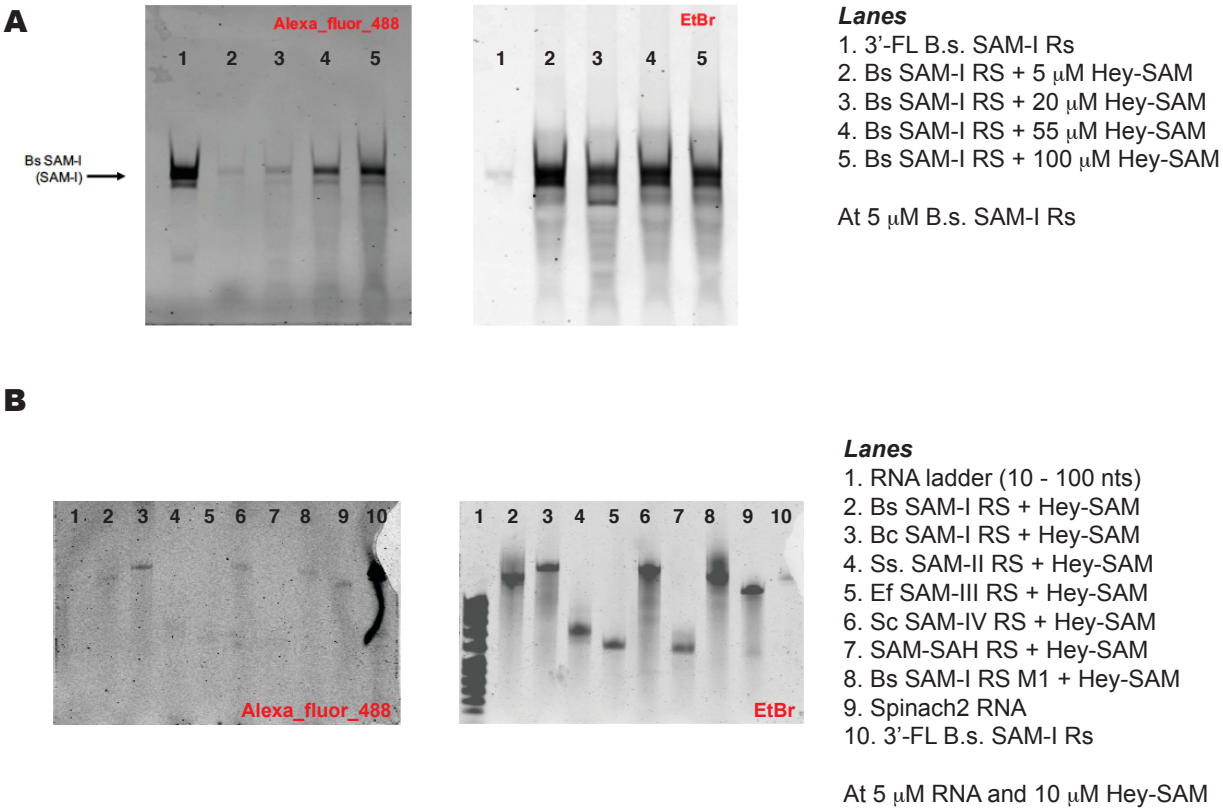


Figure 4.5. Ribozyme activity assay using fluorescent gel analysis
(a) *Bacillus subtillis* (Bs) SAM-I riboswitch is reacted with varying concentrations of Hey-SAM and after CuAAC with an azide-dye, the fluorescent products are separated on a denaturing urea-PAGE gel and analyzed with appropriate excitation light and filters corresponding to AlexaFluor488 and ethidium bromide.

Table 4.1. **Sequences from various classes of SAM riboswitches and the organisms they are derived from**. Binding affinities are presented as they were reported in the literature.

Riboswitch Family and Organism	Sequence (5' to 3')	Experimental Binding Affinity (K_d) for SAM
SAM-I (Bacillus subtilis)	GTTCTTATCAAGAGAAGCAGAGGGACTGGCCC GACGAAGCTTCAGCAACCGGTGTAATGGCGAT CAGCCATGACCAAGGTGCTAAATCCAGCAAGCT CGAACAGCTTGGAAGATAAGAAGAG	19 nM (Lu et al., 2010)
SAM-I (Bacillus clausii	AAAAACACTCTTATAACGAGAAGCGGAGGGACT GGCCCAATGAAGCTTCAGCAACCATTCATTGCG ATGAAAAGGTGCTAAATCCAGCAAAGGGAACTT TGGCAGATAAGGGGATTCAT	700 pM (Sudarsan et al., 2006)
SAM-II Sargesso Sea Env12 metX gene	TCGCGCTGATTTAACCGTATTGACAAGCCGCTG ATAAATGTAGCTAAAAAGGG	60 nM (Gilbert et al., 2008)
SAM-III Eneterococcus facalis metK	TTCCCGAAAGGATGGCGGAAACGCCAGATGCC TTGTAACCGAAAGGGGGAAT	570 nM (Lu et al., 2008)
SAM-IV Streptomyces coelicolor	GGTTTTTCGACAGGTCATGAGTGACAGTCATGA GGCCCCGGCCGACTGTCCGGCAACCCTCCGTC CGTGGCGGGGTGCCCCGGGTGAAGACCAGGT CGTGGACAGCAAGGTCCACGGCAAGCGCGGAC CCCT	150 nM (Weinberg et al., 2008)
SAM-V (Candidatus pelagibacter)	GGAATTAAGCCGGGCAGTTGAACCATATTGTGC GCCCTGCATTTGCTTAAGCACTAAAAAGGAG	1.2 μM (Poiata et al., 2009)
SAM-SAH (Dechloromonas aromatica)	GGUCUGCCGAGGAGCGCUGCGACCCUUUAAU UCGGGGGCCAGGCUCGGCAAUGAUCAACGGC GCUCGC	~25 μM (Wang et al., 2008)

