Biological Probes to Measure Transcription Dynamics in *E. coli*

Wilbert B. Copeland

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Herbert M. Sauro, Chair

Daniel L. Cook

Wendy Thomas

Program Authorized to Offer Degree:

Bioengineering

University of Washington

ABSTRACT

Biological Probes to Measure Transcription Dynamics in *E. coli*

Wilbert B. Copeland

Chair of the Supervisory Committee:
Professor Herbert M. Sauro
Bioengineering

RNA molecules exhibit dynamic behaviors that are critical to maintaining normal cell function, and their dynamic nature has been exploited to engineer biological systems. RNA dynamics processes that are often investigated include synthesis, degradation, transport, conformation, localization, and binding.

As researchers explore the nature of RNA in increasingly complex ways, it is important that they are equipped with equally complex tools that allow them to interrogate RNA function with fine detail in living systems. Several techniques currently exist to study RNA expression; however, no existing method allows for quantitative measurements of RNA concentration in living cells. As a result, precise characterization of RNA synthesis and degradation with high temporal resolution remains a challenge within the scientific community.

This dissertation seeks to demonstrate that a novel technology, fluorescence-activating aptamers, can be used to improve researchers' ability to quantify and to communicate measurements of RNA expression dynamics in both natural and engineered biological systems.

ACKNOWLEDGEMENTS

Countless mentors, family members, and friends have invested their time and effort on my behalf. I am grateful.

I feel very fortunate to have had Dr. Herbert Sauro as an advisor. Under his guidance, I have grown tremendously as a scientist, as an engineer, and as a citizen. I especially appreciate his willingness to allow me to define my own research project, and his support as I participated in many extracurricular and professional development activities that did not directly benefit the research goals of the laboratory.

I would also like to thank the rest of my supervisory committee: Drs. Wendy Thomas, Paul Wiggins, Daniel Cook, and Georg Seelig. They have provided me with direction and encouragement throughout my doctoral journey. From my project's inception through its culmination, I have listened to their advice and it has guided me past many technical roadblocks.

I have had the pleasure to work with many intelligent and friendly lab mates and collaborators. They have given me critical insights into my projects and have taught me several technical skills. From listening to me deliver a practice talk to bearing with me as I practice my latest joke, they have always made themselves available to offer a critical perspective. I would like to thank Bryan Bartley and Jamie Nunez who, for years, helped me develop experimental techniques and collect experimental data; Sean Sleight and Kyung Kim for helping me grasp concepts as varied as molecular cloning and control theory; and Aaron Miller and Bennett Ng for exposing me to continuous culture. I would also like to thank the rest of my co-workers for their thoughtful advice and assistance whenever I have needed it. Thank you, Michal, Deepak, Stanley, Kyle, Kiri, Lucian, and Frank.

During my time in Seattle, I have had the privilege to befriend many amazing people. With them I have explored the region and made great memories. I would especially like to thank Navid Farr who has been not only a great friend, but also helped me manufacture laboratory

equipment required to complete my experiments. I would also like to thank Shivang Dave for always being eager to impart his wisdom and for his help with reviewing the many fellowship and grant applications I have submitted. In addition to Seattle friends, frequent visits from San Diego friends helped to keep up my spirits.

I would like to thank family, both near and far. I will always appreciate calls from my parents and siblings; cards from my grandparents; dinners with the Grimes and Dixons; vacations with Tyrell, Wayne, and the rest of the Brown cousins; chats with those in Ohio; and lunches with those who drove across the state of Washington to see me.

Finally, I would like to thank funding agencies and previous mentors. Merck, UNCF, NSF, and UW GO-MAP provided me with the financial support that helped give me the freedom to pursue my own research interests. Past advisors Dr. Lawrence Alfred, Dr. Percy Russell, Dr. Trey Ideker, Dr. Sourav Bandyopadhyay, and Valita Jones exposed me to the exciting world of science research that initially motivated me to pursue a PhD.

