DEVELOPMENT AND EVALUATION OF A FLUORESCENT ACTIVATED DROPLET SORTING REGULATORY ASSAY FOR RIBOSOMAL CIS-REGULATORY RNAS

a dissertation

by

Elizabeth Catherine Gray

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

July 2022

ABSTRACT

Development and Evaluation of a Fluorescent Activated Droplet Sorting Regulatory Assay for Ribosomal Cis-Regulatory RNAs

Existing methods of assaying the function of cis-regulatory RNAs come with significant drawbacks when assaying large RNA libraries. Highly sensitive cell-based assays such as the β-galactosidase assay are labor intensive, difficult to scale up and may lose sensitivity with increased throughput. GFP and luciferase reporters can be used with FACS to increase assay throughput, but sorting small bacterial cells is challenging and greatly reduces assay sensitivity. Conversely, *in vitro* methods allow for fast screening of very large RNA libraries, but only select for properties of binding, not regulation. By combining the principles of classic in cell regulatory assays with modern tools, cis-regulatory RNAs can be quickly screened for regulatory activity at a large scale.

The assay under development, Fluorescent Activated Droplet Sorting Regulatory Assay (FADSRA), uses microfluidics to encapsulate single cells expressing a fluorescent protein under the control of a cis-regulatory RNA. These cells are then cultured into microcolonies within the droplets, which are subsequently sorted according to fluorescent signal. Deep amplicon sequencing of the regulatory RNAs can then reveal which sequences can regulate and which cannot. Thus, FADSRA can help bridge the gap between *in vitro* RNA binding and gene regulation assays, providing a way to answer sophisticated questions about cis-regulatory RNAs requiring high-throughput assay methods.

While many applications for FADSRA are possible, such as verifying regulatory activity of *in vitro* binders or screening synthetic regulators, one such application of

FASDRA is the creation of fitness landscapes that probe sequence-function relationships of RNA cis-regulators. This dissertation first develops and optimizes the regulatory assay for ribosomal leaders in Chapter 2, following by creating a single mutant fitness landscape of the *E. coli* S15 leader RNA in Chapter 3. Results of this fitness landscape largely support previously published mutational studies and highlight the necessity of stable hairpin formation for regulation of the *E. coli* S15 leader. Chapter 4, examining the regulation of S15 protein and leader homologs, and Chapter 5, testing the adaptability of FADSRA to other cis-regulatory RNAs, examine possible further applications of the assay.

DEDICATION

To my wife Keeva.

Toute ma vie, je suis fière de toi.

and

In memory of Chelsea “Spyder” Moller.

ACKNOWLEDGEMENTS

I have a great many people to thank for their support of this project and for all my graduate school endeavors:

My current mentor Dr. Michelle Meyer, for believing in me and for providing a model for excellence and endurance.

My committee members, for supporting me and sharing wise and temperate advice: Dr. Eric Folker, Dr. Babak Momeni, Dr. Huiqing Jane Zhou, and Dr. Katherine Berry.

My past mentors, Dr. Karl Griswold and Dr. Naomi Morrissette, for investing in me as a young scientist and helping me grow as an independent researcher.

The undergraduate researchers who assisted me with this research: Wellington Arkins and Abigail Miller.

My lab colleagues past and present, especially Arianne Babina, Daniel Beringer, Indu Warrier, Matthew Crum, Nikhil Ram Mohan, and Rebecca Korn.

My fellow graduate students who selflessly helped me when I was in need of assistance: Natalie Sandlin, Suyen Espinoza Miranda, and Bharathi Sundaresh.

My Boston College colleagues and friends, especially Andrew Denninger, Kathleen Moorhouse, Michelle Archibald, Samantha Dyckman, and Defne Surujon.

My wife, for her love and patience.

My parents, for raising me with grit and kindness.

My brother, for keeping me in trouble.

And all my friends who love and support me. I cherish you.

Select figures designed with BioRender and Serial Cloner.

TABLE OF CONTENTS

LIST OF FIGURES & TABLES

ABBREVIATIONS

CDS	coding sequence
EMSA	electrophoretic mobility shift assay
FACS	fluorescent activated cell sorting
FADS	fluorescent activated droplet sorting
FADSRA	fluorescent activated droplet sorting regulatory assay
FSC	forward scatter
GFP	green fluorescent protein
IPTG	isopropyl ß-D-1-thiogalactopyranoside
LB	lysogeny broth
M1	mutant 1: denotes a non-binding mutant leader RNA
ONPG	ortho-nitrophenyl-ß-galactoside
ONP	ortho-nitrophenol
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SOB	super optimal broth
SSC	side scatter
UTR	untranslated region
WT	wild type

CHAPTER I: INTRODUCTION

Part 1: The Bacterial Ribosome

The ability for a cell to create proteins underlies all other cell functions. To do so, mRNA is transcribed from DNA, and is then translated into protein by ribosomes. The bacterial ribosome (the 70S ribosome) is composed of two subunits, the small 30S subunit and the large 50S subunit. Contained in the 70S are three ribosomal RNAs, the 16S in the small subunit and the 5S and the 23S in the large subunit, and over 50 ribosomal proteins accounting for about one-third of the total mass. [1] Many of these proteins, as expected from proteins that bind nucleic acids, are positively charged and serve to scaffold the rRNAs into a structure, which harnesses the unique power of ribonucleic acids to position the mRNA, decode tRNAs and catalyze the peptide bond to form polypeptide chains. [2] Given its central importance to all cell functions, ribosome regulation, assembly and function are all enduring areas of research. Understanding the basic science of the ribosome also supports advances in our understanding of antibiotics and is a rich area of exploration for synthetic biology. This particular thesis focuses largely on the regulation of proteins, especially ribosomal proteins.

Mechanisms of Transcription and Translation

In order to understand autogenous ribosomal protein regulation, it is also important to understand aspects of transcription and translation in bacteria. Generally, the 5' region of bacterial mRNA contains several features critical for translation. First, there is the Shine Dalgarno sequence, also known as the ribosome binding site, which is located about 10 nucleotides upstream from the AUG start codon. This purine rich sequence can base pair with the 3' end of the 16S rRNA in the ribosome, positioning the start codon appropriately within the ribosome to begin translation. Additional structured elements are also found in

the 5' UTR and can have regulatory properties. [3] Critical to the understanding of the regulatory mechanisms in bacteria for protein synthesis is that in bacteria, transcription and translation are coupled. Furthermore, the exact degree of coupling varies by species and carries implications for regulation at the transcriptional and the translational level. In *E. coli*, transcription and translation are tightly coupled as the RNA polymerase and ribosome are physically coupled in a model called translation-coupled transcription. [4] In *Bacillus subtilis*, for example, the ribosome trails behind the RNA polymerase, leaving a gap between the two proteins in a model called runaway transcription. [5] This has implications for which regulatory techniques are favored by a species, with runaway transcription organisms favoring comparatively slow folding structures like riboswitches and translation-coupled transcription organisms favoring quick folding structures like intrinsic terminators. [6] Additionally, the favored mechanism of termination, intrinsic or rho dependent, also varies by species with some clades lacking rho entirely. [7] The rate of translation is further dependent on ribosome pausing through mechanisms like codon bias and secondary structures present in the coding region. [8, 9]

Cis-regulatory RNAs

Cis-regulatory RNAs are structured regions of the mRNA that contribute to the regulation of gene expression through modulating transcription or translation. [3,10] Although found across the domains of life, in bacteria cis-regulatory RNAs are found in the 5' untranslated region (UTR) of the mRNA. These regulatory RNAs allow for rapid changes in protein expression and frequently regulate proteins found in metabolism and stress response. Several cis-regulatory RNA leaders, a name for the 5' untranslated region of mRNAs, have been implicated in antibiotic resistance or pathogenicity. [11–13] In

bacteria, modulation of transcription and translation is frequently achieved through masking or revealing expression elements on the mRNA through the RNA secondary structure. (Figure 1.1). Cis-regulatory RNAs are also found in eukaryotic and viral mRNAs; however, these regulatory elements are rarely found in the 5' UTR and use different mechanisms than the bacterial RNA cis-regulators discussed here. [14]

Riboswitches

Riboswitches are cis-regulatory RNAs that regulate a downstream coding sequence based on small ligands. The riboswitch is typically composed of two parts: the aptamer binding domain that binds the small molecule and the expression platform that contains the coding sequence (CDS). There are two main mechanisms of action, transcriptional and translational, and riboswitches can be "on" (permissive of protein synthesis) or "off" (downregulating protein synthesis) for each type. [15] For a transcriptional riboswitch, the formation of a terminator stem loop inhibits further DNA transcription, and subsequently mRNA translation. Upon a change in ligand binding the terminator stem loop can be disrupted with the formation of an anti-terminator, permitting transcription to continue. Translational regulators typically work by sequestering the ribosome binding site or the start codon under repressive conditions (Figure 1.1). Riboswitches can also use both transcriptional and translational methods of regulation such as the TPP riboswitch, a carbohydrate metabolism cofactor sensing switch that has a transcription termination stem that also sequesters the RBS and start codon. [16]

Riboswitches are generally found in pathways requiring a fast response to changing conditions. These regulators can sense ions, small metabolites and enzyme cofactors that can help the cell turn on or off processes like metabolic or detoxification pathways.

Riboswitches are found across bacterial clades, although the specifics of which riboswitches and riboswitch location vary. [17]

Ribosomal Leaders

The ribosome can account for the plurality of the protein in the cell. Demand for ribosomes, and thus ribosome components, increases with the rate of growth. [18] Since ribosomal components are energetically expensive to produce, tight regulation of component synthesis is advantageous. One mechanism controlling ribosomal protein synthesis is the interaction between an RNA leader and a ribosomal protein. Ribosomal proteins are frequently found clustered together on operons, linking their transcription together. In some of these operons one or two of the proteins will bind the 5' untranslated region, or leader, of their own mRNA, leading to down regulation of protein expression for that operon, an elegant method of autogenous regulation that can quickly change production of ribosomal components to meet the needs of the cell. [19,20]

While there is evidence that this mechanism is conserved across bacteria and even found in archaea, the specific regulating proteins as well as the RNA leader structure to which the protein binds are variable. Some regulatory pairs like L1, L10 and S2 are broadly found, while some regulators appear specific to certain phyla like S1 (Figure 1.2). [21,22] Research in *B. subtilis* found that the L20 regulator conferred a fitness advantage. [23] Given both the ubiquity of this autogenous regulation and the degree of variability, many questions remain about how these regulatory leader sequences arise and are maintained. Early studies into ribosomal leaders largely focused on *E. coli*, thus discovering and validating regulators in gram-positive and even archaea are underexplored in comparison. [24–27]

Mechanisms of Regulation

For ribosomal leader regulation, binding of the ribosomal protein to the leader contributes to the regulation of the operon. In addition to termination hairpins or ribosome binding site sequestration one mechanism of downregulation is that of entrapment. In this mechanism the protein and the ribosome pre-initiation complex both bind the mRNA, which traps the pre-initiation complex and prevents further ribosome assembly and, thus, translation (Figure 1.3A). [28,29] Another major mechanism is displacement, where the protein and the ribosome compete for the same binding site on the mRNA (Figure 1.3B). [30,31]

The L1 Leader

The L1 leader (sometimes L11 leader) is broadly found across clades of bacteria and even found in archaea. [19,21] The L1 protein and its binding to both the rRNA and the mRNA leader have been extensively characterized, including through crystal structures. [32–35] This leader forms a single simple structure with a single stem loop and a bulge. [34] Comparisons between the 23S rRNA and the L1 leader structures show that the leader closely mimics the rRNA. [36,37] The binding between the L1 protein and the rRNA is highly conserved, although L1 is not essential for ribosome function. [32,38] L1 binds to the 3' end of the 23S rRNA and is thought to play a role in tRNA release from the ribosome through the action of a flexible stalk. [39–42] The other protein in the operon, L11, is involved in translation factor binding and is also not essential for growth in *E. coli*. [38] Similarly, the binding of L1 protein to the L1 leader is also conserved. [33] More curiously, the location of this leader changes in genomic context. [22] The L1 operon contains the genes *rplK,* which is translated into L11 protein, and *rplA,* which makes the

L1 protein. In *E. coli*, a proteobacteria, the L1 leader precedes *rplK*, while in *Actinobacteria* the leader is found between the *rplK* and *rplA* genes. Most *Firmicutes* have yet another variation where leaders of similar structure are found before both *rplK* and *rplA*. Binding studies revealed that L1 protein can bind both the L1 and L11 leaders. [43] Looking towards even more evolutionarily distant homologs, the L1 leader is also found in archaea in a very different context. [19] In *Methanococcus*, the L1 leader proceeds an operon containing *rplA* (L1), *rplJ* (10) and *rplP1* (P1), proteins not found in the bacterial L1 operon. In another case, *Sulfolobus*, the L1 leader precedes an operon containing *rplK*, *plrA*, *prlJ* and *rplP1.* In all cases, the major features of the leader structure remain the same: a stem loop with a G at the base of the stem and a GGC motif at the base of a bulge. These features are all found on the rRNA. While the similarity of the structure could suggest horizontal transfer of the leader, the simple structure and the diversity of genomic locations of the leader suggest that this leader may have arisen multiple times in evolutionary history due to an intrinsic inflexibility of L1 binding.

The S15 Leader

The S15 ribosomal protein is one of the most well studied proteins for both binding to the ribosomal RNA as well as the study of its autogenous regulatory leader. The S15 protein is a primary binder of the 16S rRNA and is important in assembly of the central domain of the small subunit, binding to S6:S18 dimer after binding the rRNA. [38,44–46] S15 is not considered an essential protein as *E. coli rpsO* knockout cells do assemble mature ribosomes, albeit at a slower rate and while showing an abnormal elongated cell phenotype. [47,48] Yet the cold sensitivity of this knockout strain may indicate that there are conditions outside of standard laboratory culture for which S15 is necessary for the

reliable formation of functional ribosomes. This cold sensitive phenotype is typical of ribosomal mutants, and for S15 binding to the rRNA was shown to be decreased at lower temperatures. [49,50]

For rRNA binding, the S15 protein was found to have two main binding sites. The first binding site is at a CGG base triple within a three helix junction that interacts with the second and third helices of the S15 protein. [51–55] The secondary binding site is a conserved CG/GU motif recognized by the S15 loop 2. [56] Although many ribosomal leaders are rRNA mimics, this is not entirely true with the S15 regulators. When comparing the structures of the S15 leader homologs to that of the rRNA, each leader contains at least one of these motifs, but often not both, and sometimes contains a conserved motif found only in that leader. For example, the *E. coli* leader has the CG/GU motif, found in the rRNA, but also contains a conserved pseudoknot, a structure completely absent in the rRNA. Extensive mutational analysis of the *E. coli* leader revealed that both the pseudoknot and the G·U/G-C motif are critical for regulation. [57–60] However, while the G·U/G-C motif is sequence specific, the pseudoknot is based on structure alone. The *T. thermophilus* leader has the three-way helical junction, but no other S15 rRNA binding motif of note. [31] The *Geobacillus* leader and the *R. radiobacter* leader each contain the G·U/G-C motif but do not preserve other sequence or structural motifs. [61,62] Beyond the sequence and structure, the mechanism of regulation is different for different homologs. Regulation in *E. coli* is by entrapment, where the small subunit partially assembles onto the S15 protein. *T. thermophilus* and *G. kaustophilus* are regulated through displacement. The relative rarity of mimics of the primary rRNA binding site (the 3-way junction) in the diversity of

structures and the favoring of the G·U/G-C helical motif is notable, as the G·U/G-C is the less critical binding site on the rRNA.

Leader sequences preceding *rpsO* (encoding S15) are found in diverse clades, including *Proteobacteria, Firmicutes and Actinobacter,* but appear to be more narrowly distributed than the L1 leader. [62] Unlike the L1 operon, the S15 operon only includes the *rpsO* (S15). However, the structure of the S15 leader is extraordinarily diverse and binding studies have identified that features important for S15 binding to rRNA are not necessarily conserved in the S15 leader. [63] To date, six different leader structures have been discovered for S15, each confined to their own clade. Each leader has at least one binding site that mimics the rRNA, but some contain only one of these binding sites or contain additional binding sites that are divergent from the rRNA structure.

Previous work examined the binding and regulation of these four distinct S15 leaders and S15 protein pairs. β-galactosidase assays comparing both protein and leader homologs found that despite the structural diversity many of the proteins have some binding and regulatory activity with leaders that are structurally different from the native leader. [63] Additional S15 leader structures were discovered in *Mycobacterium* and *Chlamydia*, unique from the other leaders as well as the rRNA. [62] All of this taken together indicates that the S15 protein has a much more flexible ability to bind to RNA than the L1 protein. Together, study of these two RNAs can reveal the binding plasticity of these ribosomal RNA regulators and answer questions about how these regulators arise and are maintained.

Part 2: Methods for Studying Cis-Regulatory RNAs

There are several aspects of cis-regulatory RNAs that can be studied to understand their action, importance and evolution. *In vitro* methods can assay for binding between an RNA of interest and a ligand or identify RNA structural changes or protein binding sites. *In vivo* studies of cis-regulatory RNAs look for changes in protein expression, assess changes to viability or virulence inside a host, or determine regulatory RNA structure inside the cell. Many methods for assaying cis-regulatory RNAs remain popular and almost unchanged from decades ago. With new tools and technologies, reimagining these older assays and expanding the scope of scientific questions asked deserves consideration. Standard assays such as the Miller (β-galactosidase) assay and SELEX are considered, as well as technologies that will be used to develop the new method described by this dissertation, including microfluidics and flow cytometry.

Miller and Derivative Assays

A foundational regulation assay in bacteria is the "Miller assay," named after the author JH Miller in his book chapter "Experiments in Molecular Genetics". [64] In this assay *lacZ*, the gene for β-galactosidase, is used as a reporter to assay regulatory activity of an element of interest. Early research into the regulation of ribosomal proteins used β-galactosidase assays, a method still used today. [62,65–67] Cells grown in bulk liquid culture are permeabilized to give access of β-galactosidase substrate ONPG, which is then cleaved to give yellow ONP. Formation of ONP is measured over time and, when compared with cell density and time, can be calculated into a single factor called a Miller unit. Comparing Miller units between conditions can then reveal if regulation is occurring. In the case of cis-regulatory RNAs, the RNA of interest is fused in frame to the *lacZ* gene.

In some cases, as with riboswitches, the ligand is then added differentially to the growth medium prior to the assay. In others, as is with the ribosomal leaders, a protein binding partner can be introduced on an inducible overexpression plasmid. Since 1972, many variations and improvements have been made to this method. Newer regulatory assays can use less noxious permeabilization methods [68] or fluorescent reporter proteins like GFP or luciferase that do not require permeabilization at all. [63,69] However, β-galactosidase remains a favorite reporter to use due to its high sensitivity. [70,71]

Traditionally these regulator assays are done in 1.5 mL microcentrifuge tubes and are both time sensitive and time consuming, thus making it quite difficult to scale up this assay. Several protocols for a 96-well format exist but require specialized liquid handling tools for accuracy and remain labor intensive. [68] In the experience of our laboratory, some precision is lost in the 96-well format. Furthermore, 96 wells at a time is certainly not enough for modern applications where RNA libraries of 10^5 or more are not uncommon.

Flow Cytometry and Cell Sorting

The first studies of bacteria using flow cytometry were published decades ago in 1977. [72,73] However, most of the utility of flow cytometry and fluorescence activated cell sorting has been realized using mammalian cells. This is unsurprising given that bacterial cells are three orders of magnitude smaller than mammalian cells and, in the early days of flow cytometry, were unable to take up dye as efficiently. [74] Additionally, small bacteria cells can be hard to distinguish from debris and show heterogeneous protein expression even within clonal lines. In fact it is the heterogeneity of bacterial growth that has made flow cytometry and cell sorting an increasingly attractive method. [75,76]

Assumptions of the homogeneity of bulk culture can lead to confounding results or miss biologically important, but rare, cell behavior. Flow cytometry is also a useful tool when wanting to characterize properties of individual cells in a mixed culture. [77,78]

However, in more recent decades flow cytometry and FACS has been an important tool for studying diverse aspects of bacterial physiology, including viability antibiotic resistance, metabolism and host-pathogen interactions. [79–82] Many of these current assays, however, rely on multiple rounds of sorting to enrich for the population of interest. For rare populations, cell enrichment may take days to weeks of cell sorting. [83] Applications for flow cytometry among bacteriologists of all stripes grow more and more diverse.

Droplets and Microfluidics

Droplet based microfluidics have a dizzying array of applications to all fields of biology, from biophysics to ecology. Microfluidics devices can encapsulate, incubate, mix, split or merge populations. [84] Encapsulating single cells into droplets and culturing them into microcolonies overcomes or improves several barriers to mutational analysis. Encapsulation in a droplet allows for cells to grow and be assayed in isolation. While competition of clonal cells will still exist, competition is prevented between genetically heterogeneous cells like mutants. Additionally, the isolation prevents exchange of community factors that may mask or confound phenotypes of a specific clonal population. Another advantage is that microcolony analysis typically occurs in smaller volumes than conventional assay, potentially reducing the amount of expensive reagents needed for an assay. [85]

Miniaturized versions of conventional assays are a popular application of microdroplets including immunoassays and pathogen diagnostics. [86–88] Microdroplets are also used for a culture system to isolate mixed populations, culture the cells, and then sort them on a property of interest. [89,90]

The small and insular environment of the droplet makes them ideal for the study of evolution due to the ability to create small sortable subpopulations. [91] Directed evolution experiments have also successfully used droplets for protein applications to select for enzymatic, protein production and secretion. [92–95] For RNA applications, droplet based microfluidics were used in the selection and optimization of RNA fluorogenic biosensors and for screening RNA aptamers. [96,97] Droplet based assays have previously also been used as an alternative to Systematic Evolution of Ligands by Exponential Enrichment (SELEX). [98,99]

Underlying many of these applications is the ability to encapsulate cells in the microdroplets and then sort the droplets using fluorescent activated droplet sorting. One notable barrier for using droplets for flow cytometry is that the sheath fluid in the flow cytometer is aqueous and thus an incompatible phase with the surfactant and carrier oil used to create the droplets. This limitation can be overcome by embedding the cells in a hydrogel to allow the droplets to maintain form while in the aqueous sheath fluid. [100–102] One example, Tn-seq, used an agarose with a low melting temperature, enabling cells to be encapsulated at 37 ˚C, gelled at 4 ˚C and then cultured at 37 ˚C as well. [85] Even if these gelled agar droplets are stripped of oil and placed into an aqueous phase, the droplets hold together. Thus, agar droplets can be run through a flow cytometer and sorted through FADS. Another strategy is to create doubly emulsified droplets where aqueous droplets in

oil are encapsulated a second time in an aqueous phase such that there is an aqueous core surrounded by an oil shell. [103,104]

In Vitro Methods

Much of the knowledge of nucleic acid and protein structure and interaction comes from *in vitro* methods, both now and historically. Such methods analyze specific components of interest by purifying them out of cells (like protein) or synthesizing them (like RNA). Components are then mixed in an artificial salt solution where properties such as binding, enzymatic activity or cleavage can be evaluated. *In vitro* methods allow for a highly controlled environment with few components, making these methods excellent for discerning biochemical interactions on the molecular level. However, the main drawback of any *in vitro* method is this isolation. These methods can determine structure or binding without cellular context. Binders may not be regulators, and *in vitro* there are few ways to truly tell the difference. This is because proteins can bind RNA in a way that does not differentially impact transcription or translation. Advances in *in vitro* techniques are closing this gap, but still can lack necessary biological context.

One commonly used method to determine binding of RNA to protein is the filter binding assay (Figure 1.4). [105] During a filter binding assay, radiolabeled RNA and purified protein are incubated together in a salt buffer to allow for binding. The mixture is then passed over a sandwich of a nitrocellulose membrane, a positively charged nylon membrane and filter paper using a vacuum apparatus. The membranes are then exposed onto a screen to assess radiation. If the protein binds the RNA, the RNA will adhere to the nitrocellulose membrane and if not, the RNA will pass through the nitrocellulose membrane and adhere to the nylon. By creating a gradient of protein concentrations and

measuring the intensity of the nitrocellulose compared to the total intensity, the K_D of the protein to the RNA can be calculated. While some optimization can be needed to find an RNA that folds as expected, this method is quick to perform. However, it requires a large quantity of highly purified protein, which can be time consuming and labor intensive to obtain.

Similar to this method, electrophoretic mobility shift assays (EMSAs) can look for protein binding to RNA or conformational changes in the RNA structure by running the RNA of interest (and any binding partner) on a native acrylamide gel. [106] RNA structure may also be analyzed by a native gel, providing information on structural changes that may occur in the presence of a ligand, by looking for the radiolabeled RNA in the gel to shift higher or lower as it binds a ligand or changes conformation.

SELEX

Systematic Evolution of Ligands by Exponential Enrichment, or SELEX, is an iterative *in vitro* method that can discover ligand binding RNAs (or single stranded DNA). SELEX has been used extensively to select small molecule binding RNA aptamers, create new riboswitches and identify health therapeutics. [107–111] SELEX can be used to evolve RNA binders to small molecules or proteins. [112] In the case of proteins, a large pool of RNAs is first counter-selected against a nitrocellulose filter or affinity column to eliminate promiscuous binders and then incubated with the protein of interest. The filter can then capture RNA-protein complexes after the counter selection step. The RNAs are then eluted off the filter for sequencing or PCR. Then the process can be repeated with the new partially selected pool of RNA under increasingly stringent conditions (for example, lowering the protein or ligand concentration each cycle). The result is recovery of RNAs

that bind strongly to the protein or ligand of interest. In the study of the evolution of RNA regulators, SELEX can be used to help define a "binding sphere" for a protein, identifying the possibilities of interaction and possible regulating structures for the protein. If we return to the examples of L1 and S15 ribosomal proteins, L1 may have a much smaller pool of potential binding and thus regulating structures than S15. S15 binding interactions to protein were previous studied for *Geobacillus kaustophilus* S15 protein, which found that *G. kaustophilus* S15 specifically bound to a variety of RNA sequences. [113] Of course, because SELEX uses purified protein and RNA, RNA binders recovered are not necessarily regulators. Thus the biological activity of the RNA binders needs to be verified after SELEX within cells. Validating RNA binders as regulators is quite time consuming and runs into the same limitations of scale that the traditional in cell assays have.

Methods for a Fitness/Regulatory Landscape

The foundation of much of our knowledge of biology rests in the power of mutagenic analysis. At first these were naturally occurring mutations, like the naturally occurring white-eyed *Drosophila* discovered by Thomas Hunt Morgan. [114] Later came tools to induce mutations and screen the resulting mutants for properties of interest. With the advent of molecular biology, mutations could be targeted very specifically whichempowered biochemists to dissect the mechanisms underpinning biology in their area of interest. All of this is to say: a common approach in biology is that understanding how something is broken is to understand how it works, and to understand how something is changed can be to understand how something was before. This is the power of fitness landscapes.

The principle behind a fitness landscape is to systematically interrogate a biological molecule for changes in function. The development of fast and accessible high throughput sequencing technologies has made generating such landscapes possible; no longer do researchers need to generate mutants a few at a time. [115] Such fitness landscapes are popular to understand the coevolution of pathogens like viruses or antibiotic resistant bacteria with a host as a study of evolutionary fitness, or in protein expression through coding region mutagenesis as molecular fitness. [81,116–118] The bacterial processes of transcription and translation have also been studied via fitness landscape, including polymerase sigma factors, intrinsic terminators, and promoters. [119–121] SELEX is another technique that can create fitness landscapes of nucleic acids. [122,123] RNA molecules provide an especially compelling case for the generation of fitness landscapes due to the diversity of RNA's biological roles.

Like DNA, RNA may have sequence specific interactions; like proteins, RNA may have enzymatic activity or structural based interactions. A fitness landscape can be made of any of these properties for RNA through a mutagenic library. For RNA, fitness landscapes were used to map small RNA regulatory pathways and create functional landscapes for ribozymes. [124,125] Of particular note is a new *in vitro* method, sequencing-based mutational analysis of RNA transcription termination (SMARTT), which was used to test single point mutations in a series of single and tandem glycine riboswitches to determine whether each nucleotide of the riboswitch contributes to riboswitch function. [126,127] This method uses *in vitro* transcription to test ligand-dependent termination of large pools of RNA sequences. However, while a subset of RNA ribosomal leaders like S15 were examined through extensive mutational analysis, no

fitness landscape of all nucleotides has been generated for a ribosomal leader to date. [59,128–130] The goal of this work is to develop an accessible method for the creation of a ribosomal leader landscape, in this case that of the classic S15 autogenous leader.

Figures

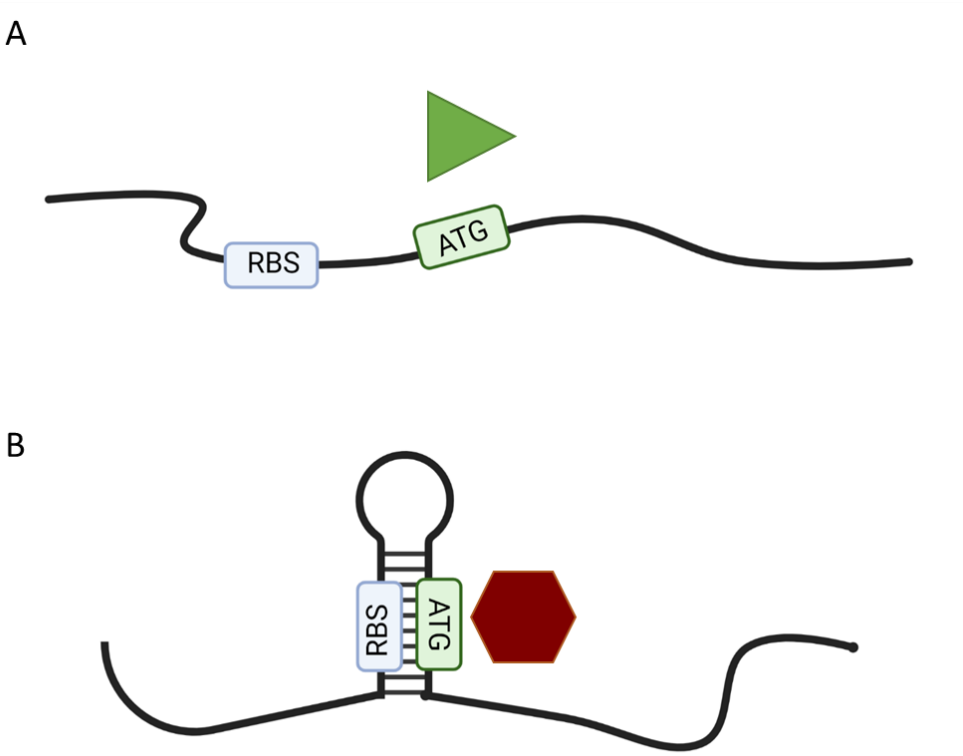


Figure 1.1 ***Translational Regulation through RNA Secondary Structure.*** *For translation to proceed, the ribosome binding site (RBS) and start codon need to be available to the ribosome (A). When the RBS and start codon are sequestered through RNA structural elements, translation is blocked.*

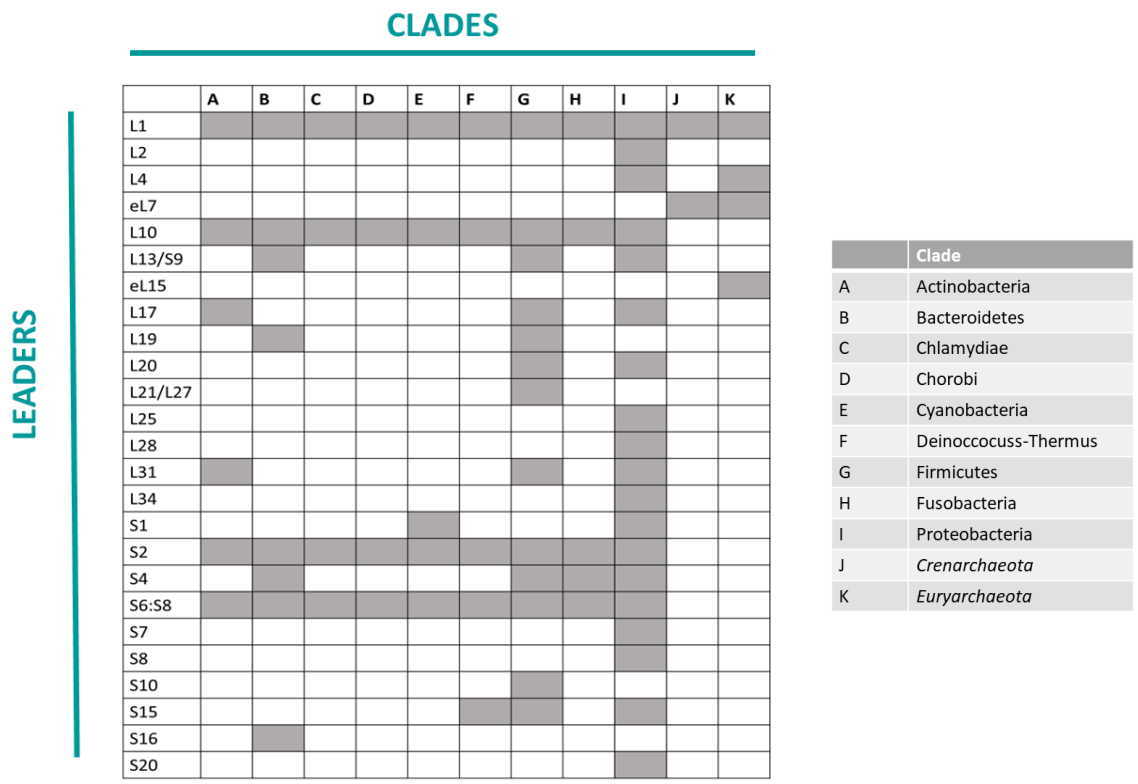


Figure 1.2 ***Distribution of Ribosomal Leaders among Prokaryotes.*** *Grey boxes indicate the presence of a leader detected computationally or experimentally identified. Italicized clades (J,K) indicate Archaea. Other clades are Bacteria.*

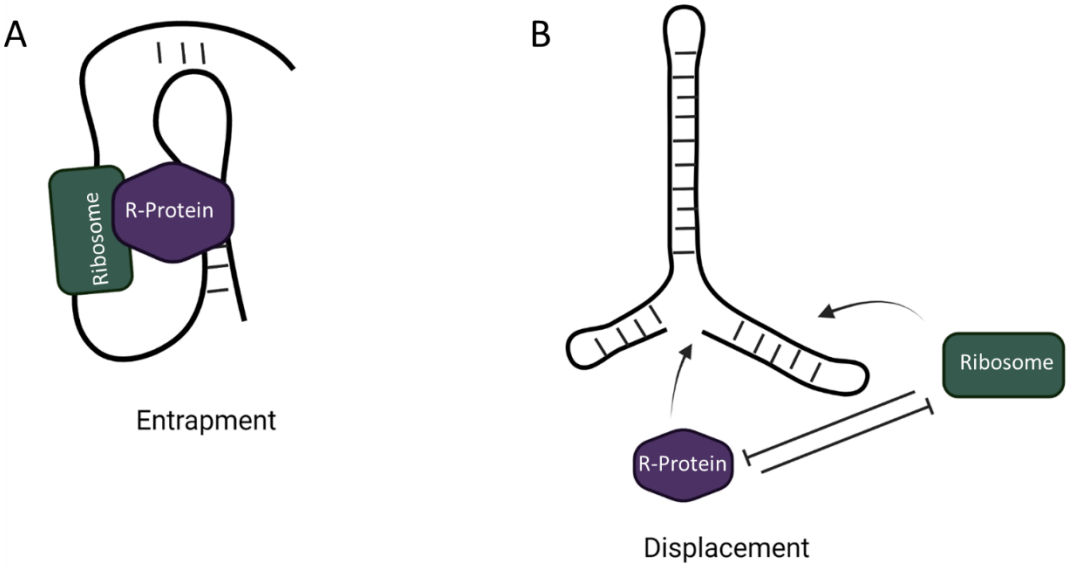


Figure 1.3 ***Mechanism of Ribosomal Leader Regulation.*** *Ribosomal proteins (r-proteins) can trap ribosome components onto the mRNA, preventing translation (A) or compete with the ribosome for binding to the mRNA (B).*

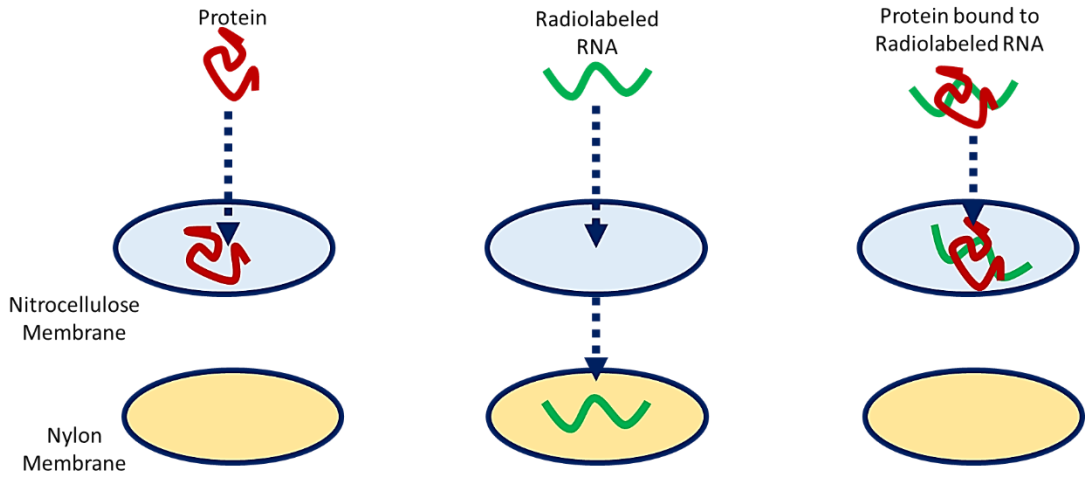


Figure 1.4 ***Filter Binding Assay.*** *In vitro RNA-protein binding can be determined with a filter binding assay. Radiolabeled RNA is incubated with protein and run through a filter apparatus via suction with a nitrocellulose membrane that traps protein and a positively charged nylon membrane that traps free RNA. The radioactivity of the two membranes is then compared to determine fraction of RNA bound to the protein.*

CHAPTER II: FADSRA ASSAY DEVELOPMENT & OPTIMIZATION

Introduction

The classic regulatory assay, the β-galactosidase or Miller assay, remains a popular, robust and sensitive assay nearly fifty years after its initial development. [64] Throughout the decades, the assay has been modestly updated, for example with different reporter proteins or detection methods, such as fluorescent proteins and microplate readers. However, with the current ubiquity and power of high throughput sequencing technologies, the scope of scientific questions asked by researchers has changed and this regulatory assay has not scaled with it. The β-galactosidase assay is frequently performed in 1.5 mL tubes or sometimes in a 96-well plate assay, making the assay low throughput and time intensive. An improved assay would allow for greatly increased throughput, while also retaining the sensitivity and breadth of application. To scale up the assay, the assay could be miniaturized such that instead of a tube or a well of clonal cells, a microdroplet could house a clonal microcolony. This effectively makes each droplet its own assay and allows for thousands of cells to be assayed for regulation at the same time. This chapter details the different aspects of design and optimization for this new assay in development called Fluorescence Activated Droplet Sorting Regulatory Assay (FADSRA).

While scaling up the throughput of the Miller assay has a number of potential applications, there remains a dearth of assays for studies of *in vivo* ribosomal leader regulation. Much research on ribosomal leaders uses either *in vitro* techniques to look at protein-RNA binding, RNA structure, or the β-galactosidase assay to look at *in vivo* function. Unlike riboswitches with small molecule ligands that can be added or taken away from defined media, assaying ribosomal regulators requires balancing the need for functioning ribosomes with the addition or subtraction of the ribosomal protein of interest.

This project endeavors to update a classic assay for use with modern tools in order to answer modern questions. To do so, aspects of assay design were carefully optimized using the *E. coli* S15 protein and the S15 ribosomal leader. The depth of understanding already present for the *E. coli* S15 leader makes for a good way to benchmark the findings of this new method and further improve assay development. For the *E. coli* S15, we know the structure and binding site of both the rRNA and the S15 leader RNA as well as a likely folding pathway for the S15 leader. [131] Furthermore, the natural diversity of S15 leader homologs make for an interesting application of this nascent assay, which is explored in Chapter 3.

Results

Cell Encapsulation and Culturing of Microcolonies

To explore viability of a method using droplet encapsulated microcolonies, colony growth from singularly encapsulated cells was investigated. Initially, XL-1 *E. coli* cells and 1A1 *B. subtilis* cells were encapsulated in molten agarose and LB medium at an OD600 of 0.005. The device used to encapsulate was made from the same design as that in Thibault *et al.* 2019 [85] (Figure 2.1) and produces droplets through a single flow focusing junction. Aperture size at the junction was about 50 µm, resulting in monodispersed droplets of about 70 µm in diameter. Droplets were then gelled to solidify the hydrogel matrix and incubated overnight on a rotator at 37 ˚C to culture the microcolony. For imaging, droplets were then gelled again, stripped of oil and resuspended in PBS for microscopy, as the oil does not have a compatible refractive index for imaging. Microscopy of the droplets revealed that both the *E. coli* and *B. subtilis* were able to grow in the microdroplets with *E. coli* forming somewhat compact ovoid colonies and *B. subtilis* growing into

snowflake-like colonies (Figure 2.2). Observable in the oil were free-swimming bacteria. Free bacteria were not noticeable after washing the unencapsulated droplets twice with PBS. For droplets gelled shortly after encapsulation, large microcolonies resulted after overnight incubation and aeration. However, for droplets with a delay in gelling, there was an increased instance of more than one microcolony for the same encapsulation OD600, suggesting that cells can divide in the molten agar and form multiple microcolonies from one cell (Figure 2.3). Thus, it is important to gel droplets shortly after cell encapsulation so that accurate droplet loading can be ascertained. While not formally measured, the total volume of cells for each droplet seemed consistent, with large single microcolonies and smaller microcolonies in droplets with two or more microcolonies (Figure 2.3). This is likely due to the limited amount of nutrients available in each droplet. Previous work has shown that once a single cell is encapsulated, it will only grow for about 5-8 generations. [85] This has the benefit of providing additional homogeny of microcolony size between cells with different growth rates, in addition to the droplets preventing resource competition.

Reporter Design

Ideally a new droplet based high throughput method of cis-RNA regulatory assay would allow the use of reporter constructs already commonly available. Considering this, β-galactosidase was the first reporter protein explored. β-galactosidase is capable of cleaving certain compounds to produce fluorescence, namely Fluorescein Di-β-D-Galactopyranoside (FDG), which is sequentially hydrolyzed to produce fluorescein, a fluorescent green product and Resorufin β-D-galactopyranoside (RFG), which produces red fluorescence resorufin. In the traditional β-galactosidase assay, cell

walls are permeabilized to allow for the protein and the substrate ONPG to interact as ONPG is not taken up by the cell. Permeabilizing cells growing in a microcolony embedded in hydrogel presents a complication. Additionally, permeabilizing the cells could prevent subsequent outgrowth or DNA recovery steps in the protocol. Thus FDG and RFD were added to the molten agarose media before encapsulation to see if the compounds could be taken up by the growing cells to produce fluorescence. Unfortunately, upon stripping the oil and adding aqueous buffer to the droplets for imaging of flow cytometry, the FDG and RDG were found to diffuse out of the droplets into the aqueous media (Figure 2.4A). Due to the auto-fluorescence of *B. subtilis*, RFD was used instead of FDG. However, the *B. subtilis* cells did not take up the compound well enough for sufficient fluorescent signal (Figure 2.4B). Imagining of *E. coli* microcolonies with FDG in the medium showed fluorescent signal in strains with β-galactosidase (Figure 2.4C).

These β-galactosidase substrates may be useful for microcolonies in double emulsion droplets (an aqueous core with an oil shell, suspended in aqueous phase for flow cytometry). The oil shell could keep the FDG and subsequent products from diffusing out of the droplet, enabling sorting of the droplets while retaining differential signal from the reporter protein. These double emulsion droplets have been previously produced [132] and I was successfully able to create devices that had hydrophilic surfaces instead of hydrophobic surfaces, but did not further pursue double emulsions.

Thus, the reporter was switched to a fluorescent protein and away from β-galactosidase. Fluorescent proteins present some drawbacks as reporter proteins in this context. Compared to β-galactosidase, an enzyme for which we measure the substrate, fluorescent proteins as reporters are less sensitive. Compared to the enzymatic reporter,

fluorescent reporters make for noisier data, and fluorescent proteins require time and oxygen to fold. However, fluorescent proteins have an advantage over the β-galactosidase as fluorescent proteins are produced inside the cell and do not require addition of a substrate or an enzymatic reaction. This means that the cells do not need to be permeabilized for the reporter to fluoresce, addressing concerns of outgrowth, diffusion and ability to permeabilize cells effectively as a microcolony in a hydrogel. Initially, a version of the β-galactosidase reporter plasmid with α-GFP instead of β-galactosidase was tested in the cells. This plasmid contains an RNA-*gfp* fusion driven by the inducible *lac* promoter and a *lacI* to regulate the *lac* promoter (Figure 2.5A). Cells carrying the GFP reporter plasmid contained fluorescent microcolonies within the droplet (Figure 2.6). The original reporter protein was α-GFP, which was previously used for bulk plate assays in the lab. However, α-GFP, compared to newly engineered fluorescent proteins, is dimmer and slow folding. Switching to a new reporter, superfolder-GFP (sfGFP), increased the fluorescent signal, had excitation and emission spectra that better fit the laser-filter sets on the flow cytometer and did not require an additional incubation period after induction to account for protein folding. [133]

Additionally, cells were co-transfected with an L-arabinose inducible protein overexpression plasmid containing the ribosomal protein of interest. This inducible plasmid allows expression of the leader binding partner to be regulated as needed for the assay, either adding additional protein to bind to the leader or, in a knock-out strain, preventing protein expression all together. An additional fluorescent protein, a DsRed expressed constitutively, was added to the reporter plasmid to improve ability to gate and check droplet loading during cell sorting (Figure 2.5B). With a constitutive fluorescent

protein, droplets that have microcolonies can be separated from empty droplets, allowing collection of "off" droplets. Another benefit of the second fluorescent protein is that droplets that have more than one colony can be partially gated out by excluding droplets that have very high red fluorescence.

