Exploring the RNA-Binding Profiles of Ribosomal Protein S15 Through *In Vitro* Selection

Daniel M. Beringer

A dissertation

submitted to the Faculty of

the department of Biology

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

May 2024

Abstract

Exploring the RNA-Binding Profiles of Ribosomal Protein S15 Through *In Vitro* Selection

Daniel M. Beringer

Advisor: Michelle M. Meyer, Ph.D.

Cis-regulatory RNA elements are structured regions of an mRNA that regulate the transcription, translational efficiency, or stability of the mRNA. These cis-regulatory RNAs are widely used across all domains of life to modulate gene expression in response to various stimuli. In bacteria, examples of these cis-regulatory RNAs include small RNAs, structured 50-500 nucleotide non-coding RNA that bind to mRNA or protein to alter expression, and riboswitches, which consist of a ligand-binding aptamer domain whose complex tertiary structure selectively responds to specific ligands to regulate downstream gene expression on the transcriptional or translational level. Ribosomal protein expression in bacteria is often controlled using an autogenous cis-regulatory mechanism, in which select ribosomal proteins (r-proteins) bind RNA structures in the 5'-untranslated of their own mRNA to regulate the expression of r-protein operons. Some of these structures, such as the RNA leaders regulating r-proteins L1, L20, and S2, have striking homology and often mimicry between the recognition motifs within their primary binding partner, ribosomal RNA (rRNA), and their secondary binding partner, the structured mRNA leader. Ribosomal protein S15 is a notable exception to this trend, as the five regulatory RNA leaders identified across various bacterial species that respond to S15 are structurally distinct, narrowly distributed to their respective phyla, and often bear little obvious homology to the rRNA. Additionally, inter-species interaction studies have shown that the S15 homologs from these species have specific recognition profiles for

the mRNA regulators, and not all interactions are reciprocal. How RNA regulators arise and are maintained in bacterial genomes is not well understood, and thus we sought to use ribosomal protein S15 as a model to study how differences in the RNA-binding profiles of the various S15 homologs may have driven the diversity of the mRNA regulators we see today.

To explore these RNA-binding profiles, I utilized an *in vitro* selection approach to enrich for aptamers (structured RNAs that bind a specific ligand) that bind the S15 homologs from *Escherichia coli* (EcS15), *Geobacillus kaustophilus* (GkS15), and *Thermus thermophilus* (TtS15) from a partially patterned RNA sequence pool. Following multiple attempts to enrich for aptamers to EcS15, I find that aptamers to this homolog are infrequent in this RNA sequence pool. I successfully enriched for Gk- and TtS15 aptamers from this sequence pool and using high-throughput sequencing and clustering analysis go on to show that these homologs have highly overlapping RNA-binding profiles, though the aptamers enriched by TtS15 exhibit slightly more sequence diversity than those enriched by GkS15. I confirm that three unique aptamers from the final RNA pools bind both homologs *in vitro*, and a single nucleotide change that differentiates two of these aptamers causes a decrease in affinity for TtS15 but not GkS15. This mutation causes a change in the predicted folding of these two aptamers, and greatly reduces its frequency in the population enriched by TtS15. Taken together, the work presented in this thesis shows overlapping but not identical RNA-binding profiles for the Gk- and TtS15 homologs to aptamers enriched from a partially patterned RNA library and represents the first comparative study of two homologous RNA-binding proteins using *in vitro* selection against an RNA library.

Dedication

To my parents, Mark and Debbie,
for their constant love and support,
And to Jeff, the Mary to my Rhoda.

—

"We did it!"
-Elle Woods

Acknowledgments

There are a multitude of friends, family, and colleagues that have supported me throughout my time at Boston College.

First and foremost, thank you to my advisor Michelle Meyer. You are truly one of the most intelligent, inquisitive, and enthusiastic scientists I have ever had the good fortune to meet, let alone work with. My journey through graduate school has had many ups and downs, but your support and encouragement has always been a constant source of inspiration and helped shape me into the scientist I am today. Thanks for everything.

To Torrey Mandigo: I could not have asked for a better partner-in-crime to be in a 2-person cohort with. Thank you for the years of fifth floor lunches, trips to the Hut, and making me an honorary uncle to Avery (and to Chelsea, too, without whom that would not have been possible).

To my former bay/bae mate Arianne Babina: thank you for your years of friendship, even after you ditched me for Europe.

To my other former bay mate Elise Gray: thank you for quite literally saving my life. I wouldn't be here without you. Thank you for being a friend.

To the many people I've had the good fortune of meeting in grad school: there are almost too many to name, but special thanks to Mary Ann Collins, Sam Dyckman, Matt Crum, Rebecca Korn, Victoria Hogan, Sara Hubbell, and Ciara Bauwens. Too many laughs and beers to count.

To my thesis committee, Babak Momeni, Eric Folker, Welkin Johnson, and Abhishek Chatterjee: thank you for your guidance and support and motivation to help me complete this work. Over the last 15 committee meetings (I counted – is that a record?), your insight and fruitful scientific discussions have been invaluable.

To my family, especially my parents Mark and Debbie Beringer: thank you for your unconditional love and support throughout my graduate school journey. I wouldn't be the person I am today without you and words cannot express how thankful I am to have such wonderful parents in my life. To my older siblings, Andy and Kristine: thanks for proving they always save the best for last.

Finally, thank you to my best friend Jeff Hinton. Thanks for the 20+ years of laughter, love, support, and visits to Boston. I don't know a better person and I'm so lucky to have you in my life. As Mary Lou Wiltberger said to Mrs. Fawcett: you are my muse.

Table of Contents

List of Figures and Tables

Abbreviations

Ec-mRNA: regulatory mRNA from *E. coli* that responds to S15

EcS15: S15 homolog from *E. coli*

GGC base triple: the S15 binding site formed at the junction of three helices in the 16S

rRNA and some mRNA regulators

Gk-mRNA: regulatory mRNA from *G. kaustophilus* that responds to S15

GkS15: S15 homolog from *G. kaustophilus*

mRNA: messenger RNA

r-proteins: ribosomal proteins

ribozyme: ribonucleic acid enzyme

rpsO: gene encoding S15 by interacting with a structured RNA in its 5'-UTR

rRNA: ribosomal RNA

SD: Shine Dalgarno sequence, part of the ribosome binding site on mRNA transcripts

SELEX: Systematic Evolution of Ligands by EXponential Enrichment

Tt-mRNA: regulatory mRNA from *T. thermophilus* that responds to S15

TtS15: S15 homolog from *T. thermophilus*

3HJ: three-helix junction

Chapter I

Introduction

Gene Regulation in Bacteria

Bacteria are constantly sensing external factors, such as nutrient availability, temperature, and cell density, as well as various intracellular stimuli, and have evolved a wide array of mechanisms to regulate gene expression in response to these stimuli. Promoter sequences are DNA elements located approximately 10 and 35 nucleotides upstream of the transcription site that direct RNA polymerase to transcribe the downstream genes (Barnard et al., 2004). Regulation of gene expression in bacteria occurs mainly at the transcriptional level, controlled by RNA polymerase. The specificity of this regulation is ensured by sigma factors, which are essential regulatory subunits of RNA polymerase that confer promoter specificity, controlling the expression of a specific set of genes (the so-called "regulon" of the corresponding sigma factor) (Helmann, 2019). σ70 is the primary sigma factor in *Escherichia coli*, and acts as the "housekeeping" sigma factor that transcribes most of the genes in dividing cells (Lal et al., 2018). There are also specialized sigma factors that respond to specific stimuli, such as σ32, which coordinates the heat shock response in *E. coli* to initiate transcription of heat shock proteins (mainly proteases and chaperones to maintain protein quality) (Nonaka et al., 2006).

Bacterial genomes are organized into arrays of operons where clusters of multiple genes are co-regulated under the control of the same promoter sequence. Regulatory sequences upstream of the operons control transcription of these genes, including enhancers/silencers and operator sequences (Bervoets & Charlier, 2019). These sequences allow for control at the transcriptional level, causing activation or repression of gene expression through binding of activator or repressor proteins, respectively. Gene regulation of the *lac* operon in *E. coli* was the first operon to be described in detail by Jacob and Monod, for which they were awarded the Nobel Prize in Physiology and Medicine 1965 (Jacob &

Monod, 1961). In the absence of lactose, the *lac* repressor protein binds an operator sequence, preventing transcription of the downstream genes used for lactose metabolism. When lactose is present, it binds to the repressor protein and prevents operator binding, activating gene expression. In contrast, the *trp* operon that encodes genes for tryptophan synthesis is inhibited by the presence of tryptophan, which binds the *trp* repressor protein, causes the repressor to bind the operator sequence, and down-regulates the operon. Beyond such mechanisms that regulate transcription initiation, there are multiple additional regulatory mechanisms that operate on the transcript itself to control gene expression.

RNA as a Regulatory Molecule in Bacteria

RNA has evolved a multitude of functions over the last 3.5 billion years of life on Earth. Due to its unique chemical structure, RNA can carry information, catalyze chemical reactions, and form complex tertiary structures that allow for finely tuned regulation of gene expression in response to various ligands in the cell (Gelfand, 2006). While genes are regulated mainly on the transcriptional level, RNA's versatility is widely exploited by bacteria, as evidenced by the diverse array of sequences, structures, and mechanisms of action they have evolved to regulate gene expression using RNA itself.

Bacteria have evolved so called "RNA thermometers" (RNATs) that act as thermosensors to control translation efficiency by occluding or exposing the ribosome binding site (Abduljalil, 2018). Some RNATs behave akin to zippers, reversibly opening and closing in response to ambient temperature changes. These RNATs play a pivotal role in controlling the expression of heat shock and virulence genes, allowing for bacterial pathogens such as *Vibrio cholerae, Listeria monocytogenes,* and *Yersinia pseudotuberculosis* to quickly turn on virulence genes upon entering a warm-blooded host or turn them off once they leave their host (Loh et al.,

2018). The wide diversity of RNA thermometer structures using the relatively simple mechanism of sequestering the ribosome binding site suggests independent derivation multiple times throughout evolutionary history. This independent derivation is also seen for another class of regulatory RNAs known as small RNAs.

Small RNAs (sRNAs) exert their effect through base pairing with target RNAs, modulating their translation and stability, or through directly binding proteins to affect their structure and function (Waters & Storz, 2009). Since their initial characterization in the 1980s, over 6,000 bacterial sRNAs have been identified and implicated in processes such as quorum sensing, stress response and virulence, biofilm formation, and metabolism (Li et al., 2013). The major families of sRNAs are true antisense RNAs, transcribed from the strand complementary to the mRNA that they regulate. sRNA-mRNA interactions are stabilized by the RNA chaperone *Hfq* to target the mRNA for degradation or to modulate translation (Vogel & Luisi, 2011). While essential for sRNAs in gram-negative bacteria, most sRNAs in gram-positive bacteria do not require *Hfq* at all to exert their effects, adding to the complexity of RNA-based regulation (Watkins & Arya, 2023). Due to their flexible structural requirements, the diverse origins of sRNAs range from *de novo* emergence to repurposing of pre-existing genetic elements from duplication events and horizontal gene transfer (Dutcher & Raghavan, 2018). Their ability to function with partial complementarity to mRNA targets enables rapid adaptation, but also complicates tracing sRNAs over long evolutionary distances.

Another major class of regulatory RNAs in bacteria are riboswitches, which are structured cis-regulatory RNA elements which act on their own mRNA transcripts to regulate gene expression. Riboswitches are located in the 5' untranslated region of some mRNAs and consist of a ligand-binding aptamer domain and a downstream expression platform. Ligand binding induces a conformational change that either up- or down-regulates downstream gene

expression and can act on either the transcriptional or translational level (Fig. 1.1A). The *glmS* riboswitch is a unique example that highlights the versatility of RNA as a regulatory molecule, as it acts as both a sensor and an effector. This riboswitch senses the metabolite glucosamine-6-phosphate, and upon ligand binding acts as an RNA enzyme (ribozyme) that turns off expression of the downstream genes by cleaving its own mRNA (Barrick et al., 2004). There are over 55 validated classes of riboswitches, many of which regulate essential processes in bacteria, making them a promising target for novel antibiotics (Olenginski et al., 2024). Their broad distribution also makes them ideal for studying important sequence and structural elements that allow them to discriminate from the myriad of molecules within the cell to selectively respond to their specific ligand. However, function does not always follow form with riboswitches, as homologous sequences that appear to be very similar may in fact have distinct ligand specificity.

Siblings or Doppelgängers? The difficulties of structure-function and homology

In contrast to DNA or protein-coding elements, the primary sequence conservation of cis-regulatory RNA elements can be extremely low because the secondary structure or folded structure is often more highly conserved than the primary sequence. However, just as with DNA or protein-coding elements, identifying apparent homology is still an integral component of the process for connecting growing sequence databases with biological functions. In recent years, improvements to computational methods have made identifying new and homologous cis-regulatory RNA elements easier. Yet, due to the unique properties of structured RNA, the use of homology still has some serious limitations. Sequence and secondary structure similarity often suggest common ligands for homologous riboswitch aptamers, but the detailed biochemical characterization and subsequent three-dimensional

structures can reveal minor sequence changes that lead to differences in ligand specificity. This phenomenon is exemplified by the *ykkC* riboswitches.

Riboswitches exhibit exquisite sensitivity for their ligands. Originally classified as a single type of riboswitch, the original *ykkC* aptamer evaded characterization for over a decade (Barrick et al., 2004). Furthermore, the discovery of non-homologous elements regulating similar sets of genes (mini-*ykkC* and *ykkC*-III) (Weinberg et al., 2007, 2010) only served to increase interest in these elements. Eventually, biochemical characterization subdivided the original *ykkC* aptamer into multiple sub-classes that recognize more than 5 distinct ligands: guanidine (*ykkC* subtype 1) (Nelson et al., 2017), guanosine-3', 5'-bisdiphosphate (ppGpp) (*ykkC* subtype 2a), and phosphoribosyl pyrophosphate (PRPP) (*ykkC* subtype 2b) (Peselis & Serganov, 2018; Sherlock et al., 2018) (Fig. 1.1B). Further validation of the *ykkC* subtype 2c expanded the ligands bound by this motif to include adenosine- and cytidine 5' diphosphates (in either their deoxyribose or ribose forms), while subtype 2d remains an orphan riboswitch whose ligand is unknown (Sherlock et al., 2019). RNA's central role in transcription and translation, coupled with the incredible diversity of structures into which an RNA can fold to interact with intracellular ligands and proteins, thus make it ideal for gene regulation.

RNA Leaders Regulate Ribosomal Protein Synthesis in Bacteria

Cis-acting regulatory RNAs can also regulate gene expression by interacting with the protein encoded by its downstream operon. Once enough of a certain protein is produced, the excess binds a structured RNA leader in the 5' untranslated region of its own mRNA. This mechanism is commonly used for the regulation of operons encoding ribosomal proteins (r-proteins). When cells are actively dividing, r-proteins preferentially bind to their primary binding partner, ribosomal RNA, to assemble and form the mature ribosome (Fig. 1.2). When

certain r-proteins are present in excess, they bind to the leader region of their own mRNA and induce structural changes to the mRNA transcript that compete with ribosome binding (displacement) or stall translation initiation (entrapment), acting as a negative feedback loop (Boehringer & Ban, 2007; Scott & Williamson, 2001). Nearly all ribosomal proteins are regulated by autogenous cis-regulatory RNAs that bind r-proteins, which allows them to maintain the correct stoichiometric amounts of ribosomal components (Nomura, 1999). The process is best described in the model organism *Escherichia coli*. More than half of the genes encoding ribosomal proteins (r- proteins) in *E. coli* are localized to twelve operons and the expression of the genes from these operons is controlled by specific autoregulatory RNAs. (Fu et al., 2013). 22 novel ribosomal leader candidates in bacteria and archaea have recently been identified, expanding the possible repertoire of ribosomal leaders even further (Weinberg et al., 2007).

For many r-proteins (e.g. L1, L20, S2), there is striking homology and often mimicry between the recognition motifs within the rRNA and the mRNA leader (Nevskaya et al., 2005). However, not all ribosomal leaders are so well-conserved. The most striking example is ribosomal protein S15, which has at least 5 distinct regulatory structures in *E. coli*, *Geobacillus stearothermophilus*, *Thermus thermophilus, Rhizobium radiobacter,* and most recently in *Mycobacterium smegmatis* (Aseev et al., 2021). These regulators all perform homologous functions, and yet their structures are unique and narrowly distributed within their respective phyla (Fig. 1.3). All the regulators appear to partially mimic their primary binding partner, the 16S rRNA, but do not have a conserved mechanism for preventing ribosome binding. The regulator from *E. coli* utilizes the entrapment mechanism while the regulator from *T. thermophilus* uses the displacement mechanism, for example. Further, inter-species *in vitro* binding and *in vivo* regulatory assays using a *LacZ* reporter indicate that all four S15 homologs studied can bind

and regulate using the mRNA regulator from *T. thermophilus* but only a subset of the homologs can bind and regulate using the mRNA regulator from *E. coli,* indicating differing requirements for S15-RNA interaction *(Slinger et al., 2015)*. The distinct mRNA binding profiles and lack of sequence and structural homology thus provide an interesting model to study how such diverse regulators came to be.

Systematic Evolution of Ligands by EXponential Enrichment (SELEX) to study structured RNAs

Systematic Evolution of Ligands by EXponential Enrichment (SELEX) is an iterative process used to select and enrich for DNA or RNA aptamers that bind a specific ligand (small molecule, protein, etc.) from a randomized sequence pool through rounds of selection. As rounds of SELEX proceed, selective pressure is increased through lowering ligand concentration or decreasing incubation time with the ligand, and structurally complex RNAs are selected and enriched (Kohlberger & Gadermaier, 2022). SELEX is thus an attractive platform to study the important structural determinants for RNA-protein interactions, especially when homologous proteins exhibit selectivity for structured RNAs, such as the case of r-protein S15 and its structurally distinct mRNA regulators. Previous *in vitro* evolution experiments using SELEX to select for RNAs from a randomized sequence pool that bind the S15 homolog from *Geobacillus kaustophilus* demonstrated that enriching for high-affinity RNAs is relatively facile, and a majority of the RNAs selected regulated *in vivo* (Slinger & Meyer, 2016). Given the S15 homolog-specific RNA-binding profiles, performing SELEX with various S15 homologs and comparing the RNAs enriched through selection could thus offer an ideal system to study if the small differences in RNA-binding profiles could have driven the evolution of the diverse extant mRNA regulators we see today.

Figures and Legends

Figure 1.1 Riboswitches in bacteria

A) Riboswitches are RNA elements located in the 5' UTR of many bacterial transcripts, regulating expression of downstream genes through conformational changes induced by ligand binding. Riboswitches can act transcriptionally or translationally and depending on the changes in RNA structure induced by ligand binding may act as ON or OFF switches. B) An overlay of the 3 riboswitches in the ligand bound state shows nearly overlapping structural scaffolds of the *ykkC* subtype 1, 2a, and 2b riboswitches. C) The cartoon schematic of the *ykkC* riboswitch motif and subtypes highlights the shared structural core of the guanidine-I (top), PRPP (middle), and ppGpp (bottom) riboswitches. Ligand selectivity for PRPP and ppGpp is conferred by an additional helical element (dashed lines). Assessing the individual active site structures shows that PRPP and ppGpp aptamers bind their ligands in overlapping positions, while the guanidine binds higher within the stem. The G96A point mutation (aqua, arrow) switches ligand specificity from PRPP to ppGpp. D) Chemical structures of the ligands bound by the ykkC subtype 1, 2a, and 2b riboswitches.