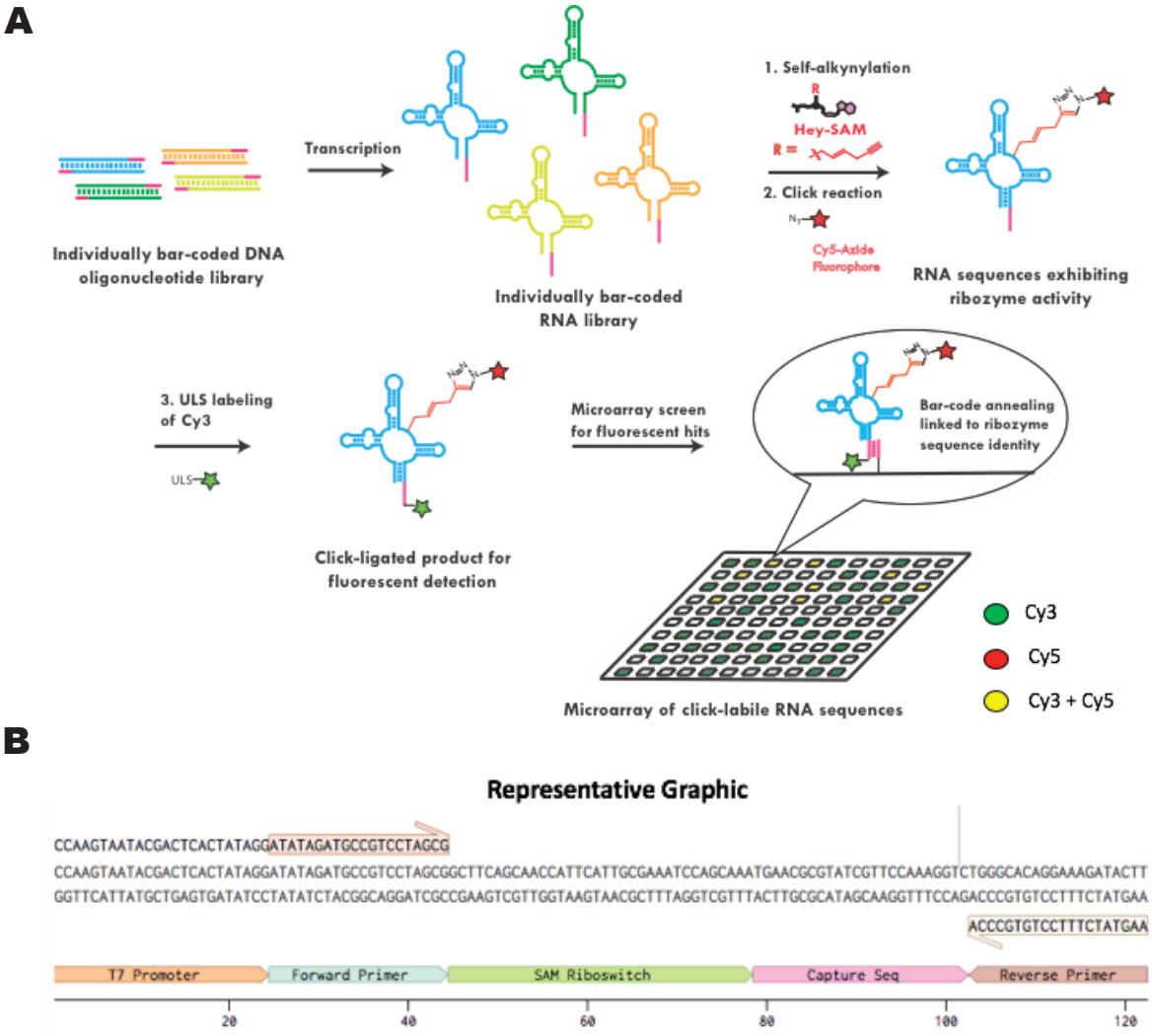
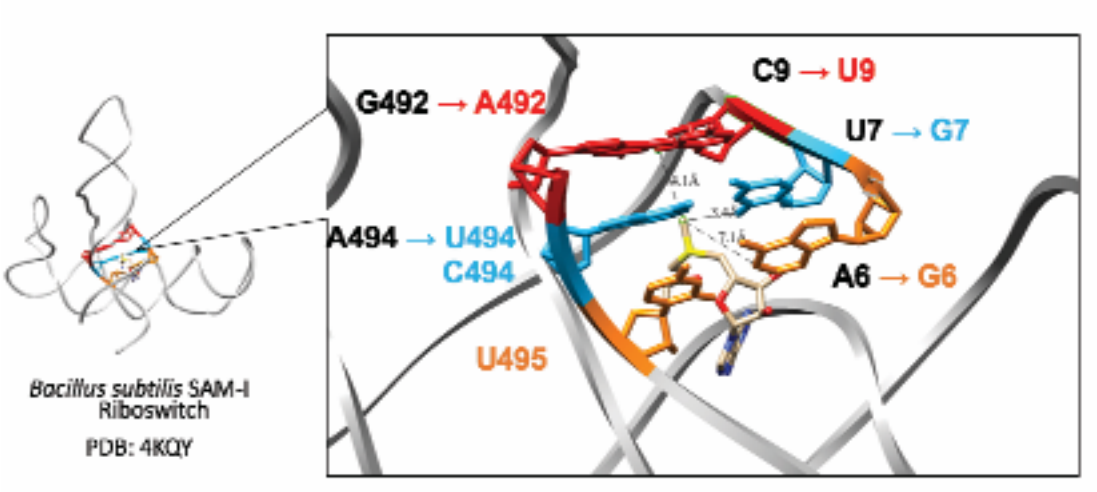


Figure 4.6. SAM-I Ribozyme library analysis using Agilent microarray platform
(a) A DNA oligonucleotide library composed of 119878 natural and engineered SAM-I sequences that is barcoded with a unique capture sequence (pink). After Hey-SAM labeling and CuAAC reaction is performed to label sequences with Cy5, the library is tagged with Cy3 using a universal labeling system (ULS) that non-specifically labels any RNA sequence. Activity of the ribozyme (Cy5) is normalized to amount of RNA (Cy3) and the efficiency of labeling is calculated. (b) Detailed graphic describing the construction of each ribozyme. The universal forward and reverse primer regions on each sequence were used for PCR amplification and to add a T7 promoter for *in vitro* transcription while the capture sequence (pink) is unique for each library member and is hybridized to a capture oligo spatially resolved and printed onto the microarray.