For cloning of the RNA, the reporter plasmid was designed with an additional restriction site between the two cloning sites so that any plasmid without insert that was still circularized would be linearized prior to transformation, helping to ensure that the plasmids transformed contained the library (Figure 2.7). Another method that was considered was to have a large insert that would allow purification of the digested backbone from both the large insert and undigested plasmid via gel electrophoresis. However, the plasmid version with the internal restriction site was sufficient to clone in the S15 *E. coli* RNA library.

Adapting a similar reporter system to a gram-positive organism like *Bacillus* would require multiple changes to this system. The main factor is that *Bacillus*, unlike *E. coli*, stably integrates linear DNA into its genome using homologous recombination during transformation instead of maintaining plasmids. A single copy of the reporter integrated into the genome means that for adequate fluorescence, the reporter would be bright, and the promoter would need to be strong. Many fluorescent proteins are poorly expressed by *Bacillus*, but through a generous gift from the Church lab, constructs with an RNA-mNeon fusion were designed to adapt FADSRA to *Bacillus*. However, due to the COVID-19 pandemic, this project was tabled.

Strain Optimization

To develop the assay, the *E. coli* S15 leader was chosen since this leader shows 15-fold repression in a β-galactosidase assay and a number of well characterized mutant sequences were available in lab, specifically the regulating wild type sequence and M1, a non-binding, non-regulating mutant that disrupts the RNA secondary structure (Figure 2.8 [63]). Comparing the fluorescence of wild type and mutant S15 leader-GFP fusions in XL-1 cells in the presence of L-arabinose (inducing the overexpression of S15) with flow cytometry; a modest two-fold change in fluorescence was observed. This does not compare favorably to the 15-fold difference seen with the β-galactosidase assay.

Since the difference in repression was likely due to the impact of native S15 protein binding to the reporter fusion, future development would ideally be in a strain without the S15 protein. A K12 Δ*rpsO* strain was transformed with the S15 overexpression plasmid and the reporter plasmid. While these cells, with and without additional plasmids, are viable, S15 plays an important role in ribosome assembly as a primary binding partner of the 16S RNA. Thus, the cells grow very slowly and are poor at making any protein due to the dearth of functional ribosomes. This is potentially problematic for a system dependent on the synthesis of a reporter protein. S15 knockout cells with no additional plasmids did grow well enough into microcolonies within the droplets, however, problems arose when attempting to use the strains in the regulatory assay as the burden of carrying the reporter and protein overexpression plasmid caused inconsistencies in strain viability in droplets and liquid culture.

DsRed Promoter Optimization

The constitutive expression of the DsRed greatly affected cell viability, with the original promoter, *rrnB*, causing cultures to grow more slowly than the parent strain or even "crash," where cell cultures would grow for a few hours but die overnight. Within the droplets, some cells were able to grow into microcolonies of the smaller end of the expected size but other droplets were seen with what appeared to be very small microcolonies or single cells (Figure 2.9). Additionally, *rrnB* codes for the 16S rRNA, the primary binding partner of S15, which could complicate use of the *rrnB* promoter in the reporter plasmid. For these reasons, weaker promoters including Anderson promoters 105 and 106 [134,135] were cloned into the plasmid to drive DsRed expression and growth of cells carrying the plasmid was tested through growth curve assay. The results show that Anderson promoter 105 provided good expression of the DsRed with a much smaller impact on growth rate (Figure 2.10, Figure 2.11). The other promoters tested also showed sufficient DsRed expression and better growth rates; however, cells carrying the Anderson 105 promoter plasmid showed the most robust growth. Another solution would be to have an inducible promoter as with the RNA-GFP fusion. Attempts were made to clone the *lac* promoter for the DsRed, but were not successful so experiments proceeded with Anderson promoter 105.

With the new DsRed promoter, the cells showed good expression of the DsRed protein as measured by flow cytometry of encapsulated microcolonies and bulk culture plate assay. However, upon IPTG induction of the RNA-*gfp* fusion, DsRed signal normalized to the OD600 diminished over time as measured through plate assay (Figure

2.12). While the original expectation was to examine droplets with both red and green fluorescent signal with no droplets expected to be green alone, after overnight growth and one hour of IPTG induction, many droplets were found to be only green. Microscopy of red and green droplets found a mixture of droplets with more than one microcolony or large single colonies expressing both red and green as measured by flow cytometry (Figure 2.13). After droplet sorting, microscopy of ten droplets per sorted population found that loaded droplets only containing one microcolony were found in GFP expressing populations, but half of the droplets with green and red fluorescence had more than one microcolony. This indicates that the original population of interest, the droplets with both red and green fluorescence, contains a higher number of confounding droplets with multiple microcolonies that may indicate GFP expression, but instead may be coexpressed. Additionally, the drop in DsRed fluorescence indicates that expression of the mature fluorescent protein is not truly constitutive. Literature detailing the regulation of the Anderson promoter is sparse, largely focused on testing expression of the promoters in different organisms since this promoter collection was originally discovered in *E. coli*. It is possible that the expression of the Anderson promoters is variable based on a yet unknown stimulus or that the *lac* promoter is preferentially transcribed over the Anderson promoters studied here.

Reporter Protein Expression

Once cell viability was improved, a larger concern was revealed: the S15 knockout cells did not express the GFP reporter as expected within the droplets. It was expected that for the wild-type leader, induction of S15 overexpression by adding L-arabinose would cause a decrease in GFP fluorescence due to the repressive binding of the S15 protein to

the S15 leader. However, comparing GFP fluorescence between droplets with and without L-arabinose (expressing S15 protein), the microcolonies with L-arabinose were always brighter (with S15 protein induction) than those without L-arabinose (Figure 2.14). This is likely because ribosome assembly greatly improves with S15 protein, even if it is overexpressed, leading to cells with overexpressed S15 able to synthesize protein far more effectively. So even if there was repression, the repression was less significant a factor than the ability to synthesize the reporter protein. Additionally, the K12 Δ*rpsO* control grew larger than the microcolonies transformed with plasmid, suggesting that the presence of the plasmids introduces a significant metabolic burden for the K12 Δ*rpsO* cells without S15 protein.

Initially, efforts were made to remedy the difference in protein synthesis by enriching the media, switching from LB to SOB and adding glucose in place of the L-arabinose for the uninduced condition. However, the difference in protein synthesis between conditions with and without S15 protein induction could not be remedied in the KO strain.

Focus then turned to determination of a better strain for this assay. MG1655 was chosen as a fast growing, commonly available and less domesticated option, and compared to cloning strains XL-1 and DH5α. In growth curve experiments and flow cytometry of droplets, MG1655 showed both the highest expression of GFP and the biggest difference between the wild type and mutant leader (Figure 2.15). MG1655 also grew faster than K12 Δ*rpsO* with S15 induction (Figure 2.16).

However, the native S15 protein expressed in MG1655 alone is sufficient to repress the wild type S15 leader. This is problematic because ideally, to screen an RNA library for

regulation, we would like to differentiate between true regulators and RNAs that are always on or always off. This problem might be somewhat unique to ribosomal leaders, since the regulating proteins are involved directly in protein translation. In order to distinguish these true regulators, we decided to use two strains. The RNA library initially screened for fluorescence in MG1655 will be able to distinguish always on RNAs from regulating and always off RNAs. Transforming the K12 Δ*rpsO* strain without any S15 protein would then be able to distinguish between true regulators and always off RNAs (Figure 2.17). Flow cytometry on droplets containing K12 Δ*rpsO* with either wild type S15 leader or leaderless plasmid (pEG10) revealed that the knockout strain can make enough of the reporter protein to distinguish it from microcolonies not expressing GFP and shows that the reporter GFP is repressed when the S15 protein is overexpressed. This repression is relieved when the leader structure is disrupted (Figure 2.18). However, the amount of protein produced is less, making differences between leaders less apparent. Instead of using two strains, one approach would be to use the K12 Δ*rpsO* strain with and without L-arabinose induced S15 expression, but the robust protein production of MG1655 over K12 Δ*rpsO* even with protein expression made growth of the strain much more reliable and easier to transform with the plasmids. Based on these properties, assay development proceeded with using MG1655 as the primary strain in the assay and with K12 Δ*rpsO* as the strain for conditions without protein expression.

Induction Optimization

GFP induction was an important parameter to optimize since the difference in fluorescence between repressive and permissive regulators is ideally maximized, but also visible to the detector. Expressing a reporter for too short a time may not result in detection

of the entire microcolony, or not allow effective sorting between microcolonies with a moderate difference in expression. Expressing a reporter for too long may result in the accumulation of fluorescent protein and diminish expression differences between microcolonies. Furthermore, when studying ribosomal protein regulation, there is added stress on the cell when overexpressing so many proteins. A long period of induction may cause the cells to crash and die in the absence of enough functional ribosomes to synthesize protein not only for the assay proteins, but also for cell division and basic cell functions.

To find the optimal reporter induction time, cells were encapsulated and cultured as previous. Droplets for the overnight condition had 1mM IPTG added to the media at encapsulation and all other induction times had the droplets grown overnight without IPTG and were stripped of oil the next morning, resuspended in ITPG containing medium and incubated for the indicated time. The results show most strikingly that despite the presence of a *lacI* in the plasmid the expression of sfGFP is quite leaky, resulting in significant GFP expression even without any addition of IPTG (Figure 2.19). A comparison with the α-GFP shows generally poor expression of the reporter (Figure 2.20). Ideally, for the reasons noted in the preceding paragraph, reporter expression would be more tightly controlled. Another common inducible and notably more tightly regulated promoter, the L-arabinose inducible promoter, is already used in this assay to control the expression of the ribosomal protein. Despite the issues with leakiness, IPTG induction does increase reporter expression, with the most difference seen between WT and M1 within the first two hours. One important point was that droplet integrity appeared to be compromised with the longer incubation times, resulting in a shift of the forward scatter profile of the droplets (Figure 2.21). Thus,

induction was kept at two hours or less for future experiments to balance differences in reporter expression with droplet integrity.

Droplet Gating

For droplet gating during FADS, the position of the droplets on the forward and side scatter plots was determined by looking for red fluorescence, indicating the presence of a microcolony. Droplets with different properties, including different side and forward scatter and different fluorescence, were sorted out for examination underneath a microscope. Droplets were then inspected for the number of encapsulated colonies per droplet, droplet size and the presence of debris (Figure 2.22, 2.23). The initial gate was set by selecting the population on the forward and side scatter containing fluorescence compared to the dark control and of a size larger than debris, which likely also contained free cells. The two main concerns with droplet gating were where to set the green gate to separate repressive leaders from permissive leaders and the number of droplets. Droplet occupancy was measured by sorting different gates based on fluorescence or side scatter and then counting the number of microcolonies per droplet (Figure 2.24). Gates were then set to minimize the number of droplets with more than one microcolony and according to non-fluorescent and wild type leader controls.

Droplet Handling

Regarding droplet handling, initial efforts to analyze and sort the microdroplets were complicated by inconsistencies resulting in machine clogs, uneven suspension of droplets, distortion of droplet shape under pressure and aggregation. Experience found that increasing the initial gelling time from ten minutes to thirty minutes, as well as gelling

droplets again before stripping them of oil helped stabilize the shape of the droplets in the flow cytometer sample stream. Additionally, resuspending the droplets immediately before flow cytometry or FADS by pipetting was the most reliable method for evenly resuspending the droplets. Vortexing or tube flicking resulted in an increased proportion of samples that were not evenly suspended, resulting in no flow cytometry events or clogs. This could be because vortexing can create droplet aggregates by subjecting the droplets to great force, and because both vortexing and tube flicking can cause droplets to stick to the sides of the tube. The concentration of droplets also had a large impact on the quality of the flow cytometry. Through trial and error, I determined that thirty minutes of droplet production was the maximum I could prepare for one flow cytometry sample. Additional droplets caused the flow cytometer to be unable to take up any droplets from the sample. Consequently, for any applications requiring more droplets multiple half hour runs were pooled after flow cytometer sample loading.

Fluorescent Activated Droplet Sorting

Taking all the previous optimization together, the focus shifted to evaluating how effective the system was at separating leaders with different regulatory properties. The plasmid pEG9-S15M1 was modified to include a kanamycin resistance cassette to allow for differentiation between WT and M1 S15 leaders by selective plating after FADS. The results show that after one hour of IPTG induction, the proportion of M1 leader in the red only gate is 0.14 and the proportion in the double positive gate is 0.9 (Figure 2.25). This indicates that the parameters used here for FADSRA can separate RNA leaders with different regulatory functions.

DNA Extraction & PCR Amplification

In order to sequence the regulatory sequence on the reporter plasmid, DNA must be extracted for amplicon amplification or subsequent transformation into a different strain. As noted previously by Thibault *et al*., the amount of gDNA retrieved from pooled samples is in the tens of nanograms range using a DNeasy kit, and recovery of plasmid DNA would presumably be lower still. [85] Thus, I tried several different methods to try to optimize plasmid DNA extraction. Some considerations I had when picking methods to try were 1) I needed to melt the agarose in order to release the cells 2) I needed to lyse the cells to release the DNA 3) I needed the DNA in a low volume for the PCR. Importantly, the melting temperature of the agarose is 65 ˚C while *E. coli* cells are known to begin lysing around 60 ˚C. [136] So any method requiring the cells separate from the agarose would ideally minimize the time at 65 ˚C to preserve cell integrity so that the cells can be lysed after the removal of the agarose. Alternatively, a method could dilute out the agarose enough to prevent gelling during subsequent DNA purification or prevent loss of DNA from any prematurely lysed cells. Other researchers have successfully used alginate instead of agarose to create a hydrogel matrix for microdroplets [137]; however, the alginate matrix is weaker than the agarose matrix and so agarose droplets are preferable for FADS.

For each method I started with 100 µL of microcolony-containing droplets suspended in PBS for a total volume of 1mL. For the first method, the droplets were centrifuged and the supernatant was removed. Three times the volume of the droplets of agarose dissolving buffer (ADB) from a Zymo purification kit was added. While the recipe is proprietary, the Zymo ADB contains sodium iodide, a chaotropic salt that melts the agarose matrix. Other chaotropic salts, like guanidinium chloride, are used in DNA

purification kits and are used to lyse cells. So by adding ADB to the pelleted droplets, I hoped to melt the agarose and lyse the cells in a single step. The melted samples were then run over silica columns to bind the DNA and washed with ethanol to remove the extra salt.

For the other methods, the droplets were heated for 10 minutes at 65 ˚C to melt the agarose to release the cells from the droplets. This appears long enough to melt the droplets, diluting out the agarose, while minimizing cell lysis. [136] The cells were then pelleted via centrifugation. For the second method, I simply removed as much of the supernatant as I could without disturbing the cell pellet, resuspended the cells in Q5 master mix, and boiled the sample at 95 ˚C for fifteen minutes to lyse the cells. For the third method, I removed the supernatant and then lysed the cells via alkaline lysis followed by silica column purification. For all DNA recovery methods, the amplicon was then generated via PCR, gel purified and the yield quantitated (Figure 2.26).

For the first trial, the alkaline lysis method had a yield almost ten times greater than that of the sodium iodide, and twenty times that of the boiling only method as measured by the DNA concentration of the excised and purified bands. For the second trial, the alkaline lysis and sodium iodide methods resulted in comparable yield, which was greater than that of the boiling method alone (Figure 2.26). Based on these results, the alkaline lysis method was chosen to extract DNA.

Another step that needed to be optimized for the creation of amplicon for Illumina sequencing was the clean up between the two rounds of PCR. To create amplicon for sequencing, a two-step PCR method is used. The first PCR is five cycles that attaches the unique molecular identifier (UMI) as well as the adapters for the barcode primers. [138] The second PCR of 30 cycles attaches the barcodes and Illumina sequencing adapters and

amplifies the target DNA for sequencing. In between these cycles, the first set of primers needs to be eliminated. The two methods for clean-up quantitated were an enzymatic product, ExoSAP-IT and column DNA purification. The results showed that column purification resulted in more amplicon produced during the second round PCR as well as fewer non-specific bands. While additional methods of both DNA purification and PCR clean-up exist, the combination of sodium iodide lysis, and column purification for template DNA for both the first and second PCRs was chosen for the final protocol.

Comparisons to Flow Cytometry

Flow cytometry has already been used for free swimming bacteria for decades, so it is important to establish what advantages droplets have over free cells for flow cytometry applications. Clones with unknown mutant S15 leaders were shown to have different GFP expression using a fluorescent plate assay were selected for flow cytometry analysis (Figure 3.10). Clones were selected to have a range of expression, from very low expression to expression higher than the S15 M1 mutant. Cells were encapsulated in droplets and cultured overnight into microcolonies. Cells from the same cultures were also grown overnight in bulk liquid culture. Wild type and M1 leader were mixed together in a 50:50 ratio after being cultured separately in droplets or in liquid culture. Samples were sorted via FACS, then checked for purity and flow profiles compared.

For purity, the results show that there is good separation between wild type and M1 leader, with comparable sorting purity between both the bacteria and droplets at above 90% for each for the green channel. Purity was not calculated for the red channel due to the difficulty of recovering enough droplets from the previous sort while keeping the droplets intact and concentrated. However, looking at the flow cytometry plots reveals that the

droplets have a distinct advantage over liquid culture. While leaders with a high degree of differential expression (WT and M1) have comparable separation between droplets and liquid culture, subtler differences in expression result in significant population overlap in free cells and distinct separation in the droplets (WT and M2) (Figure 2.27). Interestingly, the DsRed was not significantly detected in the free cells but was in the droplets, perhaps indicating a level of expression below the limit of detection in the free cells. This is good evidence of the benefit of microcolony culture: signal amplification from the increased number of clonal cells in the microcolony results in increased detectable signal, allowing better differentiation.

Discussion

While this new assay shows promise, several improvements could be made. Firstly, the design of the reporter stem could benefit from a few specific changes as the current design presents a high metabolic burden on the cell. Reporter cell lines carry two plasmids, one for overexpression of a ribosomal protein and another that has one constitutive fluorescent protein and another with an inducible protein. Each plasmid then contains a drug resistant cassette. The copy number for each plasmid is medium copy with 20 to 70 copies per cell. To address this, recombineering or linear DNA transformation (as appropriate for the particular species) could be used to integrate the reporter construct in single copy into the genome, as was the original plan for *B. subtilis* FADSRA. However, some sensitivity may be lost due to the overall reduction in fluorescence.

One observation necessary to address is the phenomenon where most of the encapsulated microcolonies show expression of only one of the fluorophores. GFP is known in *E. coli* to be less toxic than RFP, so it is possible that GFP is preferentially

expressed over the RFP when GFP expression is permissible. [139] Additionally, the *lac* promoter may be favored over the Anderson 105 promoter, resulting in a decrease in DsRed and an increase in GFP. This could be addressed by changing both promoters to be the inducible *lac* promoter, reducing toxicity for the cell by decreasing the amount of red protein synthesized. Notably, RFP was shown to be more toxic in cells than GFP, and DsRed may be more toxic still. This would also decrease the metabolic burden on the cell. While I have found the red fluorescent protein useful for observing droplet loading and also for separating droplets with microcolonies from empty droplets or droplets with cells that did not grow into microcolonies, there may be a case that having a second fluorophore is not necessary or may even inhibit the assay's function.

The next chapter, Chapter 3, uses this new assay for a specific research question, while Chapter 4 explores proof of concept applications of this assay.

Materials and Methods

Microfluidics Device Fabrication

The master for microfluidics device fabrication was provided courtesy of the van Opijnen lab. [85] The master was made by coating a silicon wafer with photoresist. The wafer was selectively etched through UV light exposure using a CAD-designed photomask. The wafer was then placed in the bottom of a plastic petri dish and the chip was cast with PDMS. The PDMS was cured overnight at 65 ˚C and the chip was excised from the dish. Ports were created with a 0.75 mm biopsy punch and cleaned with isopropanol. The PDMS chip was bound to a glass slide after 1 minute of exposure to oxygen plasma in a faraday cage at 400 sccm flow and 400 watts and subsequent baking at

65 ˚C for ten minutes. Ports were primed with Novec 7500 oil. For hydrophilic devices, the ports were flushed with 1% polyvinyl alcohol for ten minutes, flushed with nitrogen gas and then baked as the hydrophobic devices. Devices were stored at room temperature with frosted tape protecting the ports.

Strains

The K12 Δ*rpsO* was obtained previously from Gloria Culver [48,63] and the K12 Δ*rplA* strain from the Keio collection. [140] The MG1655 strain was obtained from the Babak Momeni lab. XL-1 cells were purchased from NEB. *E. coli* strains were cultured in either Lysogeny Broth or Super Optimal Broth. Antibiotics and carbon sources, when present, are in the following concentrations unless explicitly stated elsewhere: 100 μg/mL ampicillin, 34 μg/mL chloramphenicol, 15 mL L-L-arabinose, 15 mM glucose, 1 mM RDG, 1 mM FDG.

The *B. subtilis* strain used to test β-galactosidase compounds was a gly-RS *lacZ* reporter strain used by Babina *et al.* 2017 for β-galactosidase assays of the glycine riboswitch. [23] *B. subtilis* cultures were grown in 2XTY for both liquid culture and droplets.

Plasmid Design

Reporter plasmid variants were all cloned using pBS3 (*lacz* reporter) or pBS4 (GFP) as a vector. [113] RNA leaders L1 and L1M1 (used in Figure 2.6) were amplified via PCR using XL-1 gDNA template and then cloned into the vector using the EcoRI and SalI restriction sites. To swap the promoter, Anderson promoter oligos were annealed by boiling and slow cooling in duplex buffer then cloned into pEG4 digested with SphI and

KpnI. Ligations were performed using NEB Quick Ligase following the standard NEB protocol and initially transformed into XL-1 cells before plasmid purification and transformation into the strain of choice. For the insertion of the DsRed cassette, there were no compatible restriction sites so the cassette was inserted via Gibson assembly using the NEBuilder HiFi Assembly from DNA according to standard protocol. For pEG9k-S15M1, the kanamycin cassette was amplified from and cloned into pEG9-S15M1 using the NotI site. The protein overexpression plasmid pBAD33-*rpsO* was cloned previously. [63] To make pBAD33-*rplA*, the *rplA* gene from XL-1 gDNA was amplified via PCR and cloned using SacI and XbaI. To switch from αGFP to sfGFP, sfGFP was amplified from a topoclone plasmid of the confirmed sequence and cloned into pBS4 using the SalI and NsiI sites (original source of the sfGFP: [141]). To create pEG10, pEG9 was amplified and then blunt ligated. pEG10 adds an internal BlpI site flanked by SalI and EcoRI.

Encapsulation and Microcolony Culture

To encapsulate cells into agarose microdroplets, cell cultures were grown overnight in SOB +ampicillin +chloramphenicol +L-arabinose or glucose and 15 µL of overnight culture was used to inoculate 2 mL of SOB +ampicillin +chloramphenicol +L-arabinose or glucose. Cultures were grown to mid-log (OD600 of 0.4-0.8), diluted into 1% SeaPlaque agarose in SOB +ampicillin +chloramphenicol +L-arabinose or glucose carbon source. The 1% SeaPlaque was melted at 65 ˚C, syringe filter sterilized and then the temperature was reduced to 37 ˚C prior to adding cells. The molten agarose-cell mixture was loaded into a 1 mL syringe with 271/2 gauge needle and attached to the aqueous port of the microfluidics device with PE/2 tubing. 1% Picosurf I in Novec 7500 oil was similarly loaded and attached to the oil inlet of the device. Encapsulation proceeded for 30 minutes with the oil pump at

600 µL/hour and the aqueous pump at 400 µL/hour. The resulting droplets were then gelled at 4 ˚C with rotation for 30 minutes, then incubated overnight at 37 ˚C with gentle rotation. For reporter induction, agarose droplets were stripped of oil using 20% PFO and resuspended in SOB +ampicillin +chloramphenicol +L-arabinose or glucose +1mM IPTG and incubated with rotation at 37 ˚C overnight. Strains with αGFP were further incubated at 4 ˚C for an additional day to allow for extra time for the GFP to fold.

Flow Cytometry

For flow cytometry, droplets were incubated with rotation at 4 ˚C for ten minutes to solidify the agarose and washed two times into 300 µL PBS. Flow cytometry and droplet sorting was performed on a BD FACSAria III cell sorter. BD FACS Diva 8.0 software was used for instrument operation and data analysis. The FACSAria III was run by Boston College Flow Core Director Dr. Patrick Autissier. Gates were set using media only control droplet and single fluorophore encapsulated microcolonies.

Microscopy

For visualization, droplet samples were suspended in PBS and mounted on a specialized slide with coverslip spacers between the bottom slide and top coverslip and sealed with lubricating rotor grease. Droplets were imaged on an EVOS microscope and micrographs were analyzed and edited in FIJI image processing software.

Fluorescence Plate Assay

Cultures were grown overnight in 0.5 mL LB +chloramphenicol +ampicillin +L-arabinose or glucose in a 24 well plate. The following day, the cultures were diluted out 1:1000 into the same media and grown for five hours. After five hours, cultures were

induced with 1 mM IPTG for 1 hour. Cultures were then washed into PBS by centrifuging the cultures down, removing the supernatant and resuspending in PBS. 200 µL in triplicate for each sample was added to a 96 well black optical plate. The plate was read by a plate reader and the OD600, DsRed (ex:558 em:583) and αGFP (ex:397 em:506) or sfGFP (ex:485 em:510) was measured.

Growth Assay

Cultures were grown overnight in LB +chloramphenicol +ampicillin +L-arabinose or glucose. Cultures were diluted to an OD600 0.01 in the same media the following morning then grown with shaking at 37 ˚C. OD600 readings were then taken every 30 minutes using a spectrophotometer.

Primers & Oligos

Anderson Promoters	1895-pJ23105-F	C tttacggctagctcagtcctaggtactatgctagc GGTAC
	1986-pJ23105-R	C gctagcatagtacctaggactgagctagccgtaaa GCATG
	1897-pJ23106-F	C tttacggctagctcagtcctaggtatagtgctagc GGTAC
	1898-pJ23106-R	C gctagcactatacctaggactgagctagccgtaaa GCATG
	1899-pJ23119-F	C Ttgacagctagctcagtcctaggtataatgctagc GGTAC
	1900-pJ23119-R	C gctagcattatacctaggactgagctagctgtcaA GCATG
	1895-pJ23105-F	C tttacggctagctcagtcctaggtactatgctagc GGTAC
	1986-pJ23105-R	C gctagcatagtacctaggactgagctagccgtaaa GCATG
sfGFP	1961-Sfgfp-Sal-F	CAG GTCGAC TCAAAAGGAGAAGAGCTGTTCAC
	1962-Sfgfp-Nsi-R	CAG ATGCAT TTACTTATAAAGCTCATCCATGCCG
DsRed Cassette	1827_pBS4-Gib-F	AGGCCCTTTCGTCTTCAC
	1828_pBS4-Gib-R	CGTGATACGCCTATTTTTATAGG
	1829_DsRed-Gib-F	ataaaaataggcgtatcacggcatgcGGGGATCCCGGAGTTCATG
	1830_DsRed-Gib-8	aggtgaagacgaaagggcctgcggccgcAAGGCCCAGTCTTTCGAC
Sequencing Primers	1903-dsRed seq-R	gcttcacgtacaccttggag
	GH13-707-GFP-R	GTAAGTTTTCCGTATGTTGCATCACC
	1982-sfGFP seq-R	GAG AAT TTG TGT CCG TTT AC
pEG10	1997-Eco-pEG9-F	GCTGAGC GAATTC TCAAAAGGAGAAGAGCTGTTC
	1998-Sal-pEG9-R	GTCGAC AAATCGTTTTACGGGCAAGG
Kan Cassette	1957-Not-Kan-F	CAG GCGGCCGC CCGGAATTGCCAGCTG
	1958-Not-Kan-R	CAG GCGGCCGC TCAGAAGAACTCGTCAAGAA
rplA	1763_RBS-Ec-rplA-SacI-F	GACGAGCTCAGGAGGTTTTAAAATGGCTAAGAAAGTACAAGCC
	1764_Ec-rplA-XbaI-R	GACTCTAGATTAGTCCTCCACTACCAGG

L1 Leader-fusion	1765_Ec-L1-EcoRI-F	GACGAATTCGCAAATCGTTTTATTCAAGACGAGGGCCTTGCCCGTAAAAC
	1766_Ec-L1-SalI-R	GACGTCGACGACATAGGCTTGTACTTTCTTA
	1767_Ec-L1-mut1-F	TATCACGGGAAACCTCTCAG
	1768_Ec-L1-mut1-R	CTGAGAGGTTTCCCGTGATA

Figures

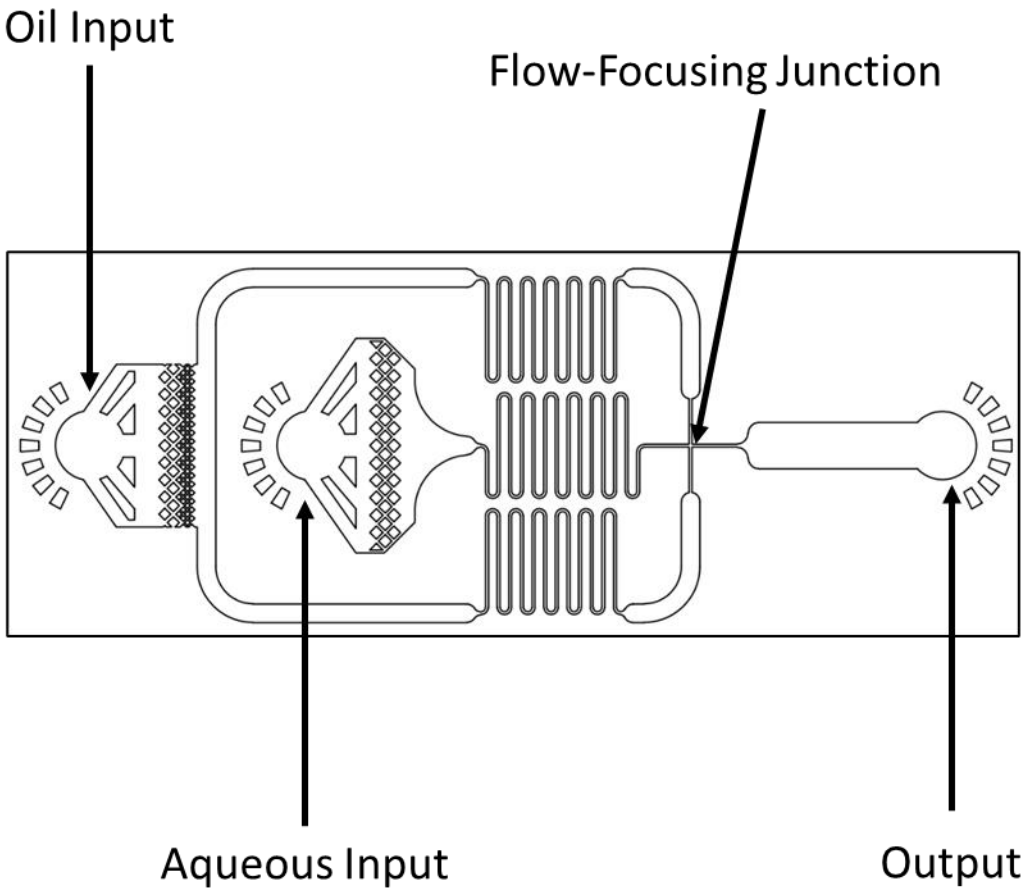


Figure 2.1. ***Schematic of Microfluidics Device for Agarose Droplets****. Growth medium with 1% molten agarose containing cells is flowed through the aqueous input. Droplets are created at the flow focusing junction where the aqueous stream is pinched on both sides by the fluorinated oil containing surfactant. Droplets then exit out through a tube in the outlet port.*

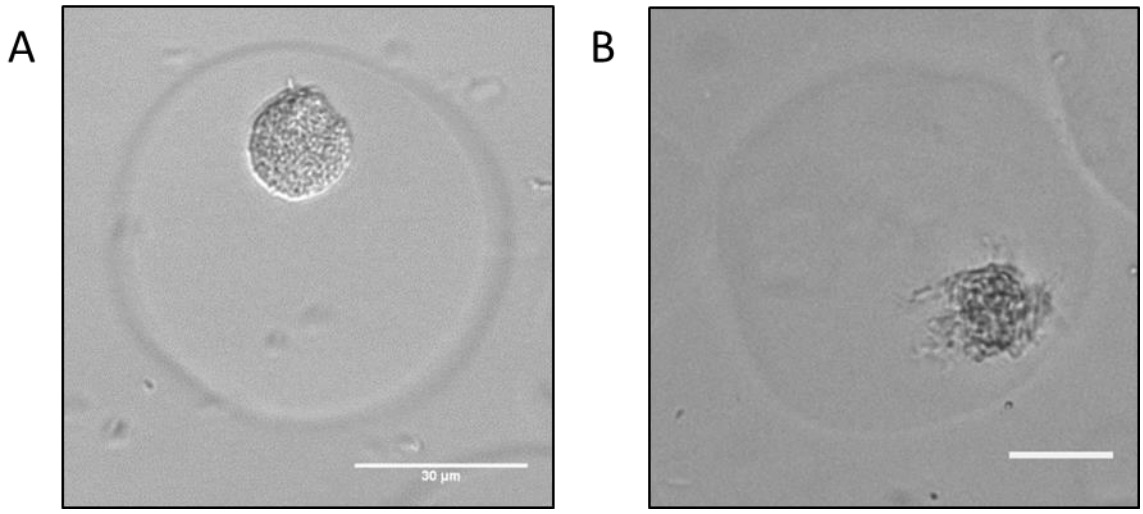


Figure 2.2. ***Microcolony Culture.*** *Microcolonies of E. coli (A) and B. subtilis (B) after 18 hours of incubation in 1% agarose LB (A) or 2XTY (B) agarose hydrogel droplets. Scale bar is 30 µm (A) and 15 µm (B).*

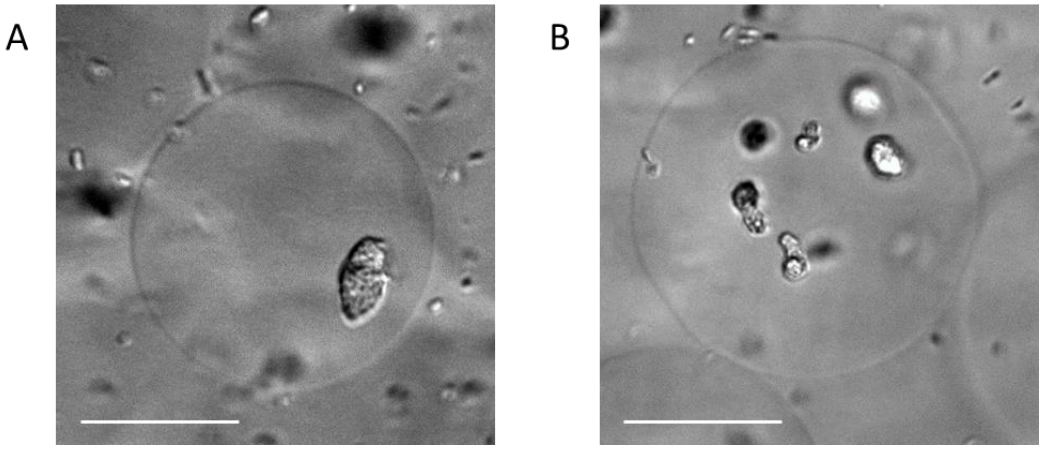


Figure 2.3 ***Microcolony Growth Limitations in Agarose Droplets****. E. coli K12 ΔrpsO cells cultured in 1% agarose SOB +chl +amp microdroplets overnight at 37 °C from two parallel encapsulations (same parent culture and same encapsulation OD600). Droplets from A and B were gelled at different times post encapsulation. A) A droplet with a single microcolony of a large size compared to B) a droplet with multiple very small microcolonies.*

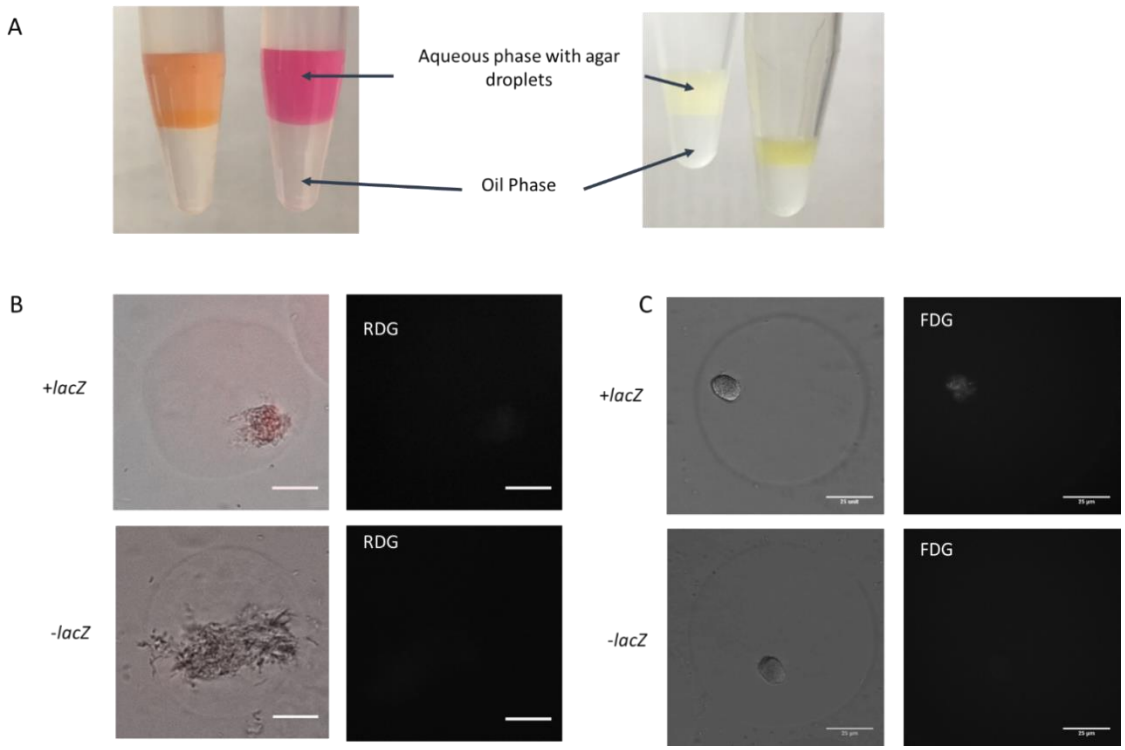


Figure 2.4 ***Fluorescent β-galactosidase Substrates.*** *A) Test tubes showing diffusion of RDG and FDG and their cleavage products into the aqueous phase during agar droplet washing in preparation for flow cytometry. B) B. subtilis cells and C) E. coli cells with and without lacZ were encapsulated with β-galactosidase substrates RFG or FDG (B), which fluoresce upon cleavage. Scale bar shows 25 μm.*

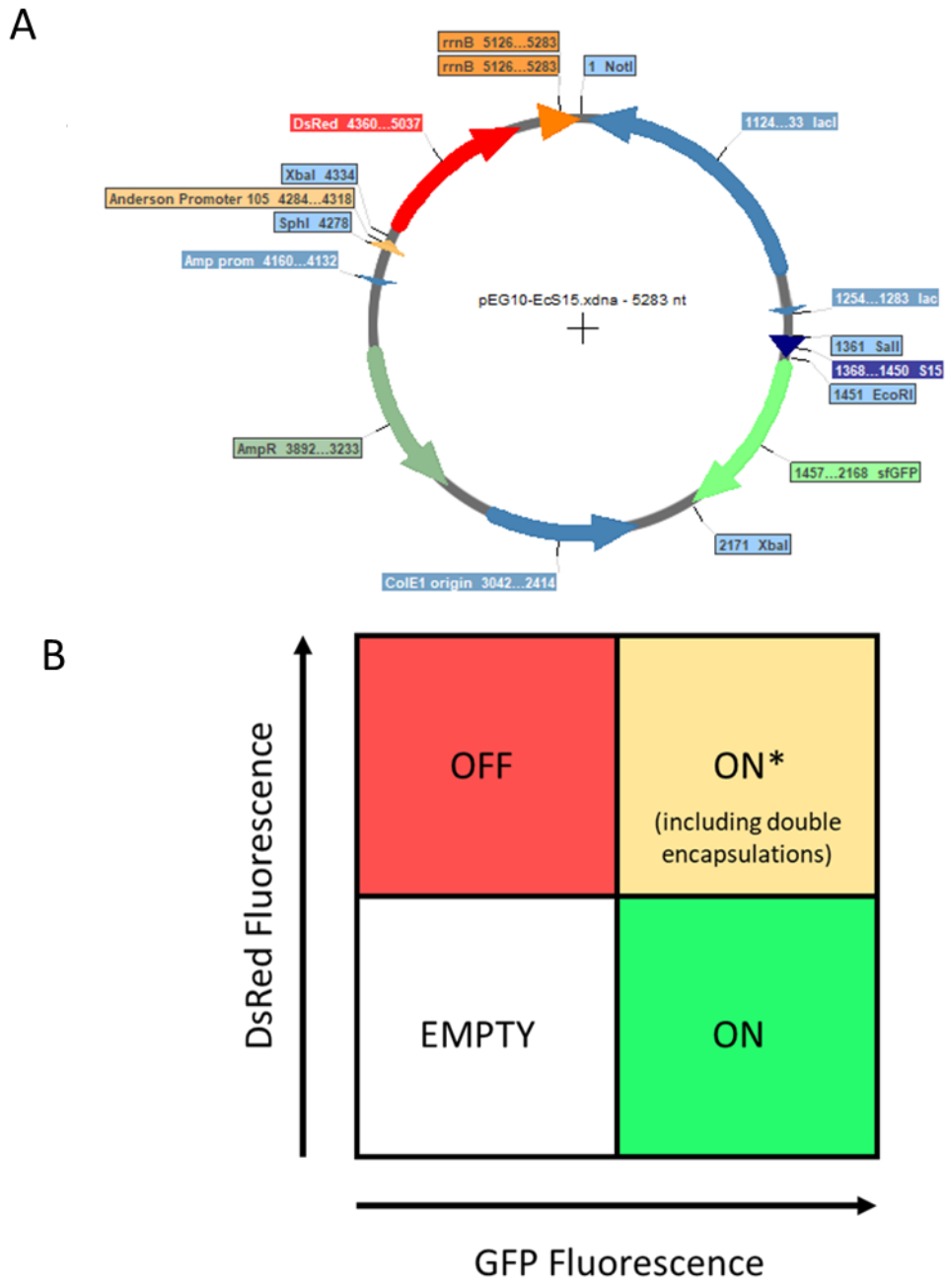


Figure 2.5 ***Reporter Plasmid Design****. A) The IPTG inducible* plac *promoter drives the RNA-*gfp *fusion. Both* α *and* sf *GFP were used in these assays. Constitutive DsRed is added as a control for droplet loading under the control of a constitutive promoter. The plasmid also contains an ampicillin resistance cassette and a ColE1 origin not shown. B) With a constitutive red fluorophore in addition to the RNA-*gfp *fusion, "OFF" RNAs can be recovered separately from "ON" and "EMPTY" droplets by collecting red fluorescing droplets. Additionally, the constitutive fluorophore can help gate out droplets with more than one microcolony.*

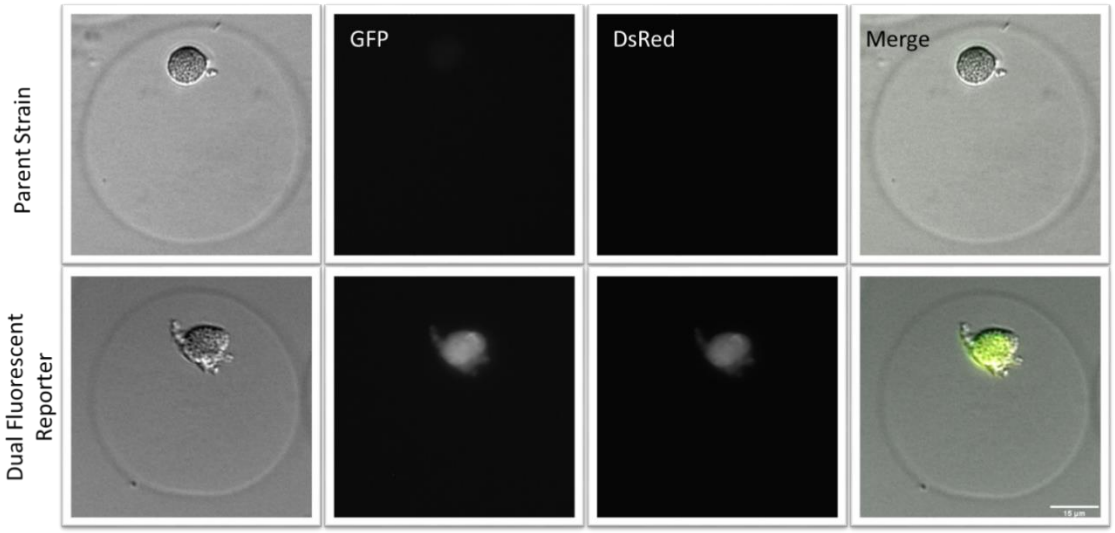


Figure 2.6 ***Fluorescence of Reporter Plasmid****. Micrographs of E. coli microcolonies without and with the dual fluorescent reporter plasmid and induced with 1 mM IPTG. Scale bar is 15 μm. Micrographs of LB 100 μg/mL amp + 34 μg/mL chl +1mM IPTG agar droplets containing either E. coli K12 ΔrplA or K12 ΔrplA with pEG4-EcL1 M1 microcolonies about 20 hours post encapsulation. Both strains contain pBAD33-RBS-EcrplA. Scale bar is 15 μM.*