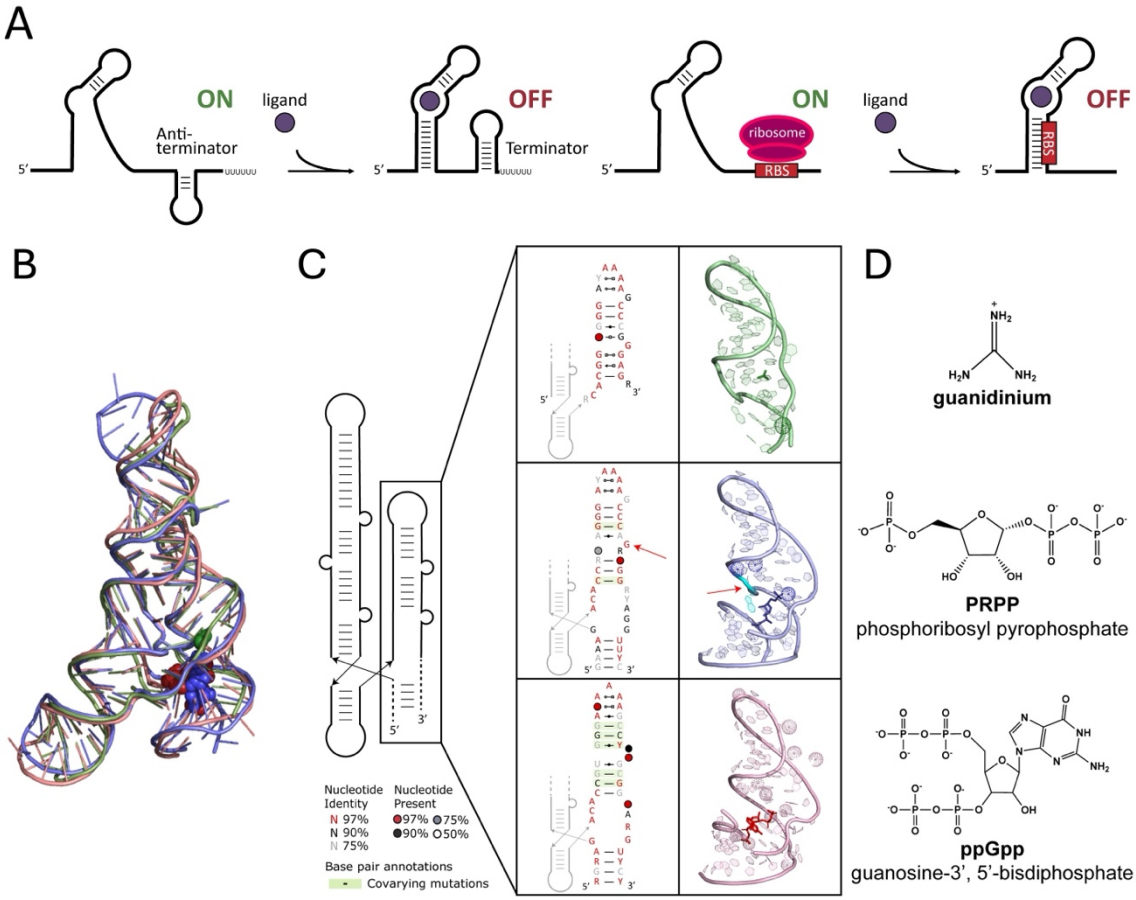


Figure 1.2 Autogenous regulation of ribosomal protein synthesis in bacteria. During ribosome assembly, ribosomal proteins typically bind to specific sites on rRNA. When rRNAs are saturated with bound proteins or when ribosomal proteins are in excess, select ribosomal proteins can interact with RNA structures located in the 5' UTR of their own mRNA transcripts to inhibit further ribosomal protein expression at the transcriptional or translational level in a negative feedback loop. Figure from (Babina, 2017.)

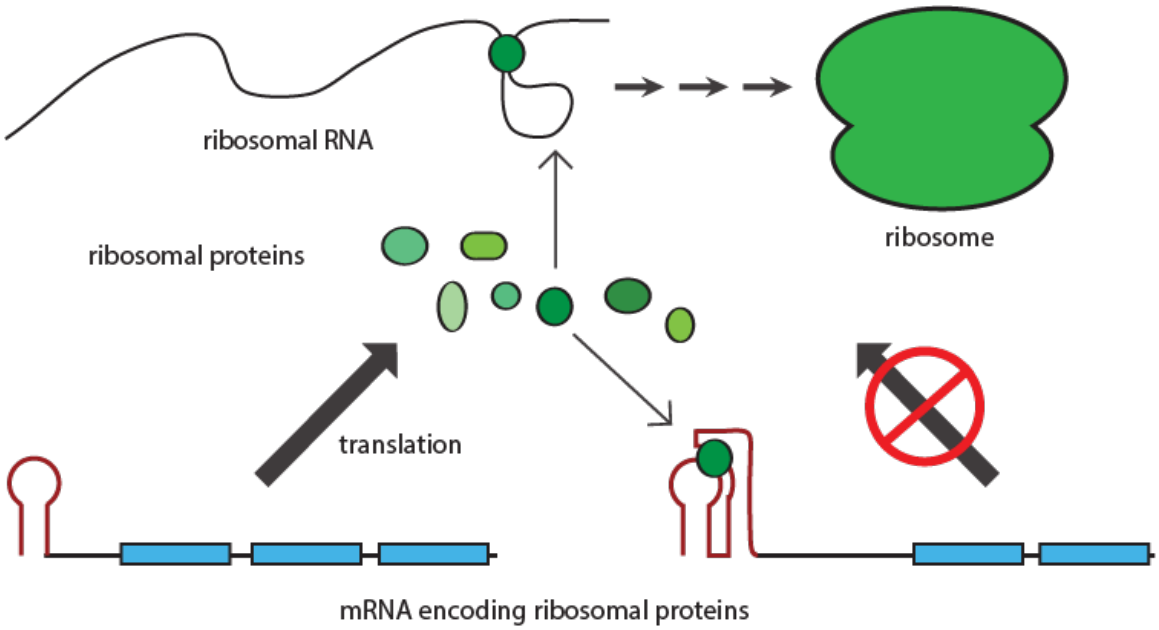
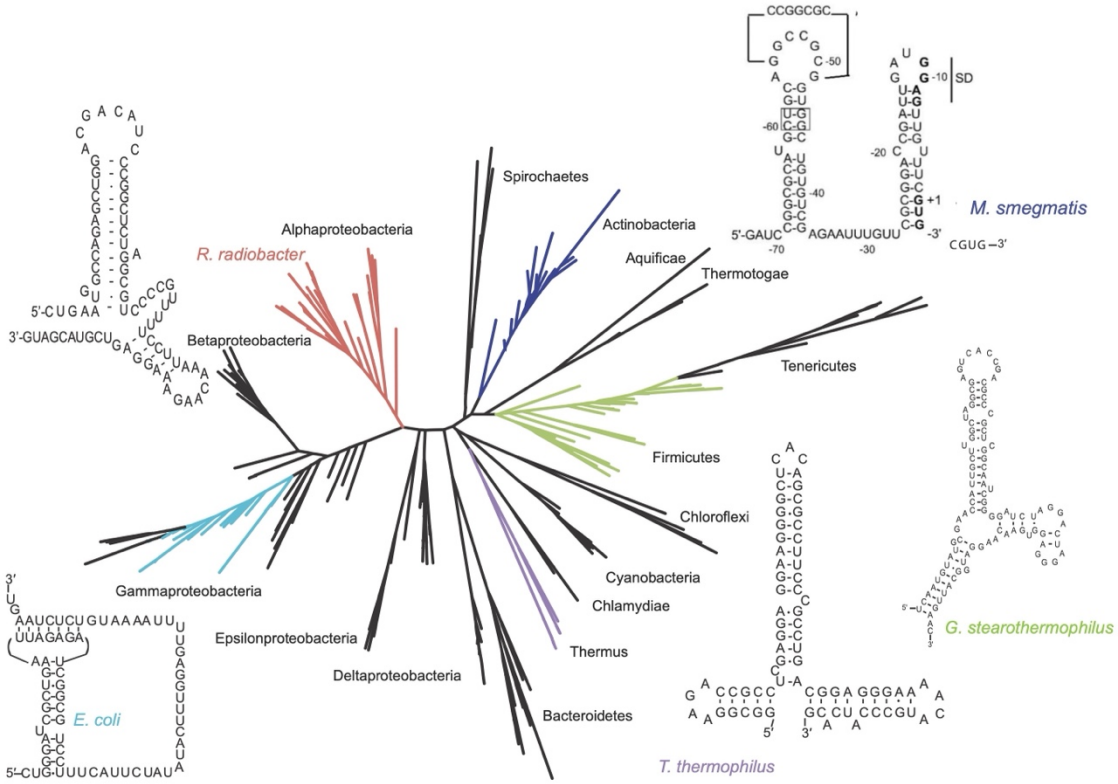


Figure 1.3 Diversity of S15-interacting RNA cis-regulators

The RNA structures that regulate gene expression in response to ribosomal protein S15 are narrowly distributed to certain bacterial phyla. The regulatory structures found in *E. coli, R. radiobacter, G. stearothermophilus, T. thermophilus,* and *M. smegmatis* have been experimentally verified. Adapted from (Slinger 2016 and Aseev 2021).



Chapter II

Library Design Considerations and Procedure Optimization for *In Vitro* Selection of S15 Aptamers

Introduction

Previous work has shown that diverse ribosomal protein S15 homologs have discrete binding profiles for the native S15 mRNA regulators both *in vitro* and *in vivo,* and reciprocal interactions are not necessarily conserved. The S15 homologs from *E. coli* and *T. thermophilus* (EcS15 and TtS15), for example, can utilize any of the mRNA regulators from diverse bacterial phyla to regulate in an *in vivo* reporter assay, whereas the S15 homolog from *G. kaustophilus* (GkS15) is more specific and only recognizes a subset of the regulators *in vivo* (Slinger et al., 2015). Notably, EcS15 can regulate using the Gk-mRNA regulator while GkS15 cannot regulate using the Ec-mRNA regulator, highlighting a lack of reciprocal interactions due to structural differences. This specificity is also reflected in *in vitro* binding assays, indicating that while all the S15 homologs tested presumably have conserved interactions with the 16S rRNA, S15-mRNA regulator interactions are not conserved. We hypothesized that differences in the RNA-binding profiles of the S15 homologs is a driver of the RNA regulator diversity seen in nature, and through comparing these RNA-binding profiles we can gain insight into how diverse sequences and structures may have evolved. It is possible that these regulators arose multiple times, or they may share a common ancestor that we are unable to detect using current RNA sequence and structure analysis. To assess the various RNA-binding profiles of the S15 homologs from *E. coli, G. kaustophilus,* and *T. thermophilus,* we utilized an *in vitro* selection approach. *In vitro* selection experiments can select for specific RNA-protein interactions and provide insights into the evolution of and requirements for binding.

Systematic Evolution of Ligands by EXponential Enrichment (SELEX) is an iterative process used to evolve RNA or DNA molecules that bind a specific target (from ions to proteins to whole cells) with high affinity, known as aptamers (Kohlberger & Gadermaier, 2022). Briefly, a fully or partially randomized RNA sequence pool is incubated with a target

molecule of interest (S15 in this case), S15 binders selected from the population via nitrocellulose binding, RNAs eluted from the nitrocellulose filter, reverse transcribed, amplified, and subjected to additional rounds of selection (Fig 2.1). As the rounds of SELEX continue, the stringency of selection can be increased by lowering protein concentration to discriminate between low and high affinity RNAs. For the SELEX process to be successful, there are several important considerations and optimization required before successfully isolating aptamers, such as library design, RNA renaturing, and binding reaction volume.

Results

Partially Patterned RNA Libraries Increase Likelihood of Structured RNAs

One of the most important considerations for SELEX is the initial library design, which will determine all the possible sequences that can be sampled during selection and ultimately the success of experiments providing insight into the potential RNA-binding pools of S15 homologs. SELEX libraries typically contain a completely randomized region of 30-50 nucleotides, which corresponds to 4^{30} to 4^{50} sequences, or $1.1x10^{18}$ to $1.3x10^{30}$ sequences. While a larger library size allows for more complex structures in the population, the sheer number of possible sequences and the limitations of current high-throughput sequencing technologies (~$4x10^{8}$ reads per run) make libraries of this size less suited for our goal to comprehensively sample the population over the rounds of *in vitro* selection. With this goal in mind, we sought to design a library of ~10^{12} sequences that would allow us to better monitor changes in the population of S15 binders with better resolution round over round. Previous work by Ruff and colleagues demonstrated that partially patterning nucleic acid libraries increases the likelihood of secondary structure formation while also reducing the total library size, resulting in functional aptamers that bind a protein target (Ruff et al., 2010). Functional DNA and RNA

aptamers selected through SELEX often have a high degree of secondary structures, like the S15 mRNA regulators.

To select a library with desirable thermodynamic properties likely to form complex structures, we designed libraries with fully and partially randomized variable regions to assess computationally. We designed 13 libraries that consisted of fully randomized (N) or partially patterned (RY, R*Y*, sz, or qx, with nucleotide ratios in Fig. 2.2A) regions followed by a Shine-Dalgarno (SD) sequence and the TTTTAAA spacer from *E. coli* or a randomized N7 spacer flanked by 2 constant primer binding regions at the 5' and 3' ends. Libraries were modeled after the EcS15 mRNA regulator, with the 3' constant region containing the first 6 amino acids of the *rpsO* gene from *E. coli* to simulate the genomic context in which these regulators would naturally evolve.

To choose the best library for our *in vitro* selection scheme, we wanted a library that had similar characteristics to the mRNA regulators that exist in nature, such as G•U/GC wobble base pairing and multi-stem structures. We used a combination of homemade scripts with RNAfold to randomly sample 100,000 sequences from each of the 13 libraries and evaluated a series of thermodynamic parameters - average Minimum Free Energy (MFE), multi-stem percentage, longest stem and longest loop, and average number of wobble base pairs (Fig. 2.2C). The average MFE of the libraries ranged from -12 to -22 kcal/mol, indicative of structured RNAs (Fig. 2.2C). All the partially patterned libraries had lower average MFEs than the completely randomized Library 1 (Fig. 2.2B, C). All of the validated natural S15 mRNA regulators have multi-stem structures, so libraries with an increased likelihood of secondary structure formation are desired. One of the key binding sites for all S15 homologs on the 16S rRNA and multiple S15 mRNA regulators includes a G•U wobble base pair, thus a library that contained potential wobble base pairing was also desired. The average number

of wobble base pairs for the libraries ranged from 2.46-3.40 per sequence. There is minimal difference in the number of wobble base pairs across the libraries, and only one wobble base pair is important for binding *in vivo* (Bénard et al., 1998). Based on our simulations, we selected Library 9 because of its low MFE (-17.04 kcal/mol), small loops (average longest loop of 7.45 nucleotides), multi-stem structures (3.36% of structures contain multiple stems), and wobble base pairing (average of 2.92 pairs per structure).

Library 9 contains a partially patterned stretch of 24 nucleotides (R*Y*24, where R*= 45%A, 45%G, 5%T, 5%C; Y*=45%C, 45%T, 5%A, 5%G), Shine-Dalgarno (SD) sequence, and 7 random bases between the SD sequence and the first 6 codons of the *E. coli rpsO* gene (Fig. 2.2D). This corresponds to a library of roughly $2.7x10^{11}$ sequences, which could be deeply sampled using *in vitro* selection, and the thermodynamic parameters reflect that partial patterning increased predicted secondary structure formation compared to a fully randomized library, which should lead to functional RNA aptamers.

SELEX Highlights EcS15's Poor *in vitro* Binding

Since our library was designed using the first 6 codons from *E. coli* in the constant region, we performed our first *in vitro* selection in duplicate with the S15 homolog from *E. coli* (EcS15) and the partially patterned Library 9 (Fig 2.1D). Throughout this chapter, we refer to Rounds of SELEX, which encompass the whole process depicted in Fig. 2.1 - removing nitrocellulose binding RNA, incubating the RNA pool with S15 at a set concentration, eluting, reverse transcribing, PCR amplifying, and transcribing the pool for the next Round. We also refer to SELEX schemes, which includes all the Rounds of SELEX using the same conditions (i.e. RNA renaturing, amount of RNA, binding reaction volume). We tested the binding affinity of the unselected library population by radiolabeling a subset of the library and

performing filter-binding assays. EcS15 showed poor binding, even at the highest protein concentration, and we were unable to calculate a binding affinity for the unselected library (Fig. 2.3A). While low binding affinity for a randomized library before selection is not surprising, we still expected to see some binding as EcS15 is an RNA-binding protein. Thus, we began SELEX at a relatively high concentration of 4 μM EcS15 in a 500 μL binding reaction with 200 pmol of library RNA renatured in water. For reference, a previous SELEX scheme with the S15 homolog from *G. kaustophilus,* GkS15, started at 1250 nM GkS15 and exhibited a K_D of 1329 nM for the fully randomized library before selection (Slinger & Meyer, 2016). As we progressed with Rounds of SELEX, we halved the EcS15 concentration every other round to increase the selection stringency. This increases competition between the RNAs in the pool for that round, and RNAs with higher affinities for EcS15 will survive to the next round. As we lowered [EcS15], we periodically measured the population's binding affinity through filter-binding assays to test for enrichment in EcS15-binding RNAs. We saw an increase in binding at the higher EcS15 concentrations (4.096 μM EcS15) over the Rounds, though there was not much shift in the fraction bound at lower concentrations (Fig. 2.3A). Following Round 11 of SELEX ([EcS15] was 15 nM in this round), the population K_D was 885 nM. Typically, the population K_D is approximately equal to the concentration of protein used in that Round of SELEX, which was 15 nM in Round 11, but we attributed this low binding affinity to EcS15 generally being a poor *in vitro* binder.

To evaluate which sequences were enriched through our first SELEX scheme, we prepared amplicons for sequencing from Rounds 1, 3, 5, 8, 10, and 11 to monitor changes in the RNA population as the Rounds of SELEX progressed. After merging and removing low-quality reads, we had ~50,000 reads per round per replicate. By Round 11, the population was still relatively diverse, though there were 9 unique high frequency sequences that ranged from

2-15% of the total reads for the final round. Looking at the sequences manually, they appeared to cluster together into roughly 4 groups that had related sequences with identical R*Y*24 randomized regions and differentiated only by their N7 regions. We selected 5 of these sequences to test individually – 1 from each of the 4 "clusters" and an additional sequence that lacked a SD sequence (Fig. 2.3B). All 5 of these sequences bound EcS15 with a better affinity than the Round 11 population, with K_D's ranging from 453 to 559 nM (Fig 2.3B). We repeated filter-binding assays with the individual sequences to get more replicates using a different EcS15 preparation but were unable to replicate our previous results. Filter-binding assays with the new EcS15 preparation and SELEX 1, the most frequent sequence from Round 11 of this SELEX scheme, showed non-specific binding and did not bind this aptamer with the same affinity (Fig. 2.3C). It is possible that other proteins co-purified with our original EcS15 preparation that were not present in the new preparation, causing non-specific binding at higher concentrations.

To determine why this may be the case, we compared the protocols for SELEX and filter-binding assays, which only differed in how we renatured the RNA. During SELEX, we renatured the RNA pool in water, then added binding buffer (Buffer A) before incubating with EcS15 in the binding reaction. During our filter-binding assays, we renatured our individual RNAs in Buffer A before incubating with EcS15. Buffer A contains magnesium, which is an important divalent cation that can affect RNA folding (Bowman et al., 2012). We hypothesized that our RNA folded differently when renatured in Buffer A versus renatured in water.

To test this hypothesis, we renatured the dominant RNA from Round 11, SELEX 1, in either water or Buffer A to determine if our renaturing protocol was affecting RNA folding and S15 binding. Using these two renaturing conditions, we then performed a series of binding

reactions from 0 to 500 nM EcS15 in Buffer A. When we ran these reactions over the nitrocellulose and nylon membranes, the RNA renatured in water bound to the nitrocellulose regardless of the EcS15 concentration (65% bound at 0 nM EcS15), indicating that binding was nonspecific and SELEX 1 RNA was not interacting with EcS15 (Fig 2.4A). SELEX 1 RNA renatured in Buffer A did not bind to the nitrocellulose at any EcS15 concentration, supporting our hypothesis. This difference in renaturing partially explains why this SELEX scheme with EcS15 was unsuccessful since we were not removing nonspecific nitrocellulose binders and thus enriched for nitrocellulose-binding aptamers in our populations (Fig 2.4B).

To correct for improper RNA folding and reduce nitrocellulose binding, we performed a second SELEX scheme where we changed our protocol to renature the RNA pool in Buffer A for every round. After 4 Rounds of SELEX, the population affinity for this Round was the same as the unselected population (Fig. 2.4C). We continued with another 8 Rounds of SELEX, halving the [EcS15] every other round, but were unable to isolate any high affinity RNAs for this homolog using this SELEX scheme (Fig 2.4C). Additional SELEX schemes to increase stringency in early rounds by starting at 2 μM or 1 μM EcS15 in the first SELEX Round also failed to enrich for EcS15 aptamers (data not shown). To determine if there was an issue with the EcS15 preparation, we performed filter-binding assays with a positive control RNA (Rr-mRNA, which binds EcS15 with a $K_D \sim 29$ nM) and found our EcS15 preparation had a similar affinity for this RNA.