Table 4.2. Statistical enrichment analysis of 30 mutations types for SAM-I riboswitches

30 mutation types were applied to ~4000 SAM-I riboswitches with each mutation category being analyzed for fold-enrichment up (higher than background labeling efficiency) and for fold-enrichment down (lower than background labeling efficiency).

#	Mutation Category	mLog(P) EnrichedUp	mLog(P) EnrichedDown
1	G7 U494	12.21	0.05
2	U9 A492	7.96	0.66
3	G7 C494	6.56	0.92
4	U7 A494	4.01	-1.68
5	C9 G492	2.49	0.23
6	G6 U495	2.41	0.21
7	U6 G495	2.16	2.5
8	A6 U495	1.75	-1.75
9	C7 G494	1.43	0.49
10	G9 U492	1.1	0.32
11	G12	0.97	0.23
12	U6 A495	0.9	2.41
13	A9 U492	0.88	0.15
14	C13	0.86	0.73
15	wt	0.8	1.79
16	A13	0.67	0.24
17	G9 C492	0.65	1.04
18	C11	0.62	1.08
19	G13	0.57	-0.3
20	U9 G492	0.53	1.02
21	A11	0.52	0.97
22	U7 G494	0.5	1.09
23	C6 G495	0.48	3.55
24	A7 U494	0.44	0.09
25	A12	0.38	1.17
26	U13	0.37	0.19
27	G11	0.26	0.32
28	G6 C495	0.15	0.87
29	U12	0.08	1.15
30	C12	0.07	0.9
31	U11	0.04	1.85



WT Base-Pair	Enriched Base-Pair
U7-A494	G7-U494 or G7-C494
C9-G492	U9-A492
A6-U495	G6-U495

Figure 4.7. Comparison of SAM-I binding pocket residues and top 4 enriched mutations.
Three key base-pairing sites were significantly enriched from the microarray analysis: 1) U7-A494 when mutated to a G-U or a G-C base-pair (blue). 2) C9-G492 when mutated to a U-A base-pair (red). 3) A6-U495 when mutated into a G-U base-pair. The ligand is also displayed in the binding with methyl group (green) of SAM with relative distances to proposed residues that are responsible for catalysis.

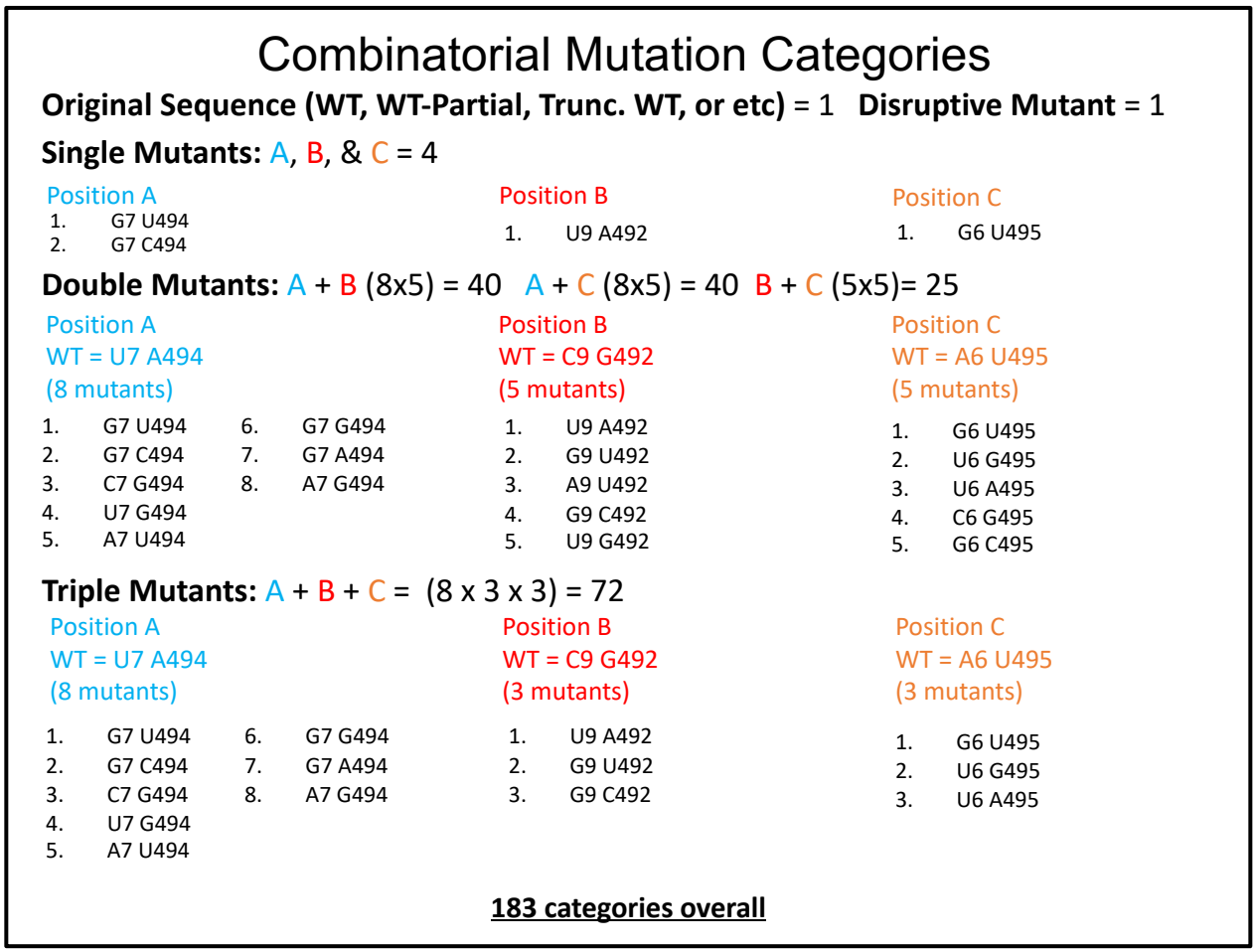
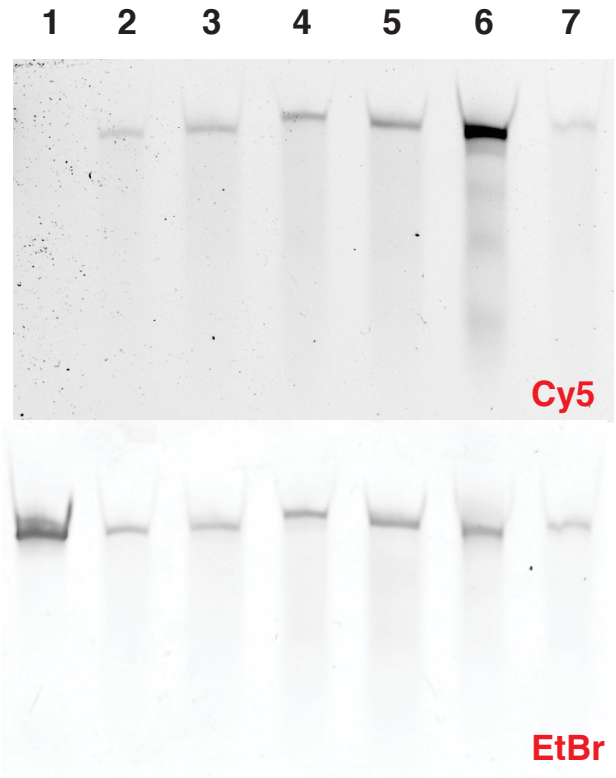