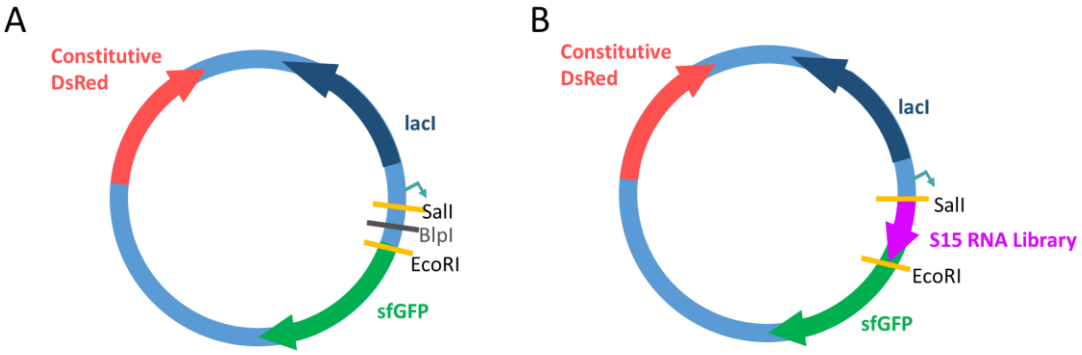


Figure 2.7 ***Ligation Strategy for Reporter Plasmid****. Vector pEG10 A) empty and B) with RNA leader insert. RNA leader is cloned into the pEG10 vector with the SalI and EcoRI sites. A BlpI site between the two cloning sites allows for linearization of remaining empty vector after RNA leader ligation.*

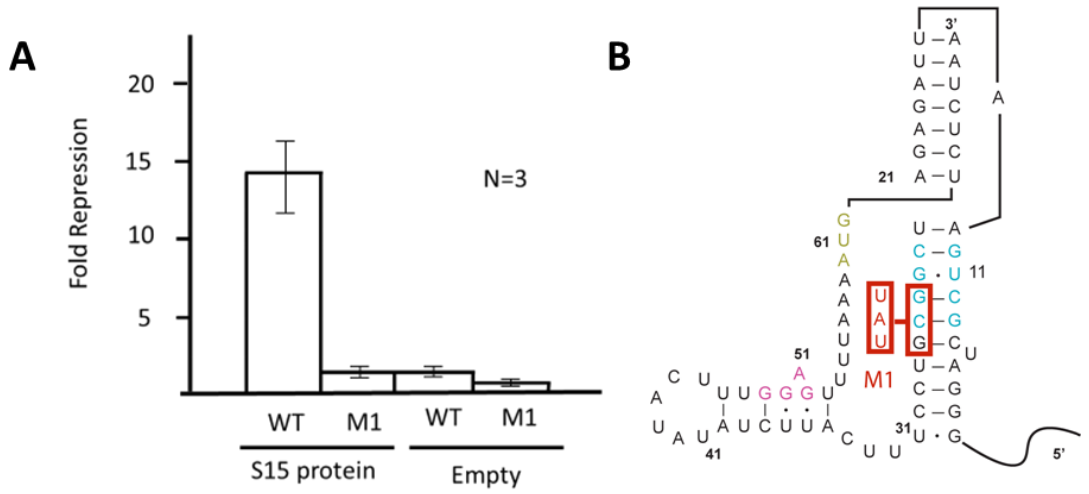


Figure 2.8 ***Regulation of the E. coli S15 Leader.*** *A) E. coli S15 β-galactosidase assay for wild type RNA and non-regulating mutant M1 with and without S15 protein expression. [63] B) Sequence and structure of E. coli S15 RNA leader, start codon in red. Always on mutation, M1, is denoted by the red box. Data from panel A from Slinger et al. 2015.*

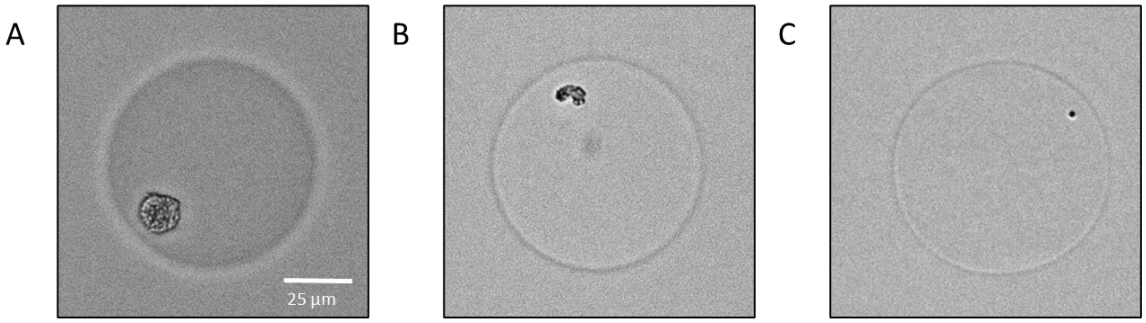


Figure 2.9 ***Growth Defects in K12 ΔrpsO Strain Microcolonies.*** *K12 ΔrpsO cells cultured overnight in 1% agarose SOB +amp +chl microdroplets. A) Microcolony of expected size and B,C) failed microcolonies are from same encapsulation, showing how lack of S15 protein leads to growth inconsistencies even within the same encapsulation.*

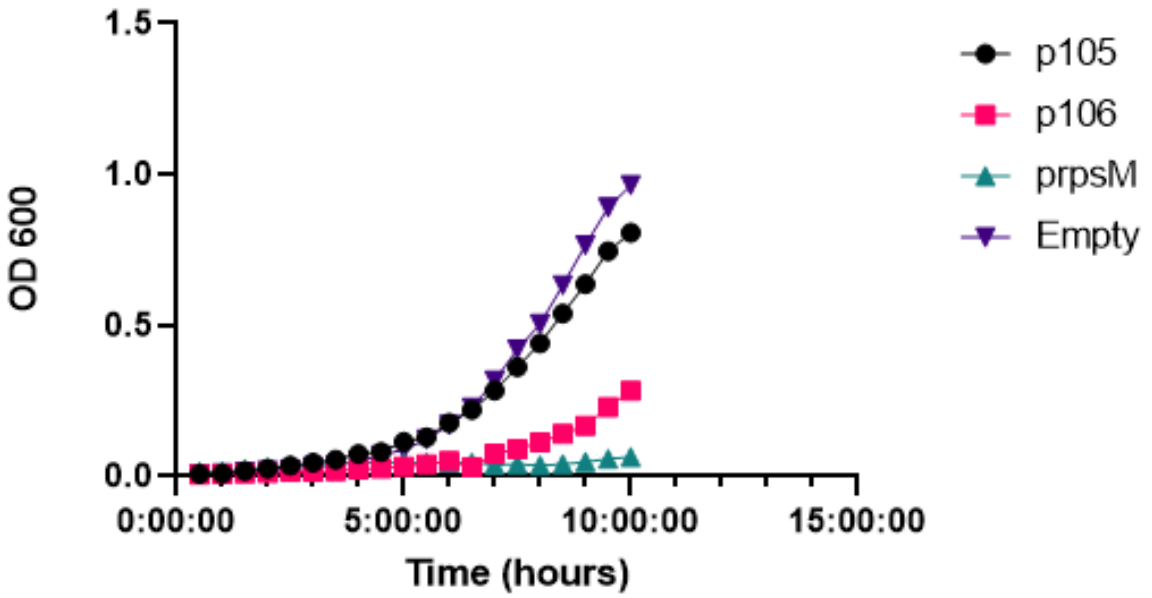


Figure 2.10 ***Promoter Growth Curve Assay.*** *K12 ΔrpsO cells transformed with reporter plasmid pEG5 with different promoters driving DsRed expression were diluted to OD600 of 0.01 in LB +chl +amp. OD600 of the bulk cultures was measured every 30 minutes.*

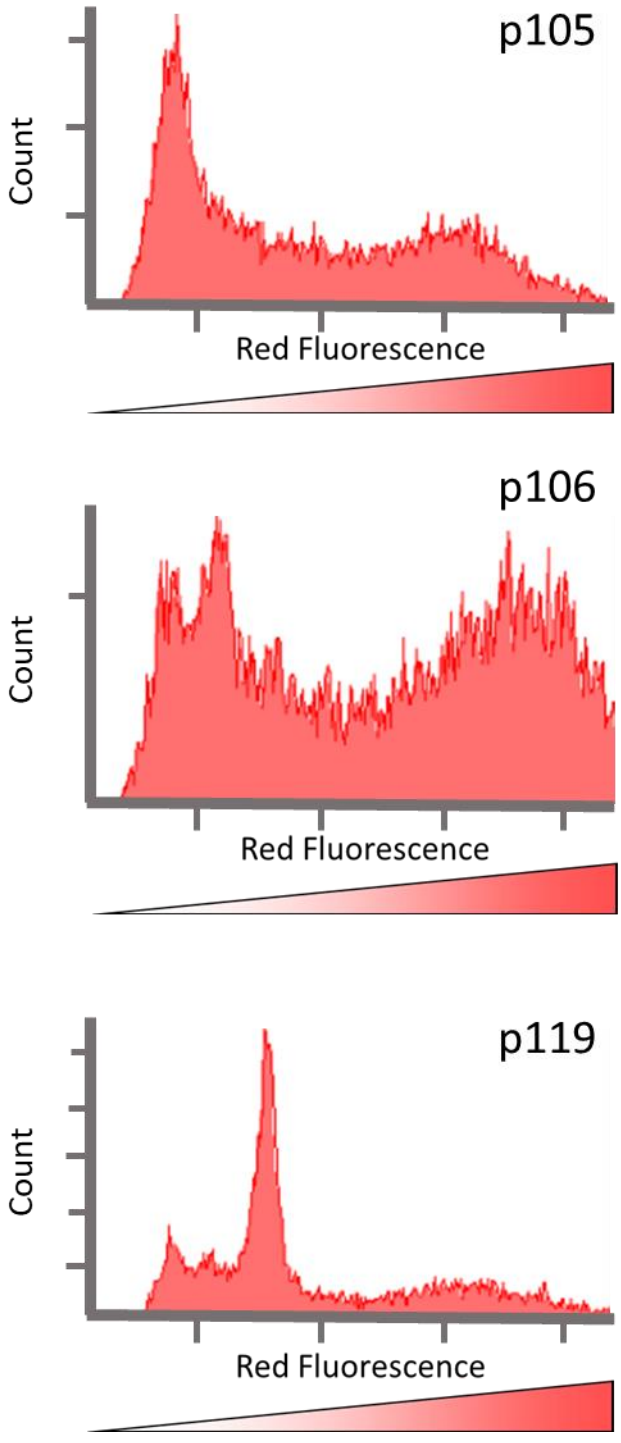


Figure 2.11 ***DsRed Expression Measured by Flow Cytometry.*** *K12 ΔrpsO cells transformed with reporter plasmid with different promoters driving DsRed were encapsulated and cultured in agarose microdroplets and analyzed via flow cytometry. Histograms show distribution of droplets with varying red fluorescence intensity.*

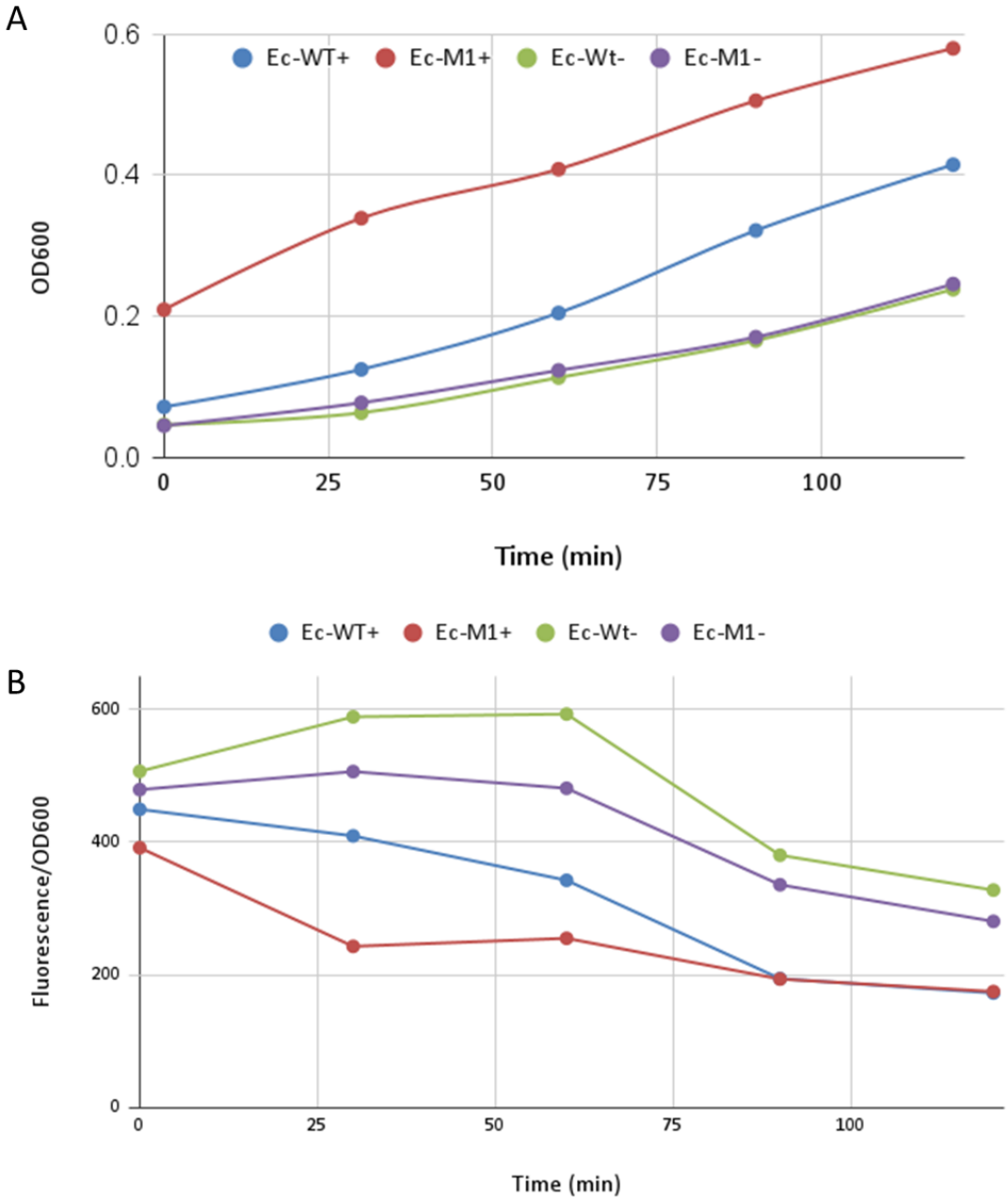


Figure 2.12 ***Fluorescence of DsRed over Induction Time.*** *MG1655 cells with pEG9-S15 WT or M1 with pBAD33-rpsO grown to mid-log in SOB +L-arabinose +chl +amp and induced at 0 minutes with 1mM IPTG. OD600 was measured (A), and fluorescence of DsRed was measured and then normalized to OD600 (B).*

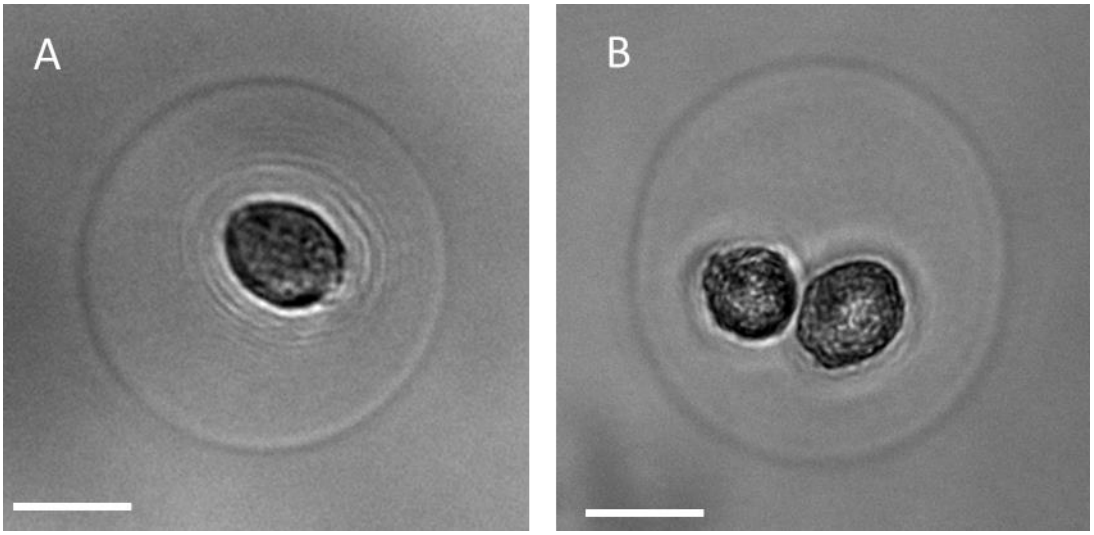


Figure 2.13 ***Comparison of Green Fluorescent and Green-Red Fluorescent Microdroplets****. MG1655 cells with pEG9-S15 WT with pBAD33-rpsO were encapsulated in 1% agarose SOB +amp +chl +L-arabinose, cultured overnight and sorted according to fluorescence. Figure shows example droplets from either the green fluorescent (A) or double fluorescent (B) sorted droplets. Scale bar is 30 μm.*

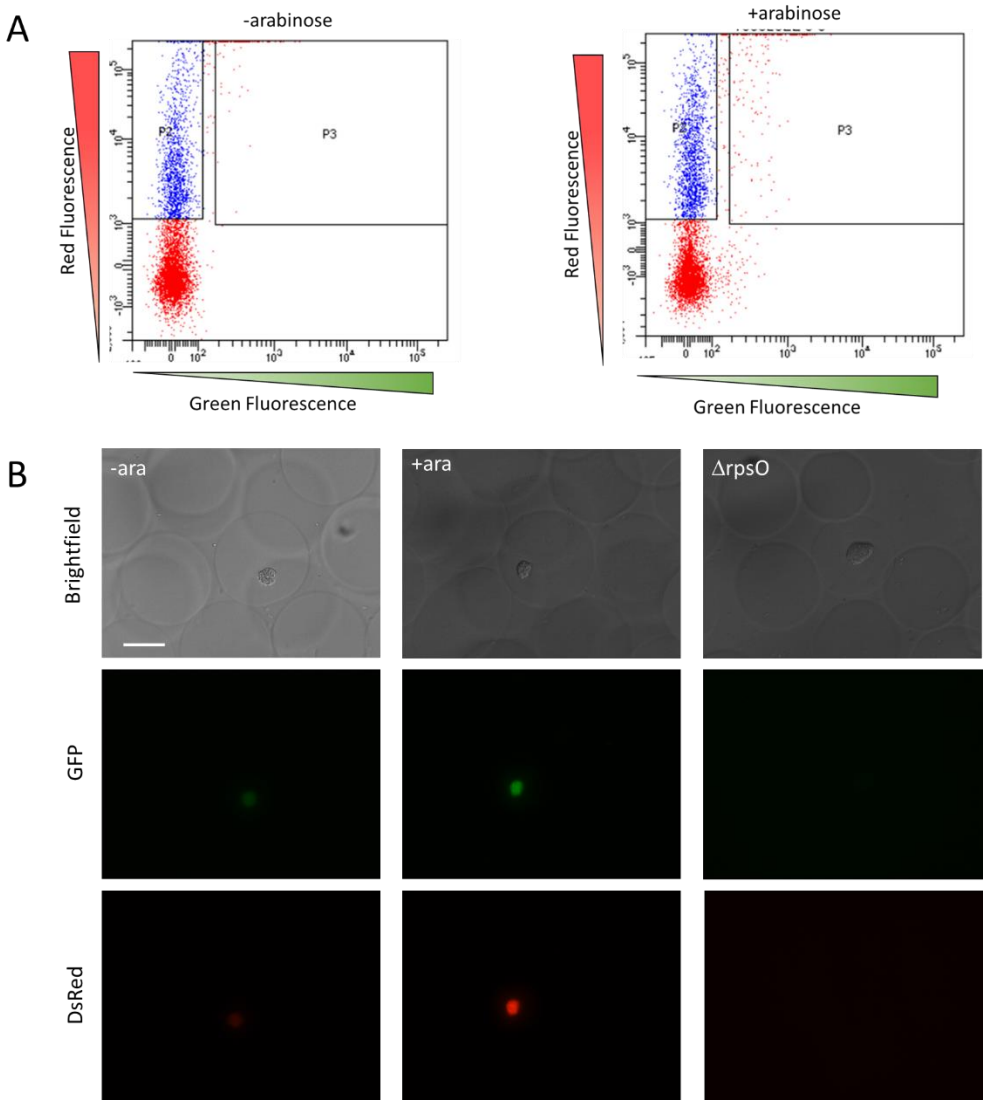


Figure 2.14 ***Lack of S15 Protein Introduces Protein Expression Defects.*** *K12 ΔrpsO cells with or without pBAD33-rpso and pEG9-S15WT were encapsulated into microdroplets (1% agarose SOB +chl +L-arabinose +amp) and analyzed with flow cytometry (A) or microscopy (B). For each method, samples with L-arabinose are brighter than samples without L-arabinose. Scale bar is 30 μm.*

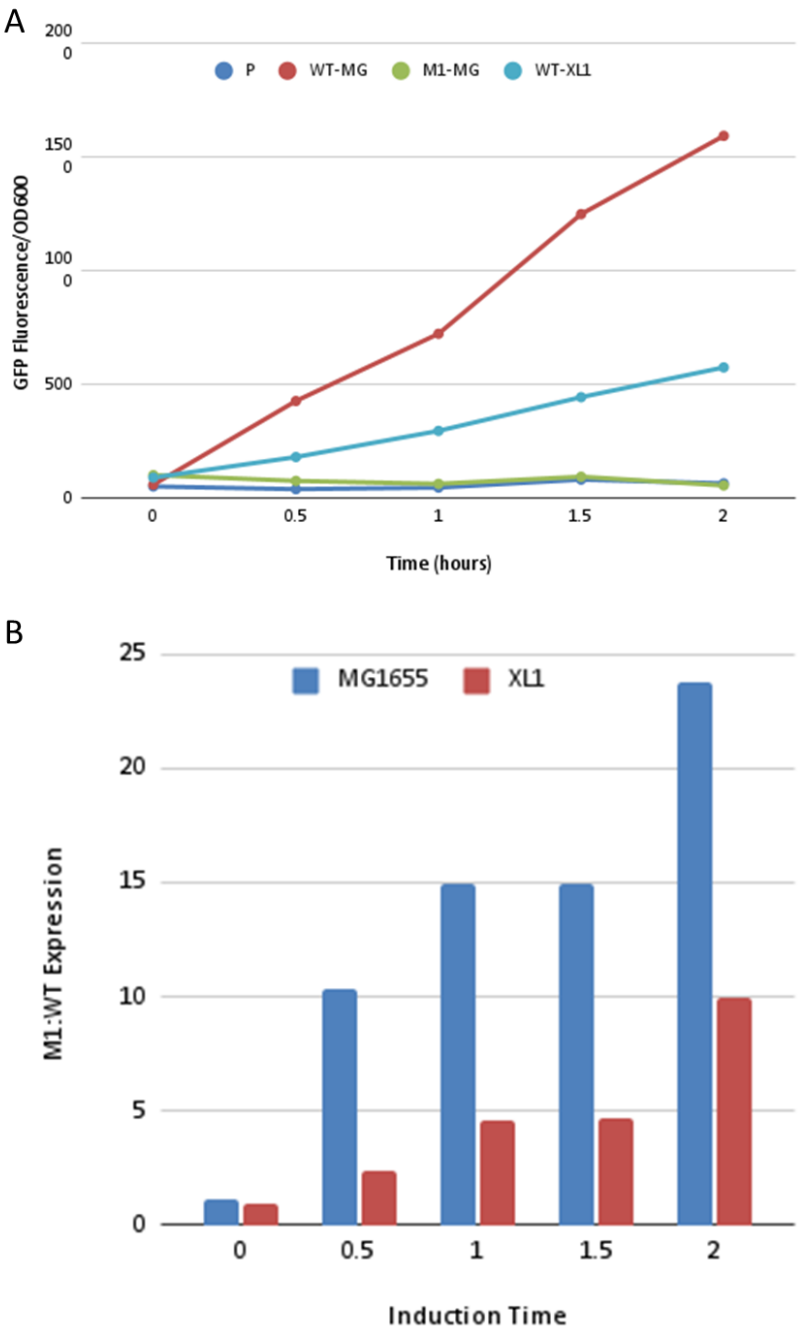


Figure 2.15 ***Strain Optimization for S15 Regulation Assay.*** *A) Fluorescent plate assay of K12 ΔrpsO, MG1655 (MG) and XL-1 cells (XL1) transformed with pBAD33-rpsO and either pEG9-S15 WT or pEG9-S15 M1 showing GFP fluorescence over time after induction of GFP expression by 1mM IPTG at 0 minutes. Graph also shows non-fluorescent MG1655 strain without plasmids (P). B) Expression ratio of the two strains over time: GFP fluorescence is normalized to OD and then M1 normalized fluorescence is divided by WT normalized fluorescence.*

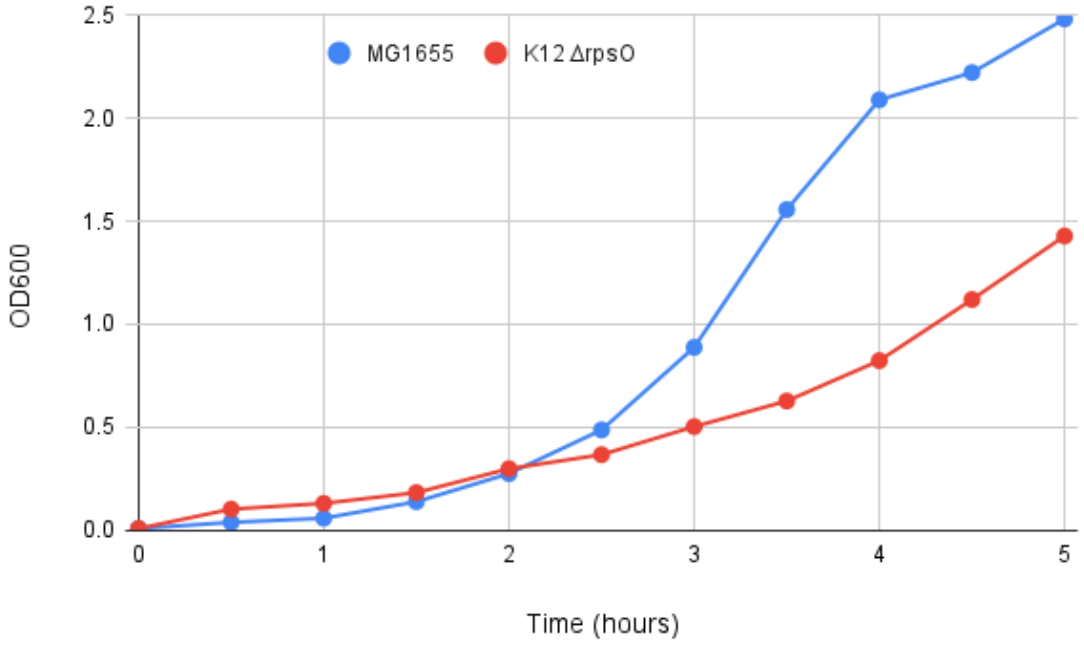


Figure 2.16 ***Growth Comparison of Strains MG1655 and K12 ΔrpsO****. Each strain was diluted to OD600 of 0.01 with arabinose-induced S15 expression, and OD600 was measured every 30 minutes.*

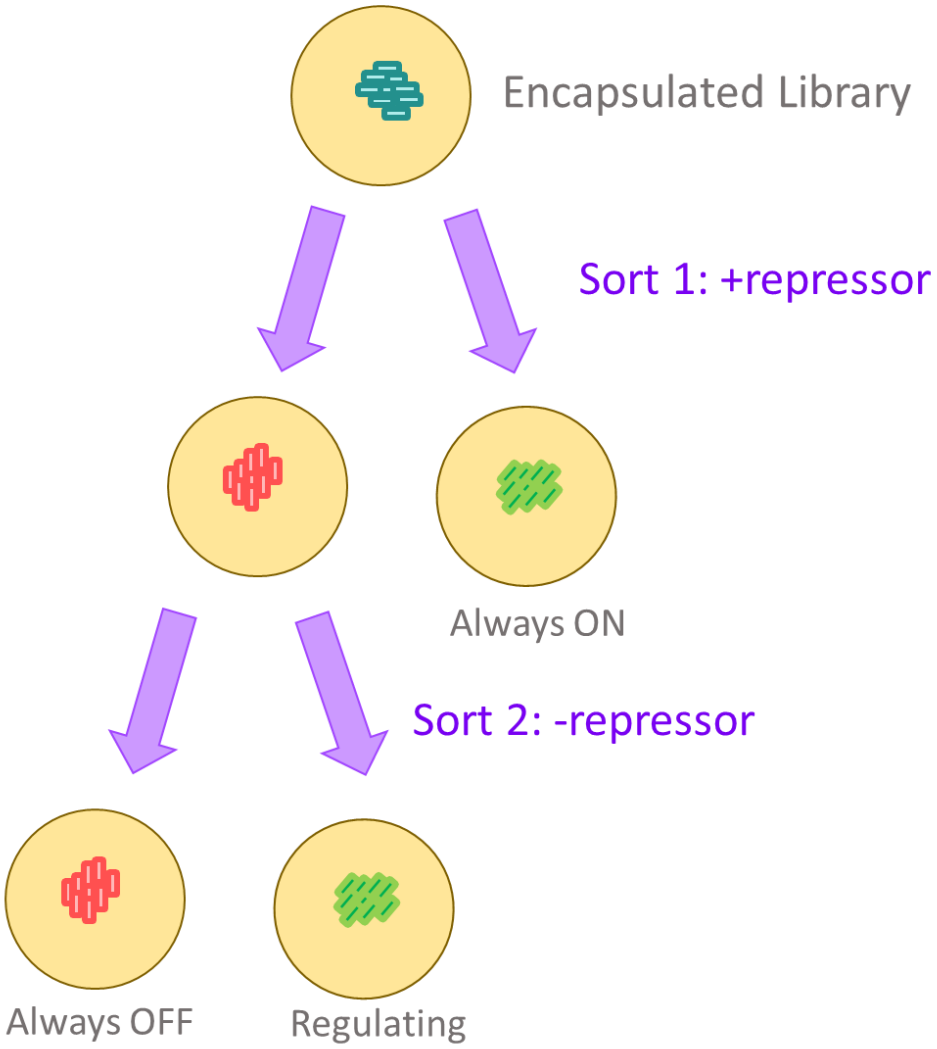


Figure 2.17 ***Droplet Sorting Schematic for Identification of Always Off, Regulating, and Always On Leader RNAs****. For study of the S15 leader, the first round of sorting is performed in MG1655 cells with induction of S15 protein, while the second round of sorting is in the K12 ΔrpsO strain without the induction of the S15 protein.*

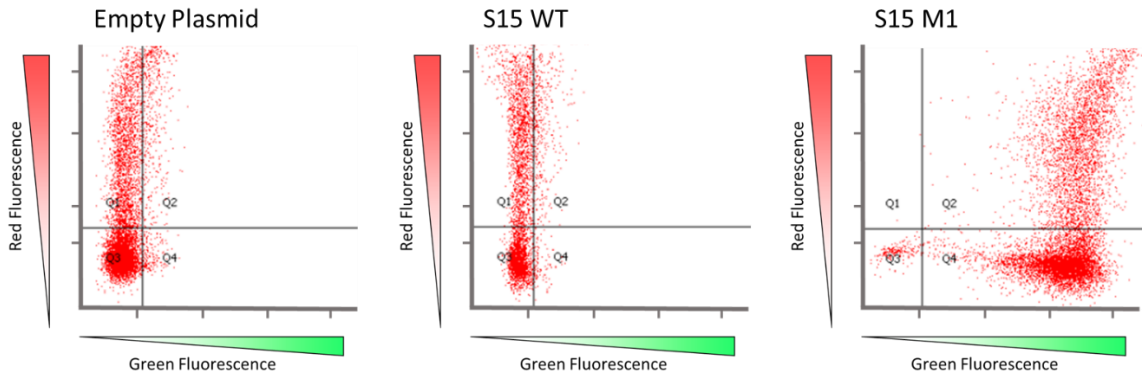


Figure 2.18 ***Repression of sfGFP Reporter****. Analytical flow of MG1655 pBAD33-rpsO pEG9-S15 WT or M1, or empty pEG10 vector (no GFP start codon) droplets. Fluorescence shows that the WT leader can effectively repress GFP. S15 was not able to regulate, as seen by the significantly greater GFP expression.*

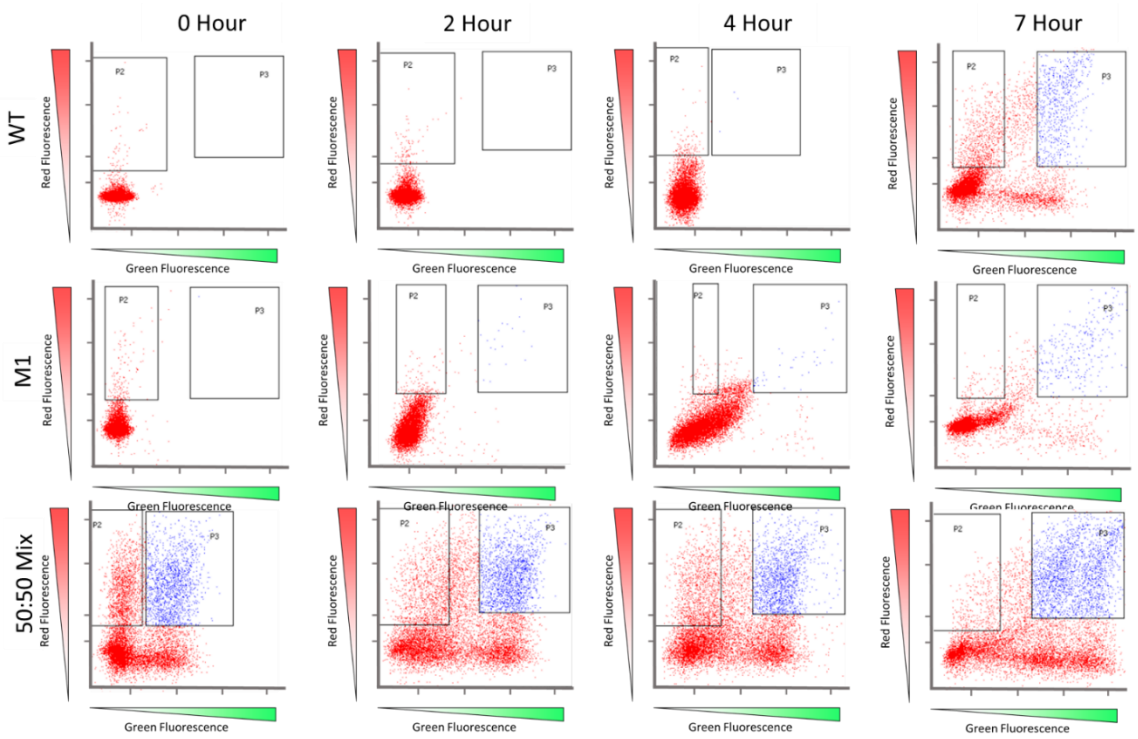


Figure 2.19 ***sfGFP Induction Time Course.*** *MG1655 pBAD33-rpsO pEG9-S15 WT, M1 or mixed 1:1 WT:M1 prior to encapsulation droplets were grown overnight, then induced with 1mM IPTG in media with L-arabinose.*

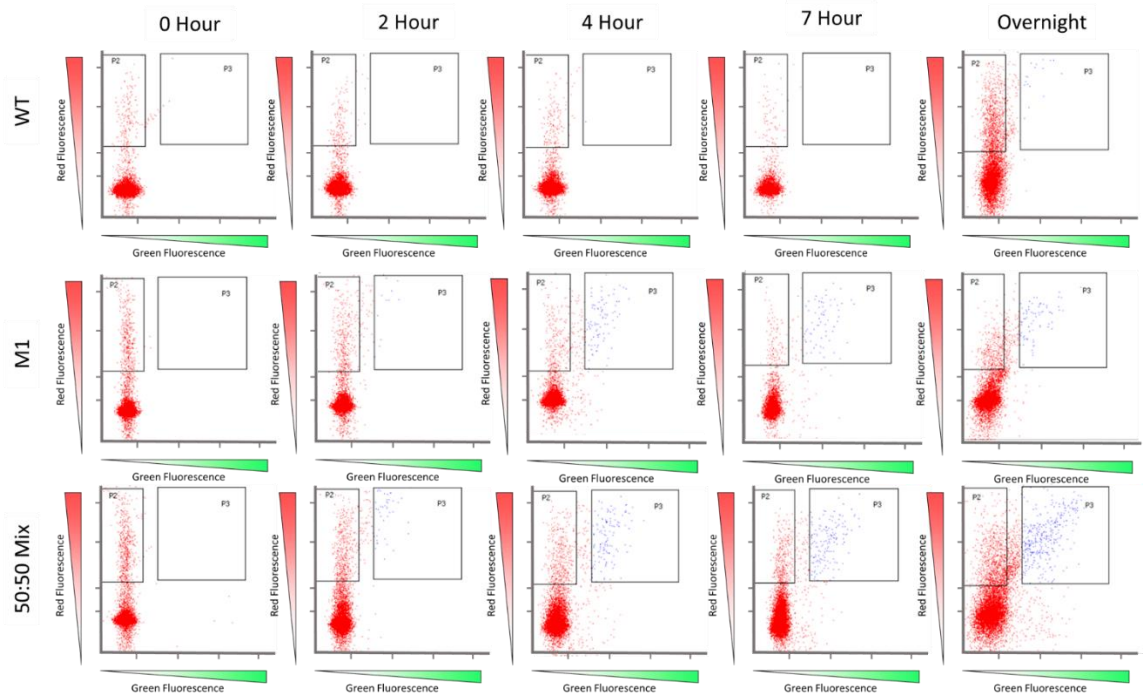


Figure 2.20 ***αGFP Induction Time Course****. MG1655 pBAD33-rpsO pEG9-S15 WT, M1 or mixed 1:1 WT:M1 prior to encapsulation droplets were grown overnight, then induced with 1mM IPTG in media with L-arabinose. For the overnight condition, 1 mM IPTG was added prior to encapsulation.*

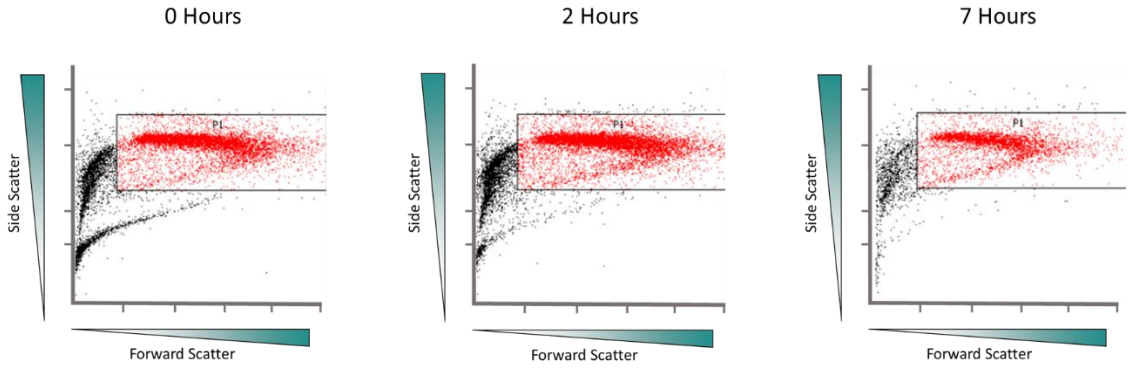


Figure 2.21 ***Shifting Forward Scatter in Induced Droplets.*** *Select forward and side scatter plots from Figure 18 showing that longer induction times result in smaller droplets. The forward scatter, representing event size, decreases with longer induction times.*

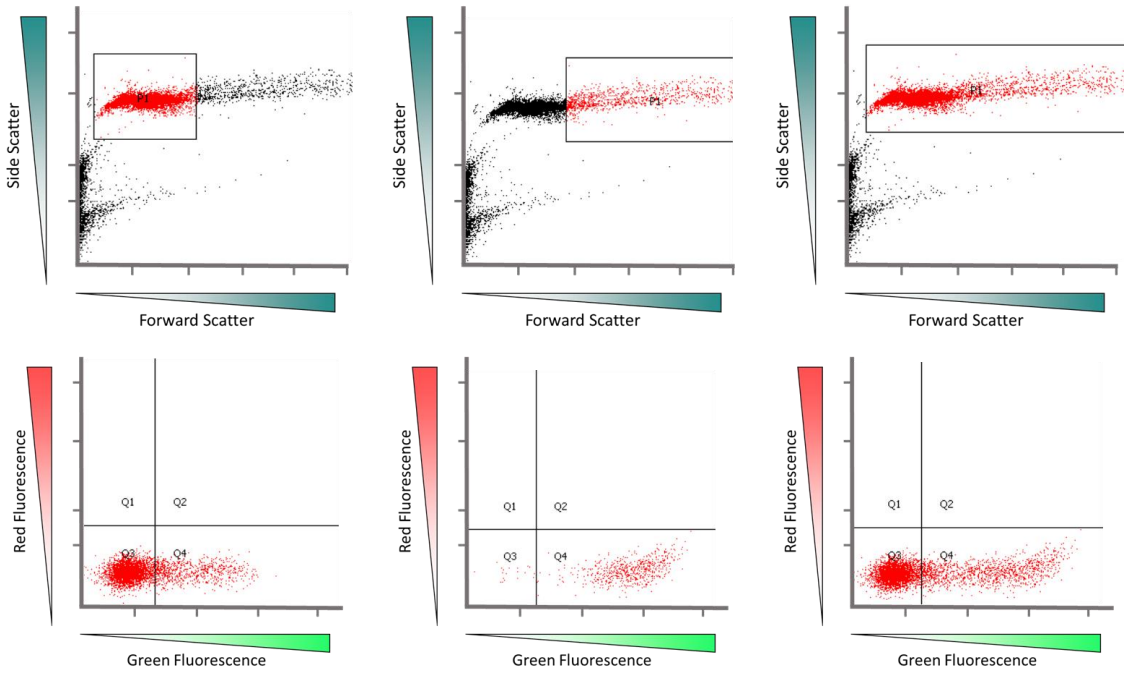


Figure 2.22 ***Flow Cytometry Gating Comparison.*** *Shown is the same sample with three different gates based on FCS v. SSC. Highly fluorescent droplets are larger than the main event population.*