Discussion

Randomized RNA libraries allow for the full complexity of sequence and structure of a population to be explored, but these libraries are limited by synthesis scale and experimental sampling *in vitro*. In a typical SELEX experiment, the unselected library has a randomized

region of 30 nucleotides (N30), which corresponds to 4^{30} possible combinations of the 4 bases, or $1.1x10^{18}$ unique sequences. This would require 7.65 μmoles of library to sample every sequence just once, and in a binding reaction with 200 picomoles of RNA there would be 0.0000261 copies of each sequence. By utilizing a partially patterned library, we sought to reduce the total library size to roughly 10^{12} sequences to both fully sample the sequence space of the library for potential S15 binders *in vitro* and more comprehensively sample the selected sequences through high-throughput sequencing following *in vitro* selection.

We designed 13 libraries containing either a fully randomized N20 region or partially patterned randomized region ranging from 24-30 nucleotides in length with A:C:G:U at 8 different ratios. We chose this range of lengths since RNAs with long stems can form helices, and all S15 homologs are known to interact with the 16S rRNA at 2 distinct but conserved sites that are separated by ~1 helical turn. We calculated thermodynamic properties of the libraries by randomly sampling 100,000 sequences from each library design and determining the average Minimum Free Energy, average longest loop and average longest stem, multi-stem percentage, and average number of wobble base pairs. The lower the minimum free energy, the more thermodynamically stable the predicted structures are. All the libraries with partially patterned regions, regardless of length, had lower average minimum free energies than the fully randomized library. This correlates with previous studies that showed partial patterning increases the likelihood of base pairing compared to randomized sequences, which is desirable for structured RNAs to form (Ruff et al., 2010). We also calculated the longest loop and longest stem for each library, as long loops are more likely to disrupt secondary structure while long stems are more likely to stabilize secondary structure and found the partially patterned libraries to have smaller longest loops and longer stems. Strikingly, the number of RNAs with multi-stem structures for the fully randomized library was 0.74% while the partially patterned

libraries ranged from ~3-12%. All the validated S15 mRNA regulators identified contain multi-stem structures, as well as the 16S rRNA, S15's primary binding partner, so we expected an increased likelihood of complex secondary structures in the partially patterned libraries. Finally, all the libraries evaluated had ~3 wobble base pairs in the sampled sequences, which was the one thermodynamic property of the libraries that did not seem to be affected by a totally randomized region versus partially patterned region, and S15 requires only 1 wobble base pair for its interaction with the 16S rRNA and is equally likely to occur in all the libraries sampled. The library we chose had a 24 base partially patterned region and fully randomized N7 spacer sequence between the Shine-Dalgarno sequence and the start codon to simulate the genomic context the S15 mRNA-regulators would have evolved in. This represents ~2.75×10^{11} sequences, or 0.45 picomole of library to sample every sequence once in a SELEX binding reaction, which corresponds to 444 copies of each sequence in the 200 picomoles used in an S15 binding reaction.

We performed our first *in vitro* selection using the S15 homolog from *E. coli* (EcS15) and the partially patterned library as a proof of concept. Following 11 Rounds of SELEX with EcS15 in our first SELEX scheme, we sequenced our populations and selected 5 individual RNAs for further study. Although it appeared that we had enriched for binders, we were unable to replicate our results with a different EcS15 preparation. Further experiments showed that due to renaturing the RNA pool in water, we selected for RNAs that bound nitrocellulose rather than EcS15 in this SELEX scheme.

Still, it is puzzling as to why the individual aptamers tested bound the old EcS15 preparation. Since we used the same EcS15 preparation for our first SELEX scheme and all the filter-binding assays, it is possible that another protein co-purified with our initial EcS15 preparation that bound our RNA pool was not present in our newer preparation. Regardless,

following repeated attempts using the EcS15 homolog using a new protein preparation and a different SELEX scheme with the RNA pool renatured in Buffer A, we were unable to enrich and select for aptamers that bind this homolog. Notably, the native Ec-mRNA interaction with EcS15 has a K_D of 231 nM, which is significantly higher than other native interactions (2.11 nM for TtS15 and 3.47 nM for GkS15), and strictly requires a G•U/G-C motif for binding (Slinger et al., 2015). Therefore, it is possible that due to our library design, high affinity aptamers for EcS15 are very infrequent in our RNA pool.

While we were unable to enrich for EcS15-binding aptamers through *in vitro* selection, our SELEX experiments did highlight some of the unique aspects of regulation with EcS15. In contrast to GkS15 and TtS15, EcS15 uses an "entrapment" mechanism to regulate expression of the *rpsO* operon When there is an excess of EcS15 in the cell compared to the 16S rRNA, EcS15 binds the structured mRNA regulator of its own transcript and the pre-initiation complex of the ribosome simultaneously, which ultimately prevents full ribosome assembly and thus inhibits translation of the *rpsO* operon (Philippe et al., 1993). This mechanism is difficult or impossible to replicate with *in vitro* experiments. In our experiments the only ribosomal component in our binding reaction is EcS15 and thus we cannot simulate this evolutionary context. To successfully isolate EcS15-binding aptamers, we could utilize an *in vivo* selection approach. Previous work in our lab has shown that it is possible to assess an RNA library for regulatory activity with S15 and a fluorescent reporter, though this method has not been fully optimized (Gray, 2022). One of the major drawbacks for this method is that forming the droplets used in this assay is a severe bottleneck (~50,000 maximum per library), which would severely limit our ability study the full RNA-binding profile of the S15 homologs. However, given the unique challenges posed by this S15 homolog, this may still be the most viable option for studying the RNA-binding profile of EcS15.

Materials and Methods

Library Folding Simulations

We computationally simulated thirteen libraries containing randomized or partially patterned sequences. Utilizing a combination of perl and clojure scripting with RNAfold, we randomly sampled 100,000 sequences from each library to assess key thermodynamic, folding, and structural parameters. We folded and calculated the average Minimum Free Energy (MFE), multi-stem percentage, average longest stem and longest loop, and average number of wobble base pairs for each library. Following these simulations, we selected Library 9 for use in our *in vitro* selection.

Protein Preparation

The *E. coli rpsO* ORF was previously cloned into the pET-HT overexpression vector and transformed into chemically competent BL-21(DE3) cells (Invitrogen). Protein was overexpressed and cells lysed by freeze-thaw lysis followed by sonication in S15 Resuspension Buffer (100 mM Tris-HCl, pH 8.0, 800 mM NaCl, 150 mM MgCl2). S15 was soluble and purified at 4°C using non-denaturing FPLC cation exchange chromatography (pH 8.0) with a linear salt gradient (20 mM – 1 M KCl). Fractions containing protein were tested for nucleases, and RNase-free protein fractions were concentrated, analyzed via SDS-PAGE, and buffer exchanged for the S15 Storage Buffer (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 0.02% sodium azide). Final protein concentration was determined by Bradford assay and stored at 4°C.

RNA Preparation and SELEX

The transcription template for the unselected library was generated through annealing, extending, and amplifying the primers Library 9-24 rev and T7 + A fwd to make double-stranded template containing a T7 promoter, as previously described (Urak et al., 2016) (Fig.

2.5). Double-stranded DNA library was gel-purified (Zymo) and used as template for transcription reactions using T7 polymerase (Milligan et al., 1987). Transcription products were purified by 6% denaturing PAGE, bands visualized using UV shadow, excised from the gel, eluted into crush-soak buffer (200 mM NaCl, 1 mM EDTA pH 8, 10 mM Tris-HCl pH 7.5), and ethanol precipitated overnight at -20º.

In the first SELEX scheme, 200 pmol of RNA was renatured in water at 42º for 15 minutes and cooled to room temperature for 10 minutes, then filtered through a 0.45 μM nitrocellulose filter. Surviving RNAs were incubated with EcS15 in Buffer A (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 5 mM dithiothreitol, 0.02% bovine serum albumin) at 25º for 30 minutes. RNA-EcS15 complexes were isolated by filtering over a second nitrocellulose filter and washed twice with Buffer A. RNAs were eluted from the filter at 95º in elution buffer (7 M Urea, 100 mM $Na_3C_6H_5O_7$, 3 mM EDTA pH 8.0), protein removed through phenol-chloroform extraction, and ethanol precipitated overnight at -20º. The selected RNA was reverse transcribed using M-MuLV and half of the cDNA amplified using standard PCR with primers complementary to the library constant region and to add the T7 promoter (Fig. 2.5). The remaining cDNA was amplified with primers to add Illumina sequencing adapters and barcodes. T7 PCR products were gel purified and used as template for T7 transcription reactions to make RNA for the next round of selection. In the subsequent SELEX scheme, 200 pmol RNA was renatured in Buffer A for 15 minutes at 42º and cooled to room temperature for 10 minutes, filtered through a 0.45 μM nitrocellulose filter, and the binding reaction and selection performed as described above.

Filter Binding Assays

EcS15-RNA binding affinity was periodically examined throughout SELEX by filter-binding assay. 10 pmole of RNA to be tested (unselected library, round, or individual

sequence) was 5'-labeled with ^{32}P-ATP. 5'-labeled RNA was renatured for 15 minutes at 42º in Buffer A then cooled to room temperature for 10 minutes. Trace amounts of RNA (1000 cpm, <1 nM per binding reaction) were then incubated with serial dilutions of EcS15 in Buffer A for 30 minutes at 25º. Nitrocellulose membrane (GE Healthcare) was used to collect EcS15-RNA complexes and nylon (GE Healthcare) to collect unbound RNA under suction. Membranes were air-dried 5 minutes and fraction bound quantified by imaging membranes following an overnight exposure to a phosphorimaging screen. Radioactivity counts per sample per membrane were measured using the GE Healthcare Typhoon™ FLA 9500 Phosphorimager and ImageQuant. The fraction bound was calculated per individual protein concentration: Fb = (counts nitrocellulose)/(counts total). To determine the K_D and the maximum fraction bound (Max %), the resulting values were fit to the equation: Fb = (Max % * [S15])/([S15] + K_D) where [S15] corresponds to the concentration of EcS15 in the reaction. The residuals were minimized using the Solver function in Microsoft Excel to find both the Max % and the K_D.

Amplicon Sequencing

2-step PCRs were used to add Illumina adapters, barcodes, and a unique molecular identifier (UMI) to the cDNA generated following each round of SELEX. Amplicon was gel purified and double-stranded DNA quantified using a Qubit fluorometer. Amplicon from rounds 1, 3, 5, 8, 10, and 11 for both replicates were sequenced through GENEWIZ. Raw FASTQ files were merged, low-quality reads removed, and frequencies for individual sequences calculated using AmpUMI (Clement et al., 2018). The top 100 most frequent sequences from Round 11 from both replicates were compared and the 5 most frequent sequences shared between the replicates were used for filter-binding assays.

Figures and Legends

Figure 2.1 Overview of the SELEX process

Diagram of the *in vitro* selection process using Systematic Evolution of Ligands by EXponential Enrichment (SELEX). A DNA pool of the library of interest containing a T7 promoter is transcribed using T7 RNA polymerase to generate the RNA pool. RNA is renatured, then non-specific nitrocellulose binders removed through filtering in "Negative Selection". The RNA pool is then incubated with the S15 homolog of interest in a binding reaction, then run over a second nitrocellulose filter in "Positive Selection" to isolate the S15-RNA complex and non-binding species remain in the flow-through. As the rounds of SELEX are run, [S15] is decreased to increase the stringency of selection. S15-RNA complexes are eluted from the filter, protein removed, RNA reverse transcribed, then amplified using PCR to add the T7 promoter for the next round of SELEX or using primers to add Illumina adapters and barcodes for sequencing. This completes a round of selection. [S15] was decreased to increase the stringency of selection over the rounds, and population binding affinity periodically measured using radiolabeled RNA.

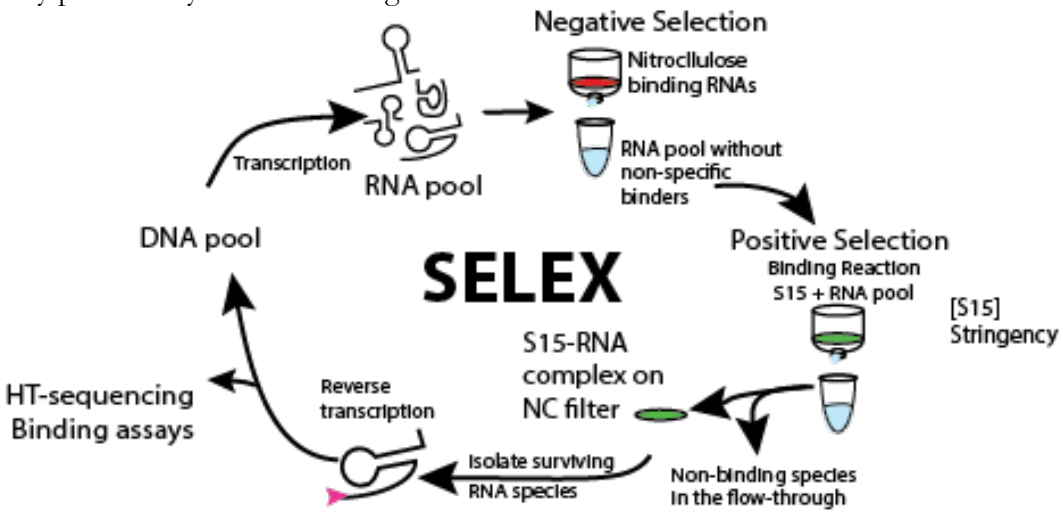


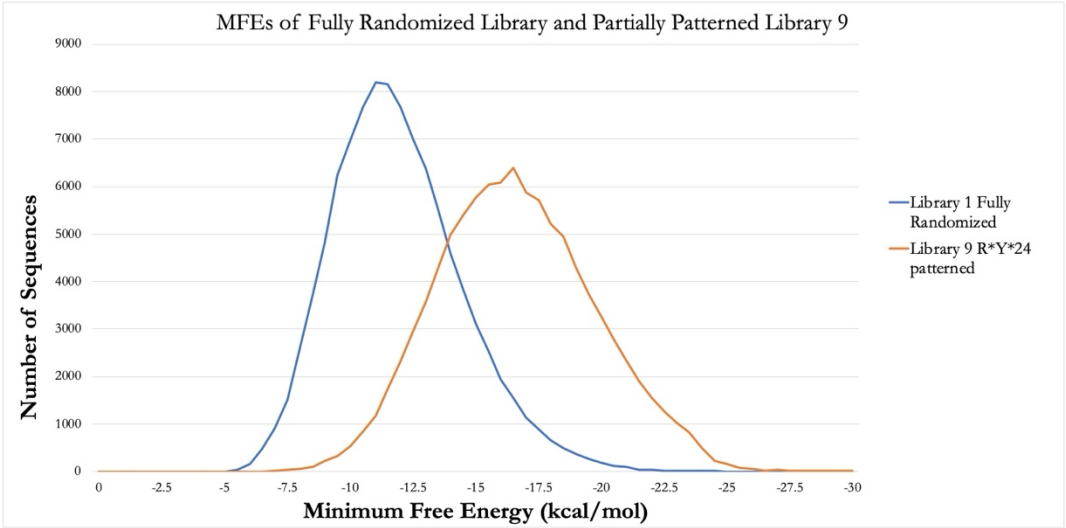
Figure 2.2 Libraries tested in folding simulations

A) Libraries with randomized and partially patterned regions of varying lengths were folded and characterized computationally. Different nucleotide ratios for the partially patterned sequences were also tested in our simulations. N = 1:1:1:1 probability of A/C/G/U, R = 1:1 probability of A/G, r = 9:1:9:1 probability of A/C/G/U, Y = 1:1 probability of C/U, y = probability of 1:9:1:9 A/C/G/U, q = probability of 42:7:43:8 A/C/G/U, x = probability of 7:42:8:43 A/C/G/U, s = probability of 47:2:48:3 A/C/G/U, and z = probability of 2:47:3:48 A/C/G/U. Constant regions are underlined, randomized regions bolded, and the Shine-Dalgarno sequence italicized. B) Histogram of the Minimum Free Energy of the 100,000 sequences sampled and folded for the fully randomized Library 1 and partially patterned Library 9. C) Thermodynamic parameters of all libraries sampled. D) Schematic of Library 9, our chosen starting pool.

A

Library_1: CGUAGUCGUAGCUGAUCGAC**NNNNNNNNNNNNNNNNNNNN***GGAG*UUUUAAAAUGUCUCUAAGUACUGAA
Library_2: CGUAGUCGUAGCUGAUCGAC**RYRYRYRYRYRYRYRYRYRYRYRYRYRY***GGAG*UUUUAAAAUGUCUCUAAGUACUGAA
Library_3: CGUAGUCGUAGCUGAUCGAC**ryryryryryryryryryryryryryry***GGAG*UUUUAAAAUGUCUCUAAGUACUGAA
Library_4: CGUAGUCGUAGCUGAUCGAC**RYRYRYRYRYRYRYRYRYRYRYRYRYRY***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_5: CGUAGUCGUAGCUGAUCGAC**ryryryryryryryryryryryryryry***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_6: CGUAGUCGUAGCUGAUCGAC**RYRYNNNRYRYNNNRYRYNNNRYRY***GGAG*UUUUAAAAUGUCUCUAAGUACUGAA
Library_7: CGUAGUCGUAGCUGAUCGAC**ryryNNNryryNNNryryNNNryry***GGAG*UUUUAAAAUGUCUCUAAGUACUGAA
Library_8: CGUAGUCGUAGCUGAUCGAC**RYRYRYRYRYRYRYRYRYRYRY***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_9: CGUAGUCGUAGCUGAUCGAC**ryryryryryryryryryryry***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_10: CGUAGUCGUAGCUGAUCGAC**szszszszszszszszszszszszszsz***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_11: CGUAGUCGUAGCUGAUCGAC**qxqxqxqxqxqxqxqxqxqxqxqxqxqx***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_12: CGUAGUCGUAGCUGAUCGAC**szszszszszszszszszszszsz***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA
Library_13: CGUAGUCGUAGCUGAUCGAC**qxqxqxqxqxqxqxqxqxqxqxqx***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA

B



C

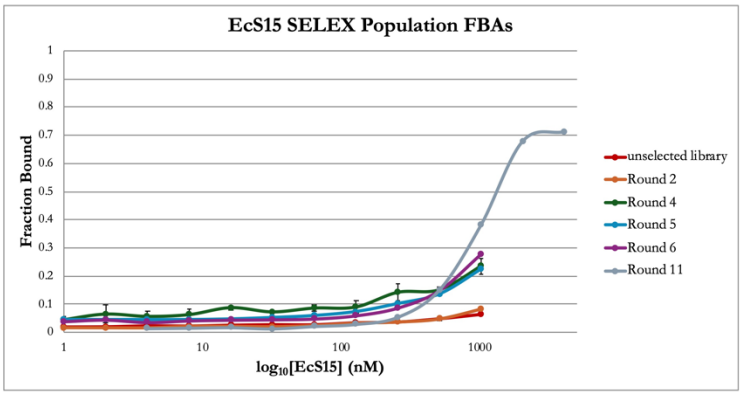
Library	1	2	3	4	5	6	7	8	9	10	11	12	13
MFE Average (kcal/mol)	-12.20	-18.82	-16.31	-22.02	-19.47	-14.15	-13.74	-18.63	-17.04	-20.63	-18.89	-17.78	-16.76
Longest Loop Average	9.91	10.56	10.20	7.25	7.55	9.81	9.90	7.14	7.45	7.39	7.62	7.30	7.52
Longest Stem Average	8.95	9.41	9.41	12.65	12.03	9.62	9.63	11.63	11.32	12.30	11.99	11.51	11.39
Multi-stem %	0.74	8.77	7.19	6.12	6.53	3.33	2.62	3.57	3.36	6.38	6.19	3.55	3.10
Wobble Base Pair Average	2.46	2.77	2.75	3.40	3.25	2.74	2.69	3.04	2.92	3.35	3.28	3.00	2.96

D

5′ CGUAGUCGUAGCUGAUCGAC**ryryryryryryryryryryry***GGAG***NNNNNNN**AUGUCUCUAAGUACUGAA

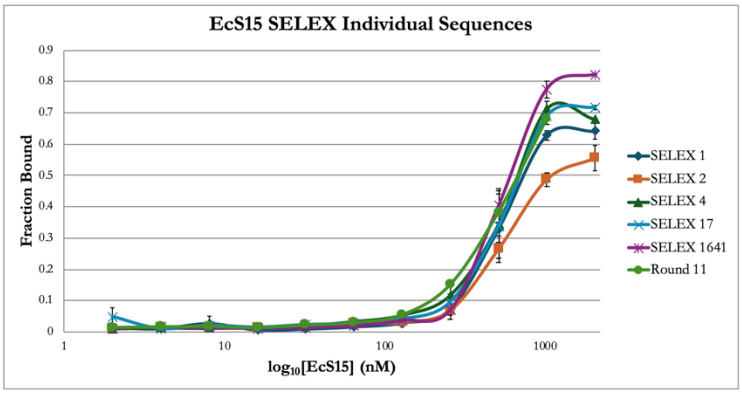
Figure 2.3 Filter-binding Assays from First SELEX Scheme with EcS15
11 Rounds of SELEX were performed with the RNA pool renatured in water using the EcS15 concentrations listed in the table. A) Filter-binding assays to determine relative binding affinity of the RNA populations in each round and the concentration of EcS15 used in each round of selection. Note that we only determined population K_D after round 11. B) Binding curves and affinities for individual sequences tested. The randomized region of the individual sequences are listed below, with the N7 underlined with the SD sequence italicized. SELEX 1 and 4 have identical N7 sequences and SELEX 1641 lacks a SD sequence. C) Binding curves with the new EcS15 preparation compared to the old preparation used in the first SELEX scheme.