DEDICATION

To my grandparents

Table of Contents

Chapter 1

INTRODUCTION

1.1 Motivation

What was once considered a molecule with a narrow biological relevance, RNA is now known to have diverse roles within cells [1]. The discovery of new classes of RNA molecules has fueled research focused on understanding the impact of RNA on cellular function. For example, studying the contributions of RNA towards phenotype [2] and regulating native RNA molecules to positively affect cellular function [3], [4] are quickly becoming prominent areas of research. Beyond basic science, many efforts to engineer novel behaviors in biological systems have been based on deliberate modification of RNA expression [5]; these engineered systems hold promise for addressing grand challenges in human health, renewable energy, and environmental sustainability.

As researchers explore the nature of RNA in increasingly complex ways, it is important that they are equipped with equally complex tools that allow them to interrogate RNA function with fine detail. The ability to accurately measure RNA expression and dynamics, and to coherently describe fundamental properties of RNA molecules will be critically important to the realization of advanced RNA-based biological engineering efforts and medical interventions.

It can be argued that the current, limited set of experimental techniques used to investigate intracellular RNA dynamics combined with a lack of canonical ways to report RNA function limits discovery and engineering efforts. **Therefore, developing strategies that improve researchers' ability to quantify and to communicate measures of RNA dynamics in both natural and engineered biological systems is a major motivation for the work presented in this dissertation.**

1.2 Specific Results

This dissertation aims: (1) to demonstrate the feasibility of using fluorescent RNA molecules in *E. coli*; (2) to validate the use of fluorescence intensity produced by these RNA as an indicator of expression; (3) to apply fluorescent RNA molecules in quantitative manner; and (4) to propose an improved strategy for measuring mRNA expression using fluorescent RNA molecules. The work presented in this thesis is conceptually separated into chapters with respect to these four major aims.

Chapter 4 provides a thorough characterization of malachite green aptamer in *E. coli*. It is important to identify whether this particular aptamer, in combination with its dye, are appropriate for use in bacterial cells. **Chapter 4 presents two major specific results.** (1) This chapter demonstrates that malachite green aptamer has favorable properties for use as a fluorescence reporter in *E. coli*. (2) This chapter establishes criteria for rigorous assessment of fluorescence-activating aptamers in *E. coli*.

Chapter 5 begins by validating the use of fluorescence from malachite green aptamer as a measure of RNA abundance. Following validation, a novel method to quantify RNA abundance from fluorescence intensity is developed by integrating mathematical models of transcription with empirical data. **Chapter 5 presents two major specific results.** (1) This chapter demonstrates that malachite green aptamer fluorescence is a valid reporter of abundance for our experimental conditions. (2) This chapter presents a novel way to predict absolute RNA concentration in living cells while using fluorescence-activating aptamers.

Chapter 6 applies malachite green aptamer to address a fundamental challenge in synthetic biology: characterization and standardization of genetic components. Specifically, we use malachite green aptamer to assess promoter strength in *E. coli*. **Chapter 6 presents two major specific results.** (1) This chapter shows that malachite green aptamer can be used for accurate, high-throughput characterization of DNA regulatory elements in *E. coli*. (2) This chapter demonstrates the importance of accounting for cell context when quantifying molecular processes in cells.

Chapter 7 explores different strategies to tag diverse mRNA sequences with malachite green aptamer in a non-obstructive manner. In this chapter, a variety of approaches to insulate malachite green aptamer and a gene-coding sequence are examined. **Chapter 7 presents two major specific results.** (1) This chapters shows that, by using a fluorescent protein in addition to malachite green aptamer, both protein and mRNA expression can be simultaneously measured in a semi-quantitative manner. (2) This chapter demonstrates that malachite green aptamer fluorescence is highly dependent upon its structure and highlights the fragile nature of malachite green aptamer fluorescence.