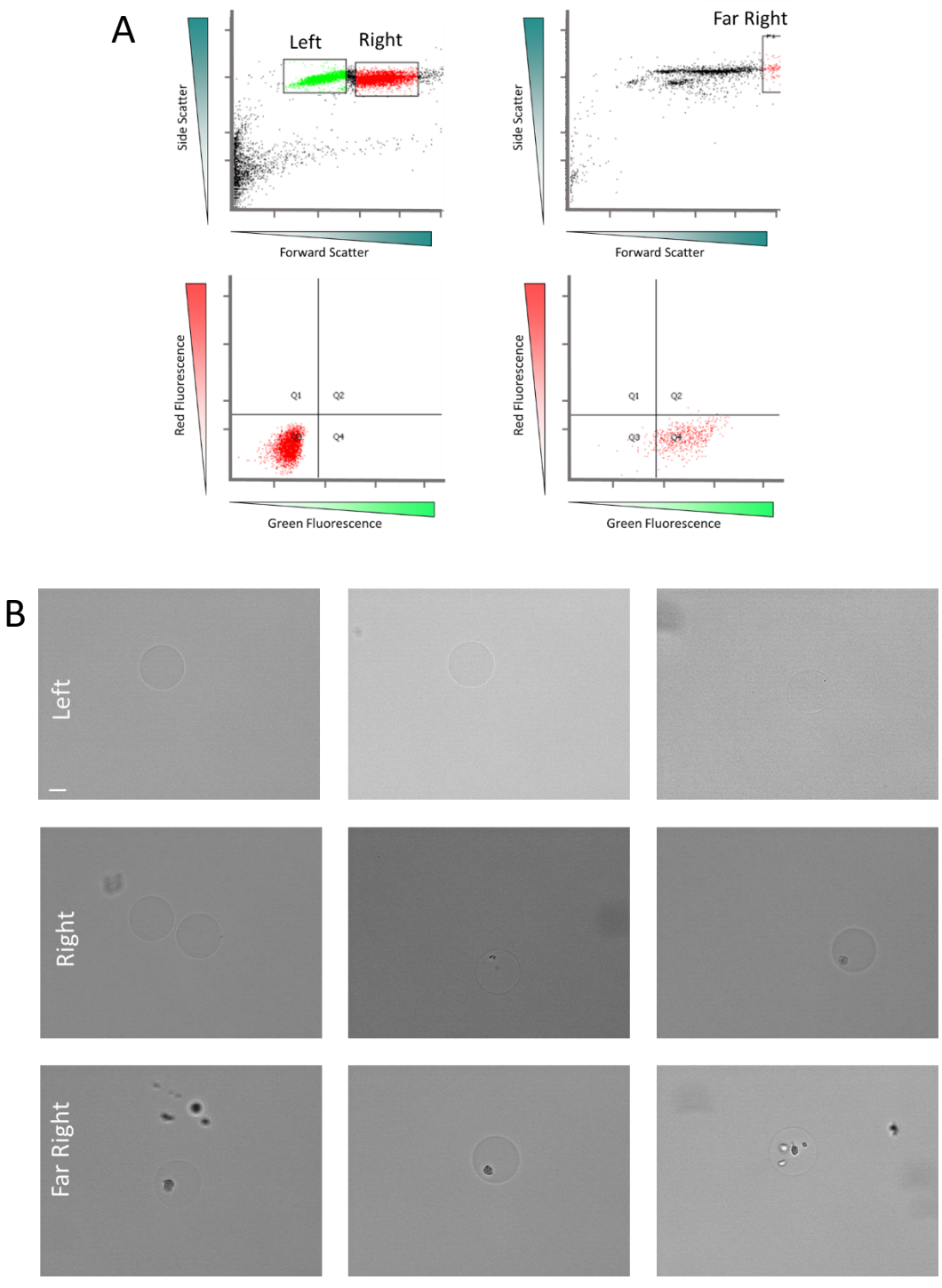


Figure 2.23 ***Microscopy of Forward Scatter Populations for Gating Optimization****. Loaded droplets (MG1655 pBAD33-rpsO pBS4-S15) were sorted according to the gates shown (A) and samples were then taken for analysis by microscopy to discern any differences in the events of different size (B). The micrographs show that empty droplets are common in smaller events and the biggest events are often droplets with more than one microcolony. Scale bar is 30 µm.*

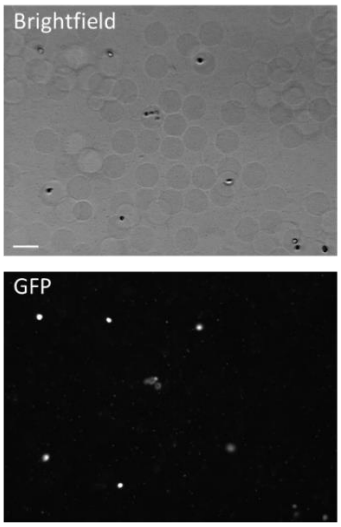


Image	Filled	Not Filled	Percent Filled
+ ara 2	11	107	10.3%
+ara 6	6	95	6.3%
+ara 3	4	85	4.7%
+ara 4	7	79	8.9%
+ara 5	8	73	10.9%
+ara 1	8	74	10.8%
-ara 2	5	47	10.6%
-ara 3	3	43	7.0%
-ara 4	2	40	5.0%
-ara 5	4	50	8.0%
Average -			7.65%
Average +			8.65%

Figure 2.24 ***Example of Droplet Occupancy Raw Data.*** *Droplets loaded at an OD600 0.005 of A) micrographs of droplets loaded and cultured overnight in the brightfield and green fluorescence channels. B) Example table generated after tallying droplet loading in FIJI. Colonies were quantified using the brightfield channel. Scale bar is 60 μm.*

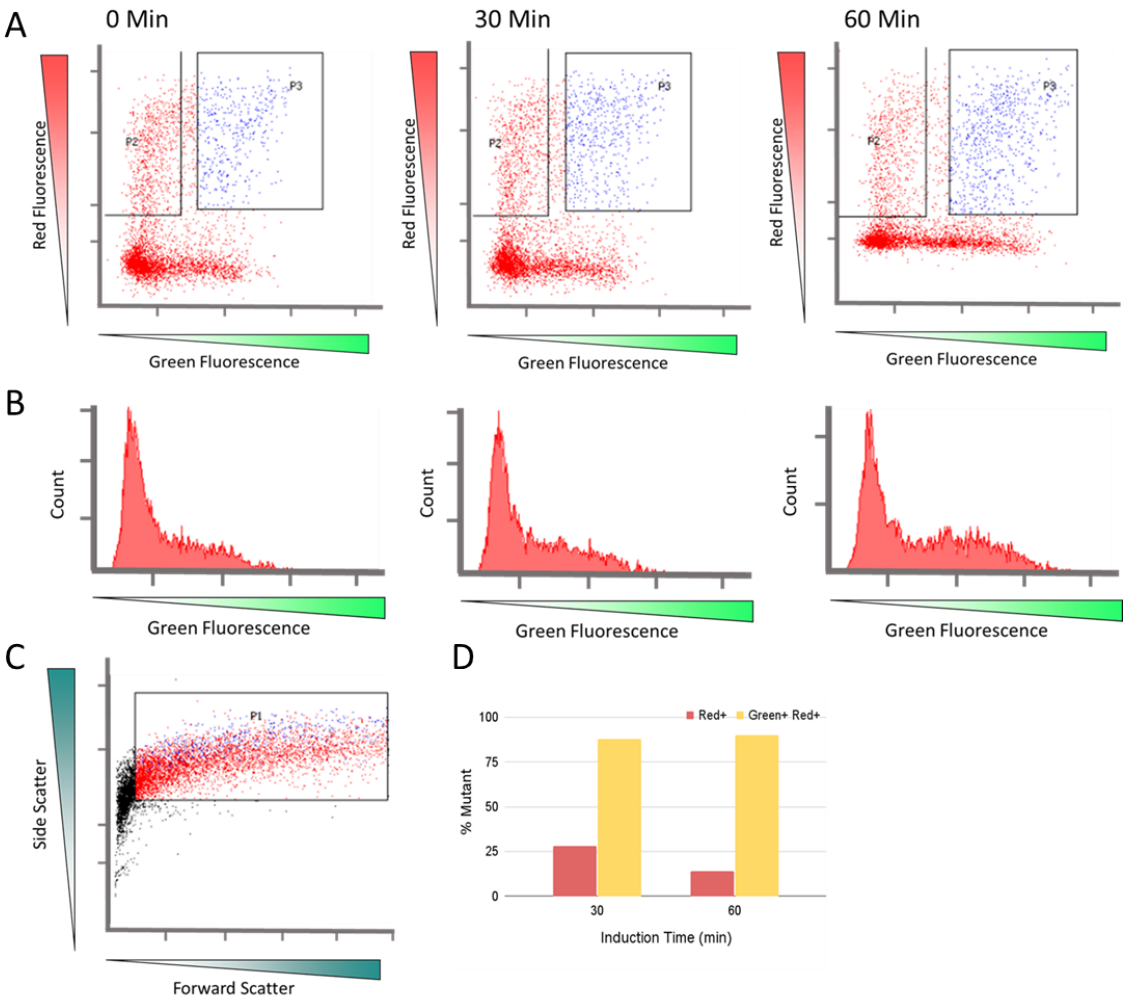


Figure 2.25 ***FACS Analysis of S15 Leader.*** *Droplets contain 1:1 mix of E. coli S15 wild type and M1 leader in the MG1655 strain. Time shows duration of IPTG induction. A) Red and green fluorescence of analyzed droplets. Gates shown were used to sort droplets into red (P2) and double positive (P3) populations to determine ability of FADS to distinguish between the always on mutant and the regulating wild type. B) Histogram of green fluorescence per droplet for P1 gate. C) Forward and side scatter of the droplet sample. Gate P1 contains the droplets. Area outside the gate are small debris or free swimming bacteria. Fluorescent axes use log10. Counts show every 50 droplets. D) Graph shows percent of M1 colonies recovered after droplet sorting for different induction times. With perfect sorting, the red droplets (P2) would have no M1 and the double positive droplets (P3) would be entirely M1. N=116 for 30 min. N=104 for 60 min.*

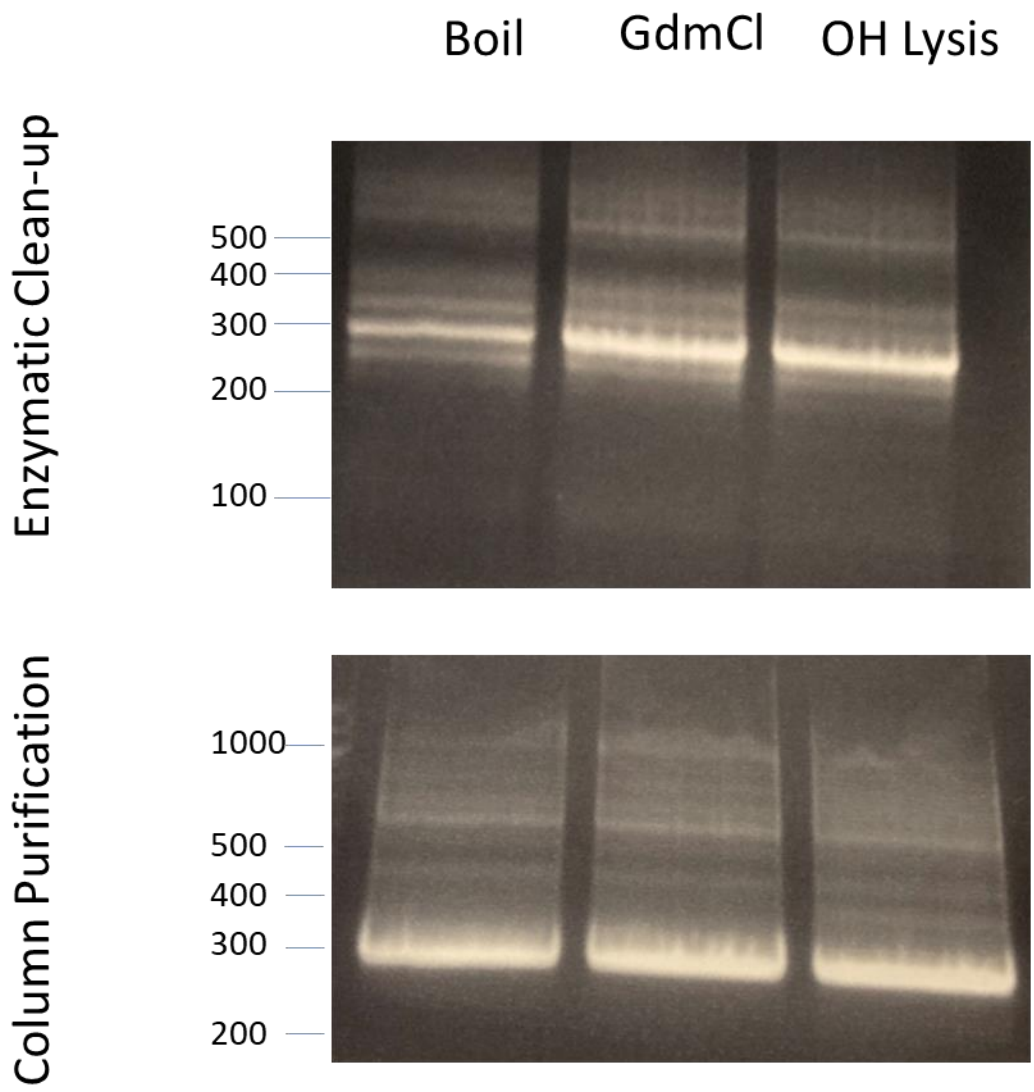


Figure 2.26 ***PCR of Extracted DNA.*** *MG1655 droplets carrying pEG-S15 droplets were cultured according to standard protocol, FADS sorted, evenly divided among conditions and the DNA recovered by boiling alone or guanidinium chloride treatment or alkaline lysis followed by column cleanup. In between the first and second PCRs, the samples were either cleaned up enzymatically with ExoSAP-IT or by column purification.*

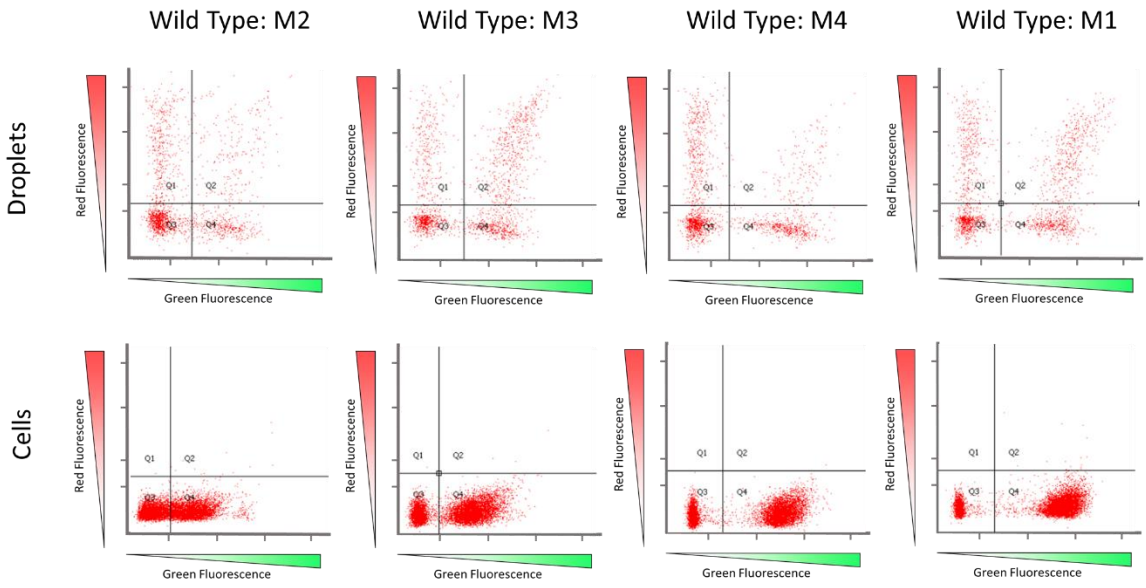


Figure 2.27 ***Comparing Droplets and Free Cells with Flow Cytometry.*** *Cells with wild type and mutant RNA leaders in the reporter plasmid were mixed in a 1:1 (droplets) or 1:2 (cells) ratio and fluorescence with the expression of the S15 repressor protein was measured with flow cytometry. M2-4 are unsequenced mutants from the S15 leader library. While strongly "always on" mutants are distinguishable in both methods, droplets can better distinguish subtler differences in expression level.*

CHAPTER III:

REGULATORY FITNESS LANDSCAPE OF THE *E. COLI* S15 LEADER

Introduction

The *E. coli* S15 leader structure provides an intriguing test case for the FADSRA assay. As discussed in Chapter 1, S15 protein binding to the rRNA is a conserved interaction important for stabilization and assembly of the central domain of the small ribosomal subunit. [45,51,54] Yet the S15 leader RNAs are often only partial mimics of the rRNA structure. [22] S15 leaders generally contain only one of the conserved rRNA binding motifs and, sometimes, contain new motifs important to binding not found in the rRNA at all. Our example leader for this study, the *E. coli* S15 leader, has one conserved motif, a G·U/G-C stem motif that is found as a secondary binding site between the S15 protein and the rRNA (Figure 3.1).

This leader also has a unique folding mechanism as revealed by Wu *et al*. [131] The RNA forms the first hairpin and then generally a second hairpin 3' to the first (Figure 3.2A, Figure 3.2B). From this double hairpin structure, a pseudoknot can form between the loop of the first hairpin and what was formerly the 3' stem of the second hairpin (Figure 3.2C). This pseudoknot structure contains an internal loop with the ribosome binding site and the start codon. When these elements are unbound and in the absence of protein, the structure is permissive to translation. However, the loop can form base pairing interactions, masking the ribosome binding site (Figure 3.2E, Figure 3.2B, [129]). Additionally, the *E. coli* S15 leader can also fold into lower stability pseudoknots not necessarily permissive of translation (Figure 3.2D).

Binding of the S15 protein to the G·U/G-C site of the S15 leader stabilizes a pseudoknot structure which allows for the small ribosomal subunit to bind the mRNA at the same time as the S15 protein, leading to subunit entrapment and preventing ribosome

assembly. [28] Other S15 leader structures can be regulated by *E. coli* S15 protein, but evidence points to multiple mechanisms for binding and regulation. Additionally, evidence of regulation is found even in the absence of strong *in vitro* binding between the RNA and the protein. [63]

To summarize, the pseudoknot structure is both a permissive structure that promotes translation in the absence of S15 protein and required for repression via S15 protein. A high throughput regulatory assay like FADSRA can be used to create a regulatory landscape of this complex regulatory structure to identify nucleotides affecting regulation, either through stabilizing a permissive or non-permissive RNA structure, or by preventing binding of S15 or ribosome entrapment. This can in turn, give more information about which areas of the leader are permissive to change and which are prone to conservation for successful *in vivo* regulation.

Results

Library Design

To build a regulatory landscape of the S15 ribosomal leader, it was necessary to assay all single mutants and with a secondary goal of assessing a subset of double mutants. The *E. coli* and *T. thermophilus* leader are each 83 nucleotides long. For a regulator of this length, 249 single mutants and 61,254 double mutants are possible. These mutants cover the region from the transcription start site to the first eight codons of the regulated gene, *rpsO* as the first four of the codons are a part of the S15 leader structure. Due to the relatively short length of these leaders, a pair of complementary doped oligos was designed that spanned the complete length of the leader, with overhangs designed for cloning into

pEG10. The oligos were doped such that each position, except the start codon, had 97% of the population with a wild type nucleotide, and 1% for each possible nucleotide mutation. To retain protein expression in the reporter assay, the ATG start site was fixed, not mutated. For an oligo library of 83 nucleotides that is 3% doped (1% of each other nucleotide) at each position a full length oligo has a 20% probability of 1 mutation and an 8% probability of retaining the wild type sequence as calculated using a binomial distribution (Figure 3.3). Unlike the *E. coli* and *T. thermophilus* leaders, the *G. kaustophilus* leader is 99 nucleotides long. This shifts the predicted distribution when doing a 1% doping, to 15% of full length sequences with one mutation and only 5% of sequences maintaining the wild type sequence (Figure 3.3).

After oligo duplexing and cloning into the reporter vector pEG10, amplicon from each of the libraries was generated and sequenced to confirm that the calculated distribution of mutations was reflected in the plasmid library. The results showed some subtle differences between the *E. coli* and *T. thermophilus* libraries, with more wild type sequences present in the *E. coli* library (Figure 3.4). Similar to calculations, the *G. kaustophilus* library had a lower percentage of WT and single mutant samples. Overall, the results of the sequence reflected the calculated distribution with modestly increased amount of WT and single mutants in the *E. coli* library.

Assay Design

For biological replicates, the complete FADSRA protocol was performed on three independent transformations of the pEG10-EcS15 library into the MG1655 + pBAD33-*rpsO* strain (Figure 2.17, Figure 3.5, Figure 3.6, Figure 3.7, Table 3.1). To collect enough droplets for DNA extraction and provide additional diversity due to the severe

bottlenecking during the encapsulation process, at least six rounds of droplets were pooled for each round for each biological replicate. This resulted in about 200,000 sorted events per round of flow cytometry. The pooling of samples also provided a sufficient amount of DNA to be extracted from the droplets for the creation of amplicon.

The amplicon improvements led to developing shorter amplicons. This made the separation of leader amplicon from empty vector amplicon using gel electrophoresis easier. One important component to the amplification strategy was the inclusion of a unique molecular identifier (UMI), which reduced sequence amplification bias. Extracted DNA is first amplified briefly (5 cycles of PCR) with a set of primers containing a UMI and partial homology to the barcoding primers. Next, the PCR reaction is purified and a second amplification step barcodes and amplifies the UMI-amplicon while attaching the primers for Illumina sequencing. Given the vast differences between the number of droplets sorted for each population, the UMI ensures that the extracted DNA can be amplified enough to be sequenced while also preserving the leader counts between samples.

Flow Cytometry and Droplet Sorting

During droplet sorting, the droplets are divided into three populations: G+, a green fluorescent population where GFP is not repressed, R+, a red fluorescent population with repressed GFP, and double positive droplets which express both green and red fluorescence. Even before any droplets were sorted and sequenced, it was striking in the first round (repressive) how the bulk of the droplets had higher GFP expression than wild type leader droplet controls, indicating a general lack of repression (Figure 3.8). This was likely caused by the high percentage of leaders with multiple mutations in the starting library (Figure 3.4). Looking at the histograms of GFP intensity, the first round sort has a

large peak representing the empty and red only droplets, and then a rapid drop-off in fluorescent intensity. This could indicate that instead of two clear populations, the leader mutants represent a gradient of binding and repression, which can be measured via GFP expression. In contrast, the histogram of the second round of sorting much more closely resembles a bimodal distribution curve. This is likely because in the absence of protein, the reporter is either being expressed because there is nothing repressing the translation (regulating sequences) or it is not, perhaps due to a change in the leader or early coding sequence prohibiting translation (always off).

Deep Sequencing & Data Processing

After droplet sorting, DNA was extracted and Illumina sequencing ready amplicon was generated via PCR. For one sample (Replicate #3, first round of sorting, R+ droplets), amplicon could not be generated due to insufficient DNA. The other samples were sequenced on an Illumina NextSeq and the resulting sequences processed for analysis (Figure 3.9). The raw FASTQ files from each flow cell were processed with fastp to filter the reads for quality. [142] While the Illumina sequencing run was a paired end run, the barcode reads for the second end were of poor quality and those samples were unable to be effectively demultiplexed. After quality control filtering, the individual flow cell files were concatenated into one file for each sample. Each of the sequenced samples had sufficient read depth for analysis with all samples having over a million reads (Figure 3.10A). However, variability between the number of reads per sample was high with a minimum of 1.5 million and a maximum of 8 million despite equimolar pooling and quantifying sample concentration with a Qubit. After deduplication, the number of reads proportionally

reflected the amount of input DNA, with lower reads from the red fluorescent positive droplets and higher reads from input samples (Figure 3.10A).

The deduplicated reads were then analyzed and further filtered using DiMSum, an analysis pipeline for analyzing the fitness of sequence variants from deep mutational scanning after a selective event. [143] In this case, fitness is defined as the natural log of the product of the variant output to input ratio and the wild type input to output ratio:

$$\log\left(\frac{Variant\ input}{Variant\ output} \times \frac{Wild\ Type\ output}{Wild\ Type\ input}\right)$$

The fitness of wild type is set to zero so a variant with similar regulatory behavior to the wild type leader would have a fitness close to zero. Since the fitness is tied to wild type and the regulation of fluorescence for wild type changes between rounds, for ease of explanation, the sign of the DiMSum fitness value has been adjusted such that a positive fitness value is correlated with increased GFP expression. This resulting value will be called the GFP expression fitness. DiMSum quality analysis found that the sequences of the input samples for two of the second round replicates had a disproportionately large number of sequences with three or more mutations (Figure 3.10B). These samples were excluded from analysis since generation of the fitness landscape requires single mutations. After quality control, the final samples analyzed for the regulatory landscape were Replicates 1 and 2 for the first round (with S15 protein) and for Replicate 1 only for the second round of sorting (without S15 protein).

Regulatory Landscapes

While every possible single mutant was found in the demultiplexed sequences, some were lost through additional quality filtering. To minimize sequences resulting from

PCR and sequencing errors, only sequences with more than 10 counts across all replicates were kept for further analysis. The remaining variants were analyzed over the two rounds of FADSRA (with (repressive) and without S15 protein (permissive)). Heat maps were made to analyze the effect of single mutations on fitness for each round. A positive fitness score on the heat map is correlated with increased GFP expression. Within each round, the GFP expressing (permissive, G+) and GFP non-expressing (repressive, R+) populations were compared to each other.

The first round heat map indicates that most sites have a neutral to negative fitness value (Figure 3.11A). This indicates that most single mutations to the leaders have low GFP expression and that repression is maintained. However, there are a couple mutations in the conserved G·U/G-C motif stem that have a high expression fitness score, indicating these mutations result in a lack of repression and thus high GFP expression. The G·U/G-C motif is the primary binding site for the S15 protein. Interestingly, not all mutations to this stem cause an increase in GFP expression, and only one mutation C26U located at the base of the G·U/G-C motif causes a large increase in GFP expression. For C26U, the first round fitness value is 3.67 ($\sigma = 0.86$) and 1.9 ($\sigma = 0.01$) for the second round indicating that this mutation consistently leads to an increase in GFP expression. Changing the C to a U at this location would change the hydrogen bonding to a wobble pairing. For the nucleotide directly adjacent to C26, the mutation G25A also shows an increase in expression in the first and second rounds with an expression fitness of 2.13 ($\sigma = 0.91$) and 4.00 ($\sigma = 0.06$).

The other important motif to regulation, the pseudoknot, is structurally required, but not sequence specific. [57,58,60] Looking at the absolute value of the fitness mapped along the structure of the leader, another site correlated with reduced regulation is A21G

(Figure 3.11B). For A21G, the first round expression fitness is 2.34 (σ = 1.32) and a second round fitness of 4.00 (σ = 0.13). These high expression fitness values indicate that this mutation is always on, independent of protein. This A begins the pseudoknot, hydrogen bonding to the nucleotide immediately downstream of the ATG start codon. It is possible that a wobble base pair here destabilizes the pseudoknot enough to prevent protein binding and decrease repression.

In the absence of protein, the majority of these mutations have a positive fitness value, indicating that most of these single mutant leaders either are expressing GFP and "regulating" or are always on mutants that made it to the second round (Figure 3.11B). The mutations most correlated with an always off phenotype are G75U (first round fitness: 0.40, σ = 0.91, second round fitness: -5.35, σ = 0.05), a mutation that creates a premature stop codon. The other mutations that result in reduced expression under permissive conditions are G3U (first round fitness: 0.30, σ = 1.23, second round fitness: -2.01, σ = 0.06) and A13C (first round fitness: 1.36, σ = 0.93, second round fitness: -3.14, σ = 0.01) near the ends of the first stem, as well as the mutation of C45U (first round fitness: -0.09, σ = 1.01, second round fitness: -3.16, σ = 0.02), located upstream of the ribosome binding site in a loop. Some of the mutants have a much higher fitness score than wild type, indicating that they may express more GFP than wild type even under permissive conditions. Most of these sites are in the first stem of the leader (Figure 3.11B, Figure 3.12).

Of note is the mutations in the coding region without fitness values (Figure 3.12). These mutations were present in the raw sequencing data, but were filtered out due to low sequencing counts. The coding sequence mutations lost during quality filtering were

present in the in roughly the same frequency as the other coding mutations in the initial pool of mutations.

Validation Assays

Because FADSRA is a new method, results were validated through analysis of select clones recovered from the assay. After each FADS, microcolonies were plated and grown overnight. A subset of the colonies was then picked into a 96-well plate for a bulk culture fluorescence assay. Such an assay has been previously used to analyze the regulation of cis-regulators in the Meyer Lab. Overnight cultures of the clones were then diluted 1:1000 and grown until late log phase. As previously discussed in Chapter 2, assaying ribosomal protein regulation requires balancing protein available for ribosomes with protein available to bind the leader-reporter fusion. Ribosomes are produced very actively during the cell's vigorous growth at mid-log, so waiting until late mid-log accounts for both fresh cultures with cells in similar metabolic states, while also assaying the cells as ribosome production decreases. This increases the amount of ribosomal protein that is able to bind to the leader-*gfp* fusion for regulation.

After 5 (or 6 hours for K12 Δ*rpsO* strains) of growth, cultures were induced with 1mM of IPTG and grown for additional time: 1 hour for MG1655 first round cultures and 2 hours for Δ*rpsO* cultures. These are the same induction times used for the droplets. After washing into PBS, fluorescence and OD600 were measured. GFP expression was normalized to the OD600 and this value was compared with that of wild type, giving the normalized fluorescence to wild type ratio. A cumulative distribution graph shows the distribution of this fluorescent ratio among the clones (Figure 3.13). While a broadly similar distribution is seen in all samples, clones selected from the GFP-positive droplets

show higher GFP expression than those from the GFP negative populations. Additionally, the difference in expression is seen between the rounds, with the MG1655 strain with S15 protein (round 1) having much higher expression than the K12 Δ*rpsO* strain without S15 protein. A relative GFP expression of 1 is the same expression as the wild type leader. For the MG1655 cell assays with S15 protein, M1 had a ratio of 7-10, while in the K12 Δ*rpsO* assay without S15 protein, M1 had a ratio of 3. This indicates that there is less dynamic range for the second round assays without S15 protein compared to the first.

As might be expected from strains with very different capabilities to translate mRNA into protein, the threshold values were quite different for each of the rounds with the first round having higher ratios than the second round. However, in each case, the clones selected from the R+ gate have a higher red fluorescence, and the clones selected from the G+ gate have higher green fluorescence. This shows that the FADSRA sorting agrees well with the results of a conventional assay, although the sensitivity is reduced when determining the regulation of specific clones.

Some of the clones from the plate assays were selected for further analysis. For each round, five colonies were selected using the normalized fluorescence to wild type ratio, spanning from below wild type to above M1 in fluorescent expression. Another four colonies were selected at random. The clone sequence was compared to the wild type sequence and any differences noted. For the clones selected from the second round green plates, all of the clones were wild type. Indeed, the sequencing suggests that wild type sequences are enriched in this populations, although this could also be partially due to fragmentation of the droplet during plating. Plating is done with beads to minimize the

amount the droplets are fragmented, but some fragmentation still is likely to occur, and the colonies from some droplets are likely overrepresented in the population.

Ten clones from the first round plates were sequenced, which revealed a mix of leaders with 1-4 mutations each (Table 3.2, Figure 3.14). One leader sequence was found twice (C4 and G7), with different expression ratios (1.29 and 0.70), indicating that the consistency of the plate assays is also less than ideal. Two mutants had changes to the Shine Dalgarno sequence (B3 and E8), which had a similar expression ratio to that of clone F3, a double mutant with a mutation in the stem adjacent to the conserved G·U/G-C motif. Most of these mutants did not have an associated fitness value. This could be because the number of reads was too low for the threshold for calculating fitness. This highlights how rare the double mutants may be in comparison to the single mutants. Three of the isolated mutants, E10, F2, and F3, had mutations to the G·U/G-C motif. For F2 and E10, the expression ratios were higher than M1 (expression ratio 7.13), a mutant that disrupts the G·U/G-C stem. The G25A single mutation has a positive fitness value, so the high expression ratio correlates with the FADSRA derived fitness score. It is unclear if there is an additional interaction from the U48A mutation. The F3 clone, the only single mutant identified in these validation experiments was E9, which has a mutation in the 3' end of the first stem. The expression ratio for this clone was 1.69 and the fitness value was -1.01 with an error of 0.98, making this value close to neutral in the FADSRA results. This difference could indicate limitations in sensitivity of the FADSRA assay.

Discussion:

This fitness landscape of the S15 leader underscores the importance of the ends of the stems for proper regulation, since that is where most of the mutation sites which cause

the biggest changes in regulation compared to wild type are located. Additionally, this landscape identified the importance of residue C26 in the G·U/G-C motif stem, this residue having the largest impact on fitness of all the nucleotides in the motif.

Overall, the regulatory fitness landscape of the *E. coli* S15 leader with its native binding partner indicates that most single nucleotide changes are tolerable and result in only small changes to the regulation of protein expression (Figure 3.12). The site most prone to breaking the regulatory activity of the leader were changes around the G·U/G-C protein binding site, as well as the starts and ends of the stems. These results indicate that the FADSRA assay identified conserved components of the leader, bolstering the credibility of this new method.

In comparing the findings of this assay to the classic mutational studies of the S15 leader using the Miller assay, the results are not synonymous, but they each identify similar areas of the switch as being critical for S15 protein binding and leader regulation. The primary binding site of S15, the G·U/G-C motif, was found in Béhard *et al.* 1998 [60] to be sequence specific and required for regulation. For bases 9 and 26, a GC base pair at the base of the motif, Béhard *et al.* found that disrupting the Watson-Crick pairing diminished regulation, while changing the pair to an A-T decreased but did not abolish regulation. For the FADSRA generated fitness landscape, the mutation C26T resulted in disrupted regulation, greatly decreasing regulation of the GFP reporter. The main G·U/G-C motif is comprised of bases C10, U11, G24 and G25. Béhard *et al.* found that any change to these bases abolished regulation. In the FADSRA landscape, C10G results in increased GPF expression, as does G25T and G24A. Each of these mutations disrupts the base pairing of the motif. The FADSRA landscape did not recognize every mutation in the G·U/G-C motif

as detrimental, but at least one mutation at all the conserved sites except for U11 were found to decrease regulation. Notably, the mutation U11C was not in the filtered and analyzed dataset. For G12 and C23, the base pair between the G·U/G-C motif and the beginning of the pseudoknot, C23 had only neutral mutations and G12A resulted in a slight decrease in repression. In Béhard *et al.*, maintenance of the WC base pairing maintained some regulation for G12-C23.

The pseudoknot structure is a secondary binding site for the S15 protein to the RNA leader. However, unlike the G·U/G-C motif, the importance of this motif is structural and is not sequence specific. Philippe *et al.* 1995 [129] used mutational analysis to extensively interrogate parts of the pseudoknot to discern the contribution of these sites to leader regulation. Of particular note, Philippe *et al.* mutated the base pairs at the extreme ends of the pseudoknot. For base pair U15-A69, at the most 3' position on the pseudoknot, disruptions to the Watson-Crick base pairing decrease repression while a C-G double mutation increases repression to over twice what is seen in wild type, perhaps due to the additional hydrogen bond. Another base pair, A13-U22, located at the end of Stem 1 before the formation of the pseudoknot, maintains regulation as long as Watson-Crick pairing is present.

In the FADSRA studies, A69 was not present in the filtered data set and U15 appeared neutral under all conditions. For the A14-U22 base pair, A14C results in less repression, but U22 was found to be a neutral mutation with (first round) and without (second round) S15 expression. One site that was not mutated in the Philippe *et al.* study was the base pair intersecting the pseudoknot, A21 and U63. The mutation A21G resulted in a loss of repression, while U63C resulted in an increase in repression. While not all of

the mutations found to change repression agreed with the classical papers, FADSRA did identify the ends of stems as having increased changes to expression compared to wild type for both the pseudo-knotted and double stem loop forms.

These results are additionally contextualized by a study of *E. coli* S15 leader folding by Wu *et al*. [131] This study used optical tweezers to assess leader folding dynamics. The study found that the mostly likely folding pathway for the leader was the formation of the first stem loop, followed by the second stem loop. The two stem loop structure can then transition to the stable pseudoknot, which is bound and further stabilized by S15. However, the single stem loop structure can also transition into weakly stable pseudoknot structures.

Without S15 binding, the two structures exist in equilibrium to each other. Changing the stability of one of the structures can have implications for downstream protein expression with and without S15 protein binding. The S15 protein binds to and stabilizes the pseudoknot, leading to gene repression. Therefore, mutations that would either destabilize a stem loop, especially the first stem loop, or stabilize an alternative pseudoknot conformation could potentially lead both to the downregulation by stabilizing a pseudoknot occluding the Shine Delgarno or start codon and to the loss of S15 binding through disruption of the G·U/G-C motif pairing necessary for binding.

The FADSRA analysis of the S15 leader spotlights the importance of stem stability for regulation of the operon. Not only were mutations deleterious in the presence of protein, but also in its absence, implying that secondary structure alone may cause changes to regulation. However, poor separation of low expression from high expression populations in the first round sort increased calculated error of the fitness scores and likely masked additional mutants of interest. Improvements could be made by increasing the stringency

of the gating or adding an additional gate altogether for near neutral mutations. Additionally, the bottleneck between rounds lead to under sampling of particular sites, which could mask additional mutant leader phenotypes. This bottleneck is, to an extent, inherent in the assay. Increasing the number of droplets assessed per round as well as increasing the replicate number would be the best way to both address the bottlenecking and increase confidence in the data.

Validation assays showed that there are changes in average fluorescence between gates with moderate selection for the wild type in the first round, results confirmed by the deep sequencing. Yet inconsistencies in the plate assays served as a reminder that even conventional methods for studying ribosomal leaders can have problems with consistency and sensitivity. For increased confidence in the data, additional replicates of clones in the plate assay would better show plate assay variability. Assessment via β-galactosidase assay would provide more sensitive results than the GFP plate assays, albeit such a step would require cloning of the leader library into a different reporter plasmid.

Materials and Methods

Library Design & Cloning

The S15 oligo libraries were ordered from IDT Oligos as doped 1% oligos with an overhang compatible with ligation into the pEG10 reporter plasmid. Oligos were annealed in duplex buffer (100 mM potassium acetate, 30 mM HEPES pH 7.5) by boiling and slow cooling to room temperature. The pEG10 plasmid was digested with SalI and EcoRI. The ligation of the S15 leader library into the pEG10 plasmid was performed with Quick Ligase

according to manufacturer's instructions. Before transformation, the plasmid was digested with BplI to linearize any plasmid without insert.

Transformation

For library transformations, MG1655 cells (first round) or K12 Δ*rpsO* (second round) +pBAD33-*rpsO* plasmid, were grown in SOB to mid-log, chilled for ten minutes on ice, then pelleted through centrifugation. Cell were washed gently 3 times with ice cold 10% glycerol. After the washing, cells were resuspended in 1:100 of the original volume in 10% glycerol. 50 µL of cells were transferred into a 1 mM gap cuvette with 5 µL of DNA for electroporation. Cells were electroporated at 1800V and recovered in warm SOB for 1 hour before plating onto LB agar plates with carbenicillin or in SOB with carbenicillin. To measure transformation effectivity, serial dilution streak plates were made of each transformation to calculate the number of transformants.

Droplet Culture and Sorting

The FADSRA steps of encapsulation, microcolony culture, induction and droplet sorting were performed as in Chapter 2. DNA was extracted using the alkaline hydrolysis method also found in Chapter 2.

Amplicon Sequencing

After DNA extraction, amplicon was generated via a two round PCR using Q5 polymerase. The 17 nucleotide UMI was added the first round of PCR, consisting of 5 cycles. The second round of PCR consisted of 30 cycles, and added the barcodes and Illumina adapters to the amplicon. Specific barcodes used can be found in Table 2. Amplicon was gel purified using a Zymo gel clean up kit according to the standard

protocol, except for adding twice as much wash buffer volume to remove more contaminating salts. Samples were quantified using a Qubit and sample purity was measured using a Nanodrop spectrophotometer. Based on the Qubit calculated concentration, the samples were pooled in equimolar ratio and purified one more time. The final sample pool was once again evaluated using the Qubit and Nanodrop. A NextSeq Mid-output kit was used to round 150 paired end cycles on a NextSeq500. Sequencing data was returned from BaseSpace as demultiplexed fastq files from the four different flow cells.

Data Analysis

The demultiplexed fastq files from each flow cell were first filtered for quality and trimmed using fastp, [142] then reads from the different flow cells for the same sample were concatenated into one file. AmpUMI [138] then deduplicated each of the samples. The resulting fastq files were input into the DiMSum pipeline with minimum input and output thresholds of 10 counts each. The G+ gate was treated as the input and the R+ was treated as the output. For first round analysis, input samples were 2 and 8 and output samples were 4 and 12. For second round analysis, sample 6 was used as input and sample 8 was used as output. The sign of the DiMSum fitness value was adjusted based on the behavior of wild type such that a positive value was always associated with increased GFP expression. Heat maps were generated using custom R scripts. Hamming distance was calculated for each variant as the number of base changes from the wild type sequence.

Plate Verification Assay

For verification of FADSRA sorting through conventional methods, 96 well fluorescent plate assays were performed. A subset of droplets was plated onto LB +ampicillin +chloramphenicol plates for validation assay. Colonies resulting from these droplets were inoculated into a deep 96-well plate in SOB + ampicillin +chloramphenicol +L-L-arabinose and grown overnight. As controls, reporter lines with WT or M1 S15 leader were grow in replicates (6 for WT and 3 for M1). The next morning, the cultures were restarted in the same media with either L-L-arabinose or glucose and grown for 5 hours (MG1655) or 6 hours (K12 Δ*rpsO*). The cultures were then induced for 1 hour with 1 mM of IPTG. The plates were centrifuged and the cell pellets washed twice with PBS. The cell pellets were resuspended in PBS and transferred to a black well optical plate. OD600 and fluorescence was measured on a plate reader. The expression ratio was calculated by normalizing OD600 to fluorescence and then dividing that by the average WT normalized fluorescence.

Table 1: Primers and Oligos

Amplicon	2073-pEG10-ill-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNGGCCTTGCCCGTAAAACGATTTG
	2074-pEG10-ill-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCAACGAGAATCGGCACAACACCTGT
Illumina barcodes	2059-Illumina-i5-D501-F	aatgatacggcgaccaccgagatctacacTATAGCCTacactctttccctacacgacgc
	2060-Illumina-i5-D502-F	aatgatacggcgaccaccgagatctacacATAGAGGCacactctttccctacacgacgc
	2061-Illumina-i5-D503-F	aatgatacggcgaccaccgagatctacacCCTATCCTacactctttccctacacgacgc
	2077-Illumina-i5-D504-F	aatgatacggcgaccaccgagatctacacGGCTCTGAacactctttccctacacgacgc
	2062-Illumina-i7-D701-R	caagcagaagacggcatacgagatCGAGTAATgtgactggagttcagacgtgtgctc
	2063-Illumina-i7-D702-R	caagcagaagacggcatacgagatTCTCCGGAgtgactggagttcagacgtgtgctc
	2064-Illumina-i7-D703-R	caagcagaagacggcatacgagatAATGAGCGgtgactggagttcagacgtgtgctc
	2065-Illumina-i7-D704-R	caagcagaagacggcatacgagatGGAATCTCgtgactggagttcagacgtgtgctc
	2066-Illumina-i7-D705-R	caagcagaagacggcatacgagatTTCTGAATgtgactggagttcagacgtgtgctc
	2078-Illumina-i7-D706-R	caagcagaagacggcatacgagatACGAATTCgtgactggagttcagacgtgtgctc
	2079-Illumina-i7-D707-R	caagcagaagacggcatacgagatAGCTTCAGgtgactggagttcagacgtgtgctc

Figures

Sample	Replicate	Round	Sort	i5	Primer #	i7	Primer #
1	1	1	Input	501	2059	701	2062
2	1	1	G+	502	2060	701	2062
3	1	1	++	503	2061	701	2062
4	1	1	R+	501	2059	702	2063
5	1	2	Input	502	2060	702	2063
6	1	2	G+	503	2061	702	2063
7	1	2	++	501	2059	703	2064
8	1	2	R+	502	2060	703	2064
9	2	1	Input	503	2061	703	2064
10	2	1	G+	501	2059	704	2065
11	2	1	++	502	2060	704	2065
12	2	1	R+	503	2061	704	2065
13	2	2	Input	501	2059	705	2066
14	2	2	G+	502	2060	705	2066
15	2	2	++	503	2061	705	2066
16	2	2	R+	501	2059	706	2078
17	3	1	Input	502	2060	706	2078
18	3	1	G+	503	2061	706	2078
19	3	1	++	501	2059	707	2079
20	3	1	R+	502	2060	707	2079
21	3	2	Input	503	2061	707	2079
22	3	2	G+	504	2077	705	2066
23	3	2	++	504	2077	706	2078
24	3	2	R+	504	2077	707	2079

Table 3.1 ***Samples Collected for FADSRA and Barcoding Strategy****.*

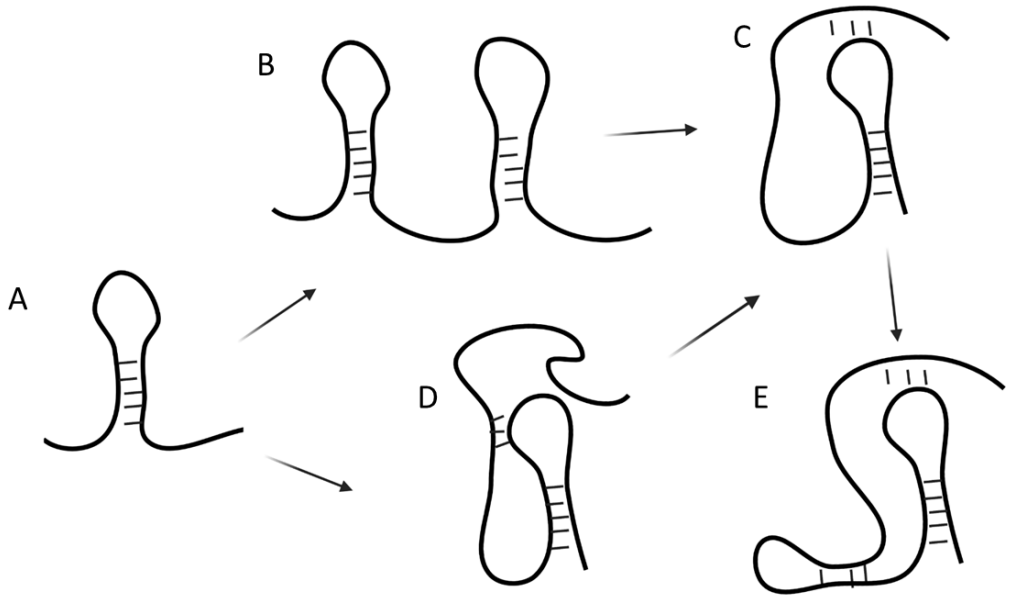


Figure 3.1 Folding of the E. coli S15 Leader. *An initial hairpin forms (A) before forming a double hairpin structure (B) that folds into a stable pseudoknot (C) that is permissive to translation without protein and the structure that binds the repressive protein. The leader can also form unstable pseudoknots (D) or have bonding between nucleotides of the stable pseudoknot's internal loop (E).*

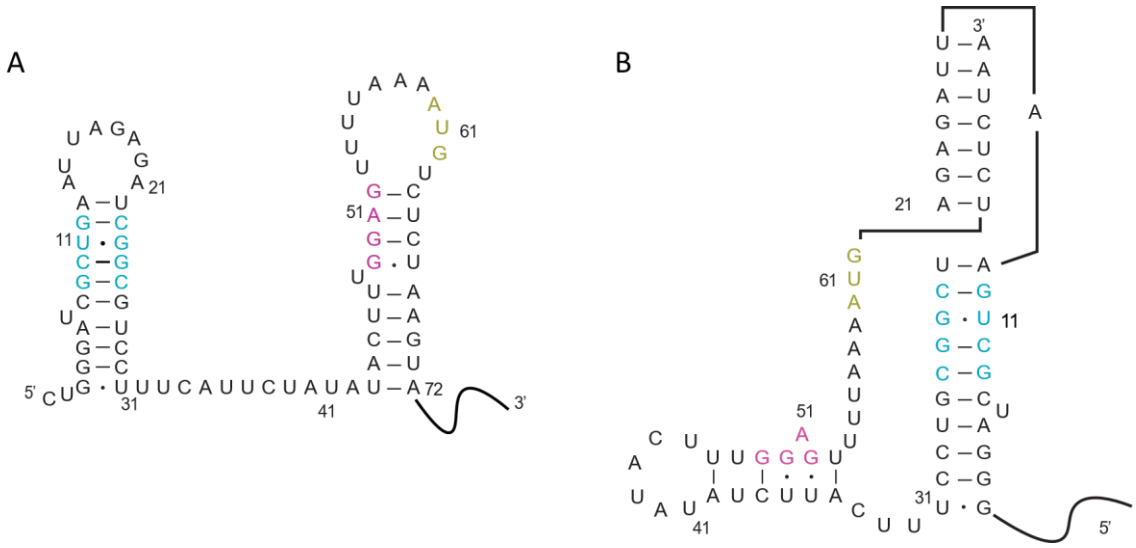


Figure 3.2 ***Structure of the*** **E. coli** ***S15 Leader*** *A) without S15 protein, and B) the structure stabilized by the S15 protein. The start codon is indicated in gold, the ribosome binding site is indicated in pink and the G·U/G-C motif where S15 protein binds is indicated in blue.*

.