A



Round	[EcS15] nM	KD (nM)
unselected	-	-
1	4000	-
2	4000	-
3	2000	-
4	2000	-
5	500	-
6	250	-
7	250	-
8	125	-
9	62.5	-
10	31.25	-
11	15	885

B



Sequence	KD (nM)
Round 11	885
SELEX 1	472
SELEX 2	453
SELEX 4	465
SELEX 17	520
SELEX 1641	559

SELEX 1 **TCGCATATGTACAGTCGTATTACGC***GGAG***TGTTGGC**
SELEX 2 **ACGTCCATATGTGAGTGTATTACGC***GGAG***CCTGGTG**
SELEX 4 **TAACGCATATCTGGTCGTATTACGC***GGAG***TGTTGGC**

SELEX 1641 **TAACGCATATCTGGTCGTATTGCGC***GGAG***TCCCTGC**

SELEX 17 **TCGTCTATAGTGAGCGTATTACGGCCCCGC**

C

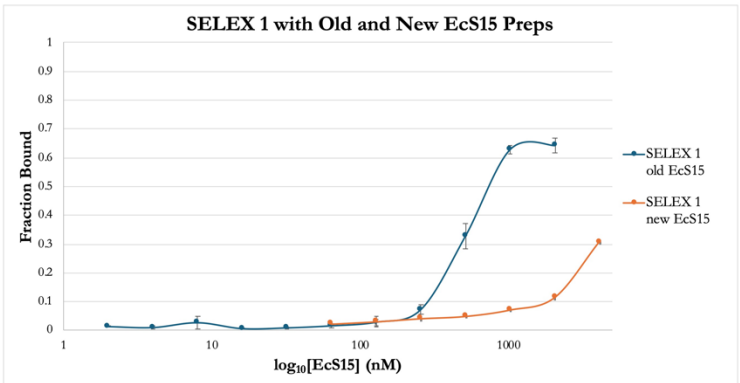
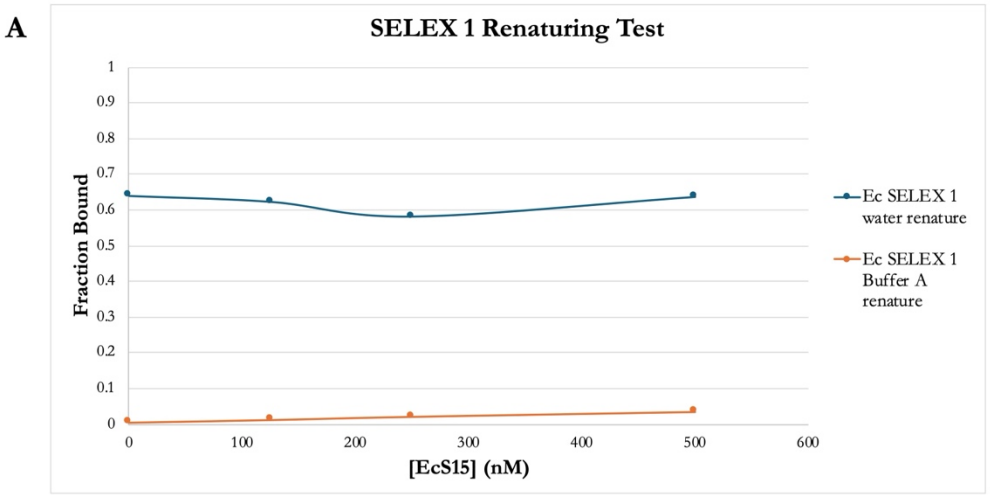
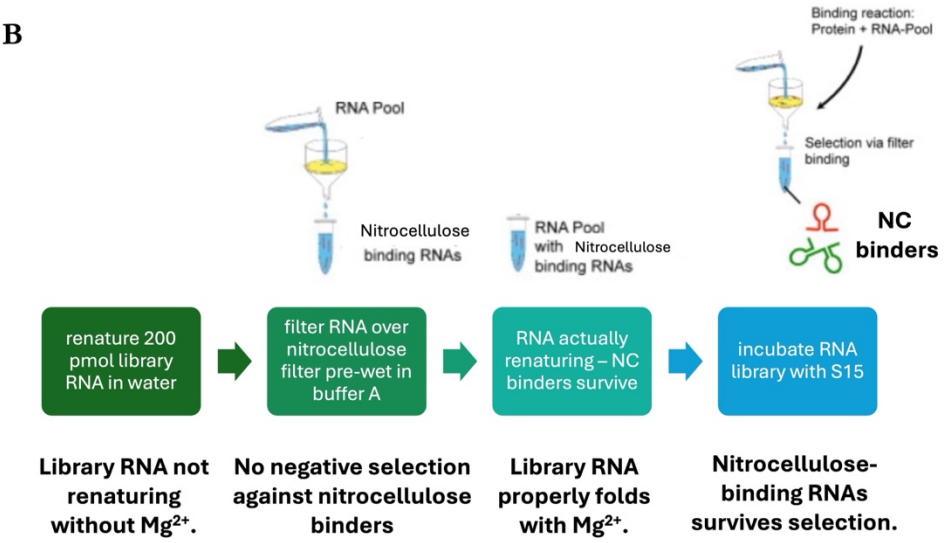
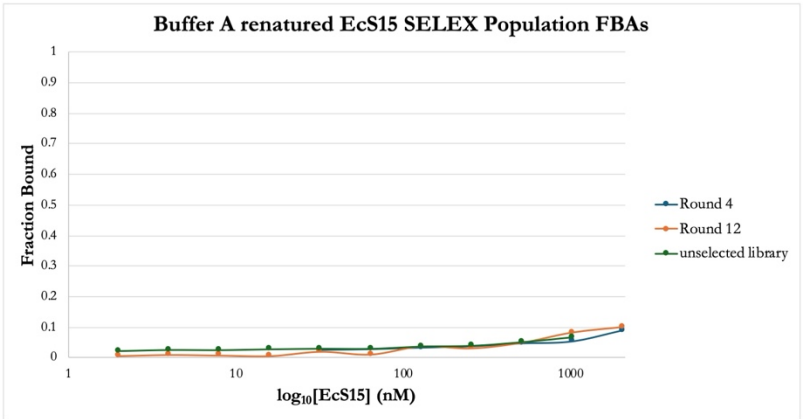


Figure 2.4 SELEX with New Renaturing Protocol Fails to Select for EcS15 Binders
A) SELEX 1 RNA, the most frequent sequence from Round 11 of the initial EcS15 SELEX, does not bind EcS15 when renatured in Buffer A. B) Schematic of why our first SELEX experiments failed to remove nitrocellulose-binding RNAs. C) Filter-binding assays of Round 4 and 12 RNA pool from our second SELEX scheme with EcS15 using new RNA renaturing protocol. We were unable to isolate high affinity EcS15-binding aptamers, as binding for the RNA pool from Round 12 of selection was identical to the unselected library.





C



Round	[EcS15] nM
1	4000
2	4000
3	2000
4	2000
5	500
6	250
7	250
8	125
9	62.5
10	31.25
11	15
12	15

Table 2.1 Table of Primers

A) To generate double-stranded DNA library with T7 promoter for transcriptions, where N=25%A, 25%C, 25%G, 25%T, N1=45%A, 45%G, 5%T, 5%C, and N2=45%C, 45%T, 5%A, 5%G.

Name	Sequence
Library 9-24 rev	5′-TTCAGTACTTAGAGACATNNNNNNNCTCC (N1)(N2)(N1)(N2)(N1)(N2)(N1)(N2)(N1)(N2) (N1)(N2)(N1)(N2)(N1)(N2)(N1)(N2)(N1)(N2)(N1)(N2) (N1)(N2)GTCGATCAGCTACGACTACG
T7 + A fwd	5′-CCAAGTAATACGACTCACTATAGGCGTAGTCGTAGCTGATCGAC

B) For reverse transcription to generate cDNA for T7 template or sequencing.

Name	Sequence
Library rev	5′-TTCAGTACTTAGAGACAT
RT primer w universal adapter cP7	5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGA TCTNNNNNNNNTTCAGTACTTAGAGACAT

C) To amplify cDNA to generate amplicon for sequencing.

Name	Sequence
Adapter cP5 BC01	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTG CTTCGTAGTCGTAGCTGATCGAC

Chapter III

In vitro selection of Gk- and TtS15 aptamers

Introduction

The work presented in Chapter II demonstrated that possibly due to its unique "entrapment" binding mechanism, the EcS15 homolog was not well-suited for *in vitro* selection experiments. While we were unsuccessful at generating aptamers that bind EcS15, the GkS15 and TtS15 homologs both readily bind both their native mRNA regulators and each other's regulators *in vitro,* displaying binding affinities in the low nanomolar range (Slinger et al., 2015). Inter-species S15-mRNA experiments demonstrated that the Gk- and TtS15 homologs have overlapping binding profiles *in vitro* and *in vivo*, and both homologs utilize a so-called "displacement" mechanism to regulate *in vivo* in which excess S15 actively competes with the ribosome to bind the mRNA and turn off the *rpsO* operon (Ehresmann et al., 2004). Since the displacement mechanism relies solely on RNA binding and does not require additional ribosomal components, we hypothesized that our SELEX scheme was more likely to successfully enrich for Gk- and TtS15 aptamers from our partially patterned library than with EcS15. Notably, we His-tagged Gk- and TtS15 to ensure that our purified protein was not contaminated with native EcS15 from the BL21 overexpression strain, which could have confounded our examination of the RNA-binding profiles of the specific homologs.

We performed *in vitro* selection experiments to identify novel Gk- and TtS15 aptamers and compare RNA-binding pools. To our knowledge the work presented in this chapter is the first comparative study of two homologous proteins using SELEX to evolve aptamers from the same starting RNA library. Our experiments indicate that aptamers that bind Gk- or TtS15 are readily evolved from our partially patterned library, though to lower affinities than a larger completely randomized library. On the population level, TtS15 had a higher affinity for the unselected library and enriched for a population of higher affinity aptamers in earlier rounds than GkS15, despite using the same selection conditions. Thus, TtS15 appears to bind and

select for higher affinity sequences from this library more frequently than GkS15 *in vitro,* highlighting the different selective pressures applied by these S15 homologs to their mRNA regulators that led to the diverse array of regulators seen in nature.

Results

In vitro selection enriches for Gk- and TtS15 aptamers

Following our studies with SELEX using EcS15, we renatured our unselected library and Round RNA in Buffer A before selection. We performed 12 Rounds of selection against our partially patterned library in duplicate with GkS15 and 14 Rounds with TtS15 using this SELEX scheme. Interestingly, the unselected library was 52% bound at 2048 nM TtS15 versus only 7% bound at 2048 nM GkS15 (Fig. 3.1A). To make more direct comparisons between the RNA-binding profiles of the homologs, we chose to use the same concentrations for both Gk- and TtS15 in their respective SELEX rounds. To enrich for higher affinity aptamers, we halved the S15 concentration every other round of SELEX to increase the selective pressure on the population (Fig. 3.1B). Filter-binding assays reflected an increase in affinity round over round for the S15 homologs (Fig. 3.1D and F).

For SELEX with GkS15, the starting affinity for the unselected library was poor. As we decreased protein concentration, there was a modest shift in binding affinity from Round 7 to Round 12 from a K_D of 1430 nM to 851 nM (Fig. 3.1D). Despite this modest shift, there was a consistent trend toward increased binding at higher GkS15 concentrations and the F_{max} reached 80% bound by Round 12. This could be due to a diverse RNA pool with relatively poor affinity whose ensemble of structures still saturated GkS15 at the highest protein concentration. Notably, even native interactions with high affinity often do not reach 100% bound at saturation due to dynamic RNA folding *in vitro*, with the most stable conformation

not necessarily saturating the protein (Hall & Kranz, 1999). When compared to Round 12 of EcS15 SELEX with an F_{max} of only 10% at the same [S15] (Fig. 2.4C), we considered our GkS15 SELEX to be modestly successful, despite the lack of a dramatic shift in affinity.

For SELEX with TtS15, the starting affinity for the unselected library was relatively high, with a K_D of 1644 nM (Fig. 3.1F). By Round 6, the population affinity (K_D of 873 nM) was already comparable with the final round of GkS15 SELEX (K_D of 851 nM). As we halved the protein concentration, we successfully enriched for TtS15 aptamers and increased the population affinity to a K_D of 435 nM following selection at 31.25 nM TtS15 in Round 12 (Fig. 3.1F). We completed two further rounds of SELEX at lower concentrations, with a comparable population affinity in Round 13 but a notable decrease in F_{max} following round 14 at 10 nM TtS15. Since our SELEX protocol does not have a mutagenic PCR step, it is possible that we selected most of all the possible binders left in our population or that our protein concentration limited our ability to effectively enrich our population further. Since many natural regulators have affinities in the hundreds of nanomolar range, we considered the population affinity to be sufficient and thus progressed to further analysis.

***Inter-species* Binding Assays Show Overlap in Gk- and TtS15 RNA-binding Profiles**

Previous work has shown that Gk- and TtS15 interact with both the Gk- and Tt-mRNA regulators (summarized in Fig. 3.2A and B). To gauge whether there was any overlap in the RNA-binding profiles of the Gk- and TtS15 homologs with our Tt SELEX population, we performed filter-binding assays with the RNA population from Round 6 of TtS15 SELEX (Tt6, since the affinity was comparable to that of Round 12 of Gk SELEX) with Gk- and TtS15 (Fig. 3.1D and F). We also tested Tt6 with EcS15 to assess the specificity of the TtS15 aptamers in our population, since EcS15 does not interact with the TtS15-mRNA regulator *in vitro* (Fig. 3.2B). TtS15 exhibited the highest affinity for Tt6 (K_D = 873 nM), as expected, and

GkS15 was also able to bind this population (K_D = 1203 nM vs >2000 nM for the unselected library and 1430 nM for Gk Round 7 RNA). EcS15 showed no binding interaction, with an F_{max} at 2048 nM equal to the fraction bound without protein.

Discussion

To explore the RNA-binding profiles of Gk- and TtS15, we used the *in vitro* SELEX scheme optimized in Chapter II and successfully enriched for aptamers to these S15 homologs from our partially patterned library. Before any *in vitro* selection, TtS15 exhibited a higher affinity for the unselected library than GkS15. Following selection, our experiments qualitatively suggest that TtS15 may have a larger RNA-binding profile than GkS15 with this library, as we isolated populations with a higher affinity from *in vitro* selection with TtS15 (Round 12 K_D = 435 nM) than we did with GkS15 (Round 12 K_D = 851 nM). This larger binding profile fits well with the previous inter-species study that showed that TtS15 was more tolerant of mutations that abolish native interactions and may generally have fewer requirements for RNA binding outside of a GGC base triple at the base of a 3-way helical junction, whereas GkS15 has stricter requirements for both the GGC base triple and G•U/GC motif to binding mRNA regulators *in vitro* and *in vivo* (Slinger et al., 2015).

TtS15 may generally have a larger RNA-binding profile than GkS15, or the library and selection conditions were not optimized for GkS15 SELEX. The starting partially patterned library may have been biased toward TtS15 binding aptamers, or the GkS15 concentrations in the early rounds of selection were not high enough to efficiently isolate aptamers from this sequence pool. Lowering the target concentration (i.e. S15) generally increases the stringency of selection, but libraries dominated by low affinity sequences actually favor weak binders over strong binders when the target concentration is too low (Kohlberger & Gadermaier, 2022).

High-affinity sequences present in fewer copies that are lost in early rounds both because they are less frequent in the population and are lost during amplification (Komarova & Kuznetsov, 2019). This may explain the difference in the final population K_D between the Gk- and TtS15 homologs, as well as why we were unable to enrich the TtS15 population for higher affinity aptamers in rounds 13 and 14 with such low [TtS15]. As selection progressed, the maximum fraction bound plateaued and the filter-binding assay curves exhibited the sigmoidal shape characteristic for specific binding for the population enriched by TtS15 but not for the population enriched by GkS15. This further supports the hypothesis that this SELEX scheme was less efficient for selecting GkS15 aptamers than TtS15 aptamers. This could be due to fewer GkS15 aptamers in the initial starting population from this library, or that the GkS15 concentration was not high enough in the early rounds of SELEX to effectively select against low-affinity aptamers.

Our inter-species filter-binding assays with the TtS15 round 6 population suggest that Tt- and GkS15 may have overlapping RNA-binding profiles in our SELEX experiments, as GkS15 was able to bind the aptamers enriched by TtS15. EcS15, which can interact with the Gk-mRNA regulator but not the Tt-mRNA regulator *in vitro*, does not bind the Tt6 population, which potentially reflects the specificity of the population enriched by TtS15. The differences between the larger RNA-binding profile of TtS15 and the smaller profile of GkS15 for the natural mRNA regulators, along with the possible overlap in binding profiles observed for the *in vitro* selected Tt6 RNA population for both homologs, thus made our SELEX experiments well-suited for comparative high-throughput sequencing analysis.

Materials and Methods

Protein Preparation

The *G. kaustophilus* and *T. thermophilus rpsO* ORFs were cloned into the pET-HT overexpression vector with a 6X-His tag, sequence verified, and transformed into chemically competent BL-21(DE3) cells (Invitrogen). Protein was overexpressed and cells lysed by freeze-thaw lysis followed by sonication in Native Binding Buffer (50 mM NaH_2PO_4, pH 8.0, 500 mM NaCl, 10 mM imidazole). Lysate was incubated with Ni-NTA Agarose resin for 1 hour at 4º and eluted with 250 mM imidazole at 4º. Following IMAC purification, fractions containing His-tagged S15 were analyzed via SDS-PAGE, concentrated, and the 6X-His tag cleaved using ProTEV Plus Protease (Promega) according to the manufacturer's protocol. Protease and 6X-His tag were removed through a second purification at 4°C using non-denaturing FPLC cation exchange chromatography (pH 8.0) with a linear salt gradient (20 mM – 1 M KCl). Fractions containing cleaved S15 were tested for nucleases, and RNase-free protein fractions were concentrated, analyzed via SDS-PAGE, and buffer exchanged for the S15 Storage Buffer (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 0.02% sodium azide). Final protein concentration was determined by Bradford assay and stored at 4°C.

RNA Preparation and SELEX

Library template was prepared and transcriptions were performed as described in Chapter II.

SELEX rounds were performed in duplicate for by renaturing 200 pmol of RNA in Buffer A at 42º for 15 minutes and cooled to room temperature for 10 minutes, then filtered through 0.45 μM nitrocellulose filter to remove nonspecific binders. Surviving RNAs were

incubated with the respective S15 homolog in Buffer A (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 5 mM dithiothreitol, 0.02% bovine serum albumin) at 25º for 30 minutes. RNA-S15 complexes were isolated by filtering over a second nitrocellulose filter and washed twice with Buffer A. RNAs were eluted from the filter at 95º in elution buffer (7 M Urea, 100 mM $Na_3C_6H_5O_7$, 3 mM EDTA pH 8.0), protein removed through phenol-chloroform extraction, and ethanol precipitated overnight at -20º. The selected RNA was reverse transcribed using M-MuLV and half of the cDNA amplified using standard PCR with primers complementary to the library constant region and to add the T7 promoter. The remaining cDNA was amplified with primers to add Illumina sequencing adapters and barcodes. T7 PCR products were gel purified and used as template for T7 transcription reactions to make RNA for the next round of selection.