In addition to the results presented within Chapters 4 – 7, a considerable amount of effort was dedicated to developing high-throughput, low-cost continuous culturing equipment. Continuous culturing allows for greater control of both environmental and intracellular factors when attempting to characterize biological systems. **Appendix 1** describes methods and results related to the development of a chemostat system that is strongly based on the research and engineering of collaborating laboratories.

1.3 Scientific Contributions

This work offers novel strategies to quantitatively measure RNA dynamics in living cells. These strategies are significantly distinct from existing methods for studying RNA in that, for the first time, researchers are presented with a tool that is both quantitative and non-destructive. While the results of this dissertation are broadly applicable, the consequence of these results towards two particular areas of research are considered.

Contribution to synthetic biology. Developing a molecular toolbox capable of probing the state of engineered biological systems is a major challenge facing the field of synthetic biology [6]. The ability to interrogate cellular events with fine resolution will allow scientists and engineers to

better understand and describe the functions of molecular components both in isolation and when combined. Additionally, this understanding should lead to the development of refined standard descriptions for the function of molecular components.

This dissertation outlines a method for using fluorescence-activating aptamers to dynamically quantify kinetic processes within cells. Moreover, an entire chapter is dedicated to demonstrating a use case for fluorescence-activating aptamers in synthetic biology: the improved characterization of DNA regulatory elements. Compared to existing characterization methods in synthetic biology, this method is relatively inexpensive and high-thoroughput, and it could easily be applied by large-scale biofabrication facilities to characterize a plethora of DNA regulatory elements. Since this technique is quantitative, as opposed to semi-quantitative or qualitative, it fosters discussions regarding standardizing measurements of promoter activity.

Contribution to RNA expression analysis. Many techniques exist to measure RNA expression, each with their distinct advantages and limitations. Through this dissertation we show that fluorescence-activating aptamers have two attractive properties: (1) they can be quantitative and (2) they are non-destructive. No existing technology can claim both of these properties; thus fluorescence-activating aptamers have the potential to satisfy an unmet need for at least a subset of RNA expression analysis studies.

Chapter 2

BACKGROUND

2.1 Measuring RNA Dynamics

RNA has diverse roles as both a protein-coding and non-coding molecule. The strict spatial and temporal control of RNA production, degradation, and transport suggest the importance of careful maintenance of intracellular RNA levels; and misregulation of RNA has been attributed to many diseases [7]. Several methods have been developed to study RNA dynamics, and these techniques can be classified according to two major categories: biochemical investigation and cellular imaging.

Biochemical investigation generally proceeds by collecting and lysing cell samples, purifying total RNA within the lysate, and probing the purified content for specific molecules. Frequently used biochemical approaches include reverse-transcriptase, quantitative PCR, RNA-Seq [8], [9], and the Nanostring nCounter system [10], [11]. All of these techniques are quantitative, offer impressive sensitivity, and can be used, in theory, to detect single copies of a transcript within a sample. Furthermore, RNA-Seq and nCounter facilitate high-throughput studies of RNA expression since hundreds to thousands of samples can be measured simultaneously [12]–[14]. Unfortunately, sample destruction is a significant drawback of these RNA expression analysis techniques. The destructive nature of these techniques mandates that only static measurements can be made for a particular sample and its limits the temporal resolution with which RNA expression can be measured.

Cellular imaging techniques typically use fluorescence microscopy to detect the abundance and localization of RNA molecules within single cells. Examples of cellular imaging methods include MS2-like systems [15], fluorescence *in situ* hybridization [16], and, most recently, fluorescence-activating aptamers. Fluorescence *in situ* hybridization is a method that quantifies transcript abundance by exposing biochemically fixed cells to dye-labeled nucleotide sequences.

These sequences hybridize to complimentary mRNA sequences, and this technique has been demonstrated to be highly quantitative when measuring RNA abundance [16]–[18]. Additionally, fluorescence *in situ* hybridization excels in its ability to resolve transcript localization. Unfortunately, like biochemical approaches, this method destroys cell samples in order to quantify expression dynamics.