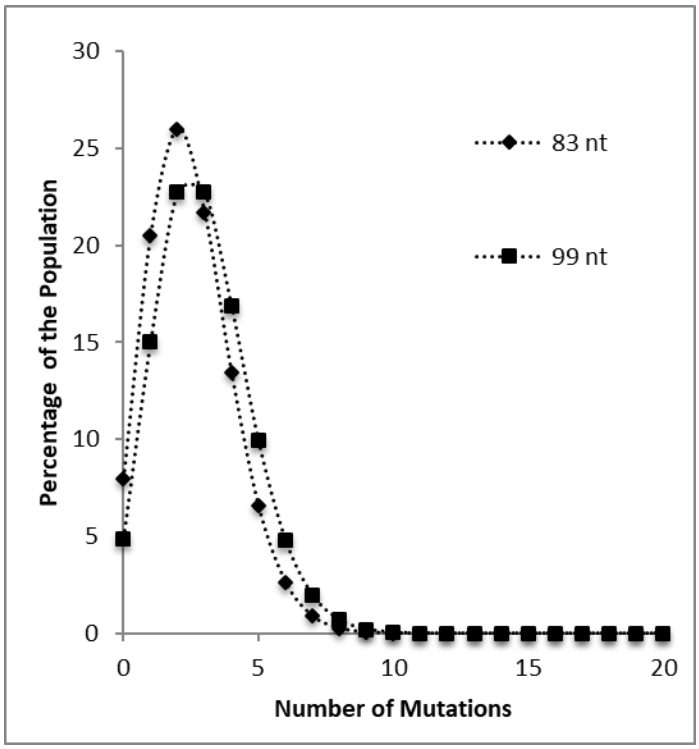


Figure 3.3 ***Calculated Mutation Distribution for Doped Oligo Library.*** *For a 97% wild type, 3% doped oligo library. Modeling for mutation population per length of sequence according to a binomial distribution.*

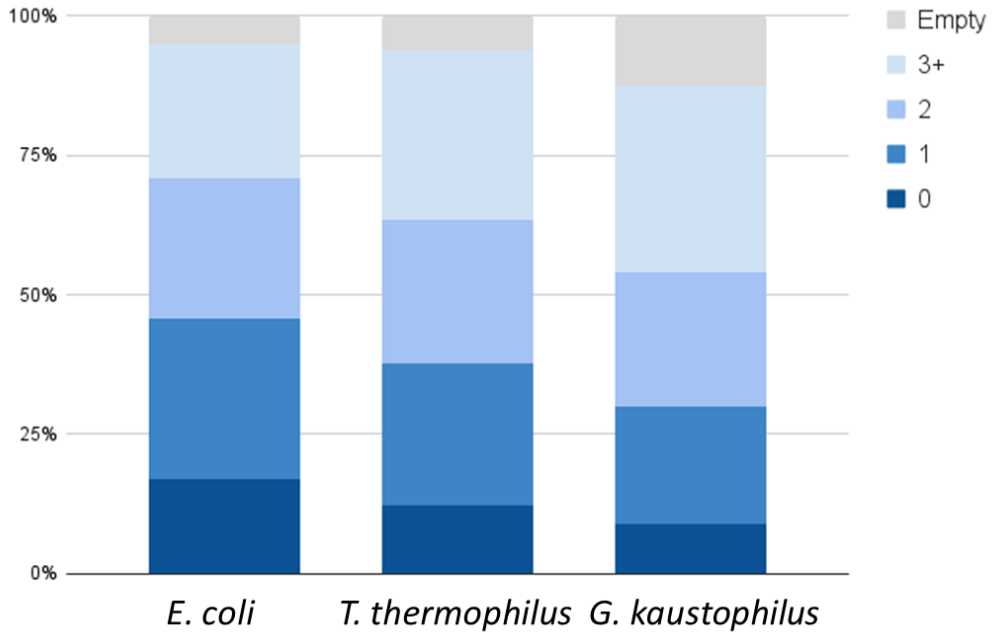


Figure 3.4 ***Mutational Distribution of S15 Leader Libraries.*** *Chart shows Hamming distance from wild type sequence for S15 leader libraries for the three species of interest.*

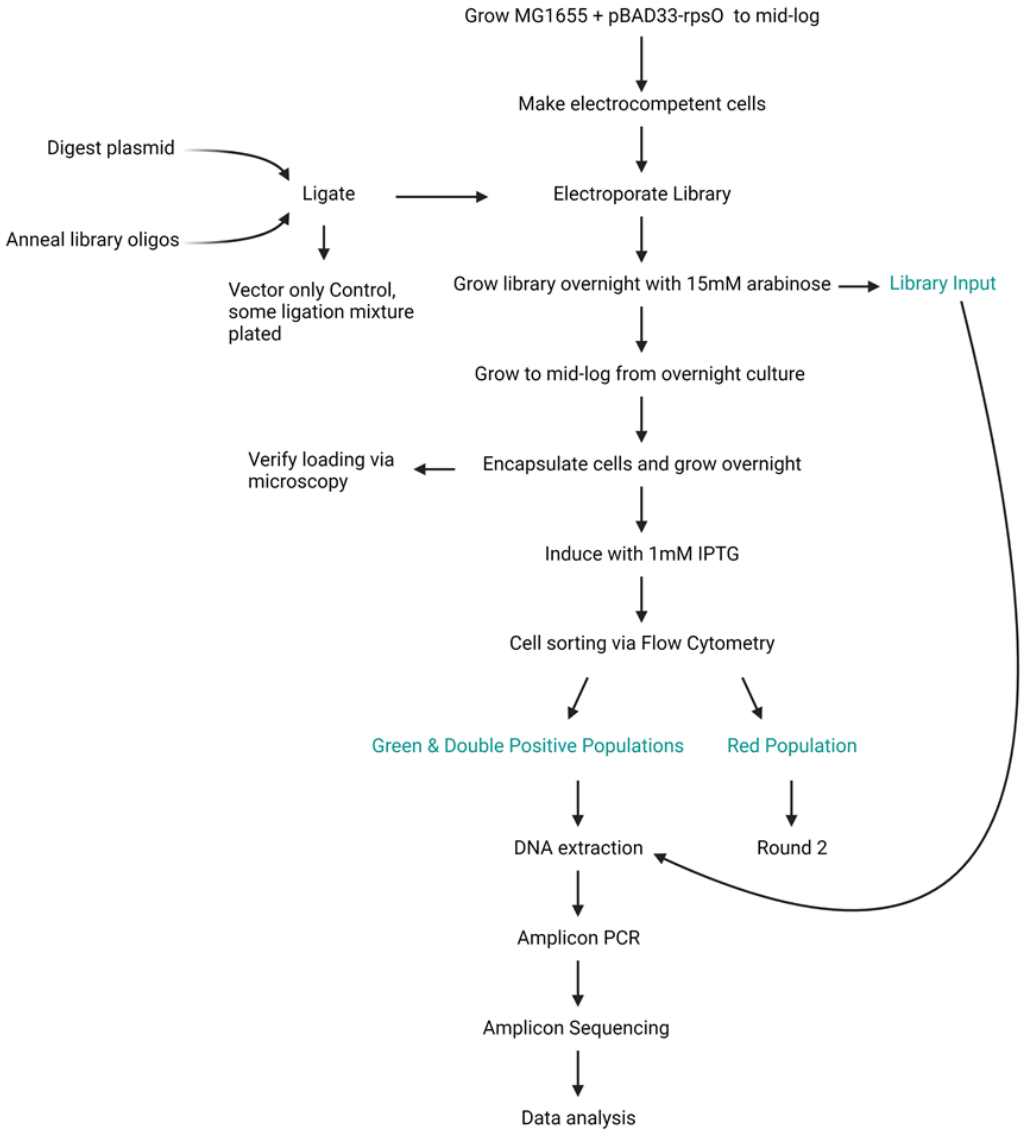


Figure 3.5 ***Flowchart of FADSRA Protocol from Library Ligation Through the First Round.***

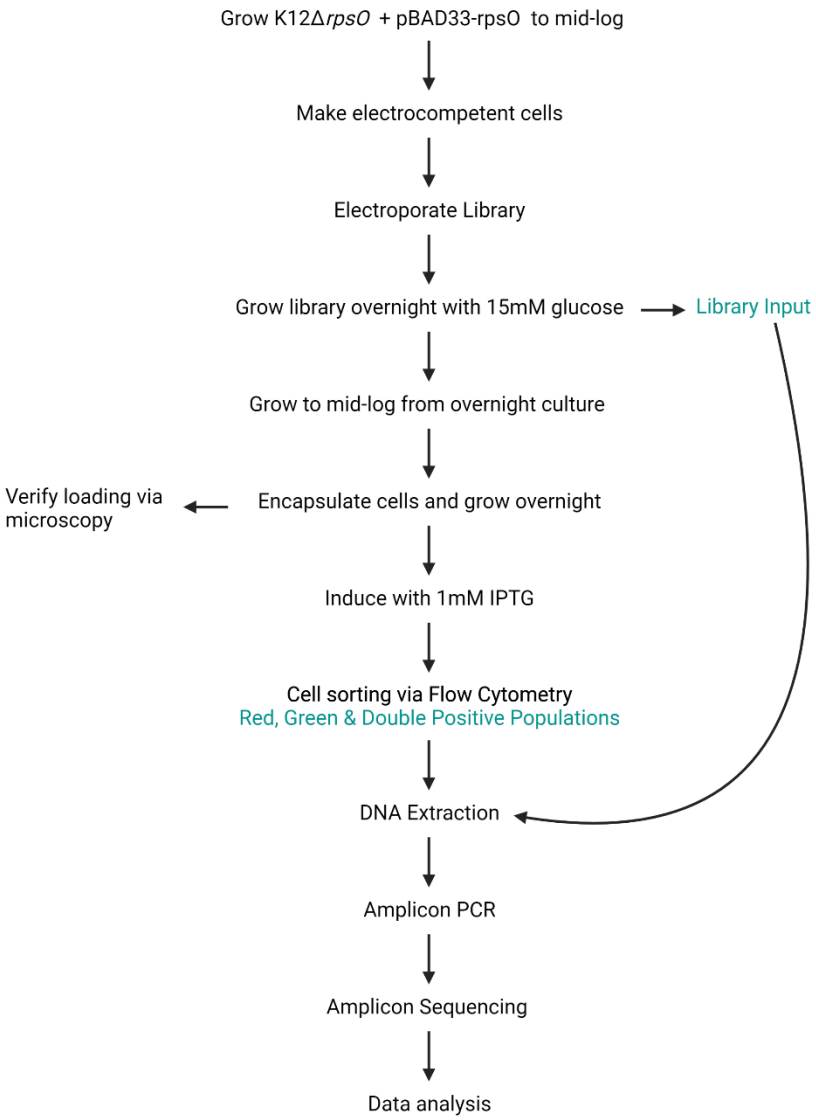


Figure 3.6 ***Flowchart of the Second Round of FADSRA***.

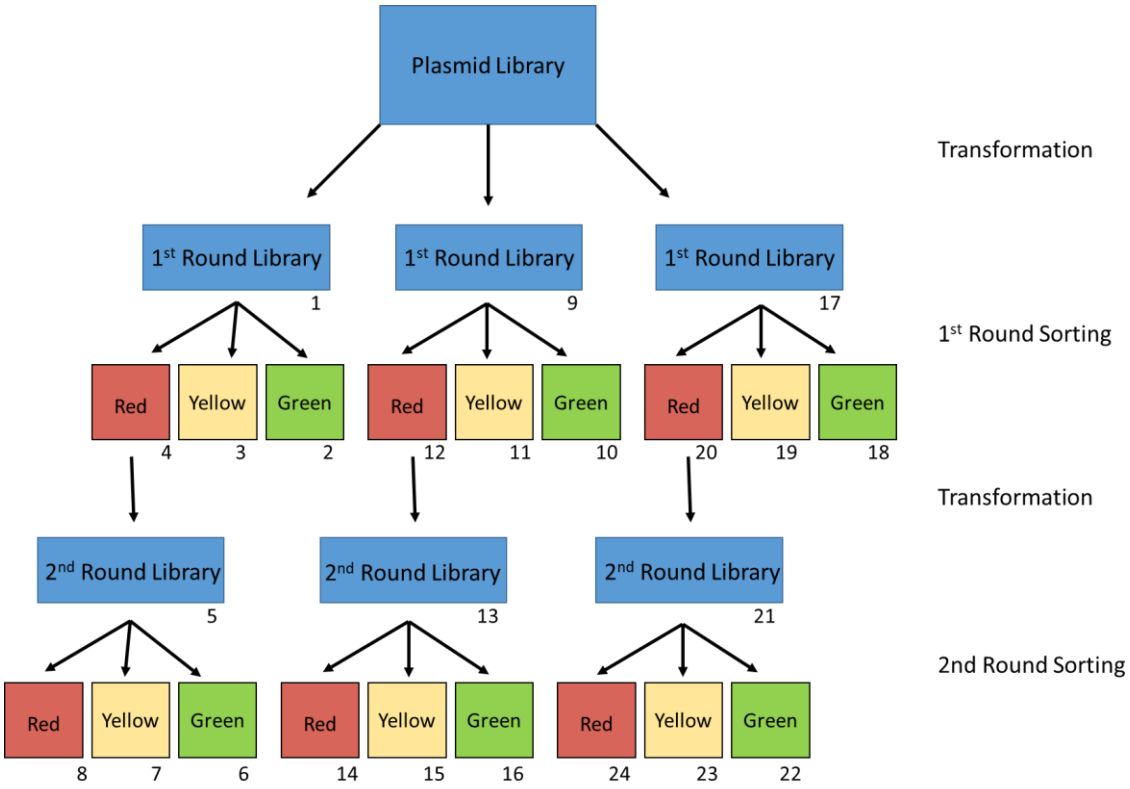


Figure 3.7 ***Sample Collection Schematic for*** **E. coli** ***FADSRA.*** *Numbers correspond to samples on Table 3.1. First round of sorting is in the presence of S15 protein, the second round is without S15. The color indicates expression of DsRed (Red), GFP (green), or both (yellow).*

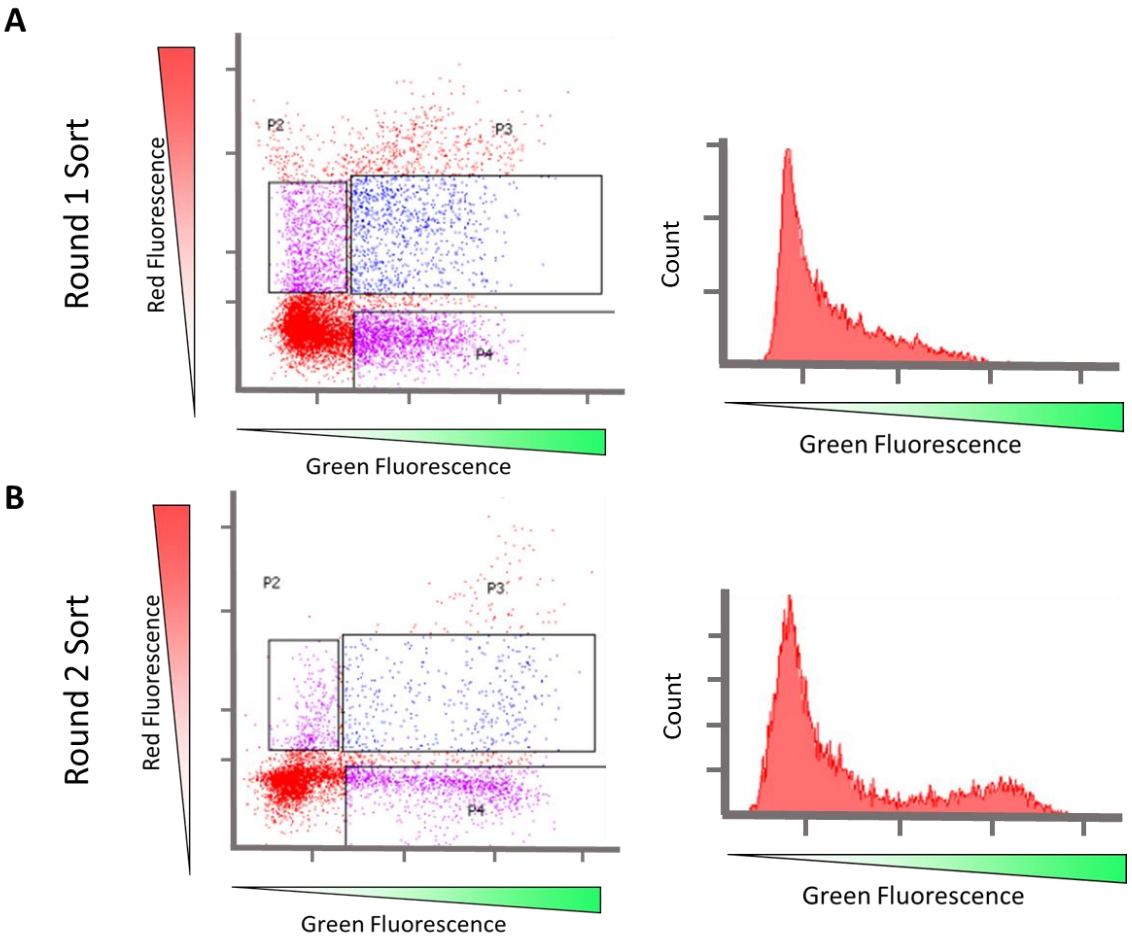


Figure 3.8 ***FADSRA S15 Library Flow Cytometry.*** *Library of* E. coli *S15 leader was cloned into the reporter plasmid and co-transformed with the protein overexpression plasmid. Droplet sorting of A) First round sort (in MG1655) with S15 protein expression and B) second round sort (in K15* ΔrpsO*) of S15 RNA leader library. Fluorescent axes use log10. Counts show every 100 (A) or 50 (B) droplets*

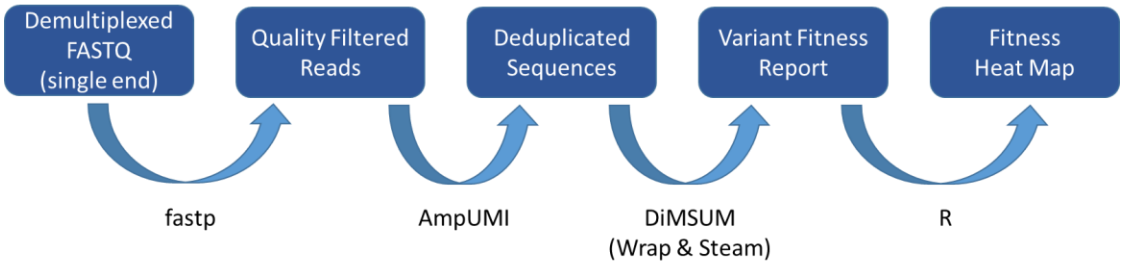


Figure 3.9 ***Illumina sequencing data analysis pipeline.***

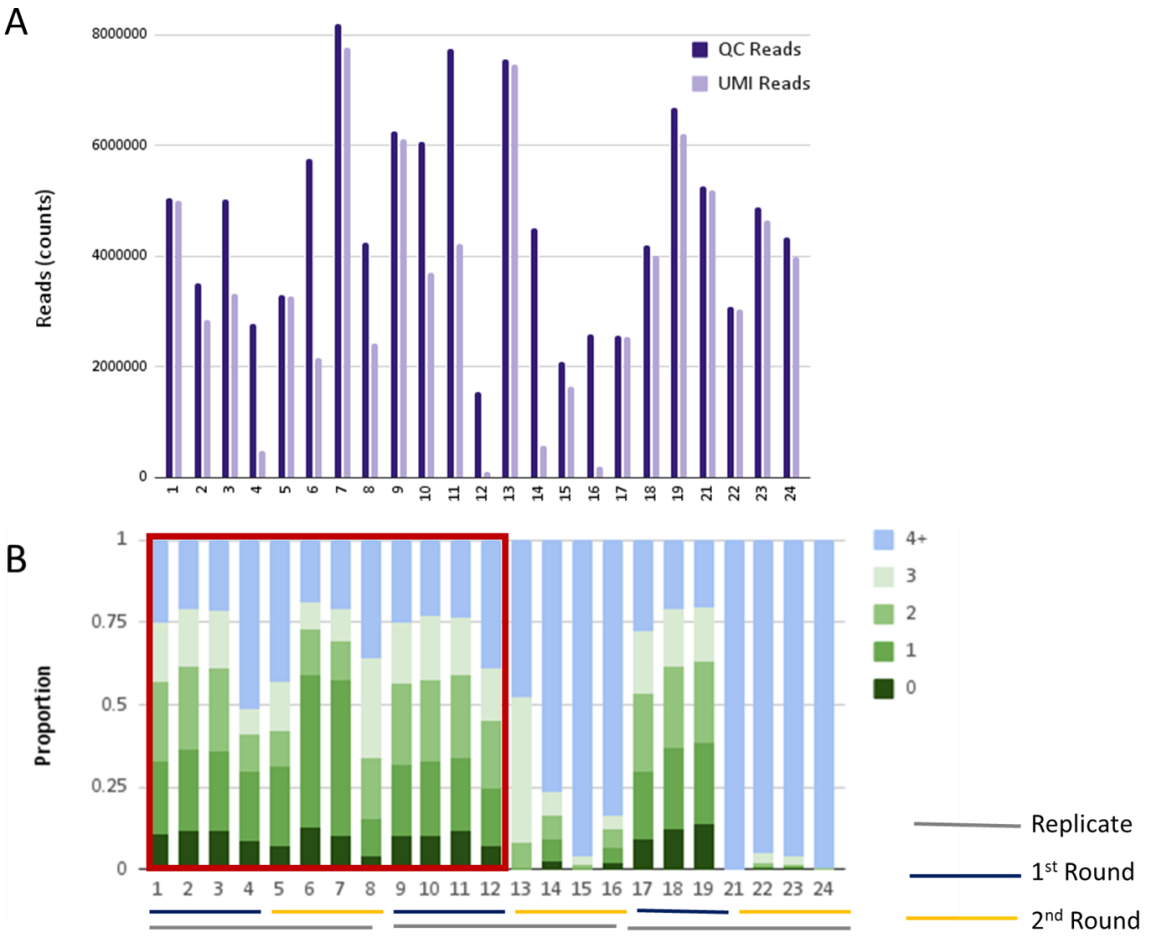


Figure 3.10 ***Illumina Sequencing Reads.*** *A) Illumina sequencing reads after fastp QC (dark purple) or after UMI deduplication (light purple). B) Hamming distance of sequences for each sample. Numbers correspond to samples in Table 2. Replicates are denoted by gray lines, 1st round samples by navy lines and 2nd round samples by yellow lines. Red box indicates samples included as input for DiMSum.*

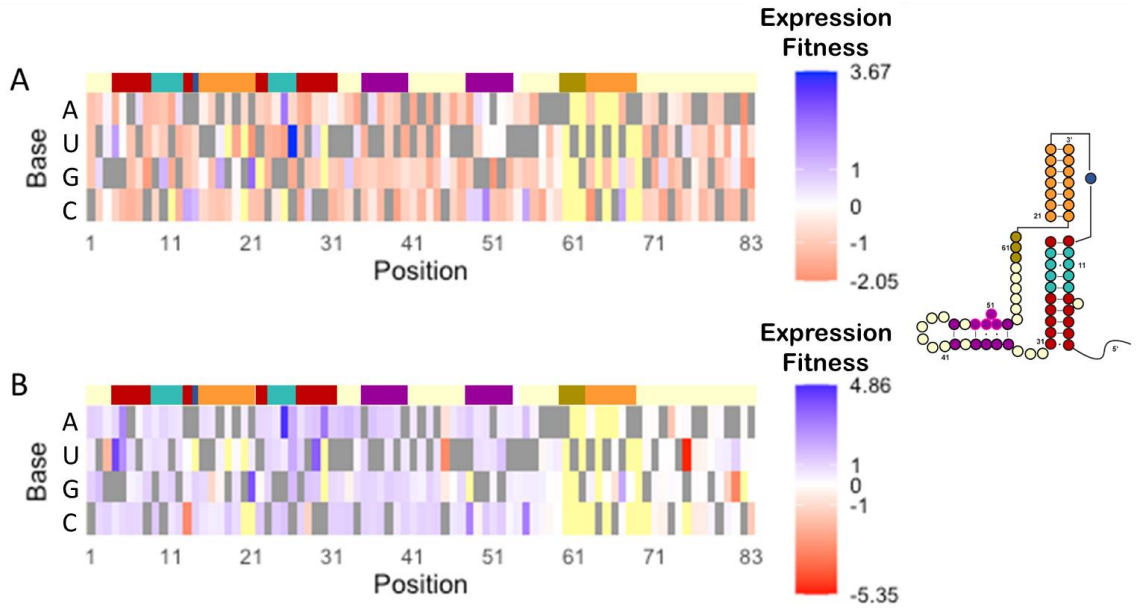


Figure 3.11 ***Regulatory Fitness Landscape of*** **E. coli** ***S15 Leader.*** *GFP expression fitness for Single Mutants in first round (+S15 protein) (A) or second round (-S15 protein) (B). WT sequence is in grey and unanalyzed positions are in cream. Expression fitness represents "GFP expression fitness" derived from the DiMSum fitness score with the sign adjusted to consistently identify positive values as associated with GFP expression.*

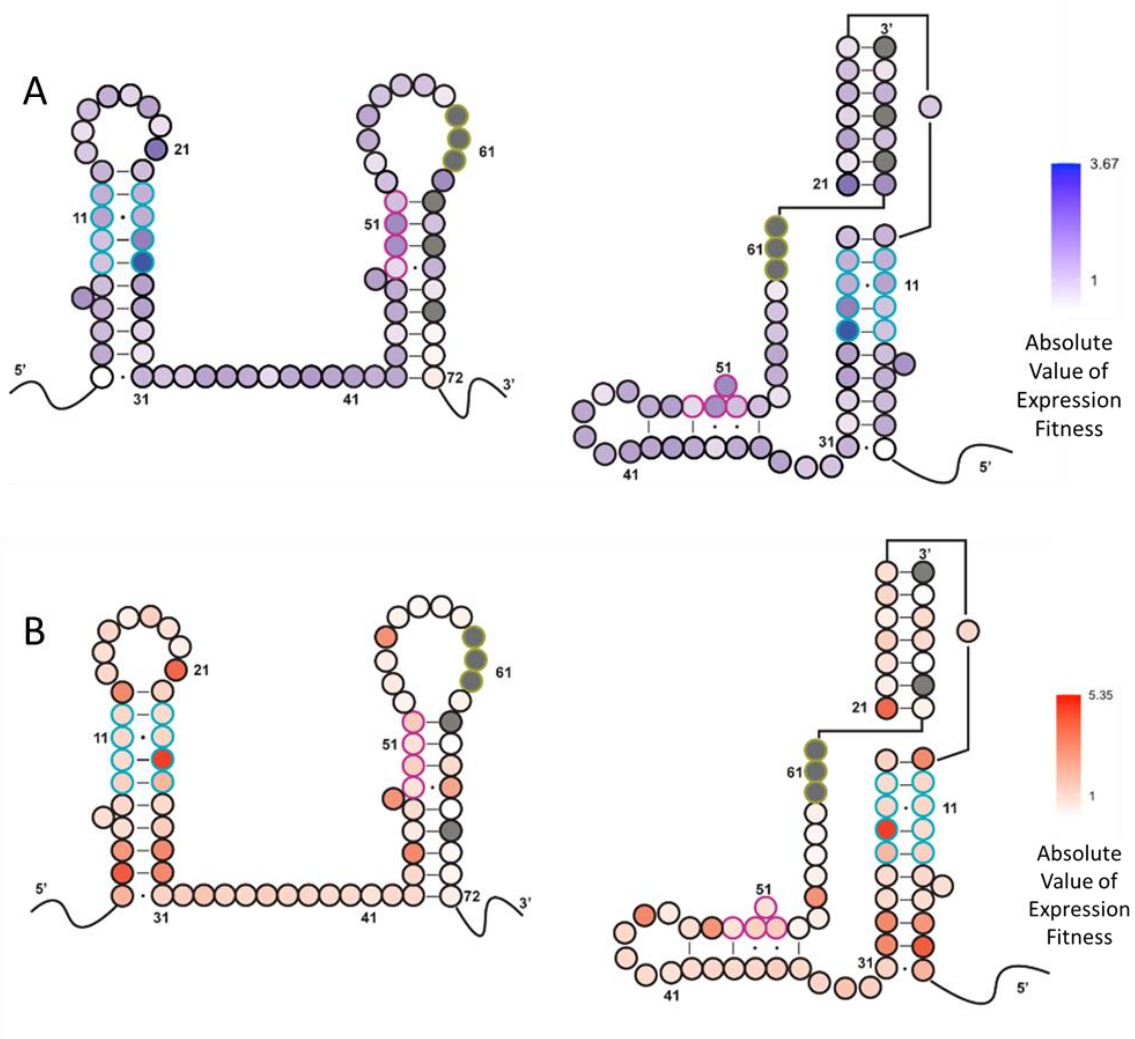


Figure 3.12 ***Absolute Value of Maximal Fitness Map.*** *Heat map of extreme absolute fitness for single mutants. First round (+S15 protein) (A) and second round (-S15 protein) (B). Unanalyzed positions are in grey. The start codon is outlined in gold, the ribosome binding site in pink and the G·U/G-C in teal. Heat map intensity corresponds to the absolute value of the largest fitness value across all mutations at each site for each condition.*

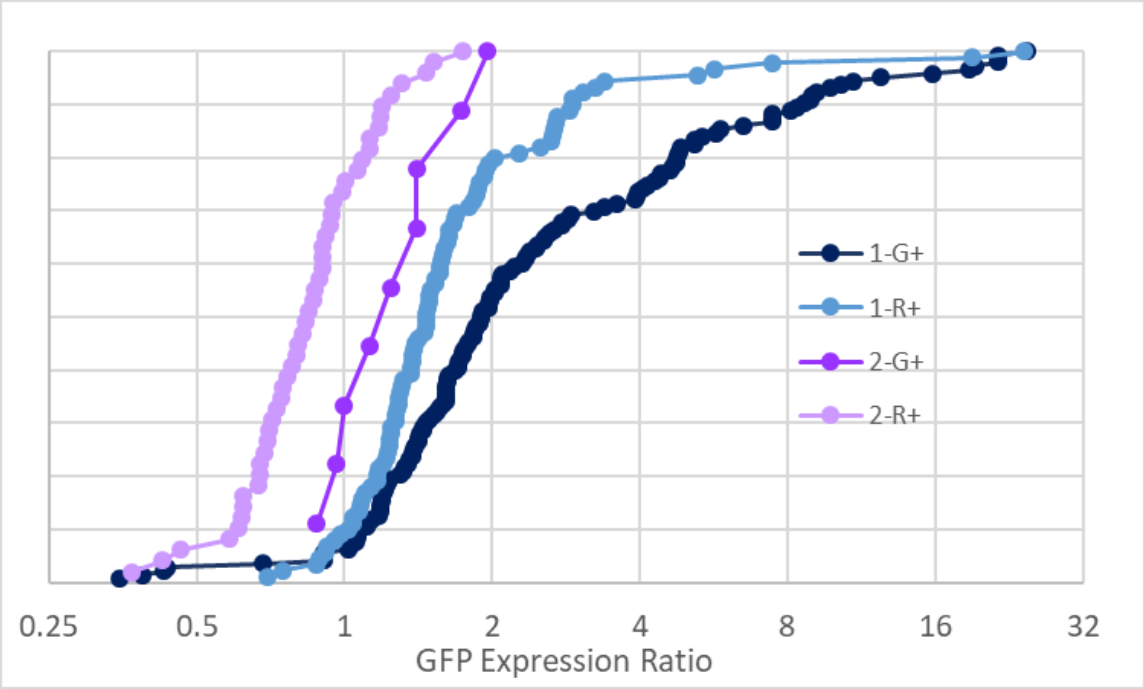


Figure 3.13 ***Fluorescent Plate Assay Validation of FADSRA.*** *Cumulative distribution graph of GFP Expression ratio for clones from indicated FADS gates as measured by bulk culture fluorescent plate assay. Expression ratio is the fluorescence normalized to OD600 divided by the average normalized fluorescence of the wild type controls. 1 and 2 indicate the FADSRA round and R+ and G+ indicate colonies from sort gates expressing GFP (G+) or DsRed (R+).*

Clone	Mutation(s)	Expression Ratio
B3	A51G, A58U	1.59
C4	A19C, U46A, AG5152CA	1.286
C11	A6C, U45A, U73A, U24C	4.04
E8	U37A, G70A	2.03
E9	U28G	1.686
E10	G25A, U48A	8.89
F2	G24C, U43G	19.08
F3	C26G, A57G	1.67
G7	A19C, U46A, A51C, G52A	0.6967

Table 3.2: ***Analysis of FADSRA Clones from First Round FADSRA Sorting.*** *Clones were Sanger sequenced and mapped to the leader with their corresponding GFP expression ratio from the fluorescence plate assays.*

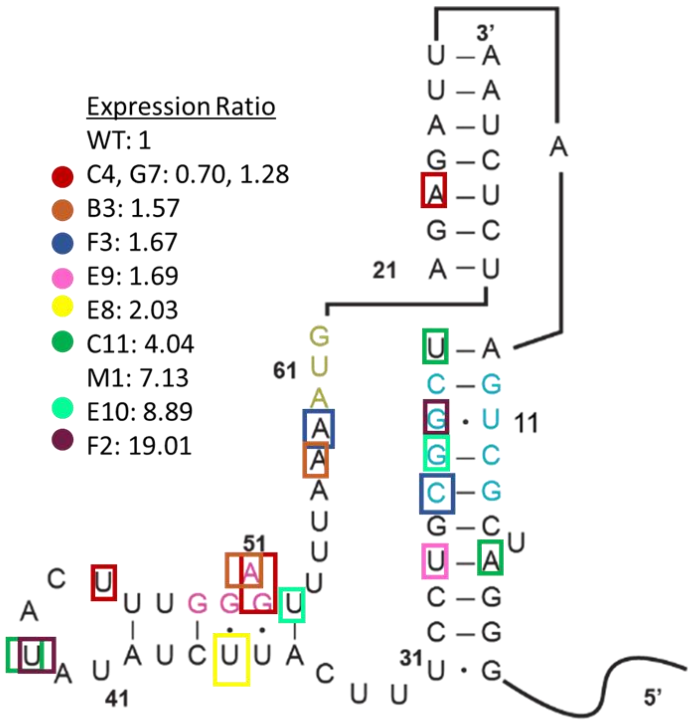


Figure 3.14 ***Sequenced Clone Mutations Mapped to the S15 Leader.*** *Colors indicate the mutation and numbers next to the mutation name are the expression ratio. Corresponds with Table 3.*

CHAPTER IV: S15 HOMOLOG ANALYSIS VIA FADSRA

Introduction

A valuable application of this work is to use FADSRA to create fitness maps of interspecies RNA-Protein parings to gain a better understanding of the evolutionary history and future of the S15 leader. Each S15 leader structure is narrowly distributed to its own clade. [62] This is in contrast to a structure like the L1 leader that is both found broadly as a regulator and also has a highly conserved structure, even between bacteria and archaea. Each of the S15 leaders has its own regulation profile and while there is overlap in protein binding and leader regulation, not all leader homologs can be regulated by each S15 protein. [63] Some structures seem more universal in function than others. Additional research found that structures capable of binding and, sometimes, regulating with S15 protein arise easily. [113] These findings were discovered through mutant analysis via β-galactosidase assay, or *in vitro* methods. FADSRA could be useful for further elucidating the plasticity of S15 RNA-protein regulation by increasing the number of mutants interrogated in a cellular context. In this chapter, the regulation of the reporter in droplet microcolonies is compared to the existing data from Slinger 2015 on the *in vivo* regulation of the parings of S15 homologs of *E. coli, G. kaustophilus, and T. thermophilus*.

Results and Discussion

In order to study the regulation of the *G. kaustophilus* and *T. thermophilus* proteins in *E. coli,* the native S15 protein must be taken into account. If native *E. coli* S15 is present in the genome, any repression observed would be a mixture of the overexpressed protein from the plasmid as well as the *E. coli* protein, leading to confounding results. Alternatively, the K12 Δ*rpsO* could be used as this strain has no genomic copy of *rpsO* and so any S15 protein in the cell would be from the pBAD33 plasmid. However, as previously

established in Chapter 2, the K12 Δ*rpsO* strain has severe growth defects compared to the more robust MG1655 strain. Therefore λ red recombineering was used to replace the native *rpsO* with an *rpsO* homolog followed by a kanamycin resistance cassette. [144] After the creation of the S15 homolog strains, it was observed that both the *T. thermophilus* and the *G. kaustophilus* recombineered strains had a growth defect compared to wild type, with the *G. kaustophilus* showing a much slower rate of growth than *the T. thermophilus* strain. *T. thermophilus* is a gram-negative organism like *E. coli*, more closely related than the gram-positive *G. kaustophilus*, so it is perhaps unsurprising that the growth rate of the *T. thermophilus* strain has less of a defect. Yet while the S15 leaders are diverse in structure, S15 protein binding to the rRNA is conserved. The defect in K12 Δ*rpsO* seems to be caused by the instability of the 16S rRNA in the absence of S15, its primary binding partner, leading to fewer functioning mature ribosomes. Thus, the growth defect is perhaps surprising, although the comparative severity of the defect is not.

To test the leader-protein pairs for FADSRA, the *T. thermophilus* and the *G. kaustophilus* reporter strains containing the same species leader in the reporter plasmid were analyzed by flow cytometry with S15 protein expression to look for GFP repression in the WT leader and GFP expression in M1, a mutation designed to disrupt the structure of each specific leader (Figure 4.1, Figure 4.2). For the *T. thermophilus* leader, there is a clear shift to higher green fluorescence in the case of the M1 leader. However, this shift is not as large as the one seen between the *E. coli* wild type leader and mutant. For the *G. kaustophilus* strain with the same species leader, the results show poor GFP expression in both the wild type and the M1 mutant compared to the other strains, but a shift to higher GFP expression was evident in the M1 compared to the wild type leader.

Interspecies pairings between the leader and the mutant were also examined. The *T. thermophilus* and the *G. kaustophilus* leaders with the *E. coli* protein resulted in some repression, but not as much repression as observed for any of the same species pairings. The results were similar for the *T. thermophilus* protein with the *G. kaustophilus* leader. The *T. thermophilus* protein was able to repress the wild type, but not the mutant *E. coli* leader. The *G. kaustophilus* protein with *E. coli* leader was as repressive as the *E. coli* same species expression, but showed elevated expression when paired with the *T. thermophilus* leader. Comparing the M1 mutant leaders, the *T. thermophilus* leader is repressed by *G. kaustophilus* S15, but the *E. coli* leader is not. The *G. kaustophilus* leader is partially repressed by *T. thermophilus* protein.

Taken together, these results are largely consistent with the previous analysis of these pairings previously established by β-galactosidase assay, [63] adding validity to FADSRA analysis. Notable differences include only partial repression of the *T. thermophilus* and *G. kaustophilus* leaders (wild type and M1) by *E. coli* protein and the partial repression of the *E. coli* leader by the *G. kaustophilus* S15. Certainly, the FADSRA analysis is less quantitative and sensitive than the β-galactosidase assay, but does provide similar information. Adjusting for changes in protein expression level and the growth defect, the *T. thermophilus* components are good candidates for FADSRA analysis. Conversely, the *G. kaustophilus* components show more differences from the previously established regulation, perhaps due to the growth defect introduced by replacing the native *rpsO* gene. However, FADSRA could still be possible with some additional optimization to compensate for the difficulty of making the protein.

Future regulatory landscapes of the *T. thermophilus* and *G. kaustophilus* leaders as well as interspecies pairings would provide additional insight into the flexibility of these leaders and how these leaders may be evolutionarily maintained. Of the known S15 structures, each has an element conserved from the rRNA binding, but additional variations. It may also provide insight into the plasticity of S15 protein binding and repression, how leaders to develop new interactions, or if interactions are relatively fixed once a regulator is present.

Materials and Methods

Library Design & Cloning

The pEG10 plasmid was digested with SalI and EcoRI. The ligation of the S15 leaders into the pEG10 plasmid was performed with Quick Ligase according to manufacturer's instructions. Before transformation, the plasmid was digested with BplI to linearize any plasmid without insert. pBAD33-*rpsO* overexpression plasmids were previous generated (Slinger 2015).

Droplet Culture and Sorting

The FADSRA steps of encapsulation, microcolony culture, induction and droplet sorting were performed as in Chapter 3.

Recombineering

To create *E. coli* strains with *rpsO* homologs from other species integrated into the genome, plasmid based λ red recombineering was used to integrate linear DNA into the genome through homologous recombination. MG1655 cells were made z-competent and

transformed with recombineering plasmid pSIM5. Unless curing the plasmid, cells with pSIM5 were grown with shaking at 30 ˚C. Linear PCR product containing 50 nucleotides of homology to integration site in the genome on each of the 5' and 3' ends, the *rpsO* homolog, and a kanamycin resistance cassette was created by amplifying the *rpsO* and the kanamycin, then assembling the two pieces using Gibson Assembly. The resulting piece of DNA was electroporated in the MG1655 strain +pSIM5. To induce recombineering, cells were grown to midlog in SOB +chloramphenicol then transferred to a shaking water bath at 42 ˚C for 15 minutes. Cells were then rapidly cooled and made electrocompetent as above. The purified linear DNA was then electroporated into the cells. Cells were recovered for 1 hour and then plated onto selective LB agar plates with kanamycin. To cure the recombineering plasmid, overnight cultures of the recombinants were diluted 1000-fold and grown for five hours with shaking at 37 ˚C then plate onto LB agar plates. Resulting colonies were patch plated to check for resistance to kanamycin and susceptibility to chloramphenicol, indicating successful integration of the kanamycin resistance cassette and successful plasmid curing.