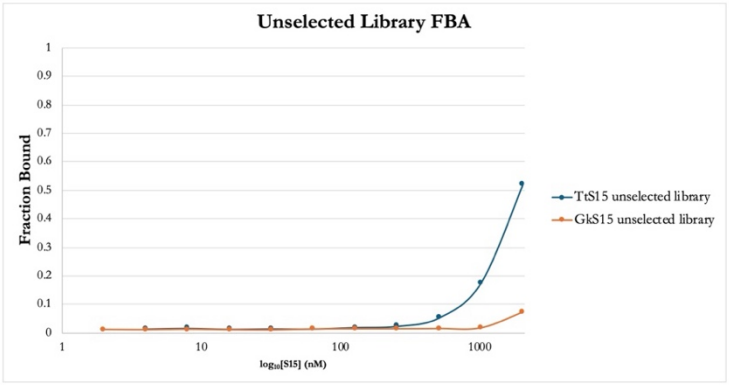
Filter Binding Assays

S15-RNA binding affinity was periodically examined throughout SELEX by filter-binding assay as described in Chapter II.

Figures and Legends

Figure 3.1 Filter-binding assays to monitor binding affinity during SELEX
Filter-binding assays were used to measure changes in affinity round over round as SELEX progressed. A) FBAs with Gk- and TtS15 and the unselected library before SELEX. B) Concentrations of S15 used for the Rounds of SELEX with each homolog. C) Compilation of all the population FBAs we performed during GkS15 SELEX. D) GkS15 round concentration and population K_D for the rounds tested. E) Compilation of all the population FBAs we performed during TtS15 SELEX. F) TtS15 round concentration and population K_D for the rounds tested. The binding affinity for the population plateaued by Round 13 of Tt SELEX.

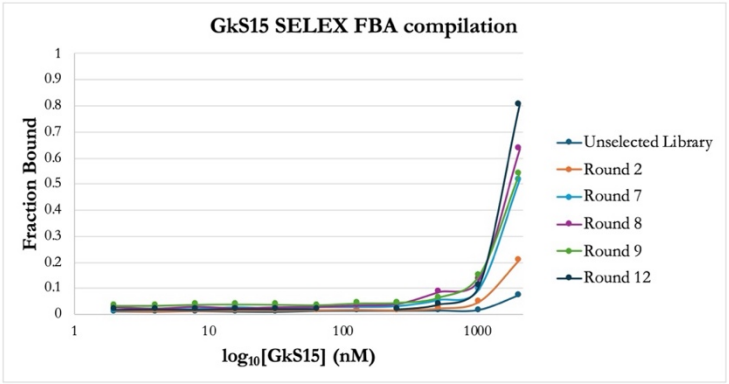
A



B

Round	[GkS15] (nM)	[TtS15] (nM)
1	1250	1250
2	1250	1250
3	500	500
4	500	500
5	250	250
6	250	250
7	125	125
8	125	125
9	62.5	62.5
10	62.5	62.5
11	31.25	31.25
12	31.25	31.25
13	-	15.625
14	-	10

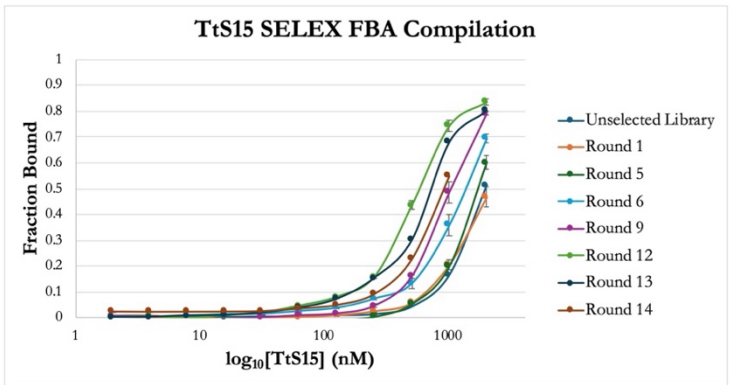
C



D

Round	[GkS15] (nM)	KD (nM)
1	1250	-
2	1250	1316
3	500	-
4	500	-
5	250	-
6	250	-
7	125	1430
8	125	1344
9	62.5	1249
10	62.5	-
11	31.25	-
12	31.25	851

E



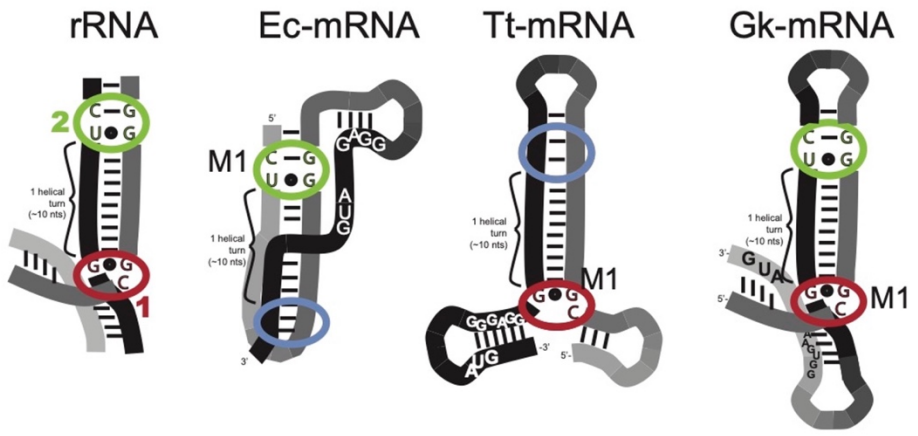
F

Round	[TtS15] (nM)	KD (nM)
0	-	1644
1	1250	979
2	1250	-
3	500	-
4	500	-
5	250	1112
6	250	873
7	125	-
8	125	-
9	62.5	708
10	62.5	-
11	31.25	-
12	31.25	435
13	15.625	432
14	10	-

Figure 3.2 Inter-species S15-mRNA interactions

Inter-species S15-mRNA interactions have been previously characterized both *in vitro* and *in vivo*. A) Cartoon representing S15 binding sites on 16s rRNA and natural mRNA regulators. Green and red circles highlight specific sequence and structural features for binding, and blue circles highlight secondary binding sites that are not sequence specific. M1 indicates mutations made to the mRNA to disrupt the native interaction. (modified from Slinger 2015). B) Inter-species interactions of S15 homologs with native mRNA regulators were previously measured *in vivo* using a *LacZ* regulatory assay and *in vitro* through filter-binding assays. Green indicates an interaction, red indicates no interaction was observed, and "–" indicates that the interaction was not measured. C) Filter-binding assays were performed with RNA from Round 6 of SELEX with TtS15 (Tt6) and Ec-, Gk-, and TtS15. Binding curves of single replicates with Tt6 and the 3 S15 homologs are shown. Inset shows the K_D for each homolog.

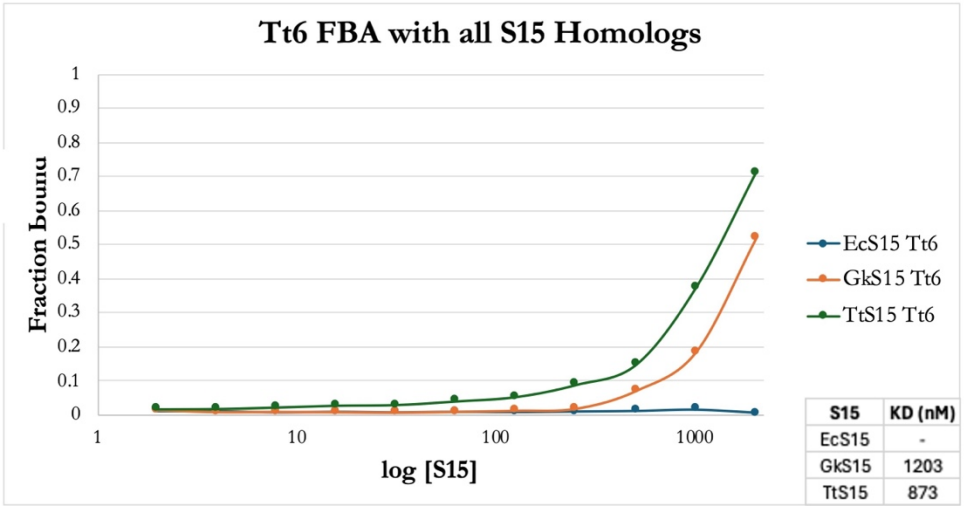
A



B

	Ec-mRNA		Ec-M1		Tt-mRNA		Tt-M1		Gk-mRNA		Gk-M1	
	In vivo	*In vitro*	*In vivo*	*In vitro*	*In vivo*	*In vitro*	*In vivo*	*In vitro*	*In vivo*	*In vitro*	*In vivo*	*In vitro*
EcS15		–		–								
TtS15		–		–								
GkS15		–		–								

C



Chapter IV

High-throughput sequencing and *in vitro* analysis of Gk- and TtS15 aptamers

Introduction

SELEX experiments have traditionally been limited by low-throughput Sanger sequencing to sample a subset of the final population. The advent of high-throughput sequencing allows us to fully interrogate the SELEX process across every round of selection, capturing the diversity of the population round over round as the selection stringency increases. High-throughput sequencing of SELEX rounds has shown that enrichment of binders often occurs early within the process. ssDNA aptamers for the bacterial protein streptavidin were enriched from a DNA library consisting of a 40 base fully randomized region (representing $1.2x10^{24}$ sequences) (Schütze et al., 2011). The authors directly compared cloning and Sanger sequencing with a high-throughput sequencing approach and showed that high affinity aptamers were easily identified in the early rounds of SELEX from simple copy number enrichment in the high-throughput sequencing data, but not Sanger sequencing.

Much work has been done to explore the specificity of DNA-protein interactions through SELEX. SELEX with the nuclear factor-κB (*NF-κB*) p50 protein and the Hox-Exd-HM complex in *Drosophila melanogaster* identified an array of DNA motifs and preferences for these transcription factors (Gu et al., 2013; Riley et al., 2014). The high-throughput sequencing data helped identify preferences for certain DNA motifs and how multi-protein complexes can modulate transcription factor specificity to help explain their preferences *in vivo*, with resolution impossible using Sanger sequencing. While many studies have looked at the specificity of DNA-protein interactions, they focus mainly on motif identification and much less work has been done to characterize RNA-protein interactions using SELEX, where RNA structure can play a much larger role than its primary sequence.

Given that Gk- and TtS15 have slightly different RNA-binding profiles for the natural mRNA regulators and appear to have somewhat overlapping binding profiles in the aptamers

isolated from *in vitro* selection, we sought to characterize their RNA-binding profiles using high-throughput sequencing and filter-binding assays. To monitor population-level changes as the stringency of selection increased, we utilized the RaptRanker algorithm to analyze our sequencing data as it easily allows for enrichment analysis of individual aptamers round over round (Ishida et al., 2020). To our knowledge the work presented in this chapter is the first comparative study of two homologous RNA-binding proteins using *in vitro* selection against an RNA library.

Results

RaptRanker analysis shows enrichment of relatively low-diversity aptamer pools

We sequenced the unselected library and rounds 1, 2, 3, 5, 7, 9, and 12 for both replicates for both homologs (~3-4 million reads per sample per replicate) to monitor changes in the population as SELEX increased population binding affinity. In the unselected library, virtually all of sequences had the constant SD sequence and consisted of mostly singletons (sequences appearing only once in the population), as expected. We utilized the RaptRanker algorithm to integrate the data from all sequenced rounds of each replicate of Gk- and Tt-SELEX (Ishida et al., 2020). One of the main advantages of this tool is that it assigns unique sequence IDs to track the frequency within rounds as well enrichment round over round. To make more direct comparisons within and between rounds, we normalized the raw reads to total reads for each sequence per round and used this relative frequency for our enrichment analysis. High-throughput sequencing data allows us to detect lower frequency sequences in our earlier rounds of selection, and simply comparing round over round enrichment of the normalized frequency can help identify high affinity aptamers in earlier rounds than is possible with traditional Sanger sequencing.

As early as Round 2, where the SELEX populations were under low selective pressure, we saw an enrichment of a subset of sequences for both Gk- and TtS15, most of which contained the SD sequence (Fig. 4.1A). Interestingly, a growing proportion of sequences lost the SD sequence as SELEX progressed. In round 2, 96% of the Gk and 98% of the Tt SELEX had a SD sequence (Fig. 4.1B). By round 12, only 12% of the Gk and 29% of the Tt SELEX had a SD sequence. Looking across rounds, the GkS15 population saw a significant and continual decline in aptamers with the SD sequence while the TtS15 population plateaued at ~30% with the SD sequence (Fig. 4.1B). Most of the sequences are the expected length, and looking at the position where the SD would be, it is often mutated from GGAG to GGTG or GCTG (bolded sequences in Fig. 4.1C and D have the SD sequence, with GGAG or the mutations underlined). Since this sequence was not within the constant region (Fig. 2.2D) and we utilized a low fidelity *Taq* polymerase, there was no selective pressure to maintain the SD sequence.

In our Round 12 populations for both replicates and both homologs, our *in vitro* selection scheme isolated highly enriched and dominant sequences leading to a relatively low-diversity final population (Fig. 4.1C and D). Our sequencing data indicate that we were successful in comprehensively sampling our populations, so we then sought to characterize any relationships between the sequences in our Round 12 populations through clustering analysis and validate their binding *in vitro* with filter-binding assays.

Clustering analysis reveals significant sequence overlap within Gk- and TtS15 SELEX replicates

To broadly assess the diversity of our Round 12 populations, we took the top 500 most frequent sequences for each replicate with each homolog and collapsed all duplicate

sequences. The top 500 individual sequences account for 56% and 52% of the total reads in Round 12 for Gk- and TtS15, respectively. For GkS15, 375 of the top 500 sequences were shared between the replicates, including all the highest frequency sequences. The sequences not shared by the replicates were much lower frequency and by visual inspection appear to differ by 1-2 nucleotides from the most frequent sequences. For TtS15, 340 sequences were shared between the replicates, also including all the highest frequency sequences. Similarly to GkS15, the sequences not shared are lower frequency and differ by 1-2 nucleotides from the most frequent sequences. We then assessed the relatedness of the sequences within the Gk and TtS15 SELEX replicates by performing clustering analysis of those 375 and 340 sequences, respectively (Crum et al., 2019). We find a high degree of relatedness within both populations of the Gk- and TtS15 SELEX (Fig. 4.2 and 4.3). As we reduce stringency and allow sequence distances up to 5 mutations, GkS15 SELEX collapses to 3 major clusters of sequences (Fig 4.2C) while TtS15 SELEX has 6 major clusters (Fig. 4.3C), anchored by high frequency sequences at the centers of the clusters. TtS15 had a higher affinity for its round 12 population than GkS15 did (Fig. 3.2D and E), and our clustering analysis highlights that TtS15 had slightly fewer shared sequences between replicates (340 v. 375 for GkS15), potentially reflective of a larger binding pool for TtS15.

Clustering analysis reveals significant sequence overlap between Gk- and TtS15 SELEX

To compare the RNA-binding profiles between the S15 homologs, we performed clustering analysis using sequences from Round 12 of Gk- and TtS15 SELEX. We combined the top 500 sequences from Round 12 from all replicates with both homologs to compare the diversity between the populations selected by the S15 homologs (Fig. 4.4). 1059 of the sequences were shared between the homologs, including all the highest frequency sequences.

Subsequent clustering analysis showed 8 clusters of related sequences, which we named Cluster A-H (Fig. 4.5A). 3 of the 8 clusters contained the SD sequence.

Cluster A, which lacks a SD sequence, contains the 2 most frequent sequences for GkS15 (sequence #273 and 51) and most of the high frequency sequences from GkS15 SELEX (Fig. 4.5A). 51 is a high frequency sequence for TtS15 as well, though 273 is not (Fig. 4.5B). Comparing the randomized regions, sequences 273 and 51 differ by only 1 base (Fig. 4.5C). Cluster B, which has a SD sequence, has most of the high frequency sequences from TtS15 SELEX. Generally, the clusters that have a SD sequence tend to have most of the high frequency sequences from TtS15 SELEX. Conversely, the clusters that lack a SD sequence tend to correlate more with the high frequency sequences from GkS15 SELEX. Cluster F, for instance, has a SD sequence but only 2 sequences from this cluster are represented in Gk SELEX (Fig. 4.5A).

Taken as a whole, TtS15 appears to have a larger RNA-binding pool, as nearly every sequence from the clusters is found in one or both Tt SELEX replicates while the same is not true for GkS15.

Individual sequences bind Gk- and TtS15

To validate that individual aptamers bind S15, we chose a subset of sequences to test *in vitro* with filter-binding assays (Fig. 4.5B). To determine if the SD sequence had any effect on *in vitro* binding, we chose 1 sequence that had the SD sequence and 2 that did not. We chose Sequence 31 from Cluster B, as it was high frequency across all replicates for both homologs, contains a SD sequence, and has the shortest distance mean distance from all other sequences in cluster B (Fig. 4.5B). We find that both homologs bind this sequence with similar affinities (Fig 4.6A). We chose sequences 51 and 273 from Cluster A (no SD sequence), as they were the highest frequency sequences for GkS15 and have the shortest distance mean

distances to all other sequences in Cluster A. We find that both homologs bind sequence 51 with similar affinities, indicating that the specific SD sequence was not required for binding (Fig. 4.7A). 51 was the most frequent sequence for GkS15 SELEX and 2nd most frequent for TtS15 SELEX. While 51 is high frequency for TtS15, 273 is not, though they differ by only one base. 273 is the 3rd most frequent sequence for GkS15 SELEX but drops to 41st for TtS15 SELEX. Interestingly, these mutations occur in the same position as the GGAG SD sequence in sequence 31, and contain GGTG for sequence 51 and GCTG for sequence 273.

To determine if the difference in relative frequency for sequence 273 in Gk- and TtS15 SELEX was due to a structural change caused by this single base change, we computationally evaluated the structures of both 51 and 273. RNAfold simulations indicate this single nucleotide change alters the predicted structure of the RNAs and disrupts a predicted stem (Fig. 4.7B and C) to form a stem loop (Fig. 4.7E and F). Thus, we hypothesized that the difference in frequency in the sequencing data is reflective of this structural change and would lower the binding affinity for TtS15 compared to GkS15. Our filter-binding assays confirm this hypothesis, as Gk- and TtS15 have similar affinities for sequence 51 while the K_D is higher for sequence 273 with TtS15 (992 nM) versus GkS15 (733 nM) (Fig. 4.7D). It is possible that the structural change disrupts a binding site for TtS15 while GkS15 is unaffected, but confirming this would require structural probing assays.

Discussion

Our results indicate that we successfully enriched for S15 binders with our *in vitro* selection experiments for the Gk- and TtS15 homologs. TtS15 was generally a better binder with this library, though we still enriched for GkS15 binders using SELEX. Using a partially patterned library allowed us to comprehensively sample our binding populations throughout SELEX, while also enriching for high frequency sequences. The top sequences in Round 12 for all SELEX replicates were on the order of thousands to tens of thousands of reads. This is in stark contrast to the GkS15 SELEX against a fully randomized N30 library, where 95.33% of the sequences in the final round were singletons, and of the 4.67% of "multitons", 69.5% of those sequences were seen fewer than 10 times in the data (Pei et al., 2017).

To better understand how our populations changed as we increased selection stringency, we utilized the RaptRanker algorithm for our sequencing analysis. This method integrates sequencing data from every round of SELEX, allowing us to track the trajectory of individual sequences throughout selection. We find significant overlap between the S15 homologs. Clustering analysis revealed that our populations converged on a few families of RNAs, both within replicates and between the S15 homologs. One limit of our clustering analysis is that it considers sequences of different lengths to be distinct, even if there is only a one base gap in the sequence along an otherwise homologous sequence, though most of our sequences were of equal length. Based on our clustering analysis, we selected representative sequences to test with our S15 homologs. We demonstrate binding for all the aptamers tested, as well as a one nucleotide change that causes a marked difference in affinity for the same sequence selected by both S15 homologs.