The MS2 system leverages the propensity of a bacteriophage protein, MS2, to target and bind to a short RNA sequence with high affinity and specificity [15], [19]. Fluorescent proteins such as GFP are tagged to MS2 and the RNA recognition motif is placed at the 3' end of an mRNA sequence. Since the discovery of MS2, additional proteins that bind specific RNA motifs have been applied in a similar manner for RNA analysis [20], [21]. In general these strategies are employed because they preserve cells during analysis; however, binding of multiple MS2 proteins to a single transcript obstructs the native function of mRNAs [22]. Favorably, MS2-like systems allow for much greater temporal resolution of RNA dynamics when compared to the previously discussed techniques. Like fluorescence *in situ* hybridization, this technique has proven useful in measuring RNA localization [23], [24]; however, unlike, fluorescence *in situ* hybridization, fluorescence from MS2-like systems only provides a semi-quantitative measure of abundance.

A final imaging strategy involves the use of RNA aptamers – short, non-coding RNA molecules - that invoke fluorescence upon binding to a small dye. Fluorescence-activating aptamers present an intriguing method for studying RNA dynamics because they allow researchers to genetically encode fluorescent RNA molecules; this strategy parallels the widely used strategy for measuring protein dynamics with fluorescent proteins. An advantage of fluorescence-activating aptamers over existing techniques is that imaging can proceed in a quantitative manner without necessitating cell destruction. This allows for quantitative measurements with potentially high temporal and spatial resolution; a combined advantage no other method can claim.

A number of fluorescence-activating aptamers have been developed, including malachite green aptamer [25], Mango [26], Spinach [2], [27], and DIR [28]. Of these aptamers, few studies

have reported using these aptamers within a cellular context. Zhang *et al*. attempted to encapsulate malachite green aptamer within a degenerate 5S rRNA sequence derived from *Vibrio protoelyticus*, and were the first group to report a potential stabilization strategy for fluorescence activating aptamers [29]. However, this study did little to characterize fluorescence and simply reported that they could detect fluorescence when excited at the expected wavelengths. Paige *et al*. reported the creation of a second fluorescence-activating aptamer, Spinach [27]. This aptamer uses DFHBI, a fluorophore that emits light in the green wavelengths, allowing it to be used in combination with the only other fluorescence-activating aptamer, malachite green aptamer. This report represented an improvement over malachite green aptamer in two major ways: (1) it demonstrated that multiple distinct aptamers could be raised to the cognate dye, DFHBI, to produce fluorescence at a variety of wavelengths from green to red; and (2) presented a second strategy for stabilizing a fluorescence-activating aptamer within cells through the use of tRNA scaffolds. The nucleotide sequence of Spinach aptamer was improved through rational mutagenesis by Strack *et al*. to produce a more thermostable aptamer; and this aptamer was then used to study RNA trafficking [2]. Spinach aptamer was used to provide semi-quantitative analysis of gene expression in *E. coli* by Pothoulokis et al. [30]. Finally, Spinach aptamer has been used to semi-quantitatively detect the concentration of intracellular metabolites and proteins through conformational changes upon binding a substrate [31], [32].

These works present a strong case for the use of fluorescence-activating aptamers to study many cellular processes in a quantitative manner; however, before these aptamer can be applied in such a manner their properties and method of action must be well-characterized. For studies relevant to synthetic biology, a particular useful starting point would involve the use of malachite green aptamer in *E. coli*, since it is a simple, model system.

2.2 Mathematical Modeling in Biology

Mathematical models are formal representations of physical systems. When provided with sufficient information about the internal state of a physical system, they describe its observable behavior. In biology, mathematical models have been used to study a wide range of phenomena, including metabolism [33], cell growth [34], and protein glycosylation [35].