Primer Table

λ Red	2030-GkrpsoGib-F	CTGAATTAGAGATCGGCGTCCTTTCATTCTATATACTT TGGAGTTTTAAA atggcattgacgcaagagcg
	2031-GkrpsoGib-R	agaaatcatccttagcgaaagctaaggatttttttttat ctg ttatcgacgtaatccaagtttctc
	2032-TtrpsoGib-F	CTGAATTAGAGATCGGCGTCCTTTCATTCTATATACTT TGGAGTTTTAAA atgcccatcacgaaggaagag
	2033-TtrpsoGib-R	agaaatcatccttagcgaaagctaaggatttttttttat ctg ttaaccccggatgcccagc
	2034-KanGib-F	agataaaaaaaatccttagctttcgctaaggatgattt ct TATGGACAGCAAGCGAACCG
	2035-KanGib-R	CAGCTTGAAAAAAGGGGCCACTCAGGCCCCCTTTTCTG AAACTCGCAAGAA TCAGAAGAACTCGTCAAGAAG
G. kaustophilus S15 leader	2055-GkS15-Eco-F	CAG GTCGAC TCAATGTATGCGAACCATTGCTTG
	2056-GkS15-Sal-R	CAG GAATTC GTTCAATGCCATCCTTGTTCAC
T. thermophilus S15 leader	2057-TtS15-Eco-F	CAG GTCGAC AGGCTTGGCGGGAGACC
	2058-TtS15-Sal-R	CAG GAATTC CTTCGTGATGGGCATGTTTTCC

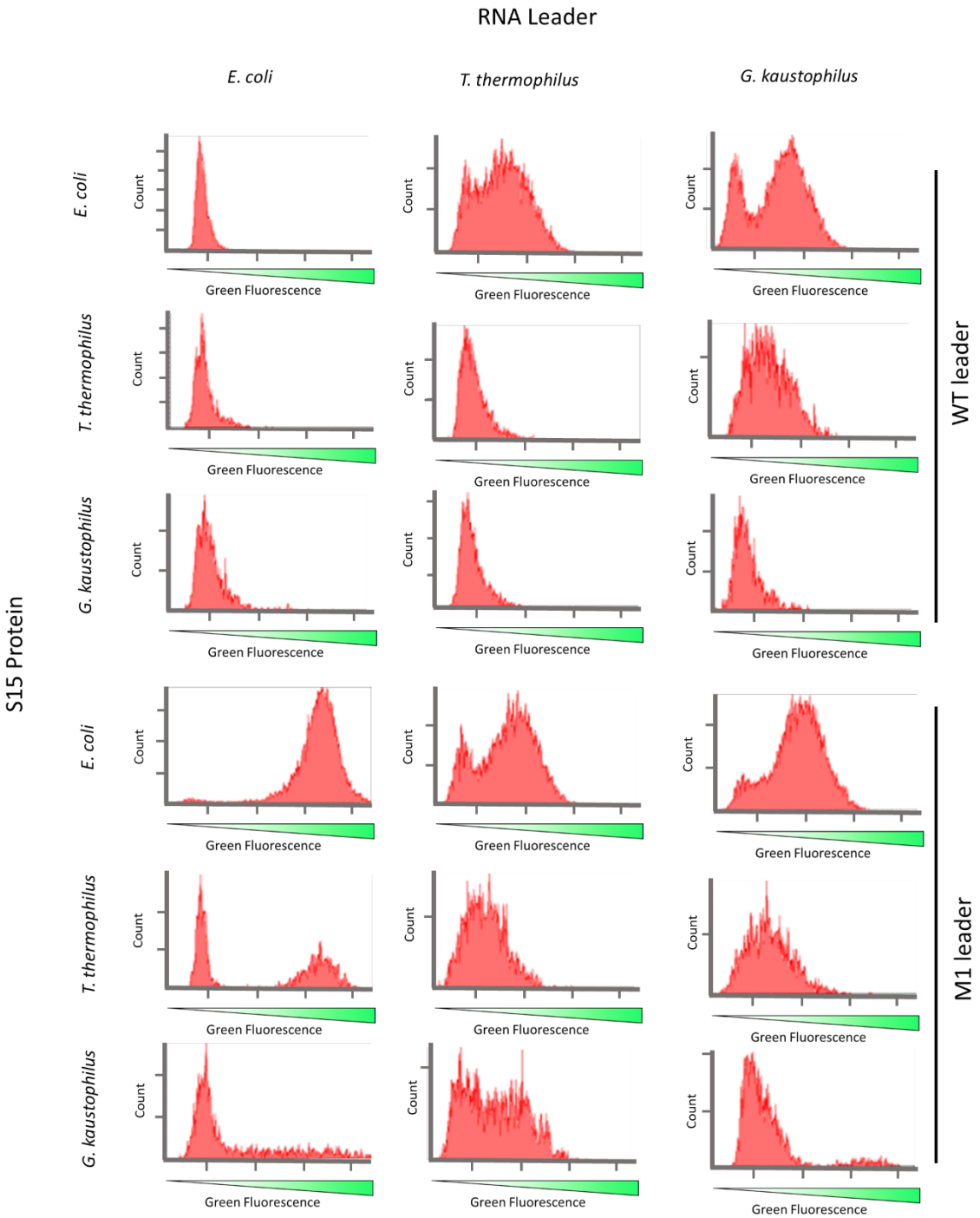


Figure 4.1 ***Flow Cytometry of Interspecies Histograms of RNA-Protein Pairs for WT and M1 Leader.*** *Droplets cultured with induction of S15 protein in the MG1655 parent strain with noted pBAD33-rpsO and pEG9-S15. MG1655 was recombineered to replace the native rpsO with* G. kaustophilus *and* T. thermophilus rpsO *where indicated. Histograms show counts of droplets per fluorescent value. Scale is every 50 for* E. coli *protein strains and every 25 for the remaining strains.*

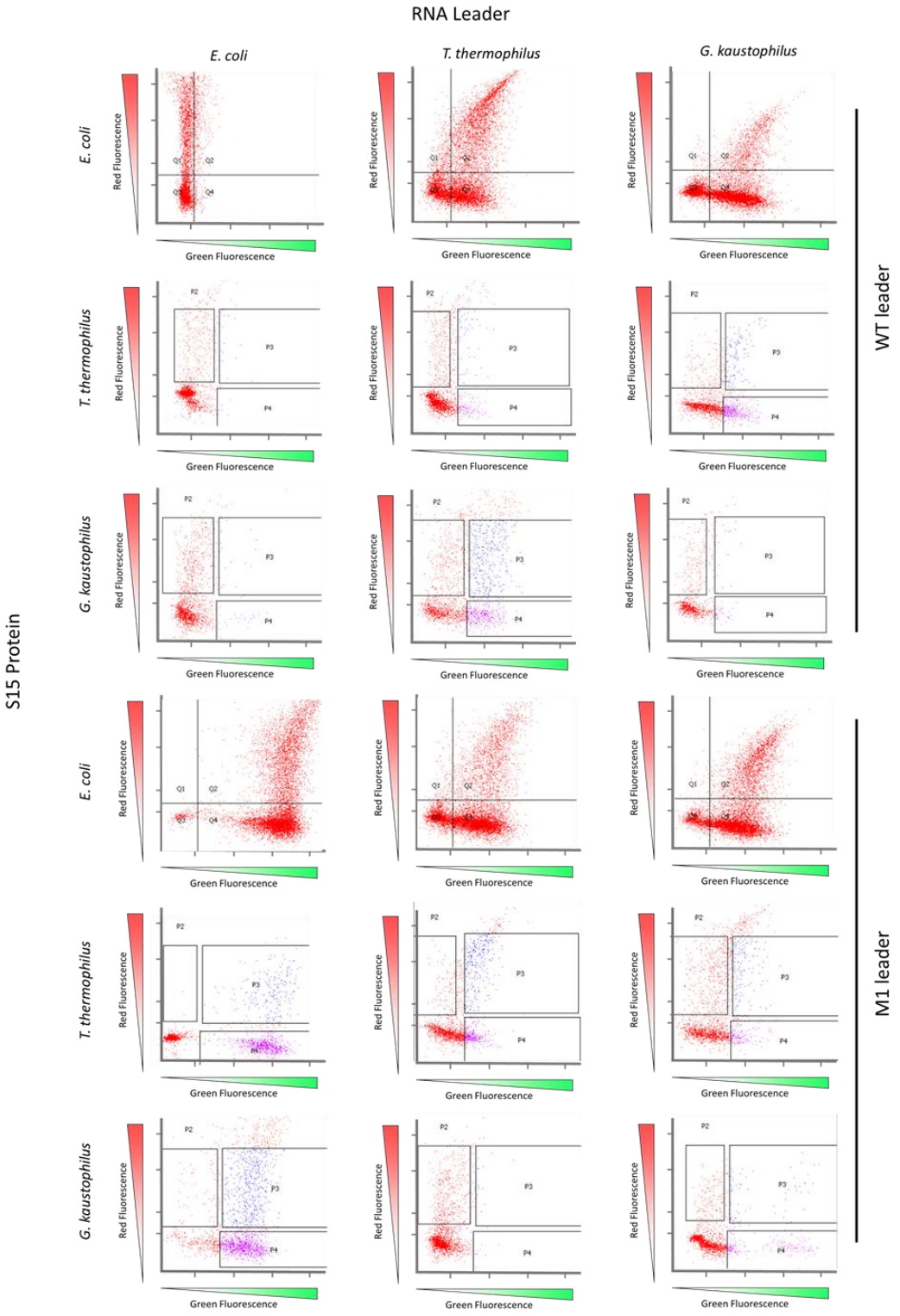


Figure 4.2 ***Flow Cytometry of Interspecies Fluorescent Plots of RNA-Protein Pairs for WT and M1 Leader.*** *Droplets cultured with induction of S15 protein in the MG1655 parent strain with noted pBAD33-rpsO and pEG9-S15. MG1655 was recombineered to replace the native rpsO with* G. kaustophilus *and* T. thermophilus rpsO *where indicated. Flow cytometry plots show fluorescence of each event.*

CHAPTER V: FADSRA BEYOND RIBOSOMAL LEADERS

Introduction

While the work in this thesis has so far focused on the autogenous regulation of the S15 leader, this method was developed with flexibility and broad applicability kept in mind. In this chapter, I demonstrate and discuss several brief, proof of concept experiments that explore applications of FADSRA including the L1 leader, another ribosomal leader, SELEX selected S15 binders, and development of the assay for use with riboswitches.

Results & Discussion

Other Ribosomal Regulators

Due to the current lack of high throughput *in vivo* assays for ribosomal regulators, it was of special interest to test FADSRA compatibility with other ribosomal leaders. The *E. coli* S15 regulator was specifically chosen because of the large difference in expression, 15-fold as measured by β-galactosidase assay, between repressive and permissive conditions. Additionally, the depth of literature available about the *E. coli* S15 leader made for an ideal example to test the proof of concept of the FADSRA assay. However, testing a different leader would allow evaluation of FADSRA utility for other ribosomal leaders.

For this purpose, I chose to focus on the *E. coli* L1 leader. As discussed previously in Chapter 1, the conservation of the L1 leader structure provides an interesting contrast to the diversity of the S15 leader homolog structures. First, the regulation of the L1 leader and non-binding mutant (M1) mutant were measured via β-galactosidase assay in a K12 Δ*rplA* strain. The β-galactosidase assay showed 6-fold repression by the wild type leader, about half of the fold-repression demonstrated by S15 in the same assay (Figure 5.1). Next, to test the L1 leader for FADSRA, the L1 and L1 M1 leaders were cloned into

reporter plasmid pEG9 and transformed into K12 Δ*rplA* cells, encapsulated with expression of the repressive protein and analyzed via flow cytometry (Figure 5.2). The results show that there is a difference in expression between the L1 WT and the M1 leader. However, the difference between the two conditions is less than the difference between S15 WT and S15 M1. Additionally, the level of expression for the L1 leader with an abundance of the repressive L1 protein is higher than that of S15. This suggests that the difference in L1 fold-repression compared to S15 is due in part to leakier repression. Notably, in the presence of the repressor protein, the S15 leader-*gfp* fusion strains are indistinguishable from empty vector with no fluorescence while for L1, GFP expression, even with overexpression of the repressor protein, is extremely high compared to S15. Considering these results, it seems like FADSRA may work for the study of ribosomal leaders. However, the higher expression under repressive conditions may cause a decrease in assay sensitivity. Furthermore, this test case demonstrates the need for optimizing gating for the particular leader of interest. Still, taken together these results indicate that FADSRA could be applicable to the study of the *in vivo* regulation of other ribosome leaders with minimal changes to the reporter plasmid or workflow.

SELEX

For *in vitro* studies, validation of binders as biologically relevant regulators is typically done with a β-galactosidase assay or similar. Considering that SELEX libraries can start on the order of 10^6 with larger libraries being common, and end with scores of effective binders, validation of SELEX libraries can be very time intensive and also risks under sampling of the population. FADSRA could be an excellent assay to quickly validate or provide an additional selection step in between SELEX rounds. Additionally, FADSRA

can not only be used to validate sequences after enrichment for verification of a final pool of binders, but could be used to monitor shifts in regulation over the rounds of selection. This is similar to how one might use a filter binding assay to monitor changes in the binding of protein to the RNA. The advantage of this approach is that the RNA population can be monitored or selected for regulation in addition to *in vitro* binding.

To test the use of FADSRA as a proof of concept for SELEX validation, I collaborated with another graduate student, Daniel Beringer, using SELEX to better understand the binding sphere of the *E. coli* S15 regulator. Briefly, this experiment started with a partially patterned library of alternating purine and pyrimidines, while preserving the Shine Delgarno and the first six codons of the *rpsO* gene. This library was input for 11 rounds of SELEX with decreasing amounts of the *E. coli* S15 protein. After 11 rounds, a specific binding curve with a biologically relevant K_D was apparent in the filter binding assay and amplicon sequencing revealed that a small number of sequences were greatly enriched after the 11^{th} round of SELEX (D. Beringer, unpublished data). SELEX sequences from both the initial pool and the 11^{th} round were cloned into the pEG10 FADSRA reporter assay. Each pool was encapsulated and examined with flow cytometry with expression of *E. coli* S15 protein to observe any shifts in GFP reporter expression (Figure 5.3). The results show that most of the leaders in the SELEX-1 library do not repress GFP, although there are a few droplets not expressing GFP. This is interesting because it could indicate a significant number of binders already present in the initial pool. The 11^{th} round SELEX library shows a marked increase in the proportion of droplets not expressing GFP, suggesting that the majority of sequences in the transformed SELEX-11 library repress GFP in some fashion. However, not all the leaders repress GFP, as there is a distinct

population of droplets with GFP expression. The libraries were not assessed under conditions without protein, so it is unknown if the droplets not expressing GFP are true regulators or always off. For this brief experiment, the potential utility of FADSRA for validation of *in vitro* binders is clear.

Riboswitches

Most of the development of this assay focused on ribosomal leaders since there is a lack of high throughput data generated for this class of cis-regulators. However, for riboswitches, high throughput methods of in cell assay are more widely used. [145–147] Even so, FADSRA, if adapted to riboswitches, could be another useful high throughput tool to study riboswitches in a cellular context. The primary constraints for riboswitch controlled regulation are much different than those of ribosomal leaders. For ribosomal leaders, the biggest concern regarding growth is balancing the need for the cell to have mature ribosomes available for protein translation with the conditional expression of the ribosomal protein of interest. Yet, for riboswitches, the concern is one of nutrition or toxicity. Most riboswitch ligands are small molecules that are produced by the cell or imported from the environment; thus to study riboswitch activity, cells are cultured in defined and sometimes minimal media. Therefore cell growth is restricted either by the minimal nutrients in the media or the toxicity of the ligand.

Given the unique growth environment of the agarose microdroplet, the first step to adapting FADSRA for the study of riboswitches was to examine microcolony growth in minimal media agarose droplets. XL-1 cells were encapsulated in either SOB medium without any amendments or M9 minimal media with iron sulfide and glucose supplementation (Figure 5.4). Likely due to the limited ability of the microcolony to

expand past a certain number of generations, the size of the microcolonies after overnight incubation in each media was very similar between the two media types.

Once the ability of cells to grow in minimal media within an agarose droplet was confirmed, my attention turned to selecting riboswitches to test. In *E. coli*, many riboswitches have a mixed mechanism with transcriptional and translational regulation. Riboswitches were selected to include diverse properties: considerations included riboswitch length, regulatory mechanism, and regulated pathway (Table 5.1). Initially, four riboswitches were tested by fluorescent plate assay for changes in *gfp* reporter expression in the presence and absence of ligand (Figure 5.5, [148–154]). The lysine (*lysC*) and the mini-ykkC (*sugE*), both on switches, did not show any changes in reporter expression regardless of ligand. However, the *thiM* TPP riboswitch, an off switch, did show a concentration dependent decrease in fluorescent reporter expression.

The TPP riboswitch is found broadly in bacterial families and regulates a pathway related to carbohydrate metabolism. It is of interest because the regulation differs between gram-positive and gram-negative organisms, likely due to global differences in regulation resulting from the degree of transcription-translation coupling between clades. This riboswitch is off when the ligand TPP is bound, sequestering the ribosome binding site and the start codon for *thiM* in a terminator stem loop. One hurdle specific to this riboswitch is that thiamine, which can be converted proportionally into the ligand TPP, is critical for cell growth and metabolism.

For the most promising candidates, the TPP and FMN riboswitches, bulk liquid culture plate assays were used to assess the reporter function prior to flow cytometry. MG1655 cells were transformed with the pEG10-EcTPP, or pEG10-EcFMN and

pBAD33-*ribU,* and grown to mid-log. Strains with the non-regulating M1 mutant were also transformed for each riboswitch. Once at mid-log, the cells were washed into media that had varying amounts of ligand, and GFP expression was induced overnight (Figure 5.6). The FMN riboswitch shows a response to the ligand, with expression decreasing as the amount of ligand is increased. However, the TPP switch shows low expression even in the absence of thiamine, compared to the TPP M1 mutant. This could relate to the requirement of thiamine for robust cell growth and protein translation.

One major concern about the analysis in riboswitches is the metabolic state of the cells in the microcolonies. Instead of liquid cultures, where cells are assumed to be metabolically homogenous, the microcolony is composed of three-dimensional spheroid structures more similar to a colony growing on a plate or a biofilm. This means that the state of individual cells will be variable, even as the droplets work to homogenize other aspects of culturation and sorting. Thus FADSRA may not be ideal for riboswitches that require the cells to be in a precise metabolic state for accurate characterization of protein expression. However, cells on the outside of the microcolony are likely to be reasonably homogenous and will still help amplify the reporter signal. This balance likely requires some optimization of growth and for the assay to accurately reflect the nature of the riboswitch.

It is clear from these results that the FADSRA system as currently designed requires additional optimization for use in riboswitches compared to ribosomal leaders other than S15. This result is perhaps unsurprising given that FADSRA was specifically designed with and for ribosomal proteins. However, the general concept of the assay remains sound and adaptable to cis-regulators beyond ribosomal leaders.

Overall, the results of these preliminary experiments indicate that FADSRA has promise to be broadly applicable to the study of ribosomal leader proteins in multiple contexts. However, applying this method to riboswitches may require more optimization or even the redesign of the reporter construct itself. Additionally, methods already available for riboswitches may be preferable even if FADSRA were optimized to work with riboswitches.

Methods

Cloning and Strains

The MG1655 strain was obtained from the Babak Momeni lab. XL-1 cells were purchased from NEB. K12 Δ*rpsO E. coli* strains were cultured in either Lysogeny Broth or Super Optimal Broth. Antibiotics and carbon sources, when present, are in the following concentrations unless explicitly stated elsewhere: 100 µg/mL ampicillin, 34 µg/mL chloramphenicol, 15mL L-L-arabinose, 15mM glucose. The strain for SELEX studies was MG1655 +pBAD33-*rpso* +pEG10-S15 SELEX. For riboswitch studies, MG1655 was the parental strain. Riboswitches were generated using PCR with XL-1 genomic DNA template, digested with EcoRI and SalI (except in the case of TPP, which was digested with EcoRI/XhoI due to an internal SalI site), and ligated into the pEG5 plasmid. For riboswitch growth studies, the minimal media used was M9 + 2.5 µg/mL FeSO4 + 0.4% glycerol. "100%" ligand concentration was considered to be 20 µM thiamine, 467 µM riboflavin, 1.25 mM guanidine chloride, 4 mM lysine. 1.4 µM thiamine was added for low thiamine conditions.

β-galactosidase Assay

E. coli were picked from plate and inoculated into 2 mL of LB +ampicillin +chloramphenicol. Cultures were grown overnight at 37 °C with shaking. The next day, cultures were diluted 1:1000 in fresh media grown to mid-log (A600 0.4-0.8) then induced with 1mM IPTG for 30 minutes. 1 mL of culture was pelleted and resuspended in z-buffer (50 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, 100 μg/mL spectinomycin) and the A600 recorded. Cells were permeabilized with toluene. ONPG was added and the time interval recorded from the time the ONPG was added to the time the solution turned yellow and was stopped with Na_2CO_3. A420 was recorded after pelleting cells. Miller units were calculated as 1000 * (A420/(Δt(min) * A600 * volume (mL))).

Encapsulation and Flow Cytometry

Cells were encapsulated, cultured in microcolonies and analyzed via flow cytometry as in Chapter 2 with the following exceptions. Protocol was the same for SELEX and L1 strains. For riboswitch strains, the agarose medium was made with M9 with 1 mM IPTG.

Plate Assay

Cultures were grown overnight in 0.5 mL LB +chloramphenicol +ampicillin +L-arabinose or glucose in a 24 well plate. The following day, the cultures were diluted out 1:100 into M9 media and grown for five hours. After five hours, cultures were induced with 1 mM IPTG and incubated overnight. Cultures were then washed into PBS and 200 μL in triplicate for each sample was added to a 96 well black optical plate. The plate

was read by a plate reader and the OD600, DsRed (ex:558 em:583) and αGFP (ex:397 em:506) or sfGFP (ex:485 em:510) was measured.

Primer Table

Riboswitches	1925-EcoRI-EcFMN-F	CAG GAATTC GCTTATTCTCAGGGCGGG
	1926-SalI-EcFMN-R	CAG GTCGAC GGAAAGTAGCGTCTGATTCAT
	1927-EcoRI-EcTPP-F	CAG GAATTC CTCTGCGATTTATCATCGCAAC
	1928-XhoI-EcTPP-R	CAG CTCGAG GTGTAACGCGTGCGCAGATT
	1929-EcoRI-Ecmykkc-F	CAG GAATTC TCCGTTCAACCATCAGCTTTG
	1930-SalI-Ecmykkc-R	CAG GTCGAC CCAGGACATATCAGGCTCC
	1931-EcoRI-EcLys-F	CAG GAATTC ACTACCTGCGCTAGCGCAG
	1932-SalI-EcLys-R	CAG GTCGAC GGAGACAACAATTTCAGACATAAC
	1948-XhoI-EcTPP R2	CAG CTCGAG GGTTTGCACCACATCATTGGTC
SELEX library	2094-SX-S15-Sal-F2	CAG GTCGAC CGTAGTCGTAGCTGATC
	2095-SX-S15-Eco-R2	CAG GAATTC TTCAGTACTTAGAGACAT

Figures

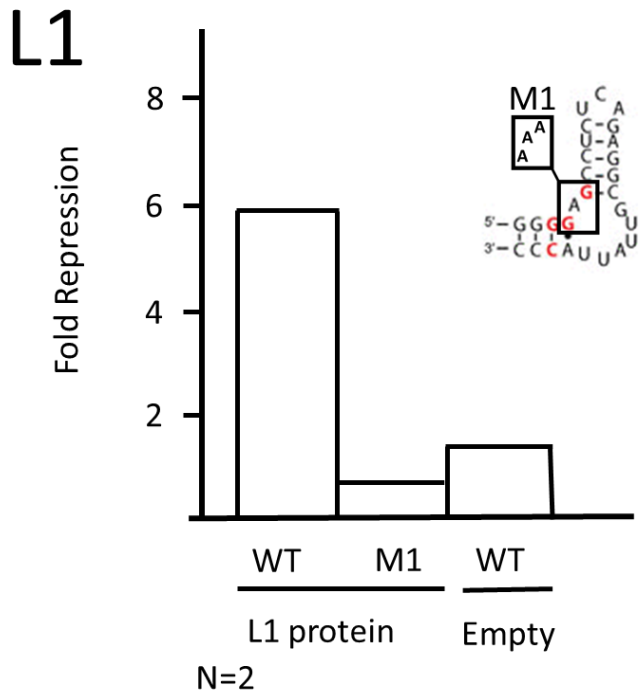


Figure 5.1 ***β-galactosidase Assay of E. coli L1 Protein and Its Leader.*** *K12* ΔrplA *cells +pBAD33-*rplA *+pEG9-L1 or pEG9-L1M1. Graph shoes the average of two replicates. Inset shows structure of M1 non-binding mutant.*

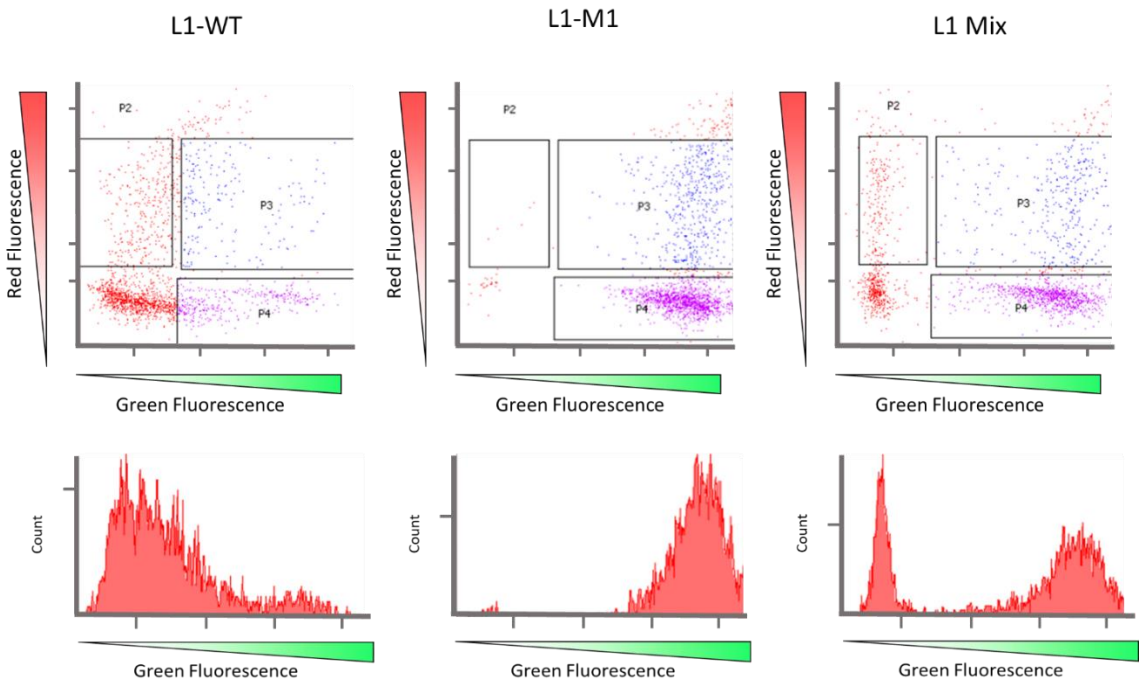


Figure 5.2 ***Analytical Flow of L1 Leader.*** *Droplets contain K12* ΔrplA *+pBAD33-L1 +pEG9-L1 WT, M1, or 50:50 Mix of each. Histogram counts are 25 events.*

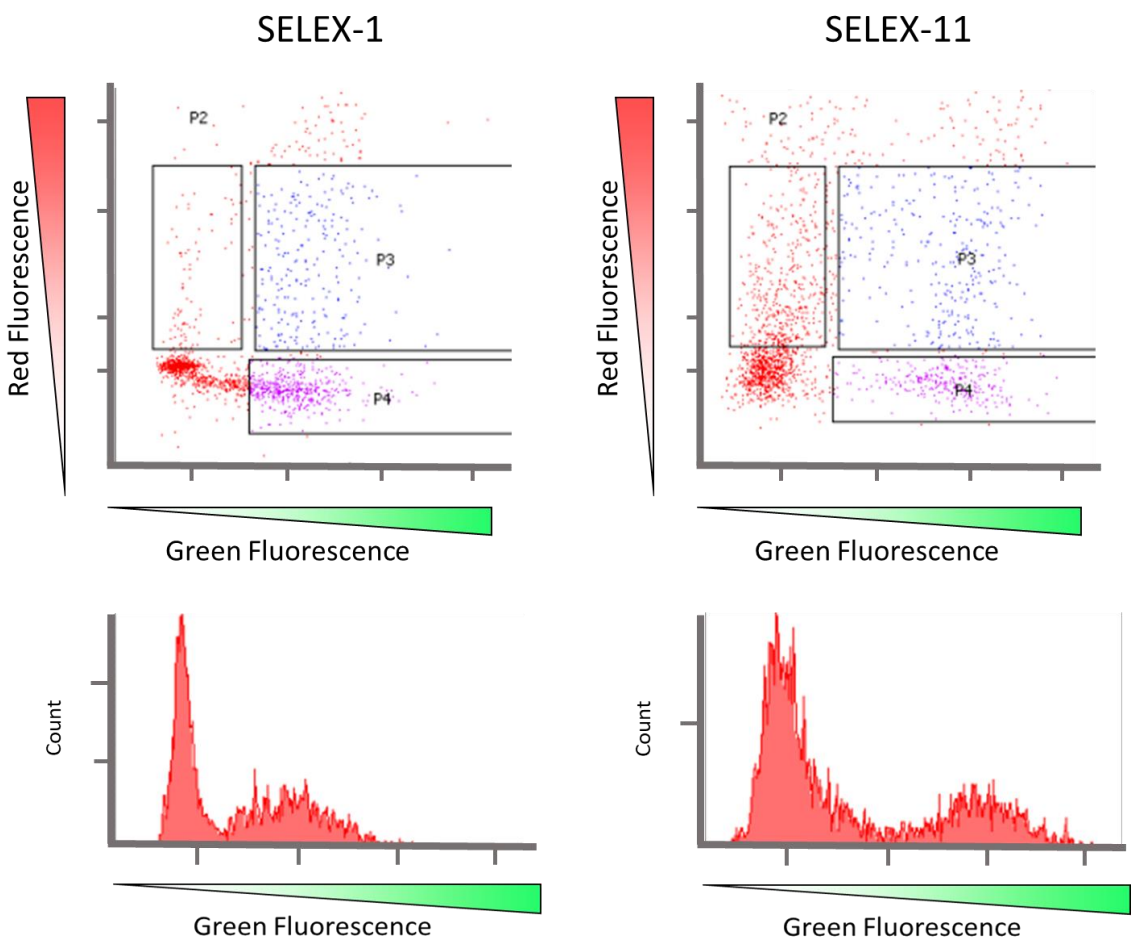


Figure 5.3 ***Analytical Flow of SELEX Binders.*** *Droplets contain microcolonies of K12* ΔrpsO *pBAD33*-rpsO *+pEG10-S15 SELEX round 1 or 11. Histogram count markers are every 25 events.*

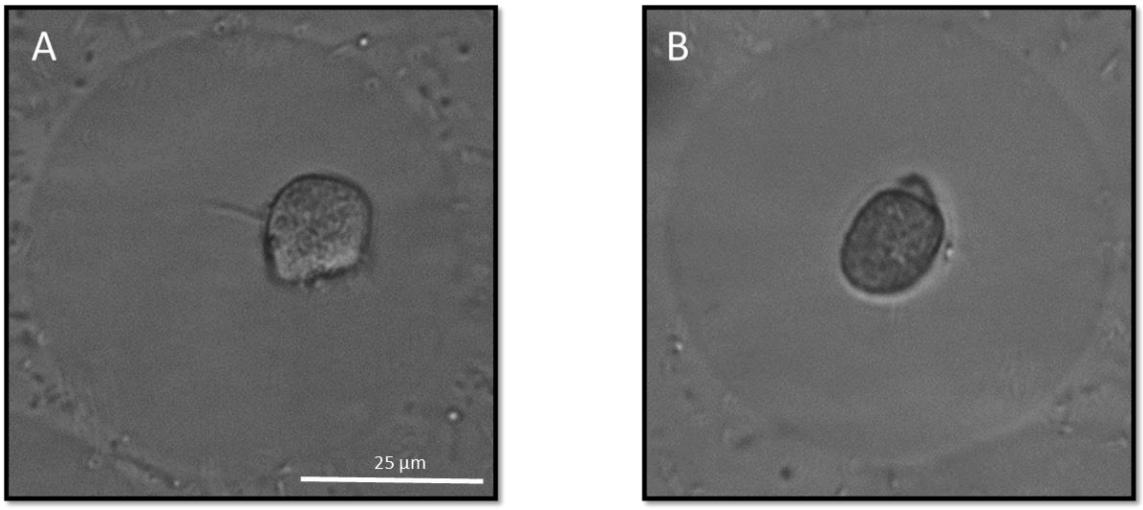


Figure 5.4 ***Microcolony Culture in Different Media.*** E. coli *XL-1 microcolonies in (A) SOB or (B) M9 +FeSO4 +glucose agarose droplets after 16 hours of growth.*

Riboswitch	Regulation	On/Off	Function	Length (nt)
FMN (*ribB*)	Transcriptional/translational	Off	Metabolism	273
TPP (*thiM*)	Transcriptional/translational	Off	Metabolism	201
Lysine (*lysC*)	translational	Off	Metabolism	325
Mini-ykkC (*sugE*)	translational	On	detoxification	90

Table 5.1: ***Riboswitch Candidates for FADSRA Pilot Experiments****.*

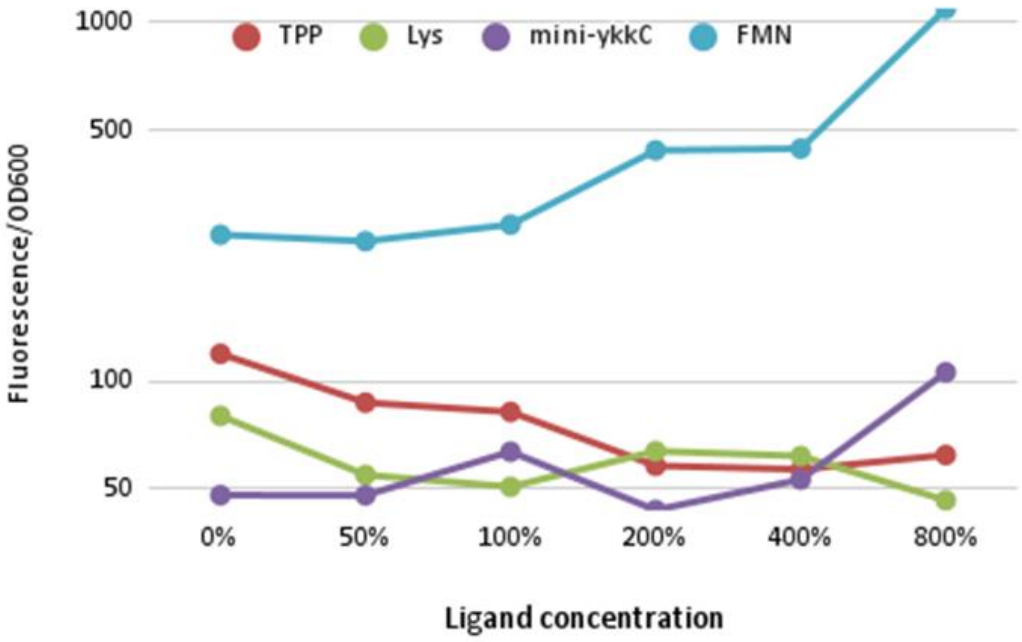


Figure 5.5 ***Fluorescent Plate Assay of Riboswitches in MG1655 +pEG5-Riboswitch with Increasing Concentrations of Ligand.*** *Cells are induced with 1 mM IPTG in mid-log phase for 4 hours. "100%" ligand concentration is 20 µM thiamine, 467 µM riboflavin, 1.25 mM guanidine chloride, 4 mM lysine.*

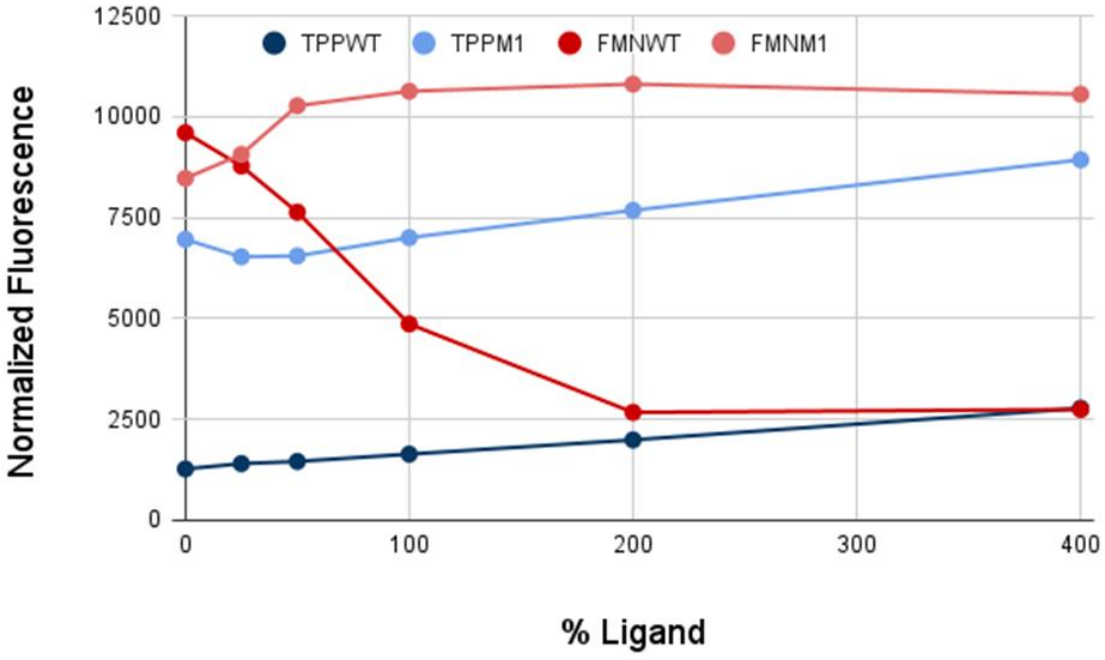


Figure 5.6 ***Fluorescent Plate Assay of Riboswitch Candidates.*** *For MG1655 +pEG10 with TPP wild type, TPP M1, FMN wild type or FMN M1 leader in M9 minimal medium. FMN strains were additionally transformed with pBAD33-ribU. GFP was induced overnight beginning at encapsulation. "100%" ligand is 20 μM thiamine or 200 μM riboflavin. GFP fluorescence is normalized to OD600.*

CHAPTER VI: DISCUSSION

Summary and Significance of FASDRA

Scientists today have access to truly astounding technologies with which to interrogate and explore the world around us. Many staple assays remain relevant, affordable, and attractive. It was my goal with this project to update a classic and widely used assay in the field of RNA regulation with a more modern innovation. An ideal assay would maintain or increase the sensitivity of the older method, but allow for greatly increased throughput. Additionally, an ideal assay would be cost effective, require commonly available reagents and equipment, and allow repurposing of constructs and reporters already used in a research groups experiments.

The assay I've designed and tested through my work is not entirely this ideal assay, but it is on its way. In a complete FADSRA run, scores of thousands of droplets can now be assayed for *in vivo* regulatory function in a matter of weeks compared to the 96 well plate regulatory assay which would take months more to evaluate the same number of mutants. That is an increase of two orders of magnitude of the number of RNAs that can be assayed at once and changes the types of questions researchers can ask about regulatory RNAs.

Previously, assaying RNA libraries of scales of 10^4 or more was largely limited to *in vitro* binding studies alone, whereas FADSRA is a direct method of analyzing the regulatory activity of RNAs within the cell. While I focused largely on cis-regulatory ribosomal leaders, the approach of miniaturizing reporter assays using agar microdroplets could be adaptable for other RNA regulatory RNAs.

Ribosomal Leaders Assay

The ubiquity and necessity of ribosomes and their components makes studying their regulation intrinsically difficult. Many reporter assays including the ones used here rely on a reporter protein to study differential conditions of regulation. Thus, the assays are sensitive to growth phase, as there is high demand by the cell for ribosomal components in mid log. In a way, this makes a droplet based method especially appropriate for studying ribosomal leaders as the growth state of the cells is homogenized by the limitation number of generations possible in a microdroplet. This means that by the time the reporter protein is induced, the cells are not actively growing and the overexpressed ribosomal binding partner is more likely to be available for binding to the reporter fusion leader than the native leader or rRNA. However, the inherent difficulty of evaluating ribosomal leaders and their protein binding partner is due to the necessity of healthy ribosomes for protein. Over- or underexpression of the ribosomal protein can lead to growth defects since ribosomal components are meticulously regulated under normal conditions for the cell. Thus by knocking-out or overexpressing the ribosomal protein of interest and then growing those cells in a resource limited environment, the ability of the cell to grow or produce protein may be compromised.

Indeed, the reporter system designed here requires a great deal of protein production due to the use of multiple drug resistance cassettes, multiple fluorescent proteins and an inducible overexpression plasmid for the protein of interest. Much of the optimization of this assay was spent balancing ribosomal protein expression: enough for the cell to make protein and grow, but not enough to bind to the leader-reporter fusion under otherwise permissive conditions. The effect of the protein expression is also not universal, requiring

some consideration when adapting this method to other leaders. While growth defects of S15 were improved with the expression of S15, overexpression of L1 lead to a slower rate of growth compared to when the L1 was repressed.

Based on the results herein, the secondary fluorophore, the DsRed, did not appear to add much benefit to the reporter system. Addition of the DsRed fluorophore was intended to improve the ability to sort for droplets containing a single microcolony, giving the ability to exclude both empty and overloaded droplets. Instead, the data show that microcolonies appeared to preferentially express the sfGFP, confounding the intended ability to sort full droplets from empty droplets as well as gating out droplets with more than one microcolony. Additionally, near constitutive expression of DsRed adds to the burden of the cells to produce protein under conditions affecting ribosome function. Future adaptation of this reporter system could likely remove the DsRed to beneficial effect.

Summary of *E. coli* Leader Analysis

The *E. coli* S15 leader has been studied extensively over many decades, providing snapshots over the years of both technological innovations as well as the changing zeitgeist of the field. However, even with significant attention, there has not been a complete fitness landscape for S15 or any other ribosomal leader This study provides the first nearly complete evaluation of all single mutants for a ribosomal leader, revealing previously unexamined nucleotides important to leader function and strengthening our understanding of the folding dynamics of the *E. coli* S15 leader. FADSRA assayed 212 of the 240 possible single mutants of the leader. The apparent importance of the stems for structural stability both in the presence and absence of S15 protein is especially notable, and is supported by previous research findings of the folding dynamics of the *E. coli* S15 leader, which indicate

that the presence of weak pseudoknots can prevent S15 binding. [131] Such pseudoknots, if stabilized by a mutation, could prevent the binding of S15 or disrupt translation through sequestering the ribosome binding site or start codon. Additionally, the G·U/G-C motif is found in the same stem in both of the main structural forms, so disruption of this structure could cause broader changes both with and without protein binding.

The *E. coli* S15 leader is extremely well studied which assisted in the analysis and validation of the FADSRA assay. But it is important to consider how a fitness landscape might be interpreted without the same level of *a priori* knowledge. Without experimental data about folding or structure, a FADSRA generated fitness landscape could still provide information about which nucleotides contribute to regulation *in vivo*. However this method would be most useful in conjunction with structural data, either computationally predicted or experimentally determined.

FADSRA Drawbacks

As a method in development, FADSRA currently has some drawbacks that would need to be addressed for future applications. One of the main drawbacks of the general FADSRA assay is the severe bottlenecking that occurs at the point of encapsulation. While this can be partially mitigated by simply increasing the number of droplets created and sorted, there are few other options to address the bottlenecking. The bottlenecking was especially severe in the specific case of the S15 leader as purifying plasmid from one round to the next resulted in low yields. The utility of the outgrowth steps is mixed. In one case, outgrowth increases the amount of DNA that can be purified, and characterization of the growth difference between strains indicates that the S15 reporter library could be grown

for up to two hours before competition became a concern. Splitting the population before the outgrowth step could lead once again to severe bottlenecking.

An additional concern with this method is the sensitivity. Droplets are sorted with a purity of 94%, and expression can vary based on plasmid copy number. Even at low loading concentrations with careful gating, droplets with more than one microcolony are likely to be present. The FADSRA pilot assay in Chapter 3 further revealed issues with sensitivity as the gating of the first sort was not stringent enough to separate the regulating wild type into the repressive bucket. While it does appear that leaders resulting in extreme expression (either a lot of expression or very little expression) were effectively separated from one another, more modest changes to expression were sorted more stochastically. Adding more than one round of sorting, or an additional gate collecting intermediately expressive populations could address this. Sensitivity could also be improved by fixing the leaky expression of the IPTG inducible promoter. If reporter expression is confined to only a short period of time, there will be more differentiation between expressing and non-expressing populations. Together, these limitations of the assay add an additional complication of maximizing the separation of populations according to the true leader behavior that is not present in bulk culture assays and must be considered.

Evaluation of Other Applications

Preliminary studies of FADSRA applications indicate that this approach may be useful beyond the version of the assay designed to study the S15 leader, but likely requires optimization for each context. The L1 leader in the FADSRA reporter plasmid optimized for use with the S15 leader resulted in much higher fluorescence even in the presence of the L1 protein. In the recent decade, riboswitches have received comparatively more

attention than ribosomal leaders and thus have a number of attractive methods *in vitro* and *in vivo* well designed for their study including methods that can produce fitness landscapes. [126,127,146] Nonetheless, an attempt was made to adapt FADSRA for use with riboswitches. This adaptation requires non-trivial changes to strains, the reporter plasmid and growth conditions. Results regarding the application of FADSRA to riboswitches are still pending while these important but basic changes are optimized. Additionally, evaluating riboswitches in microdroplets may not be an optimal strategy for regulators that require a specific stage of growth for effective evaluation.

Conclusion

Although this assay was optimized and designed with a specific application in mind, a number of potential applications was considered and preliminarily tested. It is exciting to think about how this technique could apply to other questions in fields beyond cis-regulatory RNAs. This method provides an intermediate between bulk culture assays and single cell experiments. The main benefits of FADSRA are both the isolation of each droplet, preventing competition between droplets, but also the homogenization of the populations for flow cytometry and cell sorting. With droplets, bacterial communities can be kept together as they are sorted, unlike in conventional flow cytometry where bacterial samples are treated to prevent cell interactions. Additionally, the droplet provides a consistent forward and side scatter profile independent of the droplet's contents. This could be beneficial for looking at populations with mixed cell shape or properties.