Our work shows that the SD sequence is subject to mutation for Gk- and TtS15 without in our SELEX scheme without significantly affecting binding *in vitro,* though TtS15

enriched for sequences containing the SD sequence to a greater degree than GkS15. Interestingly, while the natural Gk- and Tt-mRNA regulators are >100 nt and our aptamers are only 75 nt, deletion analysis of the natural sequences has shown that the stems containing the SD sequence are dispensable for GkS15 but necessary for TtS15 binding their respective mRNA regulators *in vitro* (Scott & Williamson, 2001; Serganov et al., 2003). It is therefore possible that there may be some selective pressure for aptamers to contain the SD sequence for TtS15 and less pressure for GkS15. Looking at the mutation that separates sequences 51 and 273, there is a GGTG for sequence 51 and GCTG for sequence 273 at the SD sequence position (Fig. 4.5B). The SD sequence for the Tt-mRNA regulator is base paired in the stem adjacent to the GGC base triple binding site for TtS15, so it is also possible that the SD sequence may stabilize a similar structure for the TtS15 aptamers in our population that a mutation, such as the one in sequence 273, would disrupt.

Clustering analysis showed overlapping RNA-binding profiles for Gk-and TtS15 (Fig. 4.4), while sequences 51 and 273 highlight that single mutations can significantly alter the trajectory of a sequence during SELEX. The binding affinities for individual sequences from our final populations are relatively weak compared to affinities for their natural regulators, which is likely due to our partially patterned library limiting the sequence space we could explore through *in vitro* selection. Further work is needed to elucidate the structures, binding sites, and whether any of our aptamers can regulate *in vivo*. Still, we were able to demonstrate that small sequence changes can lead to changes in aptamer binding affinity for Gk- and TtS15, which furthers the hypothesis that differences in the RNA-binding profiles of the S15 homologs is a driver of the diversity of natural S15 mRNA regulators.

Materials and Methods

Illumina sequencing

cDNA from rounds 1, 2, 3, 5, 7, 9, and 12 of both replicates of Gk- and Tt SELEX were amplified in 2-step PCRs to add Illumina adapters and barcodes (Table 1). Samples were pooled and sequenced on an Illumina NextSeq 2000 sequencer. Raw FASTQ files were de-multiplexed and low-quality reads removed using AmpUMI (Clement et al., 2018).

Raptranker analysis

Raw FASTQ files from rounds 1, 2, 3, 5, 7, 9, and 12 of both replicates of Gk- and TtS15 SELEX were first analyzed separately to compare sequence diversity between replicates using RaptRanker (Ishida et al., 2020). We then combined all rounds for all replicates and both homologs and repeated RaptRanker analysis for the combined Gk- and TtS15 SELEX populations. We extracted the sequence IDs, normalized the reads for each sequence based on total reads for each round, and calculated enrichment scores round over round based on the normalized reads.

Clustering analysis and choosing individual sequences

Following RaptRanker, we took the top 500 most frequent sequences in round 12 for each SELEX replicate with their respective homolog and removed duplicates in Excel. We assigned a score of 1-7 for each sequence, based on their normalized read scores for each round (with 1 being the lowest and 7 being the highest read count frequency). We then clustered the sequences using graph clustering as described (Crum et al., 2019, doi.org/10.1371/journal.pcbi.1007564) to compare the replicates for each homolog and generated plots for Gk- and TtS15 SELEX, using colored nodes as described in Figure 4.4. Distances between sequences within clusters were calculated using Hamming distance, with the maximum penalty applied to sequences that differ in length. We then combined all

replicates from both homologs and removed duplicates in Excel, and clustered and plotted the sequences to compare the RNA-binding pools between the homologs, using the same scoring as described above. Following clustering of the combined Gk- and TtS15 SELEX pools, we selected sequences 31, 51, and 273 for testing *in vitro*.

RNA preparation

T7 template for individual aptamer sequences was synthesized using assembly PCR from overlapping oligos (IDT), adding the T7-promoter sequence within the forward primer. Transcriptions were performed and RNA purified as described in Chapter II.

Filter Binding Assays

S15-RNA binding affinity was measured for sequences 31, 51, and 273 using 5'-radiolabeled RNA in filter-binding assays, as described in Chapter II.

Figures and Legends

Figure 4.1 Normalized frequencies and enrichment of sequences across SELEX
We utilized the RaptRanker algorithm to track individual sequences from the high-throughput sequencing data across the rounds of SELEX. A) Relative frequencies for the top 5 sequences in Round 2 of Gk and Tt SELEX. RaptRanker assigns the unique IDs, and frequencies are normalized to the total reads from that round. The most frequent sequences all contain the SD sequence. B) Fraction of unique sequences containing the SD sequence in the corresponding round. C) Enrichment analysis of the 20 most frequent sequences from Round 12 Gk SELEX with normalized frequencies. Enrichment scores were calculated by dividing the normalized frequency in round 7 by the frequency in round 5, round 9 by round 7, etc. Sequences that contain the SD sequence are bolded (with SD and mutations to the SD underlined), and red boxes indicate an enrichment score of >1. D) Enrichment analysis of the 20 most frequent sequences from Round 12 Tt SELEX with normalized frequencies. Bolded and colored as described in C).

A

	Gk Round 2			Tt Round 2	
seq_id	Frequency	Has SD	seq_id	Frequency	Has SD
4	0.1127	yes	31	0.06366	yes
12	0.1070	yes	95	0.05761	yes
5	0.0545	yes	274	0.03418	yes
22	0.0463	yes	126	0.02018	yes
18	0.0264	yes	154	0.01335	yes

B

	Gk SD	Tt SD
Round 2	0.9616	0.9809
Round 5	0.3659	0.5259
Round 7	0.2276	0.3550
Round 9	0.1729	0.3291
Round 12	0.1204	0.2857

C

	Gk SELEX	Round 5	Round 7	Round 9	Round 12	Round 7	Round 9	Round 12
seq_id	sequence	freq	freq	freq	freq	enrichment	enrichment	enrichment
51	TCGACTATAGTGTGCGTATTACTTGGTGCCTCGTG	0.004	0.021	0.063	0.076	5.913	2.998	1.204
5	**TAACGCATGCAAGTCGTATTACGCGGAGGGTTGTG**	0.043	0.053	0.070	0.049	1.247	1.312	0.703
273	TCGACTATAGTGTGCGTATTACTTGCTGCCTCGTG	0.003	0.010	0.047	0.042	3.383	4.869	0.883
31	**TAACGCATGCAAGTCGTATTACGCGGAGGGGTGTG**	0.098	0.111	0.095	0.041	1.135	0.857	0.430
165	TCGACTATAGTGAGCGTGTTACTTGCCCTCCCG	0.004	0.018	0.023	0.012	4.418	1.304	0.532
1830	TCGACTATAGTGTGCGTATTACTTGGTGCCCCGTG	0.000	0.002	0.005	0.009	7.703	2.399	1.746
470	TCGACTATAGTGTGCGTATTGCTTGGTGCCTCGTG	0.000	0.001	0.005	0.007	7.722	4.300	1.408
5049	TCGACTATAGTGTGCGTATTACTTGGTGCCTGGTG	0.000	0.002	0.005	0.006	7.755	2.583	1.190
1725	TAGACTATAGTGTGCGTATTACTTGGTGCCTCGTG	0.000	0.001	0.003	0.006	8.389	4.348	2.197
64	TCGTCTATAGTGAGCGTATTACTTGCCCTCCCG	0.003	0.011	0.008	0.006	3.156	0.735	0.723
1081	TCGACTATAGTGTGCGTATTACTTGCTGCCCCGTG	0.000	0.001	0.004	0.005	3.940	3.858	1.270
820	TCGACTATAGTGAGCGTGTTACTTGGCCCTCCG	0.001	0.003	0.003	0.004	4.290	0.996	1.110
2464	TCGACTATAGTGTGCGTATTACTTGCTGCCTGGTG	0.000	0.001	0.003	0.004	3.997	4.554	1.112
1828	TCGACTATAGTGTGCGTATTACTTGCCGCCTCGTG	0.000	0.001	0.002	0.003	3.897	2.533	1.516
290	TCGACTATAGTGAGCGTATTACTTGCCCTCCCG	0.001	0.004	0.004	0.003	4.635	1.043	0.779
769	TCGACTATAGTGTGCGTATTACTTGCCCTCCCG	0.000	0.002	0.004	0.003	8.269	1.821	0.881
28	TCGACTATAGTGAGCGTGTTACTTGGCCCCCCG	0.006	0.005	0.003	0.003	0.770	0.587	1.092
1240	TATAGTGTGCGTATTACTTGGTGCCTCGTG	0.000	0.001	0.002	0.003	4.246	1.561	1.780
115	TCGTCTATAGTGAGCGTATTACTTGCCCCTCCCG	0.001	0.005	0.004	0.003	4.644	0.719	0.812

D

	Tt SELEX	Round 5	Round 7	Round 9	Round 12	Round 7	Round 9	Round 12
seq_id	sequence	freq	freq	freq	freq	enrichment	enrichment	enrichment
7	**TAACGCATGCAAGTCGTATTACGCGGAGGGGTGTG**	0.099	0.077	0.089	0.039	0.772	1.159	0.438
303	TCGACTATAGTGTGCGTATTACTTGGTGCCTCGTG	0.001	0.001	0.003	0.036	1.346	3.265	11.333
9	**TCGCGTATGTGAGCGTATTGCGCGGAGTCTCGTG**	0.074	0.049	0.033	0.019	0.662	0.681	0.557
420	TCGACTATAGTGAGCGTGTTACTTGCCCTCCCG	0.002	0.005	0.010	0.018	3.060	1.834	1.886
145	**TCGCGTATGTGAGCGTATTACGCGGAGGGGTGTG**	0.005	0.006	0.009	0.013	1.078	1.599	1.412
465	TCGACTATAGTGAGCGTGTTACTTGCCCCTCCG	0.002	0.004	0.008	0.010	2.327	1.832	1.367
13	**TAACGCATGCAAGTCGTATTACGCGGAGGGTTGTG**	0.034	0.026	0.025	0.010	0.745	0.973	0.405
1473	TCGACTATAGTGAGCGTGTTACTTGCCCCTCCCG	0.001	0.002	0.005	0.009	3.497	1.986	1.930
48	TCGTCTATAGTGAGCGTATTACTTGCCCTCCCG	0.002	0.006	0.011	0.008	3.304	1.771	0.761
400	**TAACGCATACAAGTCGTATTACGCGGAGGGGTGTG**	0.003	0.003	0.005	0.008	1.161	1.612	1.473
133	**TCGCGTATGTGAGCGTATTACGCGGAGTCTCGTG**	0.005	0.005	0.006	0.007	0.997	1.159	1.259
834	TCGTCTATAGTGAGCGTATTACTTGCCCCTCCCG	0.002	0.004	0.008	0.007	2.954	1.907	0.781
91	TCGACTATAGTGAGCGTGTTACTTGCCCCCCCG	0.015	0.020	0.017	0.006	1.304	0.861	0.328
1390	TCGACTATAGTGAGCGTATTACTTGCCCTCCCG	0.000	0.001	0.003	0.006	4.361	2.065	1.977
1001	TATAGTGAGCGGTATTACTTGGTGGCTGTGG	0.001	0.002	0.004	0.005	1.650	2.205	1.434
2446	TCGACTATAGTGAGCGTATTACTTGCCCCTCCCG	0.000	0.001	0.002	0.005	5.820	2.415	1.969
2007	TCGACTATAGTGTGCGTATTACTTGGTGCCTGGTG	0.000	0.000	0.000	0.005	1.654	3.863	11.832
27	TCGTCTATAGTGAGCGTATTACTTGCCCCCTCG	0.026	0.036	0.024	0.004	1.421	0.659	0.173
11	**TAACGCATGCAGGTCGTATTACGCGGAGGGGTGTG**	0.010	0.010	0.012	0.004	0.983	1.135	0.345
833	TCGACTATAGTGAGCGTGTTACTTGGCCCTCCG	0.000	0.001	0.002	0.004	3.137	2.619	1.843

Figure 4.2 Clustering analysis of combined GkS15 SELEX replicates
The top 500 most frequent sequences from each GkS15 SELEX replicate were combined and duplicates removed. Each node represents an individual sequence and the edges between nodes represents the distance between those sequences. A) clustering of the sequences with one mutation allowed. B) clustering of the sequences with 1-2 mutations allowed. C) clustering of the sequencing with up to 5 mutations allowed, which collapses the clusters into 3 major clusters.

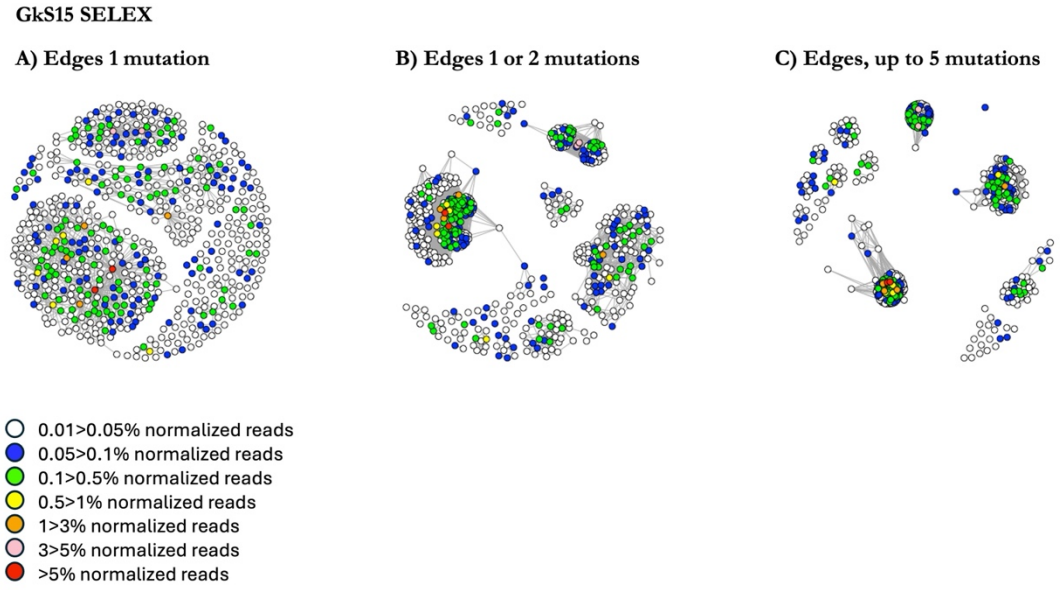


Figure 4.3 Clustering analysis of combined TtS15 SELEX replicates
The top 500 most frequent sequences from each TtS15 SELEX replicate were combined and duplicates removed. Each node represents an individual sequence and the edges between nodes represents the distance between those sequences. A) clustering of the sequences with one mutation allowed. B) clustering of the sequences with 1-2 mutations allowed. C) clustering of the sequencing with up to 5 mutations allowed, which collapses the clusters into 6 major clusters.

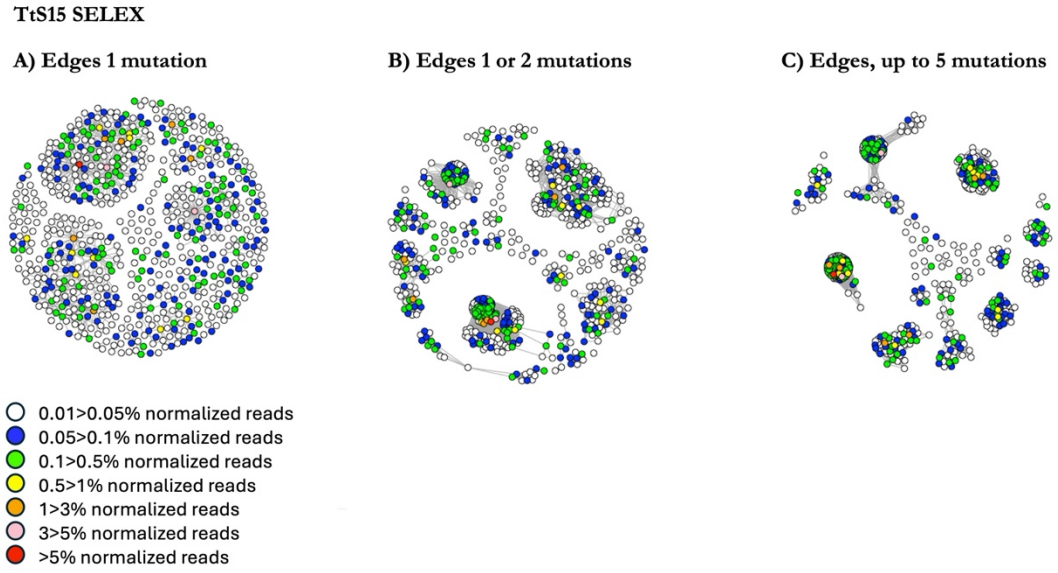


Figure 4.4 Clustering analysis of combined Gk- and TtS15 SELEX

The top 500 most frequent sequences from both replicates with both S15 homologs were combined and duplicates removed. Each node represents an individual sequence and the edges between nodes represents the distance between those sequences. A) Clustered sequences, with nodes colored based on representation in the sequencing data, with distances up to 5 mutations. B) Clustered sequences, with nodes colored based on their frequency in the sequencing data, with distances up to 5 mutations.

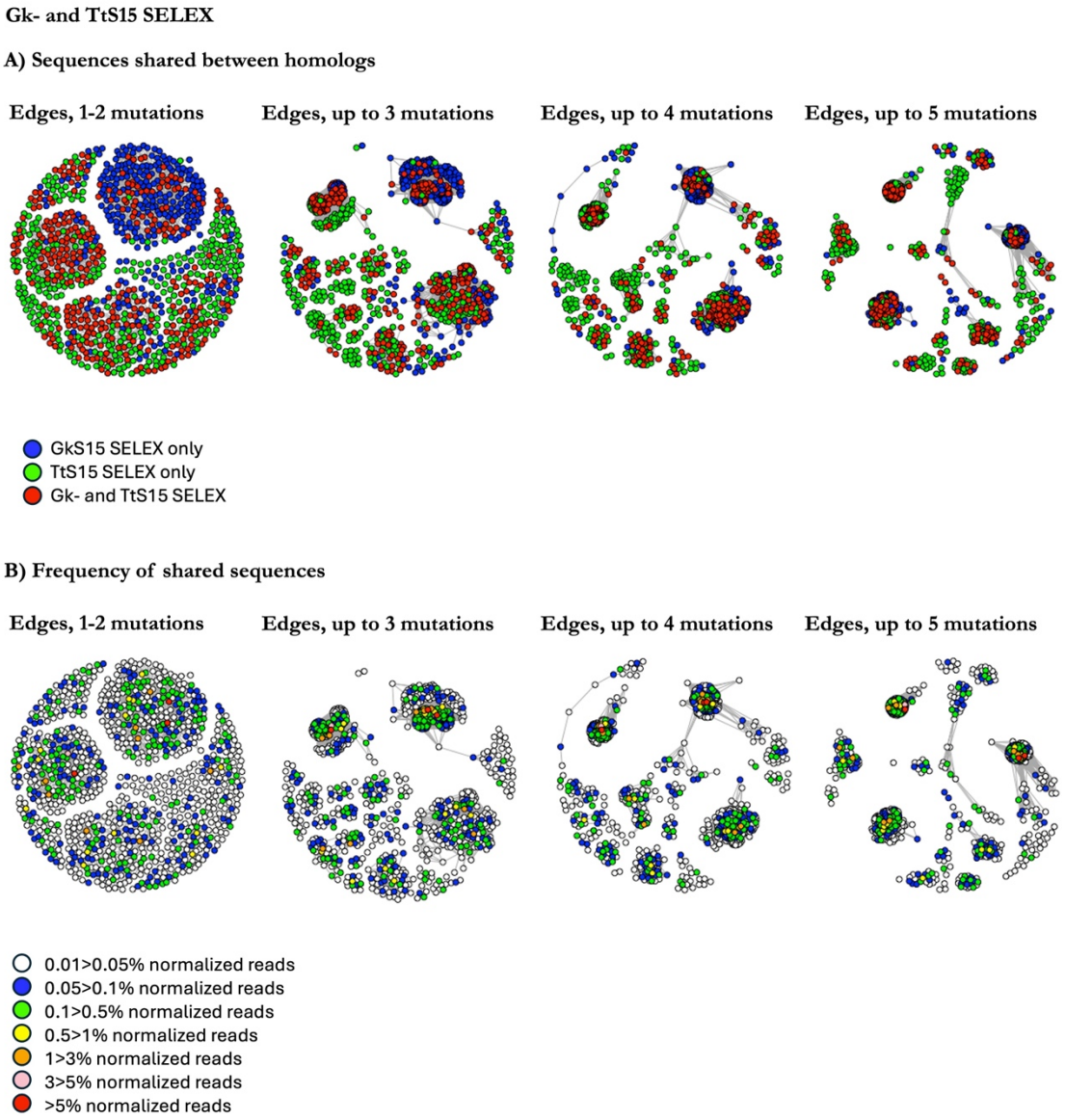


Figure 4.5 Normalized frequencies for combined Round 12 Gk- and Tt SELEX clustering

Round 12 sequencing was combined for both replicates and both S15 homologs, duplicates removed, and clustered. A) 8 clusters were identified from the combined clustering and named A-H, and presence of the SD sequence noted. The 10 most frequent sequences in each cluster are shown for clarity. Tt1 corresponds to Tt SELEX replicate 1, Tt2 to Tt SELEX replicate 2, Gk1 to Gk SELEX replicate 1, and Gk2 to Gk SELEX replicate 2. Frequency scores were assigned based on normalized reads per sequence per round, as listed. "–" indicates that the sequence was not present in that replicate. Orange boxes indicate the sequences chosen for filter-binding assays. B) sequences of the randomized regions for the chosen sequences and their distance to all other sequences in their respective cluster (A for 51 and 273, B for 31). Bold indicates SD sequence is present, with the SD and mutations to the SD underlined. C) alignment of the randomized regions for 273 and 51 using MultAlin (Corpet 1988), showing the one nucleotide difference.