Figure 4.8. Combinatorial mutations for 2nd generation ribozyme library (OLS 2.0) A 2nd generation ribozyme library was constructed combining mutations categories at the 3 most enriched positions from the OLS 1.0 microarray analysis. In total, 183 different categories were included in the OLS 2.0 with 236548 new sequences.

Table 4.3. Top 10 OLS 2.0 sequences (#1-10) with highest labeling activity from microarray analysis and their respective mutation category. Each ribozyme can then be compared respective wild-type sequence for each ribozyme (#11 – 20) to assess their improved labeling efficiency. Green shading highlights top 3 sequences and yellow shading highlights their respective counterparts. Shaded sequences were selected for biochemical follow-up.

#	Sequence Reference	Mutation Category (A / B / C)	Accession Code	Relative % Labeling
1	Hammond037442	(G7_G494)(G9_C492)(G6_U495)	ACDF01000024.1/30867-30781_MU_Partial	6.8267
2	Hammond120471	(G7_G494)(G9_U492)(G6_U495)	AEUR01000003.1/61319-61414_WT_FL	6.3684
3	Hammond213028	(U7_G494)(U9_A492)	AM990992.1/2453012-2452917_MU_Partial	4.3966
4	Hammond162226	(G7_G494)(U9_A492)(U6_G495)	CP002282.1/532630-532718_3'_MU_Partial	4.2322
5	Hammond095250	P2_mutant	ACUU01000042.1/40243-40341_MU_Partial	4.129
6	Hammond221089	(U7_G494)(G9_U492)(U6_G495)	ACWQ01000020.1/25078-25175_WT_FL	4.0407
7	Hammond028654	(U9_A492)(U6_A495)	ACSW01000146.1/28058-28156_MU_Partial	4.0192
8	Hammond087684	(G7_C494)(U9_A492)(G6_U495)	ACSV01000033.1/39999-40097_WT_FL	3.7514
9	Hammond166971	(A9_U492)(G6_U495)	BX571856.1/15939-16037_WT_FL	3.6343
10	Hammond056647	(G9_C492)(G6_C495)	ACMY01000004.1/34318-34423_MU_Partial	3.633
11	Hammond234881	Wild-Type	ACDF01000024.1/30867-30781_MU_Partial	0.6744
12	Hammond119922	Wild-Type	AEUR01000003.1/61319-61414_WT_FL	0.4934
13	Hammond223126	Wild-Type	AM990992.1/2453012-2452917_MU_Partial	0.5309
14	Hammond236469	Wild-Type	CP002282.1/532630-532718_3'_MU_Partial	0.6948
15	Hammond180987	Wild-Type	ACUU01000042.1/40243-40341_MU_Partial	0.5148
16	Hammond090192	Wild-Type	ACWQ01000020.1/25078-25175_WT_FL	1.071
17	Hammond186938	Wild-Type	ACSW01000146.1/28058-28156_MU_Partial	0.5027
18	Hammond096993	Wild-Type	ACSV01000033.1/39999-40097_WT_FL	0.3456
19	Hammond057197	Wild-Type	BX571856.1/15939-16037_WT_FL	0.8574
20	Hammond131162	Wild-Type	ACMY01000004.1/34318-34423_MU_Partial	0.6754



Lanes
1. RNA Loading Control (10 pmol)
2. OLS 2.0 Seq 1 (Hammond037442, 5 pmol)
3. OLS 2.0 Seq 2 (Hammond234881, 5 pmol)
4. OLS 2.0 Seq 3 (Hammond120471, 5 pmol)
5. OLS 2.0 Seq 4 (Hammond119922, 5 pmol)
6. OLS 2.0 Seq 5 (Hammond213028, 5 pmol)
7. OLS 2.0 Seq 6 (Hammond223126, 5 pmol)

Figure 4.9. Biochemical validation of ribozyme labeling from OLS 2.0 microarray
The 3 sequences with highest labeling activity and their respective wild-type sequences were reacted with Hey-SAM and subsequent CuAAC chemistry with a Cy5-azide. Products were purified by spin column, separated on a 6% urea-PAGE gel and stained with EtBr. Gels were scanned using an imager with appropriate excitation light and filters corresponding to Cy5 and ethidium bromide.

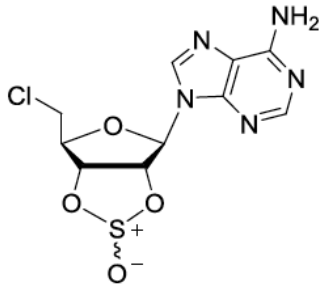
Table 4.4. Comparison of relative labeling ratios calculated from OLS 2.0 microarray and observed on a gel-based format. Green shading highlights top 3 sequences and yellow shading highlights their respective counterparts. **Bold** indicates a positive correlation between labeling ratios observed on microarray and on gel.