Given the results for the S15 leader fitness landscape, exploring the other S15 leaders and proteins can continue to refine our understanding of the plasticity of the leader and expand on the work in Slinger *et al.* 2015 and Slinger *et al.* 2016. [63,113]

APPENDIX

CHAPTER A-I:

AUTOREGULATION OF THE L21 LEADER IN *BACILLUS SUBTILIS*

Background

Cis-regulation of ribosomal protein operons is well established in *Escherichia coli,* a gram-negative bacterium (Figure A1.1), but little work has been done to assess how stoichiometry is maintained in other bacteria. In the model gram-positive organism *Bacillus subtilis* several potential mRNA leaders have been identified computationally by comparative genomics, but few have verified regulatory action (Figure A1.1). One such leader precedes the *rplU-yskB-rpmA,* also called the L21, operon encoding ribosomal proteins L21 and L27 as well as a protease involved in the maturation of L27. [155,156] The mRNA leader preceding L21 was identified in organisms across the phylum *Firmicutes* and has a well-defined and conserved structure (Figure A1.2). This project sought to demonstrate the autogenous regulation of the *rplU-yskB-rpmA* operon by L21 protein.

Results:

Regulation of *rplU-yskB-rpmA* Operon with Proteins L21 and L27

To assess regulatory activity of each protein on the L21 leader, a β-galactosidase assay was performed. The L21 leader was fused to the *lacZ* reporter gene under the control of an IPTG inducible promoter on the pAB2 reporter plasmid. Each protein was transformed on a separate plasmid, pYH213, also under an IPTG inducible promoter. Results showed two-fold repression by L21p only (Figure A1.3). There was no change in β-galactosidase activity with L27 expression. When the entire operon not including the leader, but including interstitial gene *yskB* was over expressed, reporter gene activity was partially repressed compared to L21 alone. Regulation with a leader from another

ribosomal protein operon, L20, was assessed as a control and showed no change in β-galactosidase regulation by L21, L27 or the whole operon (Figure A1.3). Together, these results suggest mild regulation of L21 protein by binding to the *rplU-yskB-rpmA* operon 5' UTR.

L21 Protein Mutations

Mutations were made to the L21 protein to determine potential RNA leader binding sites. Protein mutations were assessed using β-galactosidase assay. The mutations included a series of mutations to alanine including M1 (G8A G9A K10A), M2 (K77A P78A K79A K80A) and M3 (Q86A H88A R89A Q90A). Results showed loss of regulation of the M3 mutant, and partial but insignificant loss of repression of M1 (Figure A1.4). M2 had the same level of repression as wild type, but the strain grew very poorly overall, suggesting that the mutation may be interfering with ribosome function and making the assay results uninterpretable. Overall results from these experiments were inconclusive due to inconsistent strain growth and modest and variable results. Further analysis of the protein mutations to L21 was not pursued.

L21 RNA Mutations

Mutational analysis of the L21 leader was also performed via β-galactosidase assay. M1 (A20G) and M2 (A21G) changed residues strongly conserved in the leader consensus structure, and M3 (C49G) disrupted a conserved base pairing at the base of the stem. Results show that compared to the WT L21 RNA leader, mutations M1 and M3 show a lack of regulation, while M2 shows partial but poor regulation (Figure A1.5). This indicates

that M1 and M3 are important sites for either RNA leader structure or to protein binding. M2, though highly conserved, is by itself less important for regulation.

Mechanism of Regulation

No terminator structure was found in the L21 leader, suggesting post-transcriptional regulation. This was confirmed via qPCR. qPCR of the native L21 leader in a strain with the protein overexpression plasmid alone showed high variability between replicates, but the only significant difference was in the full operon overexpression condition, perhaps suggesting high rates of turnover, but not suggesting a transcriptional mechanism of regulation (Figure A1.6). qPCR of *lacZ* in the reporter strain showed no significant difference between protein overexpression conditions (Figure A1.7).

L21 Operon *in Vitro* Regulation

In vitro binding was assessed via filter binding assay. Purifying *B. subtilis* L21 protein was reasonably straightforward with some adaptations from the S15 purification protocol used in Slinger *et al*.[63] Since *in vitro* experiments can be sensitive to the region of RNA used, six different leader constructs were made (Table A1.1). Initial filter binding assays with *Bacillus* protein and RNA the different constructs all resulted in less than 3% of protein bound, even at high protein concentrations (data not shown). It is known that protein homologs from thermophiles can be more stable in *in vitro* assays, so assays were then attempted with *Geobacillus kaustophilus* protein and leader, both of which exhibit high similarity to the *B. subtilis* protein and leader (Figure A1.8). Comparisons between the *B. subtilis* and *G. kaustophilus* proteins and RNAs found that *G. kaustophilus* protein

could bind both the *B. subtilis* and the native leader better than the *B. subtilis* protein (Figure A1.11).

Filter binding assays showed that an RNA spanning the L21 leader consensus start through the start codon showed binding with a K_D of about 100 μM and a maximum fraction bound of 0.99 at 2 μM (Figure A1.9). In order to confirm that L21 was the regulating protein, the other ribosomal protein in the operon L27 was also purified from *Geobacillus*. A filter binding assay comparison of the two proteins shows that L21 binds the *Geobacillus* L21 leader, but L27 protein does not (Figure A1.10). Mutations made to conserved residues in the L21 leader did not appear to significantly diminish binding (Figure A1.11). However, filter binding assays between the unrelated *M. smegmatis* S15 leader RNA shows a nearly identical binding curve to the *Geobacillus* L21 leader with *Geobacillus* protein (Figure A1.12). This finding was confirmed via electrophoretic mobility shift assays, showing a similar banding pattern between both the *M. smegmatis* and the *G. kaustophilus* RNAs. The assay also revealed the formation of bands between the free RNA and the protein bound RNA – this pattern is characteristic of nonspecific binding. Although s-shaped binding curves typically indicate a specific reaction, binding to both L21 and S15 from an unrelated species may indicate that this binding is nonspecific and thus not strong evidence for regulation. A similar EMSA between *G. kaustophilus* L21 leader and *B. subtilis* L20 is also seen: notably L20 was not observed to regulate the operon (Figure A1.13).

Discussion

Since much of the work on RNA ribosomal leaders examines gram-negative *E. coli*, little attention has been paid to differences between *E. coli* ribosomal regulation and that

of other clades. Especially due to known differences between gram-negative and gram-positive organisms, it is likely that ribosomal regulation can vary as well. Discovery and validation of ribosomal leaders outside of *E. coli* can further fill in the picture into the evolution and maintenance of cis-regulatory RNAs.

For this particular leader, the data is quite inconclusive. Comparative genomics have found a clear consensus structure in *Firmicutes* before the L21 operon. β-galactosidase assays show a consistent but very modest 2-fold repression of β-galactosidase activity with induction of the L21 protein. No repression is seen with L27 protein expression. 2-fold repression is the lowest amount of repression observed among the ribosomal leader RNA-protein pairs studied in the lab. However, the repression was quite consistent, repeatable by multiple researchers and appeared specific to L21 protein. qPCR data supported the proposed mechanism of post-transcriptional regulation.

Yet *in vitro*, no specific binding was observed between the L21 leader RNA and L21 protein. Lack of *in vitro* binding between molecules known to bind and interact *in vivo* is a known problem among researchers. Similarly, non-specific *in vitro* binding is also known to occur. Both of these potentialities must be controlled for in *in vitro* experiments. While several ribosomal proteins are relatively easy to purify due to their high charge, it is possible here that the protein purified was not optimally purified, resulting in protein aggregates or misfolded protein. Yet, for now, the veracity of the *rplU-yskB-rpmA* in *Bacillus* remains elusive.

Materials and Methods

Cloning of RNA Constructs for Miller Assay

RNA mutant constructs were generated and verified previously. [157] Briefly, the mutant versions of the L21 leader were generated using PCR assembly and cloned into pAB2.

Cloning of Protein Overexpression Constructs

Protein mutants were made via PCR assembly using the following primers: M1 (1422, 1553; 1552, 1429), M2(1422, 1554; 1555, 1429), M3 (1422, 1557; 1556, 1429). Assembled and purified PCR products from each assembly were digested with EcoRI-HF and PstI in NEB 2.1 buffer and column purified. Digested products were ligated into vector pYH213 using Quick Ligase for 15 minutes at room temperature then transformed into z-competent TOP10 cells. Transformants were then plated onto LB + 12.5 µg/mL tetracycline agar medium plates and incubated overnight at 37 °C. Product insertion was verified through PCR (Primers 1429, 1357) and then sequence verified through sanger sequencing using sequencing primer 1357.

Transformation of Constructs into *Bacillus*

Bacillus 1A1 strain 168 was transformed with plasmid pYH213 to generate *Bacillus* compatible protein overexpression plasmids using the 2-step method. GM1 (0.5% glucose, 1% yeast extract, 0.025% cas amino acids, 1X Spizen salts), was inoculated from plate grown single colonies and grown overnight at 37 °C without shaking. The top culture was transfer to fresh GM1 medium the next day and incubated with shaking at 37 °C. A600 was monitored every 20 minutes to identify entrance of the culture into

stationary phase. 90 minutes past the start of stationary phase, cells were diluted into GMII medium 1:10 then grown with shaking at 37 °C for another hour. Cells were then transformed with 4 µg of plasmid by adding the DNA and incubating with shaking for 30 min at 37 °C. After 30 min, cells were recovered by addition of 2XYT medium for another hour then plated on TBAB + 12.5 µg/mL tetracycline plates and incubated overnight. DNA integration was confirmed with PCR (Primers 1429, 1357).

For transformation of pAB2 plasmid into *B. subtilis*, the one step protocol was used. Briefly, 2 mL of transformation medium + 12.5 mg/mL tetracycline was inoculated with a single colony of the protein overexpression strain of interest and incubated at 37 °C overnight with shaking. Cultures were diluted the following day to A600 of 0.25 and returned to the shaker. When cultures were about A600 of 0.5, 1 mL of cells was transformed with 3 µg of pAB2 plasmid DNA of interest. Reactions were incubated for 40 minutes shaking at 37 °C before the addition of 1 mL 2XYT + 12.5 µg/mL tetracycline and another 45 minutes of shaking at 37 °C. Cells were then plated (100 µL of neat cells and also the remaining cells concentrated to 100 µL) onto TBAB plates with 12.5 µg/mL tetracycline. Clones were screened by selective plating onto TBAB + 0.5 µg/mL erythromycin and TBAB + 1% starch. Clones without erythromycin resistance and without amylase activity were selected.

β-galactosidase Assay

Bacillus subtilis strain 168 transformants were picked from plate and inoculated into 2 mL of 2XYT + 12.5 µg/mL tetracycline + 100 µg/mL spectinomycin. Cultures were grown overnight at 37 °C with shaking. The next day, cultures were diluted 1:50 in fresh 2XYT + 12.5 µg/mL tetracycline + 100 µg/mL spectinomycin and grown to mid-log

(A600 0.4-0.8) and induced with 1mM IPTG. 1mL of culture was pelleted and resuspended in z-buffer (50 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, 100 µg/mL spectinomycin) and the A600 recorded. Cells were permeabilized with toluene. ONPG was added and the time interval recorded from the time the ONPG was added to the time the solution turned yellow and was stopped with Na_2CO_3. A420 was recorded after pelleting cells. Miller units were calculated as 1000 * (A420/(Δt(min) * A600 * volume (mL))).

qPCR

qPCR was used to evaluate transcript levels of L21 as well as *lacZ* for strains pYH213-L21 WT, pYH213-L27WT, pYH213-L21-27WT. For qPCR sample collection, 2XYT media plus 12.5 µg/mL tetracycline and 100 µg/mL spectinomycin was grown overnight. Cultures were diluted 1:50 in fresh medium the following day and grown to mid log with 1 mM IPTG. An aliquot of these cells was reserved for β-galactosidase assay and remaining culture was prepared for RNA extraction. Cells were lysed by first pelleting cells, resuspending the cells with 3 mg/mL lysozyme in TE buffer at room temperature and then undergoing three freeze thaw cycles. RNA was recovered using TRIZOL. Extracted RNA was then treated with DNase RQ1 to remove remaining DNA and then reverse transcribed using Superscript III, 1x First Strand Buffer, 1 µL 0.1 DTT, Suprase-IN, 1 µL random hexamers 50 ng/µL. Samples were also included without reverse transcriptase to confirm removal of genomic DNA. qPCR was performed on a ABI 7500 Fast Real Time PCR system. 10 µL reactions were set up with 50nM ROX, 1x SYBR Green I, 0.1 µL Phire Hot Start II, 1X Phire Buffer, 0.5 µL 10x dNTPs, and 1 µL template DNA. The qPCR cycle was 98 °C for 30 seconds, (98 °C for 5 seconds, 58 °C for

5 seconds, 72 °C for 30 seconds) for 39 cycles with a 5 minute final extension at 72 °C. Each sample was performed in triplicate. The primers were as follows: L21 (1542, 1543), *nifU* (1546, 1547), *upd* (1548, 1549), *lacZ* (1204,1205).

L21 Protein Purification

Overexpression plasmid was generated by Carolyn Larkins. Using PCR, the protein coding sequence was amplified from gDNA using the following primers: BsL21 (1532, 1533) GkL21 (1536, 1537). The PCR product was then gel purified and digested by BamHI and NdeI. The petHT vector was also digested by BamHI and NdeI then gel purified. Insert and vector were ligated using Quick Ligase at room temperature for 15 minutes. The overexpression plasmid was z-transformed into *E. coli* protein expression strain BL21. Protein expression was induced at mid-log with 1 mM IPTG for 4 hours. Cell pellets were lysed via sonication in lysis buffer (20 mM Tris pH 7.0 at 4 °C, 1 M KCl, 6 mM β-mercaptoethanol). Lysate was spun down at 10,000 rpm for 30 minutes to separate soluble and insoluble fractions. The lysate supernatant was dialyzed into wash buffer (50 mM NaAc pH5.5 at 4 °C, 150 mM NaCl). Lysate was then filtered with a 0.2 micron filter then hand loaded onto a SP FF column with a syringe. The protein was purified via ion exchange chromatography using a gradient of wash buffer and elution buffer (50 mM NaAc pH5.5 at 4 °C, 1 M NaCl). Protein fractions were tested for nuclease activity by incubating samples of the protein fraction with radiolabeled in binding buffer A (50 mM Tris-Acetate pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 5 mM DTT, 0.02% BSA) for one hour. Nuclease-free fractions were then pooled and concentrated using spin concentrators. Protein concentration was assessed via microplate Bradford Assay.

L27 Protein Purification

The overexpression plasmid was generated by Carolyn Larkins using the method as above, with the primers PCR: BsL27 (1534, 1535) Gk (1538, 1539). Overexpression plasmid was z-transformed into *E. coli* strain BL21. Protein expression was induced at mid-log with 1 mM IPTG for 4 hours. Cell pellets were lysed via sonication in lysis buffer (20 mM Tris pH 7.0 at 4 °C, 1 M KCl, 6 mM β-mercaptoethanol). Lysate was spun down first at 3,500 rpm for 10 minutes to pellet unlysed cells, and then the supernatant was spun at 10,000 rpm for 30 minutes to separate soluble and insoluble fractions. The insoluble fraction was then resuspended in wash buffer (20 mM NaAc pH 8.0 at 4 °C, 6M Urea, 20 mM KCl). Lysate was then filtered with a 0.2 micron filter then hand loaded onto a SP FF column with a syringe. Protein was purified via ion exchange chromatography using a gradient of wash buffer and elution buffer (20 mM NaAc pH 8.0 at 4 °C, 6 M Urea, 1 M KCl). Protein fractions were tested for nuclease activity by incubating radiolabeled RNA with an aliquot of the fraction in binding Buffer A (50 mM Tris-Acetate pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 5mM DTT, 0.02% BSA). Nuclease-free fractions were then pooled and concentrated using spin concentrators. Protein concentration was assessed via microplate Bradford Assay.

In Vitro Transcription

For *in vitro* transcription of RNA for radiolabeling, PCR products were generated using primers 1 (1638, 1640), 2 (1638, 1641), 3 (1638, 1642), 4 (1639, 1640), 5 (1639, 1641), 6 (1639, 1642). PCR products were then column purified and eluted in 30 µL of water. To the 30 µL of purified DNA, Transcription Buffer (200 mM Tris-HCl pH 8.0, 40 mM MgCl2, 250 mM NaCl, 10 mM Spermidine), 2 mM each NTP, and 2 µL

homemade T7 RNA polymerase were added. Reactions were incubated at 37 °C for two hours and then stopped with Urea Loading Buffer. The reactions were then purified on 6% acrylamide urea gel. Using UV shadowing for visualization, the RNA was excised and eluted in crush soak buffer (300 mM NaCl, 1 mM EDTA). RNA was then precipitated with ethanol and re-suspended in water.

Radiolabeling of RNA

RNA was radiolabeled with 5' P-32 ATP by first dephosphorylated using 1U alkaline phosphatase in alkaline phosphatase buffer for half an hour. Reactions were then heat inactivated. P-32 ATP was bonded to the RNA by incubation with T4 PNK in CHES buffer for 1 hour followed by purification on a 6% urea PAGE gel. Radiolabeled RNA was visualized by autoradiography and RNA was excised and recovered in crush soak buffer. The RNA was then precipitated using ethanol and resuspended in water.

Filter Binding Assay

The RNA was diluted in water and renatured at 42 °C for 15 min then allowed to cool to room temperature for 10 min. The RNA was diluted to 1000 counts per 25 μL water per reaction. Purified protein was diluted in Buffer A and a 1:2 protein gradient was set up across a 96 well PCR plate from 0 μM to 1.024 μM. The reaction was incubated for 30 minutes at room temperature before reactions were transferred to the filter apparatus containing from top to bottom a nitrocellulose membrane, a nylon++ membrane and two pieces of filter paper. Vacuum was applied to the apparatus to pull samples through. Wells were washed twice with 1 volume each of Buffer A. Membranes were then air dried and placed into a cassette with phosphoscreen overnight. The phosphoscreen was imaged the

next day on a Typhoon Imager. Fraction bound was calculated by dividing the nitrocellulose signal by the nylon and nitrocellulose signal.

EMSA

To test protein binding to the RNA of interest, an electronic mobility shift assay was performed. Samples were prepared by renaturing the RNA at 42 °C for 15 minutes then cooling to room temperature for 10 minutes. Serial dilutions of the protein 0 to 4 μM were made and 1000 counts of Binding Buffer A were mixed and incubated at room together for 30 minutes. A native gel was prepared with 10% acrylamide native in TBE buffer (45 mM tris-borate, 1 mM EDTA) and prerun for 1 hour at 160V. Samples were mixed with native loading buffer and loaded into the gel which was run for 4 hours at 350V at 4 °C. The gel was then dried and exposed to a phosphoscreen overnight. The screen was then imaged using a Typhoon Imager.

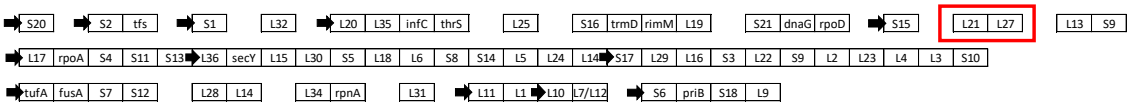
Primer Table

	Primer Name	Primer Sequence
qPCR	1204-lacZQPCRF2	TACCTGTTCCGTCATAGCGA
	1205-lacZQPCRR2	CTGTTTACCTTGTGGAGCGA
sequencing	1357-pAY132seqF	GTTGATCAGTCAACTTATCTGTATAG
L21 Leader cloning	1422-pAY-L21F1	GGCCCGAATTCAATAGGAGGTGCAGAGAATGTACGCAATCATTAAAAC
	1429-pAY-L21R	GCCGGCTGCAGTTACGCGTTGATTTTTTCGATCGTCAC
Protein overexpression	1532-BsL21petHTF	GGAATTCCATATGTACGCAATCATTAAAACAGGCGG
	1533-BsL21petHTR	CGCGGATCCTTAGTAGGGCAGAAGCAGTTCC
	1534-BsL27petHTF	GGAATTCCATATGTACGCAATCATTAAAACAGGCGG
	1535-BSL27petHTR	CGCGGATCCTTATTGAGCTACAGGATATACGCTCAC
	1536-GKL21petHTF	GGAATTCCATATGTACGCAATTATCGAAACTGGCG
	1537-GKL21petHTR	CGCGGATCCTTACGCGTTGATTTTTTCGATGACG
	1538-GKL27petHTF	GGAATTCCATATGCTGAGACTTGACCTGCAAT
	1539-GKL27petHTR	CGCGGATCCTTACGCTTCTTGGCTGACCGGGTAG
qPCR	1542-BsL21qPCRF	CACCATTGCTACAACCGCTC
	1543-BsL21qPCRR	TGCGTACATTCTCTGCACCT
	1546-BsnifUqPCRF2	TTTTACTTCGTGACGGCGGT
	1547-BsnifUqPCRR2	TTGTTGAACTTGGGCAGCTG
	1548-BsudpqPCRF	TTCCCCTACAATTCTGCGGA
	1549-BsudpqPCRR	CTGTTCTCGATCAGGTGGGT
Protein Mutants	1552-L21-GGKQ:AAAAF	CGCAATCATTAAAACAGCCGCTGCACAAATCAAAGTTGAAGAAGGCC
	1553-L21-GGKQ:AAAAR	GGCCTTCTTCAACTTTGATTGCTGCAGCGGCTGTTTTAATGATTGCG
	1554-L21-KPKK:AAAAF	CACTGTTTTCAGATACGCAGCAGCGGCAAACGTTCATAAAAAACAAGGTC
	1555-L21-KPKK:AAAAR	GACCTTGTTTTTTATGAACGTTTGCCGCTGCTGCGTATCTGAAAACAGTG
	1556-L21-QGHRQ:AGAAAF	GAAAAACGTTCATAAAAAAGCAGGTGCTGCTGCGCCTTACACTAAAGTGACG

	1557-L21-QGHRQ:AGAAAR	CGTCACTTTAGTGTAAGGCGCAGCAGCACCTGCTTTTTTATGAACGTTTTTC
In vitro L21	1638-BSL21T7F1	CCAAGTAATACGACTCACTATAGGAATGTTATGGATGTAGCACC
	1639-BSL21T7F2	CCAAGTAATACGACTCACTATAG GCACCATTGCTACAACCGCT
	1640-BSL21RNAR1	GACTCGCCAGATACCAGG
	1641-BSL21RNAR2	CCTCCTATTAGACTAAGACTC
	1642-BSL21RNAR3	CGTACATTCTCTGCACCTCC

Figures:

E. coli



B. subtilis

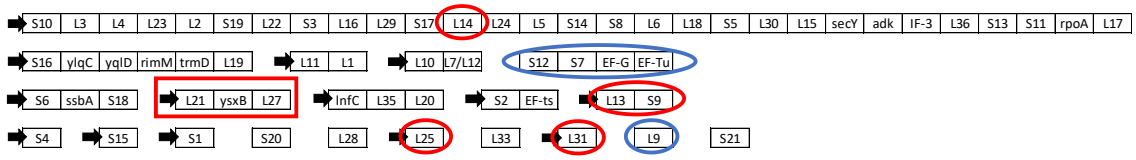


Figure A1.1 ***Ribosomal Protein Regulation in* E. coli *and* B. subtilis.** *While ribosomal protein regulation in* E. coli *is well defined, there is a dearth of information on this process in gram-positive species such as* B. subtilis*. Operons governed by cis-regulators (black arrows) in* E. coli *are not necessarily regulated by cis-regulators in* ***B. subtilis*** *(blue circles indicate cis-regulation in* E. coli*, but not* B. subtilis*; red indicates cis-regulation in* B. subtilis*, but not* E. coli*). The red square in each diagram highlights the L21/L27 operon.*

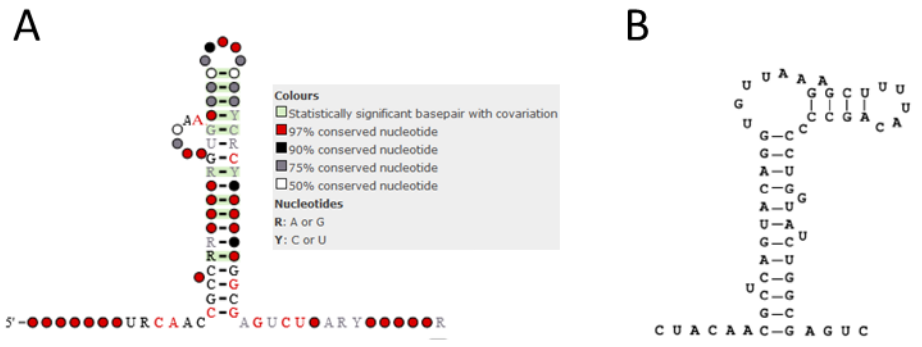


Figure A1.2 ***L21 RNA Leader Structure.*** *A) The consensus from the RFAM database of the putative structure of the L21 leader across various species compared to B) the putative structure of the* B. subtilis *L21 leader.*

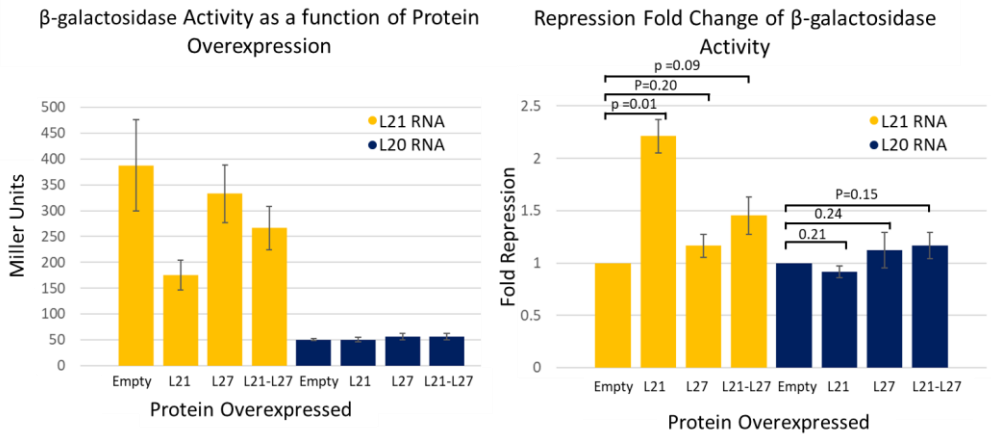


Figure A1.3 ***β-galactosidase Assay to Determine Regulation of the L21/L27 Operon.*** *Graphs show L21 or L20 leader β-galactosidase fusion activity as Miller units or fold repression compared to empty protein overexpression vector when L21p, L27p or the entire L21/L27 operon are overexpressed. Error bars represent standard error of three biological replicates. Assay shows 2-fold repression by L21, but no repression by L27. This suggests the regulator for the L21/L27 operon is L21p. Additionally, there is no observed regulation of the L20 leader, providing evidence that this regulation is specific. p-values calculated using a one-tailed Welch's t-test.*

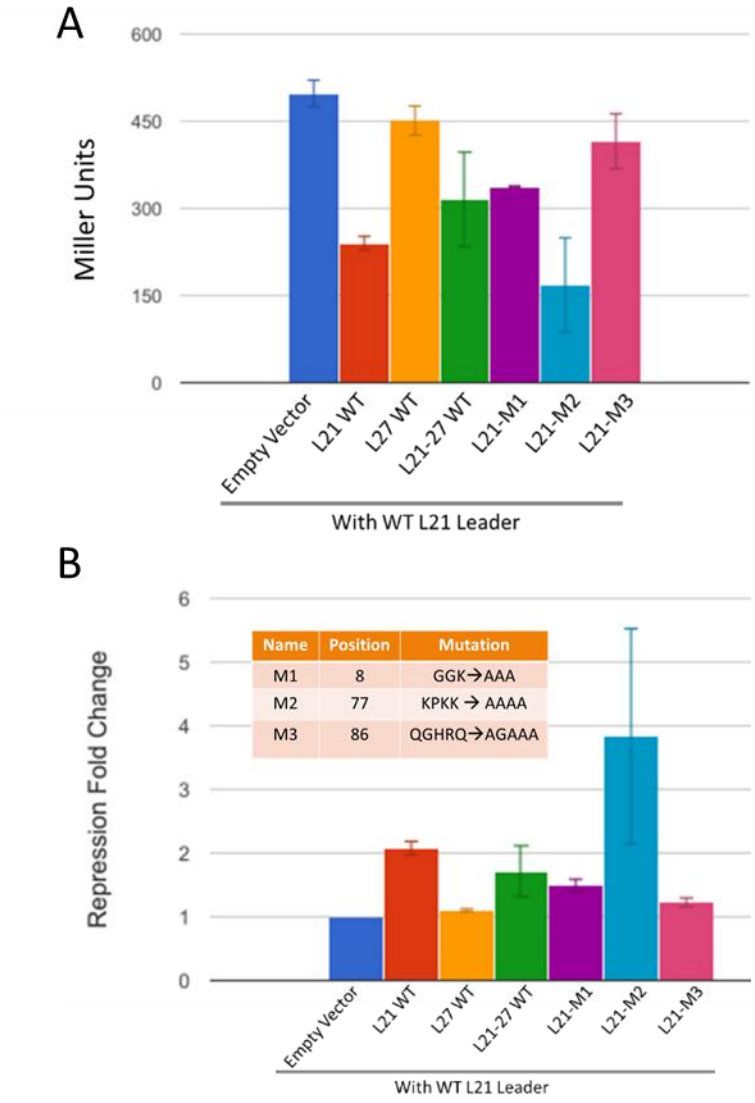


Figure A1.4 ***Repression Fold Change of β-galactosidase Assays on L21 Mutant Protein Binding to the L21 Leader RNA****. Severe mutations to L21 protein do not clearly abolish regulatory activity of L21 protein on the L21 leader compared to WT L21.*

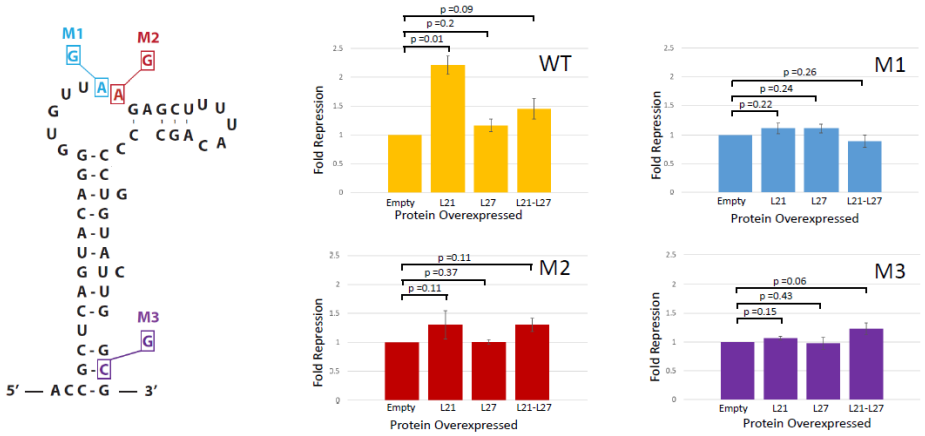


Figure A1.5 ***Repression Fold Change of β-galactosidase Activity to Determine Regulatory Efficacy of RNA Mutants.*** *Diagram shows mutations made to L21 leader sequence. Graphs show RNA leader β-galactosidase fusion activity as repression fold change compared to empty protein overexpression vector, with L21p, L27p or the entire L21/L27 operon overexpressed. Error bars represent standard error of three biological replicates. The assay shows that compared to the WT L21 RNA leader, mutations M1 (blue) and M3 (purple) show a lack of regulation, while M2 (red) shows partial, but poor regulation. p-values calculated using a one-tailed Welch's t-test.*

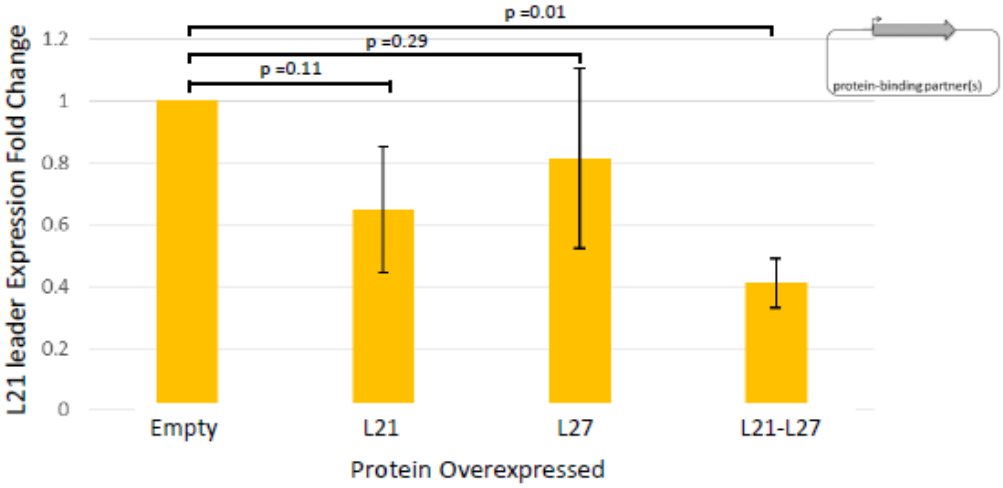


Figure A1.6 ***qPCR of Native L21 Leader in Strains with Protein Overexpression Plasmids.*** *Graph shows expression fold change of qPCR of genomic L21 RNA leader with protein overexpression plasmid only, normalized to* nifU *and compared to empty vector. Error bars represent standard error of three biological replicates. p-values calculated using a one-tailed Welch's t-test. The results show variability in transcript levels between samples.*

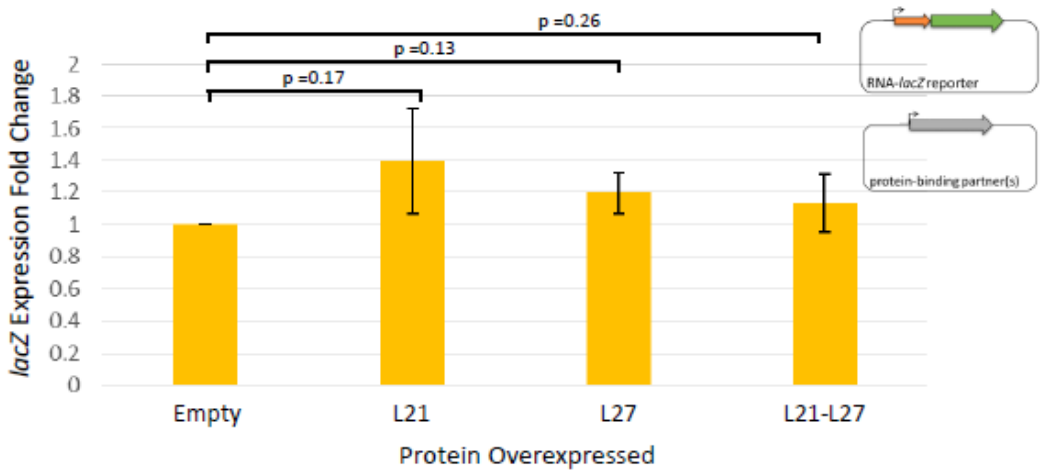


Figure A1.7 ***qPCR of* lacZ *Reporter Gene in Strains Used for β-galactosidase Assays.*** *Graph shows expression fold change, normalized to* nifU *and compared to empty protein vector of* lacZ *target in strains containing both the protein overexpression and the RNA leader lacZ fusion plasmids. Error bars represent standard error of three biological replicates. p-values calculated using a one-tailed Welch's t-test. The results show little difference in transcript levels between samples and the repression observed in the β-galactosidase assay is not seen. This suggests that L21/27 operon regulation is post-transcriptional.*

		3′ Stop		
		Consensus End	After Shine-Dalgarno	After Start Codon
5′ Start	Transcription Start	1	2	3
	Consensus Start	4	5	6

Table A1.1 ***Diagram of RNA Leader Variations Tested in in Vitro Binding Assays****.*

A

BS L21p alignment with GK L21p

Score	Expect	Method	Identities	Positives	Gaps
159 bits(401)	2e-57	Compositional matrix adjust.	78/102(76%)	89/102(87%)	0/102(0%)

```
BS L21p  1    MYAIIETGGKQLKVEEGQEIYIEKLDANEGDIVTFDKVLFVGGETVKIGNPTVEGATVTA  60
              MYAII+TGGKQ+KVEEGQ +YIEKL A  G+IVTF+ VLFVGG+ VK+GNPTVEGATVTA
GK L21p  1    MYAIIKTGGKQIKVEEGQIVYIEKLAAEAGEIVTFEDVLFVGGDNVKVGNPTVEGATVTA  60

BS L21p  61   RVQKHGRQKKIIVFKYKAKKNYRRKQGHRQPYTKVVIEKINA  102
              +V+K GR KKI VF+YK KKN  +KQGHRQPYTKV IEKINA
GK L21p  61   KVEKQGRAKKITVFRYKPKKNVHKKQGHRQPYTKVTIEKINA  102
```

B

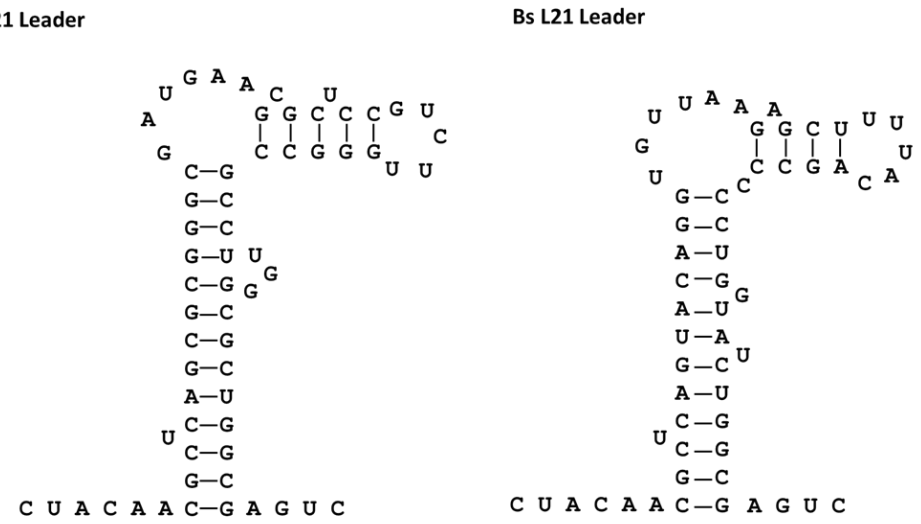


Figure A1.8 ***Comparison of Geobacillus kaustophilus (Gk) and B. subtilis (Bs) L21 Proteins and L21 RNA Leaders.*** *Bs and Gk L21p are similar; Bs and Gk leaders have a similar shape. Proteins from Gk, a thermophilic relative of Bs, often show increased stability in in vitro assays compared to proteins from Bs.*

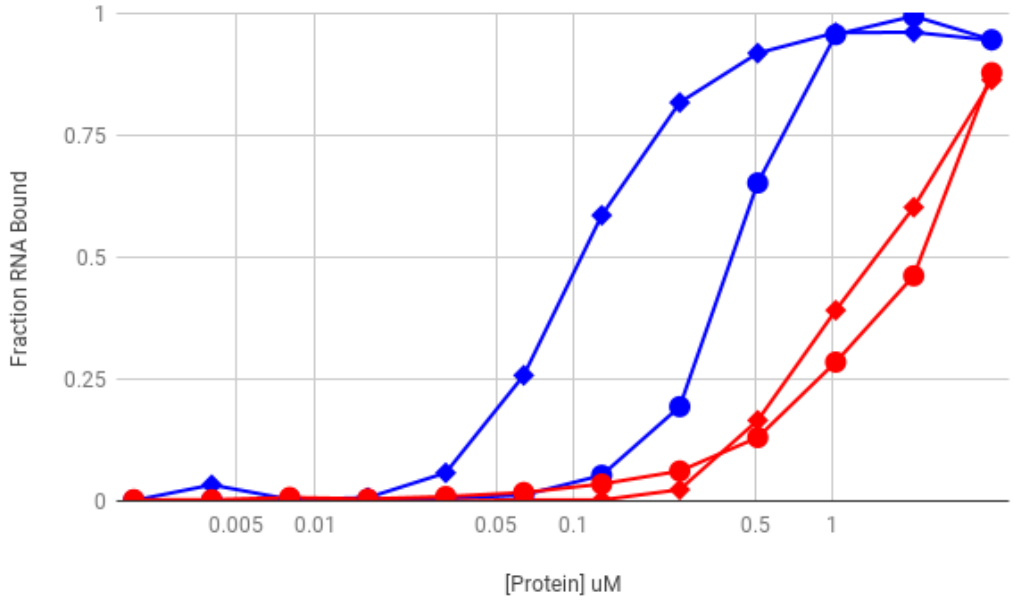


Figure A1.9 ***Filter Binding Assays of L21 RNA (Version 4) to L21 Protein.*** *Assays show that Bs L21 protein (red) binds poorly in vitro, but Gk L21 protein (blue) binds both Bs (circle) and Gk L21 RNA (diamond). Error not calculated.*

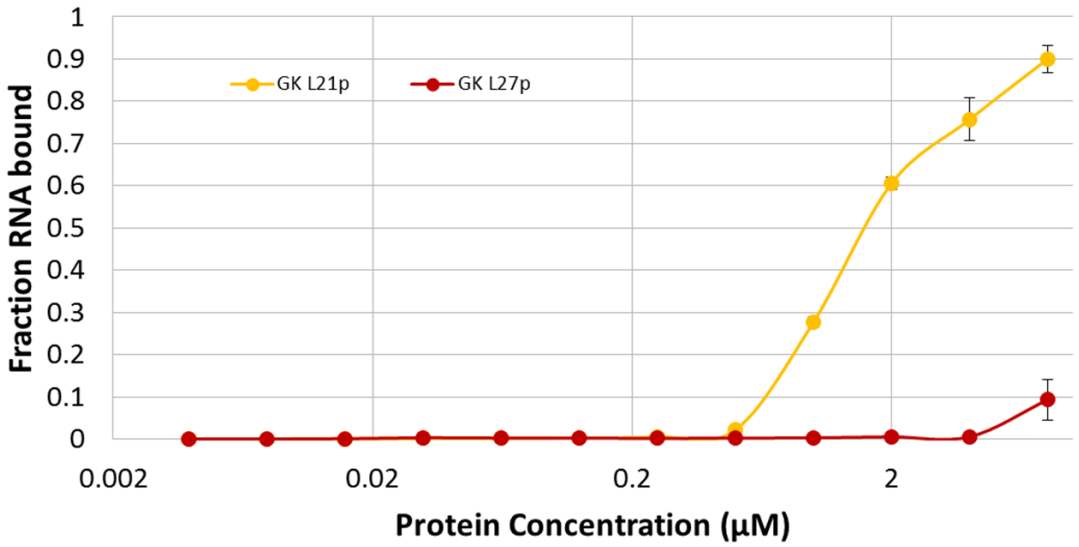


Figure A1.10 ***Filter Binding Assay for*** **G. kaustophilus** ***L21 and L27 Proteins to*** **G. kaustophilus** ***RNA Leader.*** *Gk L27p does not bind L21 RNA, supporting β-galactosidase assays results. Error bars show standard error of three technical replicates. Error bars may be within datapoint.*