In biology, three types of mathematical models are routinely used to improve the efficiency of both biological investigation and engineering efforts: qualitative, semi-quantitative, and quantitative models. A distinguishing characteristic of these different types of models is the level of accuracy and precision with which they purport system responses. Qualitative models highlight broad trends in system architecture that could produce an observed behavior; quantitative models report system response when provided specific inputs.

Qualitative and semi-quantitative models can be applied to direct investigative efforts towards experimental conditions that lead to interesting system responses. An advantage of these types of models are that they can be quite easy to construct; however, a disadvantage is that it is difficult to attribute changes in system behavior to specific biological phenomena.

Quantitative models that are well-characterized can be used to confidently predict system response without requiring physical experiments to be performed. In addition, the internal structure of quantitative models can be studied to interrogate the state of molecules and processes that are difficult to measure experimentally. An advantage of quantitative models are that they can provide a mechanistic understanding of biological systems [36]; however, a disadvantage is that it may be extremely challenging to identify appropriate values for all parameters of the model.

Chapter 3

QUESTIONS PROPOSED AND SOLUTIONS PRESENTED

This dissertation is motivated by a set of questions focused on improving the understanding of RNA expression and dynamics in *E. coli*. These questions are presented below along with a brief description of how they are addressed.

Can we quantitatively measure RNA expression in living cells? At present, researchers seeking to measure cellular RNA levels must choose between quantitative strategies that require sample destruction, and qualitative techniques that allow for repeated sampling. Developing a highly-quantitative method that avoids sample destruction would provide researchers with an obvious improvement over current offerings. Here, a fluorescence-based system that reports intracellular RNA abundance is proposed, and its feasibility is demonstrated. Fundamental properties of the reporter are assessed, and the quantitative nature of the RNA reporter is highlighted through its use to predict RNA concentration.

Is it possible to characterize and meaningfully describe the function of genetic elements? The functional properties of most genetic elements that are used in synthetic gene network designs are poorly understood. In this dissertation, we combine an RNA-based reporter with a coherent model of molecular interactions to characterize rates of RNA synthesis. Through controlled experimental conditions we relate observed rates of synthesis to specific DNA sequences that regulate RNA expression, allowing us to describe their function in a quantitative manner. This work provides evidence for the feasibility of meaningfully describing the function of genetic elements; and hopefully encourages further research within this domain.

Chapter 4

MALACHITE GREEN APTAMER EXPRESSION IN *E. COLI*

4.1 Motivation

Fluorescence-activating aptamers provide an unparalleled way to measure RNA expression and transcription dynamics. They are genetically encoded RNA sequences that allow researchers to observe intracellular events such as synthesis, degradation, and localization. Fluorescence-activating aptamers hold promise, in part, because they represent a highly-quantitative way to measure RNA expression in living cells with impressive temporal resolution; a feat that no existing methods can claim to accomplish. However, as a relatively new technology, the suitability of these aptamers for particular applications is still being investigated.

Currently, no fluorescence-activating aptamer has been extensively characterized in *E. coli*. We eventually hope to use these aptamers to quantify RNA abundance from fluorescence intensity. Before we can accomplish that task, we must ensure that the fluorescent properties of the aptamer in *E. coli*, and the effects of exposing cells to malachite green are well understood.

The work presented in this chapter is motivated by the need to better understand how malachite green aptamer and its cognate dye function in *E. coli*. First, we identify a feasible strategy to express the aptamer in cells. Then, we explore how this fluorescence-activating aptamer affects cell growth. We shows that cells exposed to malachite green produce a strong fluorescent signal that is specific and sensitive to aptamer expression. From the results of this chapter, we are confident that malachite green aptamer and malachite green, when combined in *E. coli*, provide a useful tool for quantifying RNA expression.