#	RNA sequence	% Labeling on Microarray (Cy5 / Cy3)	Relative Labeling Ratio on Microarray	Relative Labeling on Gel (Cy5 / EtBr)	Relative Labeling Ratio on Gel
1	OLS 2.0 Seq 1	6.8267	13.8	0.77	2.1
2	OLS 2.0 Seq 2	0.6744	1.4	0.89	2.5
3	OLS 2.0 Seq 3	6.3684	12.9	0.82	2.3
4	OLS 2.0 Seq 4	0.4934	1.0	0.95	2.6
5	**OLS 2.0 Seq 5**	**4.3966**	**8.9**	**2.18**	**6.1**
6	OLS 2.0 Seq 6	0.5309	1.1	0.36	1.0

4.7 Materials and Methods

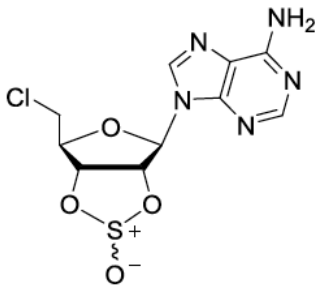
General Chemistry Methods

All reactions were performed in flame-dried glassware sealed with rubber septa under an inert atmosphere of nitrogen or argon and were agitated by magnetic stir bars unless noted otherwise. Liquid reagents were dispensed using a syringe and a standard Schlenk line for air- or moisture-sensitivity. Any anhydrous solvents were obtained freshly distilled from a Pure Solv Purification System (Innovative Technologies, Amesbury, MA) before use and other solvents were used as received. Reactions were monitored by thin layer chromatography on glass-backed silica plates visualized by UV irradiation or a denoted staining method when appropriate or on an Agilent 1260 Infinity 6120 Quadrupole LC/MS. Any column chromatography used silica gel from Sorbent Technologies and other purification methods such as HPLC purification was accomplished with an Agilent 1260 Infinity Series HPLC equipped with quaternary pump, diode array detector, and fraction collector and a Polaris 5 C18-A column (250 x 10 mm) using H2O + 0.05 % TFA (Solvent A), MeCN + 0.05 % TFA (Solvent B) as the mobile phase at 5 mL/min. HPLC solvents were filtered using a Millipore 0.2 μm nylon membrane.



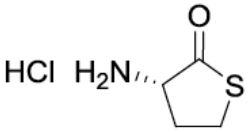
5'-Chloro-5'-deoxy-2',3'-O-sulfinyladenosine (**2**)

To a stirred suspension of adenosine (4.30 g, 15 mmol) in anhydrous pyridine (2.6 mL, 32 mmol) and CH3CN (50 mL) cooled to 0 oC in ice/H2O was added dropwise SOCl2 (5.9 mL, 9.70 g , 80 mmol) with stirring under N2. The reaction mixture was allowed to warm to rt and stirring was continued for 2 d. The resulting precipitate was collected via filtration, washed with anhydrous CH3CN under a stream of N2, and dried in vacuo to give **2** as white crystals (3.09 g, 62% yield, mixture of diastereomers 70:30). Rf 0.78 (1:10 MeOH:CH2Cl2); 1H NMR (400 MHz, DMSO) (major diastereomer) δ 8.66 (s, 1H), 8.50 (s, 1H), 6.50 (d, *J* = 2.9 Hz, 1H), 6.30 (dd, *J* = 6.5, 2.9 Hz, 1H), 5.89 (dd, *J* = 6.5, 3.8 Hz, 1H), 4.60 (td, *J* = 6.5, 3.8 Hz, 1H), 4.03 – 3.84 (m, 2H), (minor diastereomer) δ 8.70 (s, 1H), 8.49 (s, 1H), 6.70 (d, *J* = 3.0 Hz, 1H), 6.19 (dd, *J* = 7.5, 3.0 Hz, 1H), 5.78 (dd, *J* = 7.5, 3.9 Hz, 1H), 4.79 (td, *J* = 6.3, 3.9 Hz, 1H), 4.17 – 4.10 (m, 2H)



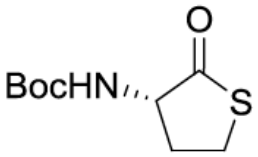
5'-Chloro-5'-deoxyadenosine (**3**)

To a stirred suspension of 30 (3.00 g, 9.04 mmol) in MeOH (45 mL) and H2O (10 mL) was added 28% NH4OH (3.76 mL, 27.1 mmol). Stirring was continued for 20 min until no starting material was observed by TLC, and MeOH was evaporated. The resulting white precipitate was dissolved in H2O at 70 oC then recrystallized by cooling to 4 oC overnight. The crystalline compound was filtered and dried in vacuo to give **3** (1.72 g, 66% yield). Rf 0.47 (1:10 MeOH:CH2Cl2); 1H NMR (400 MHz, DMSO-*d*6) δ 8.35 (s, 1H), 8.16 (s, 1H), 7.32 (s, 2H), 5.93 (d, *J* = 5.6 Hz, 1H), 5.61 (d, *J* = 6.0 Hz, 1H), 5.47 (d, *J* = 5.2 Hz, 1H), 4.76 (q, *J* = 5.4 Hz, 1H), 4.23 (q, *J* = 4.5 Hz, 1H), 4.15 – 4.01 (m, 1H), 3.95 (dd, *J* = 11.6, 5.1 Hz, 1H), 3.85 (dd, *J* = 11.6, 6.0 Hz, 1H).