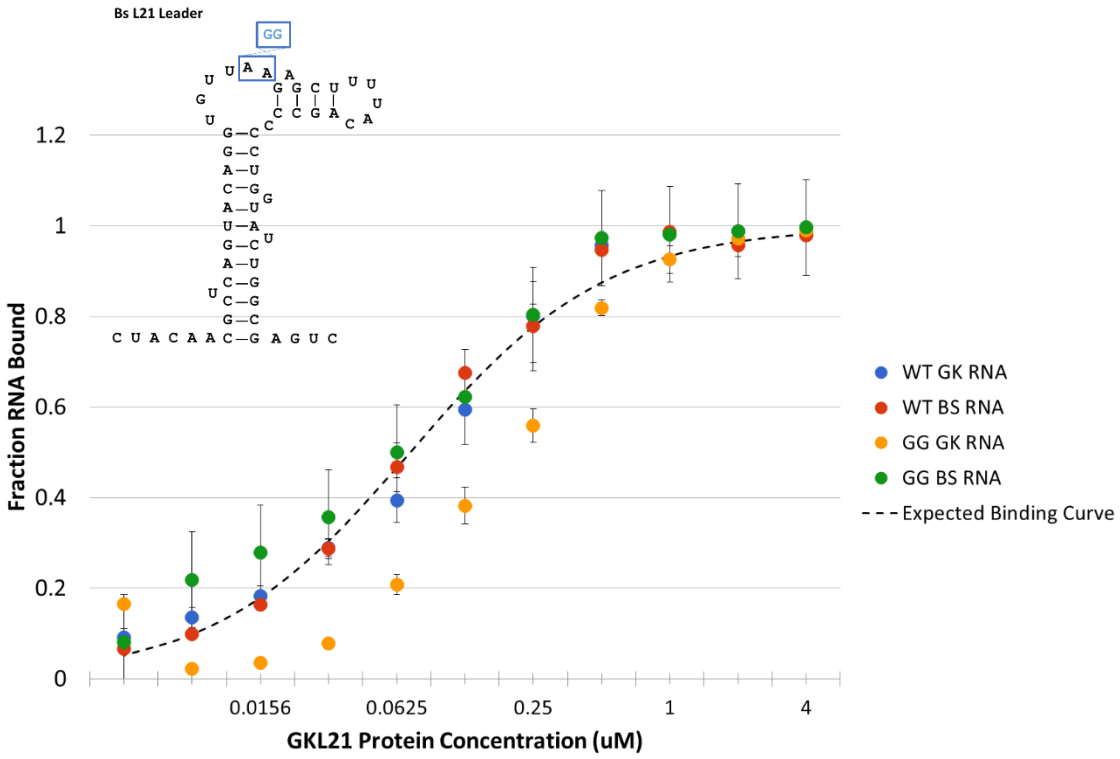
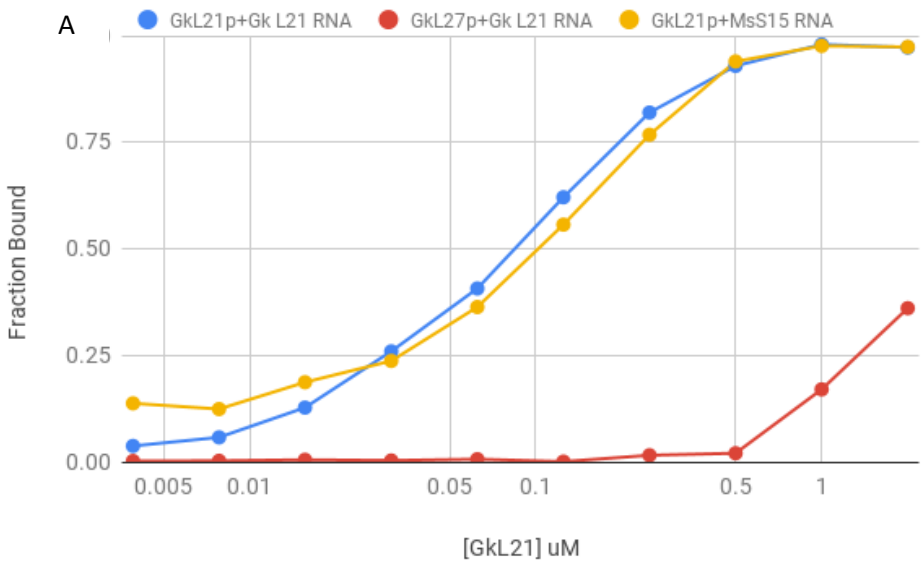


Figure A1.11 ***Filter Binding Assay of*** **G. kaustophilus** ***L21 protein with wild type and mutant RNAs.*** *For both the* B. subtilis *and the* G. kaustophilus *leaders, the AA in the loop was mutated to GG. Error bars represent standard error of technical triplicates. Although this RNA was targeted to conserved residues, there is little change in binding between species and also between wild type and mutants, suggesting that the binding observed* in vitro *may not be specific.*



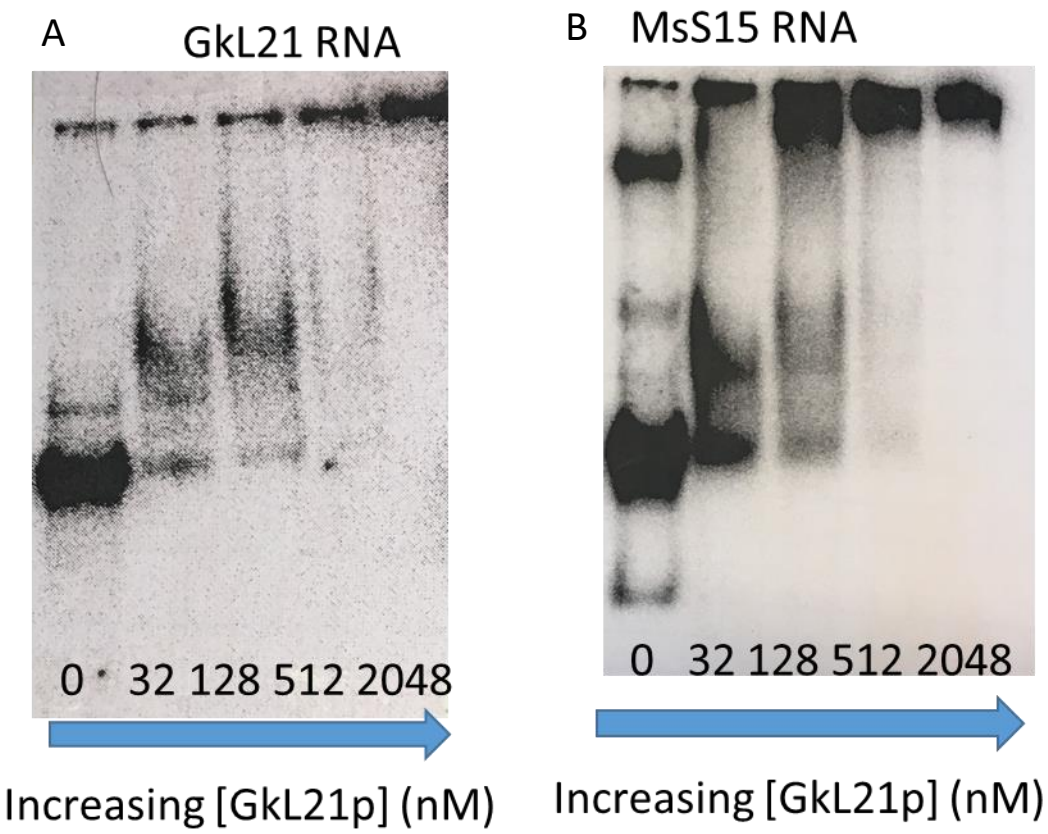


Figure A1.12 **In Vitro *Binding of* G. kaustophilus *L21 Protein to* G. kaustophilus *L21 Leader RNA and M. smegmatis S15 RNA.*** *Filter binding assay (A) shows Gk RNA and Ms RNA bind at similar concentrations to Gk L21 protein. Gk L27 protein, used as a protein control, shows no significant binding to Gk L21 leader RNA. An electrophoretic mobility shift assay (B) confirms that the binding in both cases is likely to be non-specific binding. High bands even in the absence of protein could indicate the formation of aggregates.*

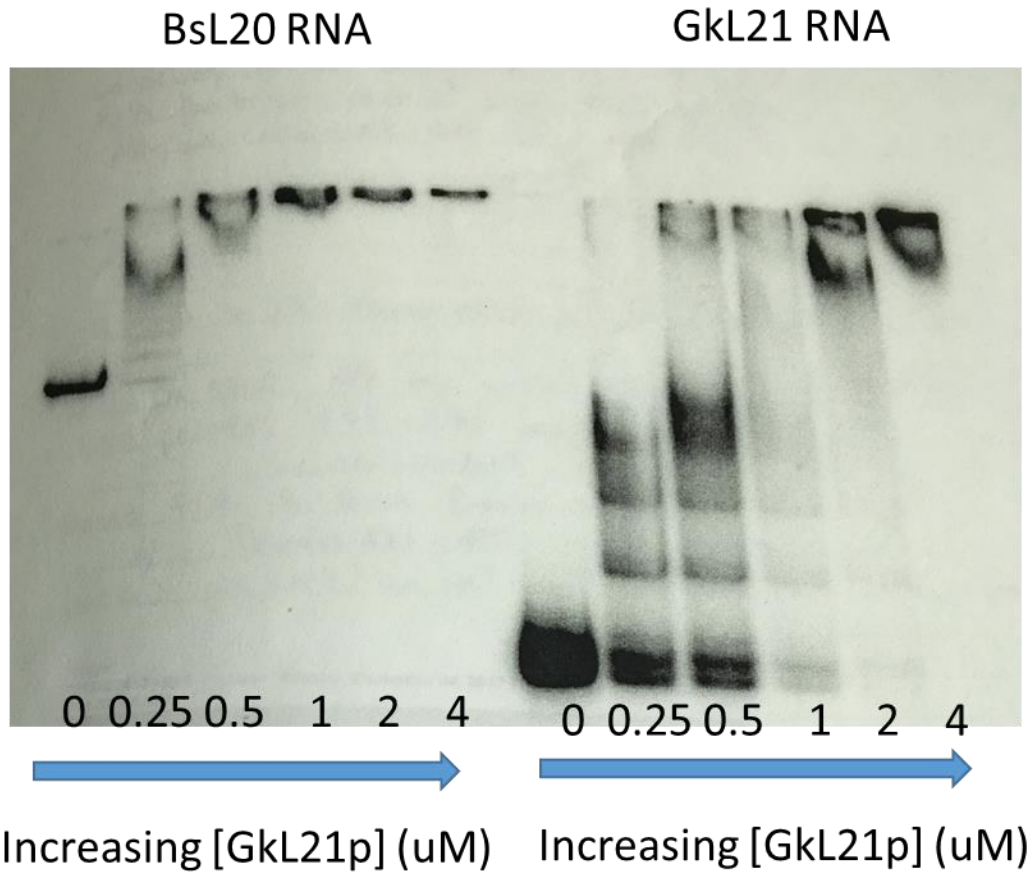


Figure A1.13 ***Electronic Mobility Shift Assay of*** **G. kaustophilus** ***L21 Protein Binding to B. subtilis L20 Leader RNA.*** *Banding pattern suggests nonspecific binding.*

REFERENCES

1 Root-Bernstein, M. and Root-Bernstein, R. (2015) The ribosome as a missing link in the evolution of life. J. Theor. Biol. **367**, 130–158.

2 Burton, B., Zimmermann, M. T., Jernigan, R. L. and Wang, Y. (2012) A Computational Investigation on the Connection between Dynamics Properties of Ribosomal Proteins and Ribosome Assembly. PLOS Comput. Biol., Public Library of Science **8**, e1002530.

3 Carrier, M.-C., Ng Kwan Lim, E., Jeannotte, G. and Massé, E. (2020) Trans-Acting Effectors Versus RNA Cis-Elements: A Tightly Knit Regulatory Mesh. Front. Microbiol. **11**, 609237.

4 Kohler, R., Mooney, R. A., Mills, D. J., Landick, R. and Cramer, P. (2017) Architecture of a transcribing-translating expressome. Science **356**, 194–197.

5 Johnson, G. E., Lalanne, J.-B., Peters, M. L. and Li, G.-W. (2020) Functionally uncoupled transcription-translation in *Bacillus subtilis*. Nature **585**, 124–128.

6 Chatterjee, S., Chauvier, A., Dandpat, S. S., Artsimovitch, I. and Walter, N. G. (2021) A translational riboswitch coordinates nascent transcription-translation coupling. Proc. Natl. Acad. Sci. U. S. A. **118**, e2023426118.

7 D'Heygère, F., Rabhi, M. and Boudvillain, M. (2013) Phyletic distribution and conservation of the bacterial transcription termination factor Rho. Microbiol. Read. Engl. **159**, 1423–1436.

8 Samatova, E., Daberger, J., Liutkute, M. and Rodnina, M. V. (2020) Translational Control by Ribosome Pausing in Bacteria: How a Non-uniform Pace of Translation Affects Protein Production and Folding. Front. Microbiol. **11**, 619430.

9 Klumpp, S., Dong, J. and Hwa, T. (2012) On Ribosome Load, Codon Bias and Protein Abundance. PLOS ONE, Public Library of Science **7**, e48542.

10 Meyer, M. M. (2017) The role of mRNA structure in bacterial translational regulation. Wiley Interdiscip. Rev. RNA **8**.

11 Lee, J.-H., Lee, E.-J. and Roe, J.-H. (2022) uORF-mediated riboregulation controls transcription of whiB7/wblC antibiotic resistance gene. Mol. Microbiol. **117**, 179–192.

12 Cai, X., Zhan, Y., Cao, Z., Yan, B. and Cai, J. (2021) Expression of ribosomal protection protein RppA is regulated by a ribosome-dependent ribo-regulator and two mistranslation products. Environ. Microbiol. **23**, 696–712.

13 Sun, Z., Zhou, N., Zhang, W., Xu, Y. and Yao, Y.-F. (2022) Dual role of CsrA in regulating the hemolytic activity of Escherichia coli O157:H7. Virulence **13**, 859–874.

14 Gray, E. C., Beringer, D. M. and Meyer, M. M. (2020) Siblings or doppelgängers? Deciphering the evolution of structured cis-regulatory RNAs beyond homology. Biochem. Soc. Trans. **48**, 1941–1951.

15 Ariza-Mateos, A., Nuthanakanti, A. and Serganov, A. (2021) Riboswitch Mechanisms: New Tricks for an Old Dog. Biochem. Biokhimiia **86**, 962–975.

16 Bastet, L., Chauvier, A., Singh, N., Lussier, A., Lamontagne, A.-M., Prévost, K., Massé, E., Wade, J. T. and Lafontaine, D. A. (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. Nucleic Acids Res. **45**, 7474–7486.

17 McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. and Breaker, R. R. (2017) Riboswitch diversity and distribution. RNA N. Y. N **23**, 995–1011.

18 Harvey, R. J. (1970) Regulation of ribosomal protein synthesis in Escherichia coli. J. Bacteriol. **101**, 574–583.

19 Kraft, A., Lutz, C., Lingenhel, A., Gröbner, P. and Piendl, W. (1999) Control of ribosomal protein L1 synthesis in mesophilic and thermophilic archaea. Genetics **152**, 1363–1372.

20 Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B. and Ehresmann, C. (1990) Target site of Escherichia coli ribosomal protein S15 on its messenger RNA. Conformation and interaction with the protein. J. Mol. Biol. **211**, 415–426.

21 Eckert, I. and Weinberg, Z. (2020) Discovery of 20 novel ribosomal leader candidates in bacteria and archaea. BMC Microbiol. **20**, 130.

22 Meyer, M. M. (2018) rRNA Mimicry in RNA Regulation of Gene Expression. Microbiol. Spectr. **6**.

23 Babina, A. M., Lea, N. E. and Meyer, M. M. (2017) *In Vivo* Behavior of the Tandem Glycine Riboswitch in *Bacillus subtilis*. mBio **8**.

24 Nomura, M. (1999) Regulation of ribosome biosynthesis in Escherichia coli and Saccharomyces cerevisiae: diversity and common principles. J. Bacteriol. **181**, 6857–6864.

25 Nomura, M., Yates, J. L., Dean, D. and Post, L. E. (1980) Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA. Proc. Natl. Acad. Sci. U. S. A. **77**, 7084–7088.

26 Worbs, M., Huber, R. and Wahl, M. C. (2000) Crystal structure of ribosomal protein L4 shows RNA-binding sites for ribosome incorporation and feedback control of the S10 operon. EMBO J. **19**, 807–818.

27 Deiorio-Haggar, K., Anthony, J. and Meyer, M. M. (2013) RNA structures regulating ribosomal protein biosynthesis in bacilli. RNA Biol. **10**, 1180–1184.

28 Philippe, C., Eyermann, F., Bénard, L., Portier, C., Ehresmann, B. and Ehresmann, C. (1993) Ribosomal protein S15 from Escherichia coli modulates its own translation by trapping the ribosome on the mRNA initiation loading site. Proc. Natl. Acad. Sci. U. S. A. **90**, 4394–4398.

29 Boehringer, D. and Ban, N. (2007) Trapping the ribosome to control gene expression. Cell **130**, 983–985.

30 Haentjens-Sitri, J., Allemand, F., Springer, M. and Chiaruttini, C. (2008) A competition mechanism regulates the translation of the Escherichia coli operon encoding ribosomal proteins L35 and L20. J. Mol. Biol. **375**, 612–625.

31 Serganov, A., Polonskaia, A., Ehresmann, B., Ehresmann, C. and Patel, D. J. (2003) Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. EMBO J. **22**, 1898–1908.

32 Nevskaya, N., Tishchenko, S., Gabdoulkhakov, A., Nikonova, E., Nikonov, O., Nikulin, A., Platonova, O., Garber, M., Nikonov, S. and 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 Res. **33**, 478–485.

33 Korepanov, A. P., Kostareva, O. S., Bazhenova, M. V., Bubunenko, M. G., Garber, M. B. and Tishchenko, S. V. (2015) Studying the properties of domain I of the ribosomal protein l1: incorporation into ribosome and regulation of the l1 operon expression. Protein J. **34**, 103–110.

34 Tishchenko, S., Nikonova, E., Nikulin, A., Nevskaya, N., Volchkov, S., Piendl, W., Garber, M. and Nikonov, S. (2006) Structure of the ribosomal protein L1-mRNA complex at 2.1 A resolution: common features of crystal packing of L1-RNA complexes. Acta Crystallogr. D Biol. Crystallogr. **62**, 1545–1554.

35 Kostareva, O., Tishchenko, S., Nikonova, E., Kljashtorny, V., Nevskaya, N., Nikulin, A., Sycheva, A., Moshkovskii, S., Piendl, W., Garber, M., *et al.* (2011) Disruption of shape complementarity in the ribosomal protein L1-RNA contact region does not hinder specific recognition of the RNA target site. J. Mol. Recognit. JMR **24**, 524–532.

36 Branlant, C., Krol, A., Sriwdada, J., Ebel, J. P., Sloof, P. and Garrett, R. A. (1976) The binding site of protein L1 ON 23-S ribosomal RNA of Escherichia coli. 2. Identification of the rna region contained in the L1 ribonucleoproteins and determination of the order of the RNA subfragments within this region. Eur. J. Biochem. **70**, 457–469.

37 Branlant, C., Krol, A. and Ebel, J. P. (1980) Characterization of the Escherichia coli 23S Ribosomal RNA region associated with ribosomal protein L1. Evidence for homologies with the 5'-end region of the L11 operon. Nucleic Acids Res. **8**, 5567–5577.

38 Wilson, D. N. and Nierhaus, K. H. (2005) Ribosomal proteins in the spotlight. Crit. Rev. Biochem. Mol. Biol. **40**, 243–267.

39 Tishchenko, S., Nikonova, E., Kljashtorny, V., Kostareva, O., Nevskaya, N., Piendl, W., Davydova, N., Streltsov, V., Garber, M. and Nikonov, S. (2007) Domain I of ribosomal protein L1 is sufficient for specific RNA binding. Nucleic Acids Res. **35**, 7389–7395.

40 Ameres, S. L., Shcherbakov, D., Nikonova, E., Piendl, W., Schroeder, R. and Semrad, K. (2007) RNA chaperone activity of L1 ribosomal proteins: phylogenetic conservation and splicing inhibition. Nucleic Acids Res. **35**, 3752–3763.

41 Mohan, S. and Noller, H. F. (2017) Recurring RNA structural motifs underlie the mechanics of L1 stalk movement. Nat. Commun. **8**, 14285.

42 Nikulin, A., Eliseikina, I., Tishchenko, S., Nevskaya, N., Davydova, N., Platonova, O., Piendl, W., Selmer, M., Liljas, A., Drygin, D., *et al.* (2003) Structure of the L1 protuberance in the ribosome. Nat. Struct. Biol. **10**, 104–108.

43 Fu, Y., Deiorio-Haggar, K., Anthony, J. and Meyer, M. M. (2013) Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. Nucleic Acids Res. **41**, 3491–3503.

44 Gregory, R. J., Zeller, M. L., Thurlow, D. L., Gourse, R. L., Stark, M. J., Dahlberg, A. E. and Zimmermann, R. A. (1984) Interaction of ribosomal proteins S6, S8, S15 and S18 with the central domain of 16 S ribosomal RNA from Escherichia coli. J. Mol. Biol. **178**, 287–302.

45 Shajani, Z., Sykes, M. T. and Williamson, J. R. (2011) Assembly of bacterial ribosomes. Annu. Rev. Biochem. **80**, 501–526.

46 Jagannathan, I. and Culver, G. M. (2003) Assembly of the central domain of the 30S ribosomal subunit: roles for the primary binding ribosomal proteins S15 and S8. J. Mol. Biol. **330**, 373–383.

47 Naganathan, A., Keltz, R., Lyon, H. and Culver, G. M. (2021) Uncovering a delicate balance between endonuclease RNase III and ribosomal protein S15 in *E. coli* ribosome assembly. Biochimie **191**, 104–117.

48 Bubunenko, M., Korepanov, A., Court, D. L., Jagannathan, I., Dickinson, D., Chaudhuri, B. R., Garber, M. B. and Culver, G. M. (2006) 30S ribosomal subunits can be assembled *in vivo* without primary binding ribosomal protein S15. RNA N. Y. N **12**, 1229–1239.

49 Guthrie, C., Nashimoto, H. and Nomura, M. (1969) Structure and function of *E. coli* ribosomes. 8. Cold-sensitive mutants defective in ribosome assembly. Proc. Natl. Acad. Sci. U. S. A. **63**, 384–391.

50 Schnier, J. and Isono, K. (1984) Insertion of IS1 into the rpsE gene for ribosomal protein S5 causes cold-sensitivity in Escherichia coli. Mol. Gen. Genet. MGG **195**, 364–366.

51 Wiener, L., Schüler, D. and Brimacombe, R. (1988) Protein binding sites on Escherichia coli 16S ribosomal RNA; RNA regions that are protected by proteins S7, S9 and S19, and by proteins S8, S15 and S17. Nucleic Acids Res. **16**, 1233–1250.

52 Clemons, W. M., Davies, C., White, S. W. and Ramakrishnan, V. (1998) Conformational variability of the N-terminal helix in the structure of ribosomal protein S15. Struct. Lond. Engl. 1993 **6**, 429–438.

53 Serganov, A., Rak, A., Garber, M., Reinbolt, J., Ehresmann, B., Ehresmann, C., Grunberg-Manago, M. and Portier, C. (1997) Ribosomal protein S15 from Thermus thermophilus--cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein. Eur. J. Biochem. **246**, 291–300.

54 Baker, K. A., Lamichhane, R., Lamichhane, T., Rueda, D. and Cunningham, P. R. (2016) Protein-RNA Dynamics in the Central Junction Control 30S Ribosome Assembly. J. Mol. Biol. **428**, 3615–3631.

55 Zimmermann, R. A., Muto, A., Fellner, P., Ehresmann, C. and Branlant, C. (1972) Location of Ribosomal Protein Binding Sites on 16S Ribosomal RNA. Proc. Natl. Acad. Sci., Proceedings of the National Academy of Sciences **69**, 1282–1286.

56 Nikulin, A., Serganov, A., Ennifar, E., Tishchenko, S., Nevskaya, N., Shepard, W., Portier, C., Garber, M., Ehresmann, B., Ehresmann, C., *et al.* (2000) Crystal structure of the S15-rRNA complex. Nat. Struct. Biol. **7**, 273–277.

57 Bénard, L., Philippe, C., Dondon, L., Grunberg-Manago, M., Ehresmann, B., Ehresmann, C. and Portier, C. (1994) Mutational analysis of the pseudoknot structure of the S15 translational operator from Escherichia coli. Mol. Microbiol. **14**, 31–40.

58 Bénard, L., Philippe, C., Ehresmann, B., Ehresmann, C. and Portier, C. (1996) Pseudoknot and translational control in the expression of the S15 ribosomal protein. Biochimie **78**, 568–576.

59 Philippe, C., Bénard, L., Eyermann, F., Cachia, C., Kirillov, S. V., Portier, C., Ehresmann, B. and Ehresmann, C. (1994) Structural elements of rps0 mRNA involved in the modulation of

translational initiation and regulation of *E. coli* ribosomal protein S15. Nucleic Acids Res. **22**, 2538–2546.

60 Bénard, L., Mathy, N., Grunberg-Manago, M., Ehresmann, B., Ehresmann, C. and Portier, C. (1998) Identification in a pseudoknot of a U.G motif essential for the regulation of the expression of ribosomal protein S15. Proc. Natl. Acad. Sci. U. S. A. **95**, 2564–2567.

61 Scott, L. G. and Williamson, J. R. (2005) The binding interface between Bacillus stearothermophilus ribosomal protein S15 and its 5'-translational operator mRNA. J. Mol. Biol. **351**, 280–290.

62 Slinger, B. L., Deiorio-Haggar, K., Anthony, J. S., Gilligan, M. M. and Meyer, M. M. (2014) Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 in diverse bacterial phyla. BMC Genomics **15**, 657.

63 Slinger, B. L., Newman, H., Lee, Y., Pei, S. and Meyer, M. M. (2015) Co-evolution of Bacterial Ribosomal Protein S15 with Diverse mRNA Regulatory Structures. PLoS Genet. **11**, e1005720.

64 Miller, J. H. (1972) Experiments in molecular genetics, N.Y.] Cold Spring Harbor Laboratory.

65 Aseev, L. V., Koledinskaya, L. S. and Boni, I. V. (2016) Regulation of Ribosomal Protein Operons rplM-rpsI, rpmB-rpmG, and rplU-rpmA at the Transcriptional and Translational Levels. J. Bacteriol., American Society for Microbiology **198**, 2494–2502.

66 Aseev, L. V., Levandovskaya, A. A., Tchufistova, L. S., Scaptsova, N. V. and Boni, I. V. (2008) A new regulatory circuit in ribosomal protein operons: S2-mediated control of the rpsB-tsf expression *in vivo*. RNA **14**, 1882–1894.

67 Mikael Wikström, P. and Björk, G. R. (1989) A regulatory element within a gene of a ribosomal protein operon of Escherichia coli negatively controls expression by decreasing the translational efficiency. Mol. Gen. Genet. MGG **219**, 381–389.

68 Schaefer, J., Jovanovic, G., Kotta-Loizou, I. and Buck, M. (2016) Single-step method for β-galactosidase assays in Escherichia coli using a 96-well microplate reader. Anal. Biochem. **503**, 56–57.

69 Olsson, O., Koncz, C. and Szalay, A. A. (1988) The use of the luxA gene of the bacterial luciferase operon as a reporter gene. Mol. Gen. Genet. MGG **215**, 1–9.

70 Wang, C. D., Mansky, R., LeBlanc, H., Gravel, C. M. and Berry, K. E. (2021) Optimization of a bacterial three-hybrid assay through *in vivo* titration of an RNA-DNA adapter protein. RNA N. Y. N **27**, 513–526.

71 Stockert, O. M., Gravel, C. M. and Berry, K. E. (2022) A bacterial three-hybrid assay for forward and reverse genetic analysis of RNA-protein interactions. Nat. Protoc. **17**, 941–961.

72 Bailey, J. E., Fazel-Makjlessi, J., McQuitty, D. N., Lee, Y. N., Allred, J. C. and Oro, J. A. (1977) Characterization of bacterial growth by means of flow microfluorometry. Science **198**, 1175–1176.

73 Paau, A. S., Cowles, J. R. and Oro, J. (1977) Flow-microfluorometric analysis of Escherichia coli, Rhizobium meliloti, and Rhizobium japonicum at different stages of the growth cycle. Can. J. Microbiol. **23**, 1165–1169.

74 Steen, H. B. (2000) Flow cytometry of bacteria: glimpses from the past with a view to the future. J. Microbiol. Methods **42**, 65–74.

75 Ambriz-Aviña, V., Contreras-Garduño, J. A. and Pedraza-Reyes, M. (2014) Applications of flow cytometry to characterize bacterial physiological responses. BioMed Res. Int. **2014**, 461941.

76 Davey, H. M. and Kell, D. B. (1996) Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. Microbiol. Rev., American Society for Microbiology **60**, 641–696.

77 Vives-Rego, J., Lebaron, P. and Nebe-von Caron, G. (2000) Current and future applications of flow cytometry in aquatic microbiology. FEMS Microbiol. Rev. **24**, 429–448.

78 Nebe-von-Caron, G., Stephens, P. J., Hewitt, C. J., Powell, J. R. and Badley, R. A. (2000) Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. J. Microbiol. Methods **42**, 97–114.

79 Diaper, J. P. and Edwards, C. (1994) Survival of Staphylococcus aureus in lakewater monitored by flow cytometry. Microbiol. Read. Engl. **140 (Pt 1)**, 35–42.

80 Amor, K. B., Breeuwer, P., Verbaarschot, P., Rombouts, F. M., Akkermans, A. D. L., De Vos, W. M. and Abee, T. (2002) Multiparametric flow cytometry and cell sorting for the assessment of viable, injured, and dead bifidobacterium cells during bile salt stress. Appl. Environ. Microbiol. **68**, 5209–5216.

81 Saini, K. and Gupta, R. (2022) Cell surface expression of γ-CGTase from Evansella caseinilytica on *E. coli*: Application in the enzymatic conversion of starch to γ-cyclodextrin. Enzyme Microb. Technol. **159**, 110066.

82 Saint-Ruf, C., Crussard, S., Franceschi, C., Orenga, S., Ouattara, J., Ramjeet, M., Surre, J. and Matic, I. (2016) Antibiotic Susceptibility Testing of the Gram-Negative Bacteria Based on Flow Cytometry. Front. Microbiol. **7**, 1121.

83 Sycuro, L. K., Rule, C. S., Petersen, T. W., Wyckoff, T. J., Sessler, T., Nagarkar, D. B., Khalid, F., Pincus, Z., Biboy, J., Vollmer, W., *et al.* (2013) Flow cytometry-based enrichment for cell shape mutants identifies multiple genes that influence Helicobacter pylori morphology. Mol. Microbiol. **90**, 869–883.

84 Suea-Ngam, A., Howes, P. D., Srisa-Art, M. and deMello, A. J. (2019) Droplet microfluidics: from proof-of-concept to real-world utility? Chem. Commun. Camb. Engl. **55**, 9895–9903.

85 Thibault, D., Jensen, P. A., Wood, S., Qabar, C., Clark, S., Shainheit, M. G., Isberg, R. R. and van Opijnen, T. (2019) Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. Nat. Commun. **10**, 5729.

86 Liu, D., Sun, M., Zhang, J., Hu, R., Fu, W., Xuanyuan, T. and Liu, W. (2022) Single-cell droplet microfluidics for biomedical applications. Analyst, Royal Society of Chemistry **147**, 2294–2316.

87 Hsieh, K., Mach, K. E., Zhang, P., Liao, J. C. and Wang, T.-H. (2022) Combating Antimicrobial Resistance via Single-Cell Diagnostic Technologies Powered by Droplet Microfluidics. Acc. Chem. Res. **55**, 123–133.

88 Zhang, P., Kaushik, A. M., Hsieh, K., Li, S., Lewis, S., Mach, K. E., Liao, J. C., Carroll, K. C. and Wang, T.-H. (2022) A Cascaded Droplet Microfluidic Platform Enables High-Throughput Single Cell Antibiotic Susceptibility Testing at Scale. Small Methods **6**, 2101254.

89 Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M. and Keller, M. (2002) Cultivating the uncultured. Proc. Natl. Acad. Sci. U. S. A. **99**, 15681–15686.

90 Xu, A., Zhang, X., Cao, S., Zhou, X., Yu, Z., Qian, X., Zhou, J., Dong, W. and Jiang, M. (2022) Transcription-Associated Fluorescence-Activated Droplet Sorting for Di-rhamnolipid Hyperproducers. ACS Synth. Biol. **11**, 1992–2000.

91 Rotem, A., Serohijos, A. W. R., Chang, C. B., Wolfe, J. T., Fischer, A. E., Mehoke, T. S., Zhang, H., Tao, Y., Lloyd Ung, W., Choi, J.-M., *et al.* (2018) Evolution on the Biophysical Fitness Landscape of an RNA Virus. Mol. Biol. Evol. **35**, 2390–2400.

92 Balasubramanian, S., Chen, J., Wigneswaran, V., Bang-Berthelsen, C. H. and Jensen, P. R. (2021) Droplet-Based Microfluidic High Throughput Screening of Corynebacterium

glutamicum for Efficient Heterologous Protein Production and Secretion. Front. Bioeng. Biotechnol. **9**, 668513.

93 Chiu, F. W. Y. and Stavrakis, S. (2019) High-throughput droplet-based microfluidics for directed evolution of enzymes. Electrophoresis **40**, 2860–2872.

94 Vallejo, D., Nikoomanzar, A., Paegel, B. M. and Chaput, J. C. (2019) Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. ACS Synth. Biol. **8**, 1430–1440.

95 Holstein, J. M., Gylstorff, C. and Hollfelder, F. (2021) Cell-free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. ACS Synth. Biol. **10**, 252–257.

96 Abatemarco, J., Sarhan, M. F., Wagner, J. M., Lin, J.-L., Liu, L., Hassouneh, W., Yuan, S.-F., Alper, H. S. and Abate, A. R. (2017) RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. Nat. Commun. **8**, 332.

97 Autour, A., Bouhedda, F., Cubi, R. and Ryckelynck, M. (2019) Optimization of fluorogenic RNA-based biosensors using droplet-based microfluidic ultrahigh-throughput screening. Methods San Diego Calif **161**, 46–53.

98 Ryckelynck, M., Baudrey, S., Rick, C., Marin, A., Coldren, F., Westhof, E. and Griffiths, A. D. (2015) Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. RNA N. Y. N **21**, 458–469.

99 Bouhedda, F., Cubi, R., Baudrey, S. and Ryckelynck, M. (2021) µIVC-Seq: A Method for Ultrahigh-Throughput Development and Functional Characterization of Small RNAs. Methods Mol. Biol. Clifton NJ **2300**, 203–237.

100 Fischlechner, M., Schaerli, Y., Mohamed, M. F., Patil, S., Abell, C. and Hollfelder, F. (2014) Evolution of enzyme catalysts caged in biomimetic gel-shell beads. Nat. Chem. **6**, 791–796.

101 Zhu, Z. and Yang, C. J. (2017) Hydrogel Droplet Microfluidics for High-Throughput Single Molecule/Cell Analysis. Acc. Chem. Res. **50**, 22–31.

102 Ma, C., Li, J., Zhang, B., Liu, C., Zhang, J. and Liu, Y. (2019) Hydrogel Microparticles Functionalized with Engineered Escherichia coli as Living Lactam Biosensors. Sensors **19**, E5556.

103 Zinchenko, A., Devenish, S. R. A., Kintses, B., Colin, P.-Y., Fischlechner, M. and Hollfelder, F. (2014) One in a million: flow cytometric sorting of single cell-lysate assays in monodisperse picolitre double emulsion droplets for directed evolution. Anal. Chem. **86**, 2526–2533.

104 Larsen, A. C., Dunn, M. R., Hatch, A., Sau, S. P., Youngbull, C. and Chaput, J. C. (2016) A general strategy for expanding polymerase function by droplet microfluidics. Nat. Commun. **7**, 11235.

105 Schwarzbauer, J. and Craven, G. R. (1981) Apparent association constants for *E. coli* ribosomal proteins S4, S7, S8, S15, S17 and S20 binding to 16S RNA. Nucleic Acids Res. **9**, 2223–2237.

106 Orr, J. W., Hagerman, P. J. and Williamson, J. R. (1998) Protein and Mg(2+)-induced conformational changes in the S15 binding site of 16 S ribosomal RNA. J. Mol. Biol. **275**, 453–464.

107 Boussebayle, A., Torka, D., Ollivaud, S., Braun, J., Bofill-Bosch, C., Dombrowski, M., Groher, F., Hamacher, K. and Suess, B. (2019) Next-level riboswitch development-implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. Nucleic Acids Res. **47**, 4883–4895.

108 Groher, F., Bofill-Bosch, C., Schneider, C., Braun, J., Jager, S., Geißler, K., Hamacher, K. and Suess, B. (2018) Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator. Nucleic Acids Res. **46**, 2121–2132.

109 Urak, K. T., Shore, S., Rockey, W. M., Chen, S.-J., McCaffrey, A. P. and Giangrande, P. H. (2016) *In vitro* RNA SELEX for the generation of chemically-optimized therapeutic RNA drugs. Methods San Diego Calif **103**, 167–174.

110 Schneider, C. and Suess, B. (2016) Identification of RNA aptamers with riboswitching properties. Methods San Diego Calif **97**, 44–50.

111 Kramat, J. and Suess, B. (2022) Efficient Method to Identify Synthetic Riboswitches Using RNA-Based Capture-SELEX Combined with *in Vivo* Screening. Methods Mol. Biol. Clifton NJ **2518**, 157–177.

112 Darmostuk, M., Rimpelova, S., Gbelcova, H. and Ruml, T. (2015) Current approaches in SELEX: An update to aptamer selection technology. Biotechnol. Adv. **33**, 1141–1161.

113 Slinger, B. L. and Meyer, M. M. (2016) RNA regulators responding to ribosomal protein S15 are frequent in sequence space. Nucleic Acids Res. **44**, 9331–9341.

114 Thomas Hunt Morgan and the Discovery of Sex Linkage | Learn Science at Scitable.

115 Blanco, C., Janzen, E., Pressman, A., Saha, R. and Chen, I. A. (2019) Molecular Fitness Landscapes from High-Coverage Sequence Profiling. Annu. Rev. Biophys. **48**, 1–18.

116 Patrick, C., Upadhyay, V., Lucas, A. and Mallela, K. M. G. (2022) Biophysical Fitness Landscape of the SARS-CoV-2 Delta Variant Receptor Binding Domain. J. Mol. Biol. **434**, 167622.

117 Lind, P. A., Arvidsson, L., Berg, O. G. and Andersson, D. I. (2017) Variation in Mutational Robustness between Different Proteins and the Predictability of Fitness Effects. Mol. Biol. Evol. **34**, 408–418.

118 Wu, Z., Cai, X., Zhang, X., Liu, Y., Tian, G.-B., Yang, J.-R. and Chen, X. (2022) Expression level is a major modifier of the fitness landscape of a protein coding gene. Nat. Ecol. Evol. **6**, 103–115.

119 Park, J. and Wang, H. H. (2021) Systematic dissection of σ70 sequence diversity and function in bacteria. Cell Rep. **36**, 109590.

120 Otwinowski, J. and Nemenman, I. (2013) Genotype to phenotype mapping and the fitness landscape of the *E. coli lac* promoter. PloS One **8**, e61570.

121 Safina, K. R., Mironov, A. A. and Bazykin, G. A. (2017) Compensatory Evolution of Intrinsic Transcription Terminators in Bacillus Cereus. Genome Biol. Evol. **9**, 340–349.

122 Knight, C. G., Platt, M., Rowe, W., Wedge, D. C., Khan, F., Day, P. J. R., McShea, A., Knowles, J. and Kell, D. B. (2009) Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape. Nucleic Acids Res. **37**, e6.

123 Zhou, Q., Sun, X., Xia, X., Fan, Z., Luo, Z., Zhao, S., Shakhnovich, E. and Liang, H. (2017) Exploring the Mutational Robustness of Nucleic Acids by Searching Genotype Neighborhoods in Sequence Space. J. Phys. Chem. Lett. **8**, 407–414.

124 Peri, G., Gibard, C., Shults, N. H., Crossin, K. and Hayden, E. J. (2022) Dynamic RNA Fitness Landscapes of a Group I Ribozyme during Changes to the Experimental Environment. Mol. Biol. Evol. **39**, msab373.

125 Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. and Chen, I. A. (2013) Comprehensive experimental fitness landscape and evolutionary network for small RNA. Proc. Natl. Acad. Sci. U. S. A. **110**, 14984–14989.

126 Torgerson, C. D., Hiller, D. A., Stav, S. and Strobel, S. A. (2018) Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. RNA rna.067884.118.

127 Torgerson, C. D., Hiller, D. A. and Strobel, S. A. (2020) The asymmetry and cooperativity of tandem glycine riboswitch aptamers. RNA rna.073577.119.

128 Ehresmann, C., Ehresmann, B., Ennifar, E., Dumas, P., Garber, M., Mathy, N., Nikulin, A., Portier, C., Patel, D. and Serganov, A. (2004) Molecular mimicry in translational regulation: the case of ribosomal protein S15. RNA Biol. **1**, 66–73.

129 Philippe, C., Bénard, L., Portier, C., Westhof, E., Ehresmann, B. and Ehresmann, C. (1995) Molecular dissection of the pseudoknot governing the translational regulation of Escherichia coli ribosomal protein S15. Nucleic Acids Res. **23**, 18–28.

130 Serganov, A., Ennifar, E., Portier, C., Ehresmann, B. and Ehresmann, C. (2002) Do mRNA and rRNA binding sites of *E. coli* ribosomal protein S15 share common structural determinants? J. Mol. Biol. **320**, 963–978.

131 Wu, Y.-J., Wu, C.-H., Yeh, A. Y.-C. and Wen, J.-D. (2014) Folding a stable RNA pseudoknot through rearrangement of two hairpin structures. Nucleic Acids Res. **42**, 4505–4515.

132 Sukovich, D. J., Kim, S. C., Ahmed, N. and Abate, A. R. (2017) Bulk double emulsification for flow cytometric analysis of microfluidic droplets. The Analyst **142**, 4618–4622.

133 Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. **24**, 79–88.

134 Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D. and Endy, D. (2009) Measuring the activity of BioBrick promoters using an *in vivo* reference standard. J. Biol. Eng. **3**, 4.

135 iGEM.

136 Ren, X., Yu, D., Yu, L., Gao, G., Han, S. and Feng, Y. (2007) A new study of cell disruption to release recombinant thermostable enzyme from Escherichia coli by thermolysis. J. Biotechnol. **129**, 668–673.

137 Shao, F., Yu, L., Zhang, Y., An, C., Zhang, H., Zhang, Y., Xiong, Y. and Wang, H. (2020) Microfluidic Encapsulation of Single Cells by Alginate Microgels Using a Trigger-Gellified Strategy. Front. Bioeng. Biotechnol. **8**, 583065.

138 Clement, K., Farouni, R., Bauer, D. E. and Pinello, L. (2018) AmpUMI: design and analysis of unique molecular identifiers for deep amplicon sequencing. Bioinforma. Oxf. Engl. **34**, i202–i210.

139 Abdelaal, A. S. and Yazdani, S. S. (2021) A genetic toolkit for co-expression of multiple proteins of diverse physiological implication. Biotechnol. Rep. Amst. Neth. **32**, e00692.

140 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. **2**, 2006.0008.

141 Overkamp, W., Beilharz, K., Weme, R. D. O., Solopova, A., Karsens, H., Kovács, Á. T., Kok, J., Kuipers, O. P. and Veening, J.-W. (2013) Benchmarking Various Green Fluorescent Protein Variants in *Bacillus subtilis*, *Streptococcus pneumoniae*, and *Lactococcus lactis* for Live Cell Imaging. Appl Env. Microbiol **79**, 6481–6490.

142 Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinforma. Oxf. Engl. **34**, i884–i890.

143 Faure, A. J., Schmiedel, J. M., Baeza-Centurion, P. and Lehner, B. (2020) DiMSum: an error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. Genome Biol. **21**, 207.

144 Thomason, L., Court, D. L., Bubunenko, M., Costantino, N., Wilson, H., Datta, S. and Oppenheim, A. (2007) Recombineering: genetic engineering in bacteria using homologous recombination. Curr. Protoc. Mol. Biol. **Chapter 1**, Unit 1.16.

145 Martin, S., Blankenship, C., Rausch, J. W. and Sztuba-Solinska, J. (2019) Using SHAPE-MaP to probe small molecule-RNA interactions. Methods San Diego Calif **167**, 105–116.

146 Watters, K. E. and Lucks, J. B. (2016) Mapping RNA Structure *in Vitro* with SHAPE Chemistry and Next-Generation Sequencing (SHAPE-Seq). Methods Mol. Biol. Clifton NJ **1490**, 135–162.

147 Zeller, M. J., Favorov, O., Li, K., Nuthanakanti, A., Hussein, D., Michaud, A., Lafontaine, D. A., Busan, S., Serganov, A., Aubé, J., *et al.* (2022) SHAPE-enabled fragment-based ligand discovery for RNA. Proc. Natl. Acad. Sci. U. S. A. **119**, e2122660119.

148 Pedrolli, D., Langer, S., Hobl, B., Schwarz, J., Hashimoto, M. and Mack, M. (2015) The ribB FMN riboswitch from Escherichia coli operates at the transcriptional and translational level and regulates riboflavin biosynthesis. FEBS J. **282**, 3230–3242.

149 Howe, J. A., Xiao, L., Fischmann, T. O., Wang, H., Tang, H., Villafania, A., Zhang, R., Barbieri, C. M. and Roemer, T. (2016) Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the *E. coli* FMN riboswitch. RNA Biol., Taylor & Francis **13**, 946–954.

150 Nelson, J. W., Atilho, R. M., Sherlock, M. E., Stockbridge, R. B. and Breaker, R. R. (2017) Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. Mol. Cell **65**, 220–230.

151 Sherlock, M. E., Malkowski, S. N. and Breaker, R. R. (2017) Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. Biochemistry **56**, 352–358.

152 Grundy, F. J., Lehman, S. C. and Henkin, T. M. (2003) The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. Proc. Natl. Acad. Sci. U. S. A. **100**, 12057–12062.

153 Patte, J.-C., Akrim, M. and Méjean, V. (1998) The leader sequence of the Escherichia coli lysC gene is involved in the regulation of LysC synthesis. FEMS Microbiol. Lett. **169**, 165–170.

154 Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. and Patel, D. J. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature **441**, 1167–1171.

155 Jeong, J. H., Kitakawa, M., Isono, S. and Isono, K. (1993) Cloning and nucleotide sequencing of the genes, rplU and rpmA, for ribosomal proteins L21 and L27 of Escherichia coli. DNA Seq. J. DNA Seq. Mapp. **4**, 59–67.

156 Schwedler, G., Albrecht-Ehrlich, R. and Rak, K. H. (1993) Immunoelectron microscopic localization of ribosomal proteins BS8, BS9, BS20, BL3 and BL21 on the surface of 30S and 50S subunits from Bacillus stearothermophilus. Eur. J. Biochem. **217**, 361–369.

157 Prew, M. (2014, May) The RNA-Based Regulation of Ribosomal Protein Operons L21 and S10 in *Bacillus subtilis*.