4.2 Background

4.2.1 Malachite green and malachite green aptamer

Malachite green is a commercially available triphenylmethane dye that strongly absorbs light energy at 618 nm. Due to its molecular structure, energy absorbed from light is normally dissipated as heat through rational movement of its phenyl groups. Restricting the dye's molecular motion forces it to fluoresce as it releases the absorbed light energy at a longer wavelength [37].

Malachite green has been found to readily traverse bacterial cell membranes; however, the molecule must be modified with an ester linkage to enable it to cross eukaryotic membranes [38]. The fluorophore is a known antifungal agent [39]; however, it does not exhibit antibacterial properties.

Malachite green aptamer is one of the first fluorescence-activating aptamers to be described in literature [25], and is extensively used throughout this dissertation. The aptamer was identified through *in vitro* selection for high binding affinity and strong fluorescence enhancement. It is a 38-nucleotide, non-coding RNA sequence that strongly bindings malachite green to greatly enhance the fluorophore's brightness.

When in complex, malachite green and malachite green aptamer are maximally excited at approximately 625 nm and fluorescence emission is strongest near 655 nm. Quantum yield of the dye increases from 7.9e-5 to 0.187, a 2,360-fold change; and the *in vitro* dissociation rate constant was found to be 117 nM [25].

In contrast to the cognate dyes for other fluorescence-activating aptamers, such as Mango and Spinach, malachite green is a widely-available and inexpensive fluorophore.

4.2.2. RNA stability

RNA degradation plays an important role in controlling gene expression in cells [40], [41]. In prokaryotes, most RNA molecules have relatively short lifespans with half-life measurements on the order of 2 – 8 minutes [42], [43]. It is known that certain environmental conditions can alter the rate of RNA decay [44]. Furthermore, mRNA degradation is affected by nucleases [45], secondary structure [46], translation influences [47], and transcription [48].

Studies seeking to increase RNA half-life have utilized a variety of stabilization strategies. A few notable strategies include ribosome loading [49], flanking sequences with 5′ and 3′ stem loops [4], [50], and capping sequences with paired termini [51]. Fluorescence-activating aptamers, in particular, have been stabilized for intracellular expression in *E. coli* by appending the aptamers to 5S rRNA [29] or 6S rRNA [26], or by incorporating them into tRNA scaffolds [27], [52].

4.3 Results

4.3.1 Stable expression and intracellular fluorescence detection

To begin, we investigated ways to achieve stable expression of malachite green aptamer in *E. coli*. We expressed the 38-nucleotide aptamer under the control of a constitutive promoter, BBa_J23101 (parts.igem.org), on a p15A plasmid as shown in **Figure 4.1**. Cells expressing the aptamer were exposed to 5 μM malachite green; however, we did not detect fluorescence.

Figure 4.1. SBOL diagram for an initial attempt to express malachite green aptamer in *E. coli*.

We hypothesized that poor intracellular stability of the aptamer was the primary reason that we did not observe fluorescence in *E. coli*. Previous research has demonstrated that the brightness of fluoresce-activating aptamers inside of cells can be significantly reduced due to rapid turnover [29] and improper folding [2]. To enhance fluorescence, we expressed malachite green aptamer in conjunction with known RNA stability elements towards the goal of extending it half-life and reducing its potential to misfold. We constructed several plasmids that either integrated malachite green aptamer into an RNA scaffold of appended the aptamer onto an mRNA sequence. All scaffold sequences were identified in literature, and the predicted secondary structure of malachite green and the scaffolds are shown in **Figure 4.2**.