A

Cluster A – no SD sequence						Cluster B – SD sequence						Cluster C – no SD sequence						Cluster D – SD sequence						Cluster E – no SD sequence						Cluster F – SD sequence						Cluster G – no SD sequence						Cluster H – no SD sequence					
Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max	Seq ID	Tt1	Tt2	Gk1	Gk2	max
273	3	3	7	7	7	31	7	7	6	6	7	130	5	4	3	3	5	276	5	5	4	4	5	115	4	3	3	2	4	12	5	5	3	2	5	121	2	1	3	3	3	4233	3	1	1	–	3
51	6	5	7	7	7	5	6	6	6	6	6	165	5	5	5	4	5	18	3	3	3	3	3	162	4	3	3	3	4	188	4	4	2	2	4	152	3	3	3	3	3	432	3	1	1	–	3
1830	3	3	5	5	5	190	5	5	3	3	5	290	4	3	3	3	4	103	3	2	1	1	3	1225	3	3	3	1	3	6	3	3	–	–	3	161	2	–	2	1	2	47	3	1	2	–	3
470	3	3	5	5	5	257	5	5	3	3	5	30	4	3	3	3	4	395	3	3	1	1	3	230	3	2	2	2	3	184	3	3	–	–	3	1234	1	–	1	–	1	524	3	1	2	–	3
5049	3	3	5	5	5	182	4	4	3	3	4	64	4	4	4	3	4	880	3	3	1	1	3	235	2	2	3	2	3	214	3	3	–	–	3	268	1	–	1	1	1	859	3	1	2	–	3
1081	1	1	4	4	4	210	4	4	3	3	4	3	3	2	2	2	3	1212	2	2	1	1	2	32674	3	2	1	–	3	200	2	2	–	–	2	27210	1	1	1	1	1	926	3	1	2	1	3
1725	3	3	4	4	4	237	4	4	3	3	4	8	3	3	3	3	3	1269	2	2	1	1	2	5810	3	2	1	–	3	106	1	1	–	–	1	2794	1	1	1	1	1	1815	2	1	1	–	2
2464	2	2	4	4	4	1061	3	3	1	1	3	1035	3	2	2	2	3	3512	2	2	1	1	2	7009	3	2	3	2	3	1060	1	1	–	–	1	33093	–	–	1	–	1	33	1	–	2	–	2
7518	3	3	4	4	4	1161	3	3	1	1	3	11422	3	2	1	1	3	108	2	1	–	–	2	11645	2	1	1	–	2	1121	1	1	–	–	1	383	–	–	1	1	1	169	1	–	1	–	1
12552	–	–	3	3	3	1319	3	3	3	3	3	12451	3	1	2	1	3	624	1	1	1	1	1	13242	1	–	2	1	2	1362	1	–	–	–	1	4684	1	1	1	–	1	87	–	–	1	–	1

7	>5% normalized reads
6	3>5% normalized reads
5	1>3% normalized reads
4	0.5>1% normalized reads
3	0.1>0.5% normalized reads
2	0.05>0.1% normalized reads
1	0.01>0.05% normalized reads

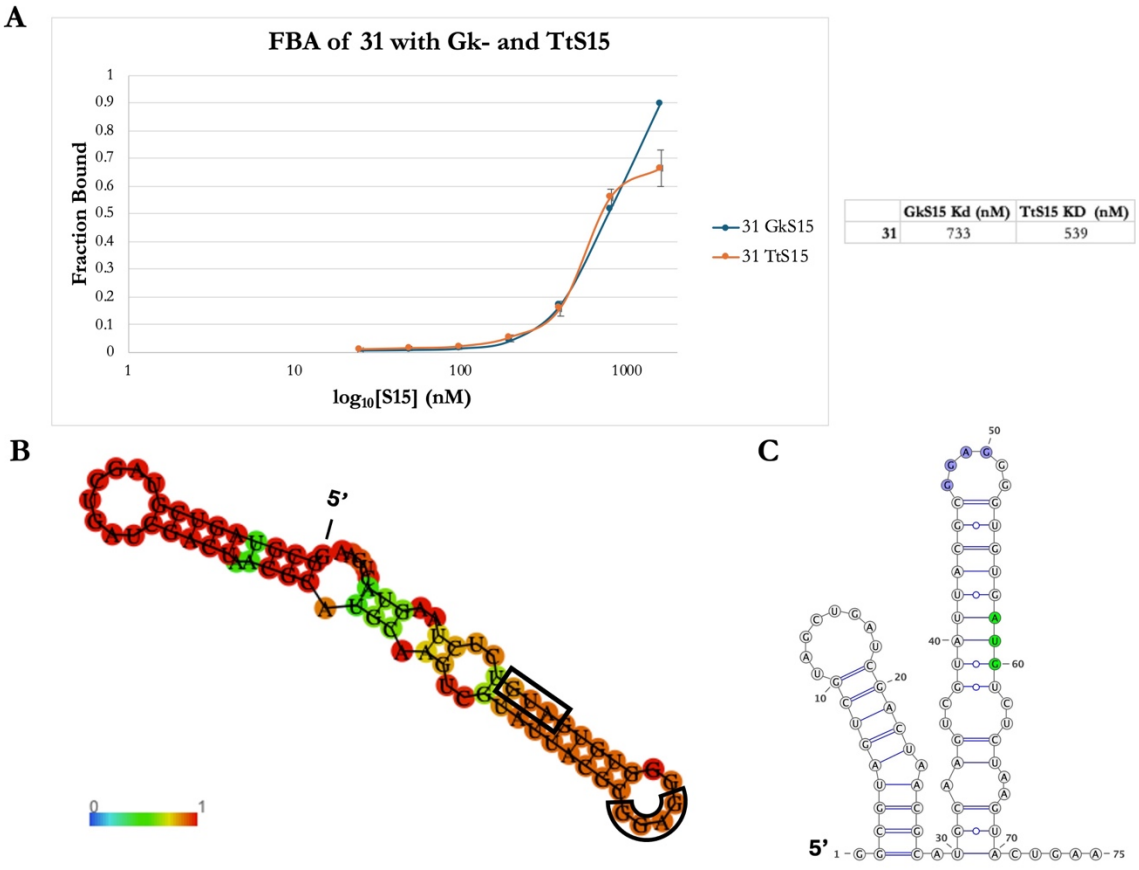
B

Seq ID	Sequence	Cluster distance	Notes
273	TCGACTATAGTGTGCGTATTACTTGCTGCCTCGTG	2.08	High frequency for Gk, but not Tt - no SD
51	TCGACTATAGTGTGCGTATTACTTGGTGCCTCGTG	1.84	High frequency for both - no SD
31	**TAACGCATGCAAGTCGTATTACGCGGAGGGGTGTG**	1.62	High frequency for both - has SD

C

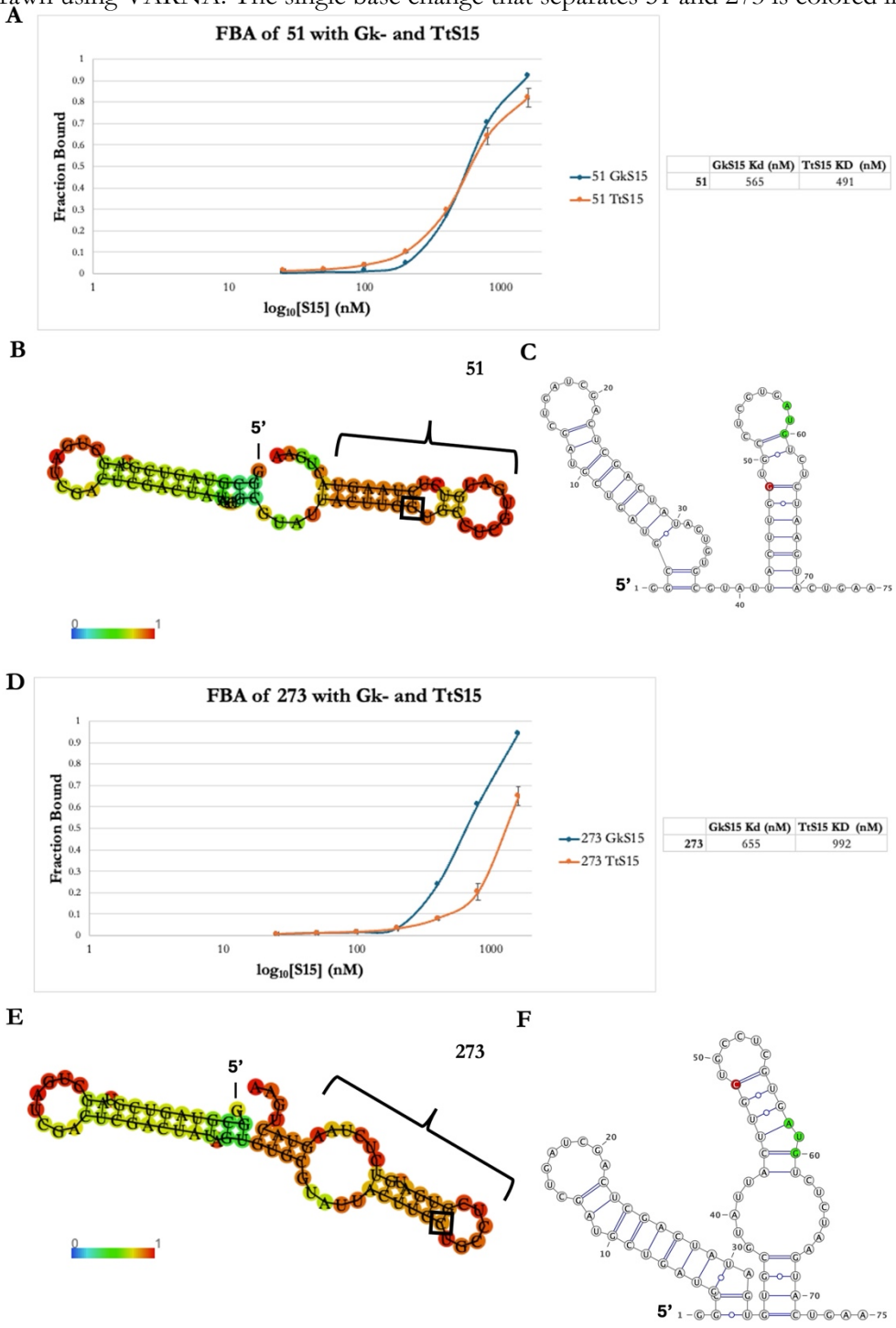
```
           1        10        20        30   35
           |--------+---------+---------+----|
273        TCGACTATAGTGTGCGTATTACTTGCTGCCTCGTG
51         TCGACTATAGTGTGCGTATTACTTGGTGCCTCGTG
Consensus  TCGACTATAGTGTGCGTATTACTTGcTGCCTCGTG
```

Figure 4.6 Filter-binding assay confirms binding for 31 with Gk- and TtS15
Filter-binding assays were performed with sequence 31 with Gk- and TtS15. A) Curves for the filter-binding assays, with the binding affinity for each S15 to the right. B) The predicted structure (minimum free energy) for 31 using RNAfold (Lorenz et al., 2011), with energy parameters scaled for 22º, as binding assays occurred at room temperature. Colors correspond to base-pair probabilities according to the scale, where red indicates a base-pairing probability of 1. The start codon and SD sequence are boxed. C) Folded RNA from B) drawn using VARNA (Visualization Applet for RNA, (Darty et al., 2009)), with the SD in lavender and start codon in green.



	GkS15 Kd (nM)	TtS15 KD (nM)
31	733	539

Figure 4.7 Differences in binding affinity for 51 and 273 with Gk- and TtS15
Filter-binding assays were performed with sequence 51 and 273 with Gk- and TtS15. A) Curves for the filter-binding assays with 51, with the binding affinity for each S15 to the right. B) The predicted structure (minimum free energy) for 51 using RNAfold, with energy parameters scaled for 22º, base that differentiates it from 273 boxed, and affected structure bracketed. C) Folded RNA from B) drawn using VARNA. The single base change that separates 51 and 273 is colored in red and start codon in green. D) Curves for the filter-binding assays with 273, with the binding affinity for each S15 to the right. E) The predicted structure for 273 using RNAfold, with the base that differentiates it from 51 boxed, and affected structure bracketed. F) Folded RNA from E) drawn using VARNA. The single base change that separates 51 and 273 is colored in red.



	GkS15 Kd (nM)	TtS15 KD (nM)
51	565	491

	GkS15 Kd (nM)	TtS15 KD (nM)
273	655	992

Table 4.1 Primers used to generate amplicon for Illumina sequencing

Name	Sequence
RT primer w universal adapter cP7	5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNN TTCAGTACTTAGAGACAT
P5 + universal adapter	5′-AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCTA
Adapter P7 + BC_01	5′- CAAGCAGAAGACGGCATACGAGATTCGTGTGCGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Adapter P7 + BC_05	5′- CAAGCAGAAGACGGCATACGAGATATCCGACAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Adapter P7 + BC_06	5′- CAAGCAGAAGACGGCATACGAGATAACATAATGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Adapter P7 + BC_08	5′- CAAGCAGAAGACGGCATACGAGATGCTAAGTAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Adapter cP5 BC01	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TATTGCTTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC02	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TATGAATTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC03	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAACTTGTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC04	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAGGCTGTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC05	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TATCAGGTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC06	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAATTAGTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC07	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TATTGAGTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC08	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TATGGTCTCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC09	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAGTTTATCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC10	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAATGTATCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC11	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAATTGATCGTAGTCGTAGCTGATCGAC
Adapter cP5 BC12	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC TAGTGGATCGTAGTCGTAGCTGATCGAC

Chapter V

Discussion

Summary and significance

This thesis offers novel insights into the RNA-binding profiles of two homologous proteins using *in vitro* selection against a partially patterned RNA library, highlighting that seemingly inconsequential sequence changes can lead to real changes in affinity which may affect the evolution of natural mRNA regulators.

Discussion

In Chapter II, I demonstrate the limitations of using a partially patterned RNA library for an *in vitro* selection scheme to analyze the RNA-binding profiles of the EcS15 homolog. I show that while partial patterning leads to desirable thermodynamic properties for the starting library *in silico*, it ultimately limited our ability to study the RNA-binding profile of EcS15 due to this homolog's unique mechanism of action compared to the Gk- and TtS15 homologs. I show that altering patterning and randomization of potential RNA libraries has a significant effect on the minimum free energy (reflective of the ability of the library to fold into stable structures) and aspects of structural complexity (stem length, multi-stem percentage, wobble base-pairing), as measured through computational modeling of library thermodynamic properties. Partial patterning reduces library size to allow for more comprehensive sampling of selection rounds via high-throughput sequencing and has been shown to enrich for functional aptamers, which is why I chose a partially patterned library for SELEX (Ruff et al., 2010). However, this choice significantly limited the number of possible sequences available for EcS15 aptamer enrichment. While partially patterned libraries increase the likelihood of structured RNA formation compared to fully randomized libraries, this also limited the sequence space used to study the RNA-binding profile of the EcS15 homolog and potentially led to our inability to isolate EcS15 aptamers.

I made several attempts to enrich for EcS15 aptamers with SELEX, altering selective pressure by starting at different EcS15 concentrations, changing the amount of library in the binding reaction, and optimizing the RNA renaturing protocol. Ultimately, I was unsuccessful in isolating aptamers for this homolog. Though we could not successfully select for EcS15 binders, this is still a result, indicating that high affinity aptamers for EcS15 are very infrequent in our partially patterned RNA pool. The desirable thermodynamic properties of the library did not translate into successfully generating EcS15 aptamers, and a fully randomized library might be a better starting point for future EcS15 SELEX experiments. The autogenous regulation of the *rpsO* operon in *E. coli* has been known for over 40 years and the structured mRNA regulator in the 5' UTR has been extensively studied and characterized. EcS15 utilizes an "entrapment" mechanism to simultaneously bind the mRNA regulator and 30S ribosomal subunit in the cell, stalling the ribosome and preventing translation of the downstream gene. The Ec-mRNA regulator exists in equilibrium as two stem-loops and a pseudo-knotted structure, which is stabilized by EcS15 binding, and is thus structurally distinct from the Gk- and Tt-mRNA regulators, whose respective S15 homologs utilize a "displacement" / competitive binding mechanism for regulation. Since we did not include any other components of the ribosome besides EcS15 in our SELEX, it would be difficult or impossible to replicate the entrapment mechanism in our experiments. Intriguingly, the recently described regulator from *Mycobacterium smegmatis* also contains a putative pseudoknot within its 5' leader region upstream of *rpsO* and may also utilize an entrapment mechanism, though this has yet to be experimentally validated (Aseev et al., 2021). EcS15 and other S15 homologs that utilize the entrapment mechanism may apply different selective pressures to their RNA leaders than Gk- and TtS15 and S15 homologs that utilize displacement, and EcS15's binding profile may be better studied using a fully randomized library or in an *in vivo* context using a method like

FADSRA (Fluorescent Activated Droplet Sorting Regulatory Assay, (Gray, 2022)), where entrapment could be replicated.

In Chapter III, I apply the optimizations learned from the EcS15 SELEX scheme in Chapter II to isolate aptamers for Gk- and TtS15. These homologs both utilize a displacement mechanism to regulate their *rpsO* operons and have evolved mRNA regulators that partially mimic the 16S rRNA with a GGC base triple at the base of a 3-way helical junction. Inter-species interaction studies have shown that the homologs can bind each other's regulator *in vitro* with comparable affinities and regulate using either mRNA regulator, as shown in an *in vivo* reporter assay. Mutational analysis revealed that the G•U/GC motif in the Gk-mRNA regulator is essential for GkS15 binding *in vitro* and *in vivo*, while these mutations have no effect on the interaction with TtS15, suggesting it interacts with the Gk-mRNA regulator differently than the native interaction. The inter-species interaction studies generally showed TtS15 to be a more promiscuous binder than GkS15, which indicates it might have a larger RNA-binding profile while GkS15 is more selective in its interactions.

To investigate potential differences in Gk- and TtS15 RNA-binding profiles, I performed *in vitro* selection experiments to isolate aptamers for these S15 homologs from the same starting library. Since these S15 homologs do not require any other ribosomal components to bind their mRNA regulators *in vitro* and *in vivo* and rely only on S15-mRNA binding, they were more amenable to our *in vitro* SELEX with our partially patterned library than the EcS15 homolog. Nitrocellulose filter-binding assays indicated that TtS15 bound the unselected RNA library with a higher affinity than GkS15, and I find that TtS15 is generally a better *in vitro* binder throughout SELEX with this library than GkS15. While EcS15 is a poor *in vitro* binder, it selectively interacts with the natural Gk-mRNA but not Tt-mRNA regulator. I demonstrate that GkS15 and TtS15 have overlapping binding profiles, as GkS15 is able to

interact with the round 6 population RNA from Tt SELEX, while EcS15 is unable to bind this population. This mirrors the inter-species interaction studies, as EcS15 can interact with the Gk-mRNA regulator but not the Tt-mRNA regulator and does not interact with the RNA pool enriched using TtS15 in our SELEX experiments. However, following 12 rounds of SELEX, neither population enriched by Gk- or TtS15 reached the low nanomolar affinity that these homologs have for their natural mRNA regulators. The benefits of using a partially patterned library with theoretically desirable thermodynamic properties that could be comprehensively sampled with high-throughput sequencing did not effectively translate into enriching for high affinity aptamers. I then go on to characterize the RNA-binding profiles using high-throughput sequencing and clustering analysis.