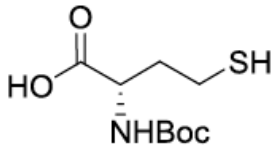
L-Homocysteine thiolactone hydrochloride (**5**)

To a stirred suspension of D, L-homocysteinethiolactone hydrochloride (10.0 g, 65.1 mmol) in acetone (50 mL) was added triethylamine (10.0 mL, 7.25 g, 71.6 mmol) and the mixture was stirred at ambient temperature for 2 h. Triethylamine hydrochloride precipitate was removed via filtration and the filtrate was saved. (S)-(+)-Mandelic acid (10.9 g, 71.6 mmol) was added to the reaction mixture in acetone (100 mL) followed by salicylaldehyde (0.9 mL,1.0 g, 8.5 mmol), which immediately turned the solution yellow. A white solid precipitate was recovered via filtration after 17 h. The solid was resuspended in acetone (100 mL) and cooled to 4 oC. A solution of 37% HCl (4.4 mL) was added with stirring for 90 min. The solid precipitate was recovered via filtration, washed with acetone and dried *in vacuo* to give **5** (4.74 g, 47% yield). 1H NMR (400 MHz, MeOD) δ 4.25 (m, 1H), 3.57 – 3.40 (m, 2H), 2.79 (m, 1H), 2.20 (m, 1H).
/



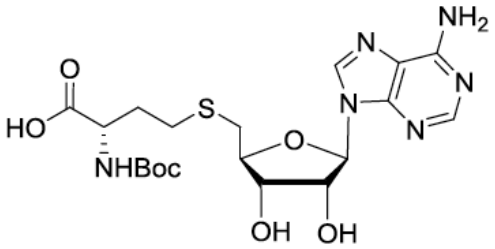
N-Boc-L-homocysteine thiolactone (**6**)

To a stirred suspension of **5** (5.00 g, 32.5 mmol) in CHCl3 (65.5 mL) was added NaHCO3 (2.73 g, 32.5 mmol) in H2O (50 mL) and NaCl (6.51 g, 110 mmol). Di-tert-butyl dicarbonate (7.5 mL, 7.1 g, 33 mmol) was dissolved in CHCl3 (5 mL) and added to the stirred reaction solution. The mixture was heated at reflux for 24 h. The organic layer was saved, and the aqueous layer was extracted with three additional 20 mL portions of CHCl3. The combined organic layers were dried over MgSO4. The solution was filtered, concentrated, and dried *in vacuo* to give **6** (6.33 g, 90% yield). Rf 0.88 (1:20 MeOH:CH2Cl2); 1H NMR (400 MHz, CDCl3) δ 5.01 (bs, 1H), 4.38-4.22 (bm, 1H), 3.37 – 3.27 (m, 1H), 3.27 – 3.19 (m, 1H), 2.98 – 2.68 (bm, 1H), 2.06 – 1.91 (m, 1H), 1.46 (s, 9H);



N-Boc-L-homocysteine (**7**)

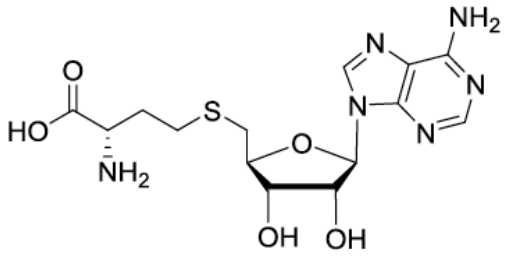
Compound **6** (5.02 g, 23.0 mmol) was dissolved in minimal MeOH and stirred at rt. A solution of 5 N NaOH (70 mL) was added, and after the disappearance of **6** was observed by TLC the reaction mixture was neutralized to pH 8 with 1 M KH2PO4 (345 mL). The MeOH was evaporated and the resulting aqueous solution was acidified to pH 2 with 12 N HCl. The cloudy white solution was extracted with three 100 mL portions of CH2Cl2 and the combined organic layers were dried over MgSO4. The solution was filtered, concentrated and dried *in vacuo* to give the product (5.35 g, 99% yield) 1H NMR (400 MHz, CDCl3) δ 6.78 (s, 0.25H), 5.17 (d, *J* = 8.3 Hz, 1H), 4.53 (bm, 0.6H) , 2.66 (m, 2H), 2.19 (m, 1H), 2.04 (m, 1H), 1.65 (s, 1H), 1.49 (s, 9H).



N-Boc-*S*-adenosyl-L-homocysteine (**8**)

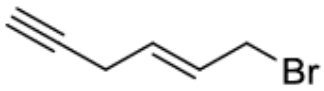
A solution of **7** (0.545 g, 2.32 mmol) in water (1.3 mL) and 10 N NaOH (0.7 mL, 7.0 mmol) was added to **3** and stirred at 40 oC overnight. The solution was neutralized with acetic acid and concentrated under reduced pressure. The crude mixture was purified by silica gel column chromatography (1:5:45 AcOH:MeOH:CH2Cl2) to give the

product as a yellow oil (1.07 g, 46% yield). Rf 0.22 (1:5:45 AcOH:MeOH:CH2Cl2); 1H NMR (400 MHz, MeOD) δ 8.33 (s, 1H), 8.19 (s, 1H), 5.99 (d, *J* = 5.2 Hz, 1H), 4.74 (t, *J* = 4.8 Hz, 1H), 4.31 (q, *J* = 4.8 Hz, 1H), 4.20 (m, 1H), 4.09 – 3.95 (m, 1H), 3.06 – 2.80 (m, 2H), 2.61 (m, 2H), 2.03 (m, 1H), 1.87 (m, 1H), 1.40 (s, 9H).



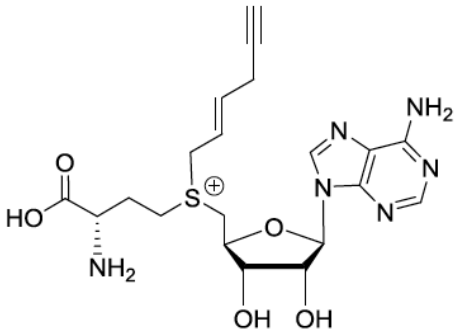
S-Adenosyl-L-homocysteine (**9**)

A 1:1 mixture of trifluoroacetic acid and dichloromethane was added to **8** (0.050 g, 0.089 mmol) and stirred for 5 min at ambient temperature. The mixture was diluted with H2O and the aqueous fraction was concentrated under reduced pressure. The crude mixture was purified by HPLC (5 % to 95 % solvent B over 10 min, ret. time 5.0 min) to give **9** (0.019 g, 40% yield). 1H NMR (400 MHz, D2O) δ 8.40 (s, 1H), 8.32 (s, 1H), 6.02 (d, *J* = 4.8 Hz, 1H), 4.74 (t, *J* = 5.0 Hz, 1H), 4.32 (t, *J* = 5.1 Hz, 1H), 4.22 (m,1H), 4.03 (m, 1H), 3.03 – 2.81 (m, 2H), 2.63 (t, *J* = 7.4 Hz, 2H), 2.08 (m, 2H). 13C NMR (101 MHz, D2O) δ 174.18, 152.48, 150.82, 147.01, 145.44, 121.46, 91.05, 86.21, 76.13, 74.84, 54.22, 35.87, 32.16, 29.85. HRMS calcd for C14H21O5N6S [M+H]+ 385.1289, found 328.1289.



(E)-1-bromo-hex-2-en-5-yne (**10**)

To a stirred solution of trans-1,4-dibromo-2-butene (**10**) (1.07 g, 5.0 mmol) and CuCl (0.060 g, 0.60 mmol) in anhydrous THF was added a 0.5 M ethynylmagnesium bromide solution (2.0 mL, 2.5 mmol) dropwise under N2. The solution was then heated to 60 oC for 2 h and allowed to stir at rt overnight. The reaction mixture was quenched with water (10 mL) and then 1 M HCl (10 mL) and was extracted with ether (3 x 15 mL). The organic layer was washed successively with 1M HCl, saturated sodium chloride solution, and water (3 x 5 mL each) The crude extract was dried and concentrated *in vacuo* to obtain a dark brown oil (0.853 g) that was used further without purification.



(E)-hex-2-en-5-ynyl-(*S*)-adenosyl-L-methionine (**11**)

Crude **11** was dissolved in 0.4 mL of a 1:1 formic acid and acetic acid at 4 oC and solution of **9** (0.005 g, 0.013 mmol) in water and AgClO4 (0.010 mg, 0.052 mmol) was added. The solution was stirred for 5 min and then allowed warm to r.t. and stir overnight. The reaction was monitored by LC-MS ((0 to 10% solvent B over 25 min and then 10 to 70% solvent B over 5 min) until completion then quenched with 2.5 mL of H2O + 0.01% TFA. LR-ESI displays [M+H]+ = 463.1 indicative of **11**. Purification is currently ongoing. The aqueous layer will be washed with ether (3 x 2 mL) and the purified extract will undergo purification by HPLC (0 to 10% solvent B over 25 min and then 10 to 70% solvent B over 5 min).

DNA oligonucleotides and PCR materials

DNA oligonucleotides for riboswitch constructs were purchased as Ultramers from Integrated DNA Technologies (Coralville, IA) and other DNA oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA). Mutant DNA libraries (OLS 1.0 and OLS 2.0) were constructed and gifted to us by Agilent Technologies for subsequent library amplification by PCR.

***In vitro* transcription**

DNA templates for in vitro transcription were prepared by PCR amplification using Phusion DNA polymerase (NEB) from sequence-confirmed plasmids, Ultramer oligonucleotides, or Oligonucleotide Library Synthesis products (Agilent Technologies) using primers that added the T7 polymerase promoter sequence at the 5' end. PCR products were purified either by QIAquick PCR purification kit (Qiagen) for subsequent transcription. RNA was transcribed from DNA templates using T7 RNA polymerase in 40 mM Tris- HCl, pH 8.0, 6 mM $MgCl_2$, 2 mM spermidine, and 10 mM DTT. RNAs were either purified by RNA-25 Clean & Concentrator spin columns (Zymo Research) or by denaturing (7.5 M urea) 6% PAGE. RNAs purified by PAGE were visualized by UV shadowing and extracted from gel pieces using Crush Soak buffer (10 mM Tris-HCl, pH 7.5, 200 mM NaCl and 1 mM EDTA, pH 8.0). Purified RNAs were precipitated with ethanol, dried, and then resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Accurate RNA concentrations were determined by measuring the absorbance at 260 nm after performing a hydrolysis assay to eliminate the hypochromic effect due to RNA secondary structure (Wilson et al., 2014).

Riboswitch reaction with Hey-SAM for alkyne labeling

Riboswitch/ribozyme transcript (5 μM, final concentration) was refolded in reaction buffer 1 (50 mM HEPES, pH 7.4, 300 mM NaCl, and 10 mM $MgCl_2$) by incubating to 72 °C for 3 min and allowing to cool to RT gradually. Hey-SAM was then added (10 μM, final concentration) and sample was incubated at 37 °C for 16 h. After incubation, the reacted transcript was EtOH precipitated and re-dissolved in 100 mM K_2PO_4, pH 7.4.

CuAAC click labeling with azide-dye

Modified ribozyme was subjected to coupling with AlexaFluor 488 azide or Cy5 azide (Invitrogen) by CuAAC click chemistry. In brief, modified ribozyme (25 μM , 14.4 mM) was incubated at 37 °C for 0.5 h in a mixture containing 50 μM dye azide, 10 mM Na-ascorbate, 5 mM CuSO4 and 10 mM THPTA ligand. Unreacted dye and Cu–THPTA complex were removed by gel-filtration spin column (Illustra G25, GE Healthcare) and elution products were subsequently analyzed on a 10% urea-PAGE gel.

Analysis of labeling efficiency by PAGE and fluorescence scan

Eluted products were quantitated using Beer's Law absorbance values at 260 nm and calculated molar extinction coefficients and approximately 5 pmol of product was loaded onto a 10% urea-PAGE gel. After electrophoresis, the gel was scanned with settings appropriate for the dye: AlexaFluor-488 with excitation at 488 nm and emission at 525 nm (520 BP 40 filter) and Cy5 fluorescence) with excitation at 532 nm and emission at 670nm (670 BP 30 filter) on a Typhoon Trio+ (GE Healthcare). As loading control, the unlabeled and unreacted ribozyme amount per lane was quantified after subsequent staining with ethidium bromide and a second fluorescence scan with EtBr settings (532 excitation, 580nm emission: (610 BP 30). All quantitation of band intensities were done by ImageJ analysis software.