In Chapter IV, I utilize a combination of high-throughput sequencing and clustering analysis to characterize the Gk- and TtS15 RNA-binding profiles, and filter-binding assays to confirm individual Gk- and TtS15 aptamers isolated from *in vitro* selection. I find significant overlap in the RNA pools enriched through SELEX for the two replicates with GkS15, with 75% of the top 500 most frequent sequences shared between the replicates. I find 68% of the top 500 sequences are shared between the TtS15 SELEX replicates, which supports the hypothesis that TtS15 has a larger RNA-binding profile than GkS15 with this library.

I also find significant overlap between the RNA pools enriched by the Gk- and TtS15 homologs. Some overlap is not totally unexpected, as these S15 homologs are able to bind and regulate using each other's mRNA regulators. The extent of the overlap may partially be due to the patterning of our library, which decreased the total number of possible sequences in our starting RNA pool. As with the individual replicates, TtS15 exhibits a higher diversity of sequences in the final population than GkS15, further supporting TtS15's larger RNA-binding profile. Clustering analysis showed a larger number of lower frequency sequences for TtS15

than GkS15, also indicative of a potentially larger RNA-binding profile, though these low frequency sequences have not been validated as binders. One notable trend seen across both populations was the mutation of the Shine-Dalgarno sequence included in the original library design, indicating that while this sequence is important for regulating *in vivo*, it is not required for binding *in vitro*. There was no selective pressure to maintain the specific "GGAG" sequence, and as such, this sequence was frequently mutated across rounds of selection.

I identify 8 clusters of related sequences shared between the S15 homologs and confirm binding of 3 individual sequences through filter-binding assays. Sequence 31 from Cluster B, which has a SD sequence, was the most frequent sequence for Tt and 4th most frequent for GkS15 and binds both homologs with a similar affinity. Sequences 51 and 273 from Cluster A, which lack a SD sequence, differ by only one nucleotide in their randomized region. Despite this small difference, they have strikingly different frequencies in the Gk- and Tt SELEX data. Sequence 51 is the most frequent sequence for Gk and 2nd most frequent for TtS15 while sequence 273 is the 3rd most frequent sequence for Gk but 41st most frequent for TtS15. Folding simulations indicate that this single nucleotide change significantly alters the structure for half of the aptamer, which I hypothesize affects Tt but not GkS15 binding for this aptamer. The SD sequence for the Tt-mRNA regulator is base paired in the stem adjacent to the GGC base triple binding site for TtS15, so it is also possible that the SD sequence may stabilize similar structures for the TtS15 aptamers in our selected population. I confirm that both homologs bind sequence 51 with similar affinities, while GkS15 binds sequence 273 with a higher affinity than TtS15, which may explain its higher frequency in the GkS15 population in Round 12. Taken together, our results highlight overlapping but not identical binding profiles for Gk- and TtS15 and support the hypothesis that there are differences in RNA-

binding profiles for S15 homologs that may have driven the diversity of extant mRNA regulators.

Concluding remarks and future directions

Since its inception over 30 years ago, SELEX and its many derivatives has been widely utilized to select for aptamers that bind ions, small molecules, proteins, and even whole cells. While often used to select for functional aptamers to use as therapeutics or for diagnostic purposes, SELEX also allows us to explore more basic science questions. SELEX has been commonly used to explore DNA-protein interactions to determine binding motifs for transcription factors, but RNA-protein interactions are not well-characterized through SELEX, especially with structured RNAs. The work in this thesis aimed to fill in this gap using the diversity of the structured mRNA regulators that interact with r-protein S15 as a model. High-throughput sequencing and algorithms designed specifically for SELEX data allow us analyze aptamers with greater resolution than ever before. The work presented in this thesis provides the first comparative study of the RNA-binding profiles of two homologous proteins using SELEX.

Partially patterning our library allowed us to deeply sample our populations with high-throughput sequencing, but we were unable to isolate aptamers with the high affinity of the natural mRNA regulator we modeled our library after. Our partially patterned library was predicted to have more features of highly structured RNAs (long stems, short loops, multiple stems) compared to a fully randomized library, but this did not translate to enriching for high affinity S15 aptamers in our experiments. Despite this, we were able to enrich for multiple families of RNA aptamers that bind our S15 homologs and demonstrate that Gk- and TtS15 have overlapping but not identical binding profiles. Future work to elucidate the structures of the aptamers and identify any Gk- and TtS15-specific interactions will further our

understanding of the specific requirements for S15 binding. Our work highlights that due to RNA's incredible capacity for folding into complex and dynamic structures, even seemingly inconsequential changes in sequence can have a very real effect on the evolutionary trajectory of a sequence, which may also be shaped by the mechanism of action used by the protein to exert its regulatory effects. The natural diversity of the S15 mRNA regulators highlight that there are often multiple solutions to the same biological problem, and our comparative method can help to interrogate how this diversity came to be.

To better elucidate the RNA-binding profiles of the S15 homologs, future work using SELEX with fully randomized libraries may improve our understanding of the selective pressures imposed by S15 on mRNA cis-regulators. In our work, the partially patterned library significantly reduced our ability to explore the RNA structures able to interact with various S15 homologs *in vitro*. The presumed benefits of the partially patterned library, such as the increased likelihood of secondary structure formation and reduced number of SELEX rounds needed to enrich for aptamers, did not lead to the isolation of high affinity aptamers. Compared to the natural mRNA regulators, the dissociation constant of the Gk- and TtS15 aptamers isolated were several orders of magnitude higher than their native interactions, and we were unable to isolate any aptamers to EcS15 using this library and thus a fully randomized library may be better suited to assess all 3 S15 homologs.

Another possible alternative is to combine *in vitro* SELEX with an *in vivo* method in later rounds to add a more stringent selective pressure for S15-interacting RNAs. Following several rounds of SELEX, the regulatory activity of the aptamers can be assessed with various S15 homologs using a fluorescent reporter in FADSRA (Fluorescent Activated Droplet Sorting Regulatory Assay). This has the benefit of simulating the context of how these regulators would have evolved in the cell, as well as allowing for the possibility of the aptamers

to utilize an entrapment mechanism. While droplet formation is a severe bottleneck for this method, performing FADSRA in later rounds after the population has narrowed through *in vitro* SELEX would allow for better sampling of the population and monitor the selection of the regulators. Efforts to transition this droplet-based microfluidic assay into a more straightforward method using flow cytometry and fluorescence activated cell sorting are also ongoing and may enable *in vivo* selection and sampling of a much larger range of regulatory RNAs than FADSRA. Additionally, by including *in vivo* rounds of selection that allow for an aptamer to utilize ribosome entrapment, we can interrogate if the specific mechanism of action (displacement vs. entrapment) is intrinsic to the S15 homologs themselves, or if either mechanism is equally likely to have evolved.

Results presented in this thesis demonstrate that GkS15 and TtS15 exhibit overlapping but not identical RNA-binding profiles following *in vitro* selection against the same partially patterned RNA library. Despite their shared use of the "displacement" regulation mechanism *in vivo*, the small differences in the RNA-binding profiles of these homologs seen in our *in vitro* selection experiments may be indicative of the distinct selective pressures that shaped the homologs' diverse mRNA regulators. We further demonstrated that while partially patterned RNA libraries may produce desirable thermodynamic properties *in silico,* it does not necessarily translate to the selection of high affinity aptamers *in vitro*. Future *in vitro* selection experiments using S15 homologs may benefit from utilizing a fully randomized library to provide a more complete view of their RNA-binding profiles. Whether diverse RNA regulators arise independently or are so diverged from a common ancestor that we are unable to detect them as distant homologs remains an open question in the RNA genomics field. The evolutionary processes contributing to the rise and maintenance of RNA-based regulatory mechanisms in bacteria are complex, as both the RNA and protein can co-evolve over time. S15 continues to

serve as a useful model for the selective pressures driving the diversity of RNA cis-regulators, and the work presented in this thesis highlights that the RNA itself, the protein it interacts with, and the mechanism of interaction may all contribute in distinct ways to shape the regulators we see today.

References

Abduljalil, J. M. (2018). Bacterial riboswitches and RNA thermometers: Nature and contributions to pathogenesis. *Non-Coding RNA Research*, *3*(2), 54–63. https://doi.org/10.1016/j.ncrna.2018.04.003

Aseev, L. V., Koledinskaya, L. S., Bychenko, O. S., & Boni, I. V. (2021). Regulation of Ribosomal Protein Synthesis in Mycobacteria: The Autogenous Control of rpsO. *International Journal of Molecular Sciences*, *22*(18), Article 18. https://doi.org/10.3390/ijms22189679

Babina, A. M. (2017). *In vivo characterization of RNA cis-regulators in bacteria*.

Barnard, A., Wolfe, A., & Busby, S. (2004). Regulation at complex bacterial promoters: How bacteria use different promoter organizations to produce different regulatory outcomes. *Current Opinion in Microbiology*, *7*(2), 102–108. https://doi.org/10.1016/j.mib.2004.02.011

Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J. K., & Breaker, R. R. (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences*, *101*(17), 6421–6426. https://doi.org/10.1073/pnas.0308014101

Bénard, L., Mathy, N., Grunberg-Manago, M., Ehresmann, B., Ehresmann, C., & Portier, C. (1998). Identification in a pseudoknot of a U·G motif essential for the regulation of the expression of ribosomal protein S15. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(5), 2564–2567.

Bervoets, I., & Charlier, D. (2019). Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews*, *43*(3), 304–339. https://doi.org/10.1093/femsre/fuz001

Boehringer, D., & Ban, N. (2007). Trapping the Ribosome to Control Gene Expression. *Cell*, *130*(6), 983–985. https://doi.org/10.1016/j.cell.2007.09.002

Bowman, J. C., Lenz, T. K., Hud, N. V., & Williams, L. D. (2012). Cations in charge: Magnesium ions in RNA folding and catalysis. *Current Opinion in Structural Biology*, *22*(3), 262–272. https://doi.org/10.1016/j.sbi.2012.04.006

Clement, K., Farouni, R., Bauer, D. E., & Pinello, L. (2018). AmpUMI: Design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics*, *34*(13), i202–i210. https://doi.org/10.1093/bioinformatics/bty264

Crum, M., Ram-Mohan, N., & Meyer, M. M. (2019). Regulatory context drives conservation of glycine riboswitch aptamers. *PLOS Computational Biology*, *15*(12), e1007564. https://doi.org/10.1371/journal.pcbi.1007564

Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, *25*(15), 1974–1975. https://doi.org/10.1093/bioinformatics/btp250

Dutcher, H. A., & Raghavan, R. (2018). Origin, evolution, and loss of bacterial small RNAs. *Microbiology Spectrum*, *6*(2), 10.1128/microbiolspec.RWR-0004–2017. https://doi.org/10.1128/microbiolspec.RWR-0004-2017

Ehresmann, C., Ehresmann, B., Ennifar, E., Dumas, P., Garber, M., Mathy, N., Nikulin, A., Portier, C., Patel, D., & Serganov, A. (2004). Molecular Mimicry in Translational

Regulation: The Case of Ribosomal Protein S15. *RNA Biology*, *1*(1), 65–72. https://doi.org/10.4161/rna.1.1.958

Fu, Y., Deiorio-Haggar, K., Anthony, J., & Meyer, M. M. (2013). Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. *Nucleic Acids Research*, *41*(6), 3491–3503. https://doi.org/10.1093/nar/gkt055

Gelfand, M. S. (2006). Bacterial cis-regulatory RNA structures. *Molecular Biology*, *40*(4), 541–550. https://doi.org/10.1134/S0026893306040066

Gray, E. C. (2022). *Development and Evaluation of a Fluorescent Activated Droplet Sorting Regulatory Assay for Ribosomal Cis-Regulatory RNAs* [Boston College]. http://dlib.bc.edu/islandora/object/bc-ir:109524

Gu, G., Wang, T., Yang, Y., Xu, X., & Wang, J. (2013). An Improved SELEX-Seq Strategy for Characterizing DNA-Binding Specificity of Transcription Factor: NF-*κ*B as an Example. *PLOS ONE*, *8*(10), e76109. https://doi.org/10.1371/journal.pone.0076109

Hall, K. B., & Kranz, J. K. (1999). Nitrocellulose Filter Binding for Determination of Dissociation Constants. In S. R. Haynes, *RNA-Protein Interaction Protocols* (Vol. 118, pp. 105–114). Humana Press. https://doi.org/10.1385/1-59259-676-2:105

Helmann, J. D. (2019). Where to begin? Sigma factors and the selectivity of transcription initiation in bacteria. *Molecular Microbiology*, *112*(2), 335–347. https://doi.org/10.1111/mmi.14309

Ishida, R., Adachi, T., Yokota, A., Yoshihara, H., Aoki, K., Nakamura, Y., & Hamada, M. (2020). RaptRanker: In silico RNA aptamer selection from HT-SELEX experiment based on local sequence and structure information. *Nucleic Acids Research*, *48*(14), e82. https://doi.org/10.1093/nar/gkaa484

Jacob, F., & Monod, J. (1961). On the Regulation of Gene Activity. *Cold Spring Harbor Symposia on Quantitative Biology*, *26*, 193–211. https://doi.org/10.1101/SQB.1961.026.01.024

Kohlberger, M., & Gadermaier, G. (2022). SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry*, *69*(5), 1771–1792. https://doi.org/10.1002/bab.2244

Komarova, N., & Kuznetsov, A. (2019). Inside the Black Box: What Makes SELEX Better? *Molecules (Basel, Switzerland)*, *24*(19), 3598. https://doi.org/10.3390/molecules24193598

Lal, A., Krishna, S., & Seshasayee, A. S. N. (2018). Regulation of Global Transcription in Escherichia coli by Rsd and 6S RNA. *G3: Genes|Genomes|Genetics*, *8*(6), 2079–2089. https://doi.org/10.1534/g3.118.200265

Li, L., Huang, D., Cheung, M. K., Nong, W., Huang, Q., & Kwan, H. S. (2013). BSRD: A repository for bacterial small regulatory RNA. *Nucleic Acids Research*, *41*(Database issue), D233–D238. https://doi.org/10.1093/nar/gks1264

Loh, E., Righetti, F., Eichner, H., Twittenhoff, C., & Narberhaus, F. (2018). RNA Thermometers in Bacterial Pathogens. *Microbiology Spectrum*, *6*(2), 10.1128/microbiolspec.rwr-0012–2017. https://doi.org/10.1128/microbiolspec.rwr-0012-2017

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 26. https://doi.org/10.1186/1748-7188-6-26

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research*, *15*(21), 8783–8798. https://doi.org/10.1093/nar/15.21.8783

Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B., & Breaker, R. R. (2017). Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Molecular Cell*, *65*(2), 220–230. https://doi.org/10.1016/j.molcel.2016.11.019

Nevskaya, N., Tishchenko, S., Gabdoulkhakov, A., Nikonova, E., Nikonov, O., Nikulin, A., Platonova, O., Garber, M., Nikonov, S., & Piendl, W. (2005). Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Research*, *33*(2), 478–485. https://doi.org/10.1093/nar/gki194

Nomura, M. (1999). Regulation of Ribosome Biosynthesis in Escherichia coli and Saccharomyces cerevisiae: Diversity and Common Principles. *Journal of Bacteriology*, *181*(22), 6857–6864. https://doi.org/10.1128/jb.181.22.6857-6864.1999

Nonaka, G., Blankschien, M., Herman, C., Gross, C. A., & Rhodius, V. A. (2006). Regulon and promoter analysis of the E. coli heat-shock factor, σ32, reveals a multifaceted cellular response to heat stress. *Genes & Development*, *20*(13), 1776–1789. https://doi.org/10.1101/gad.1428206

Olenginski, L. T., Spradlin, S. F., & Batey, R. T. (2024). Flipping the script: Understanding riboswitches from an alternative perspective. *Journal of Biological Chemistry*, *300*(3), 105730. https://doi.org/10.1016/j.jbc.2024.105730

Pei, S., Slinger, B. L., & Meyer, M. M. (2017). Recognizing RNA structural motifs in HT-SELEX data for ribosomal protein S15. *BMC Bioinformatics*, *18*(1), 298. https://doi.org/10.1186/s12859-017-1704-y

Peselis, A., & Serganov, A. (2018). ykkC riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. *Nature Chemical Biology*, *14*(9), 887–894. https://doi.org/10.1038/s41589-018-0114-4

Philippe, C., Eyermann, F., Bénard, L., Portier, C., Ehresmann, B., & Ehresmann, C. (1993). Ribosomal protein S15 from Escherichia coli modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(10), 4394–4398.

Riley, T. R., Slattery, M., Abe, N., Rastogi, C., Mann, R., & Bussemaker, H. (2014). SELEX-seq, a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods in Molecular Biology (Clifton, N.J.)*, *1196*, 255–278. https://doi.org/10.1007/978-1-4939-1242-1_16

Ruff, K. M., Snyder, T. M., & Liu, D. R. (2010). Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure. *Journal of the American Chemical Society*, *132*(27), 9453–9464. https://doi.org/10.1021/ja103023m

Schütze, T., Wilhelm, B., Greiner, N., Braun, H., Peter, F., Mörl, M., Erdmann, V. A., Lehrach, H., Konthur, Z., Menger, M., Arndt, P. F., & Glökler, J. (2011). Probing the SELEX Process with Next-Generation Sequencing. *PLOS ONE*, *6*(12), e29604. https://doi.org/10.1371/journal.pone.0029604

Scott, L. G., & Williamson, J. R. (2001). Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with its 5'-translational operator mRNA1. *Journal of Molecular Biology*, *314*(3), 413–422. https://doi.org/10.1006/jmbi.2001.5165

Serganov, A., Polonskaia, A., Ehresmann, B., Ehresmann, C., & Patel, D. (2003). Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within

its mRNA. *The EMBO Journal*, *22*, 1898–1908. https://doi.org/10.1093/emboj/cdg170

Sherlock, M. E., Sadeeshkumar, H., & Breaker, R. R. (2019). Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. *Biochemistry*, *58*(5), 401–410. https://doi.org/10.1021/acs.biochem.8b00617

Sherlock, M. E., Sudarsan, N., & Breaker, R. R. (2018). Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 6052–6057. https://doi.org/10.1073/pnas.1720406115

Sherlock, M. E., Sudarsan, N., Stav, S., & Breaker, R. R. (2018). Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife*, *7*, e33908. https://doi.org/10.7554/eLife.33908

Slinger, B. L., & Meyer, M. M. (2016). RNA regulators responding to ribosomal protein S15 are frequent in sequence space. *Nucleic Acids Research*, *44*(19), 9331–9341. https://doi.org/10.1093/nar/gkw754

Slinger, B. L., Newman, H., Lee, Y., Pei, S., & Meyer, M. M. (2015). Co-evolution of Bacterial Ribosomal Protein S15 with Diverse mRNA Regulatory Structures. *PLOS Genetics*, *11*(12), e1005720. https://doi.org/10.1371/journal.pgen.1005720

Urak, K. T., Shore, S., Rockey, W. M., Chen, S.-J., McCaffrey, A. P., & Giangrande, P. H. (2016). In vitro RNA SELEX for the generation of chemically-optimized therapeutic RNA drugs. *Methods (San Diego, Calif.)*, *103*, 167–174. https://doi.org/10.1016/j.ymeth.2016.03.003

Vogel, J., & Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nature Reviews Microbiology*, *9*(8), 578–589. https://doi.org/10.1038/nrmicro2615

Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, *136*(4), 615–628. https://doi.org/10.1016/j.cell.2009.01.043

Watkins, D., & Arya, D. (2023). Models of Hfq interactions with small non-coding RNA in Gram-negative and Gram-positive bacteria. *Frontiers in Cellular and Infection Microbiology*, *13*. https://doi.org/10.3389/fcimb.2023.1282258

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., Neph, S., Tompa, M., Ruzzo, W. L., & Breaker, R. R. (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Research*, *35*(14), 4809–4819. https://doi.org/10.1093/nar/gkm487

Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H., & Breaker, R. R. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology*, *11*(3), R31. https://doi.org/10.1186/gb-2010-11-3-r31