4.8 References

Bennett, M. R.; Shepherd, S. A.; Cronin, V. A.; Micklefield, J.; Hauer, B.; Lutz, S. Recent Advances in Methyltransferase Biocatalysis. *Curr. Opin. Chem. Biol.* **2017**, *37*, 97–106.

Breaker, R. R. Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (2), a003566–a003566.

Cochrane, J. C.; Strobel, S. A. Riboswitch Effectors as Protein Enzyme Cofactors. *RNA* **2008**, *14* (6), 993–1002.

Forster, A. C.; Symons, R. H. Self-Cleavage of Virusoid RNA Is Performed by the Proposed 55-Nucleotide Active Site. *Cell* **1987**, *50* (1), 9–16.

Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T. Structure of the SAM-II Riboswitch Bound to S-Adenosylmethionine. *Nat. Struct. Mol. Biol.* **2008**, *15* (2), 177–182.

Gražvydas Lukinavičius, §; Vidmantas Lapienė, §; Zdislav Staševskij, §; Christian Dalhoff, #; Elmar Weinhold, # and; Saulius Klimašauskas*, §. Targeted Labeling of DNA by Methyltransferase-Directed Transfer of Activated Groups (MTAG). **2007**.

Islam, K.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Expanding Cofactor Repertoire of Protein Lysine Methyltransferase for Substrate Labeling. *ACS Chem. Biol.* **2011**, *6* (7), 679–684.

Jimenez, R. M.; Polanco, J. A.; Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem. Sci.* **2015**, *40* (11), 648–661.

Kim, S.; Gottfried, A.; Lin, R. R.; Dertinger, T.; Kim, A. S.; Chung, S.; Colyer, R. A.; Weinhold, E.; Weiss, S.; Ebenstein, Y. Enzymatically Incorporated Genomic Tags for Optical Mapping of DNA-Binding Proteins. *Angew. Chemie - Int. Ed.* **2012**, *51* (15), 3578–3581.

Labeling, C.; Peters, W.; Willnow, S.; Duisken, M.; Kleine, H.; Macherey, T. Enzymatic Site-Specific Functionalization of Protein Methyltransferase. **2010**.

Lu, C.; Ding, F.; Chowdhury, A.; Pradhan, V.; Tomsic, J.; Holmes, W. M.; Henkin, T. M.; Ke, A. SAM Recognition and Conformational Switching Mechanism in the Bacillus Subtilis YitJ S Box/SAM-I Riboswitch. *J. Mol. Biol.* **2010**, *404* (5), 803–818.

Lu, C.; Smith, A. M.; Fuchs, R. T.; Ding, F.; Rajashankar, K.; Henkin, T. M.; Ke, A. Crystal Structures of the SAM-III/S(MK) Riboswitch Reveal the SAM-Dependent Translation Inhibition Mechanism. *Nat. Struct. Mol. Biol.* **2008**, *15* (10), 1076–1083.

McDonald, R. I.; Guilinger, J. P.; Mukherji, S.; Curtis, E. a; Lee, W. I.; Liu, D. R. Electrophilic Activity-Based RNA Probes Reveal a Self-Alkylating RNA for RNA Labeling. *Nat. Chem. Biol.* **2014**, *10*, 1049–1054.

Motorin, Y.; Burhenne, J.; Teimer, R.; Koynov, K.; Willnow, S.; Weinhold, E.; Helm, M. Expanding the Chemical Scope of RNA:Methyltransferases to Site-Specific Alkynylation of RNA for Click Labeling. *Nucleic Acids Res.* **2011**, *39* (5), 1943–1952.

Paredes, E.; Das, S. R. Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **2011**, *12* (1), 125–131.

Poiata, E.; Meyer, M. M.; Ames, T. D.; Breaker, R. R. A Variant Riboswitch Aptamer Class for S-Adenosylmethionine Common in Marine Bacteria. *RNA* **2009**, *15* (11), 2046–2056.

Pressman, A.; Blanco, C.; Chen, I. A. The RNA World as a Model System to Study the Origin of Life. *Curr. Biol.* **2015**, *25* (19), R953–R963.

Roth, A.; Weinberg, Z.; Chen, A. G. Y.; Kim, P. B.; Ames, T. D.; Breaker, R. R. A Widespread Self-Cleaving Ribozyme Class Is Revealed by Bioinformatics. *Nat. Chem. Biol.* **2014**, *10* (1), 56–60.

Sharma, A. K.; Plant, J. J.; Rangel, A. E.; Meek, K. N.; Anamisis, A. J.; Hollien, J.; Heemstra, J. M. Fluorescent RNA Labeling Using Self-Alkylating Ribozymes. *ACS Chem. Biol.* **2014**, 1680–1684.

Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R. Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. *Science (80-.).* **2006**, *314* (5797).

Wang, J. X.; Lee, E. R.; Morales, D. R.; Lim, J.; Breaker, R. R. Riboswitches That Sense S-Adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling. *Mol. Cell* **2008**, *29* (6), 691–702.

Wang, R.; Islam, K.; Liu, Y.; Zheng, W.; Tang, H.; Lailler, N.; Blum, G.; Deng, H.; Luo, M. Profiling Genome-Wide Chromatin Methylation with Engineered Posttranslation Apparatus within Living Cells. *J. Am. Chem. Soc.* **2013**, *135* (3), 1048–1056.

Weinberg, Z.; Regulski, E. E.; Hammond, M. C.; Barrick, J. E.; Yao, Z.; Ruzzo, W. L.; Breaker, R. R. The Aptamer Core of SAM-IV Riboswitches Mimics the Ligand-Binding Site of SAM-I Riboswitches. *RNA* **2008**, *14* (5), 822–828.

Wilson, C.; Szostak, J. W. In Vitro Evolution of a Self-Alkylatlng Ribozyme. *Nature* **1995**, *374* (6525), 777–782.

Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. Control of Gene Expression by a Natural Metabolite-Responsive Ribozyme. *Nature* **2004**, *428* (6980), 281–286.