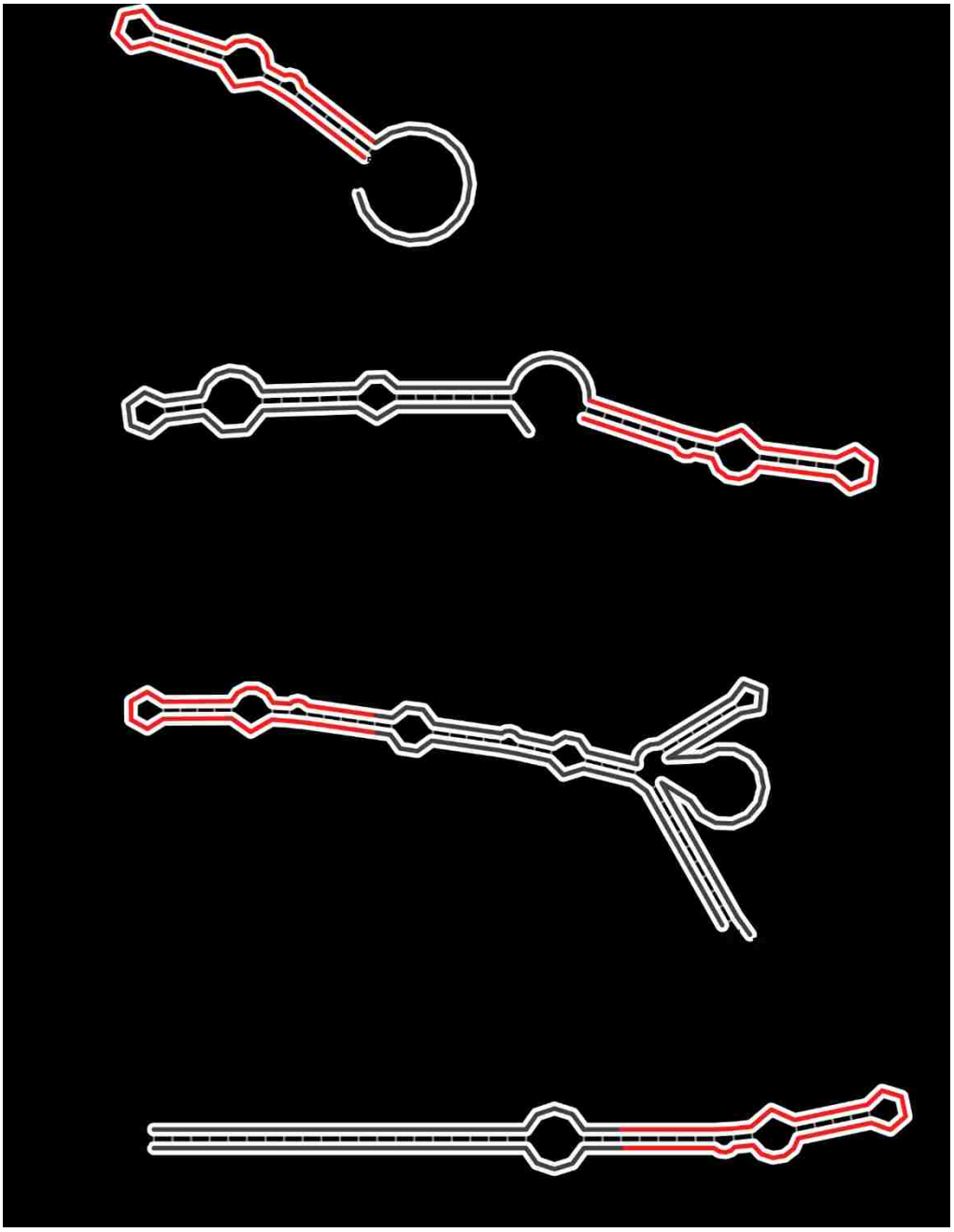


Figure 4.2. The predicted secondary structures for subsequent attempts to stably express malachite green in *E. coli*. The 38-nucleotide aptamer was flanked by previously described RNA stability elements in an attempt to enhance the brightness of the reporter. Stabilization strategies included appending the aptamer to a protein coding sequence (**A)**, placing the aptamer behind a synthetic secondary structure (**B**), placing it within a degenerate ribosomal RNA sequence (**C**), and incorporating it into a paried termini stem (**D**). 38-nt malachite green aptamer highlighted in red.

We screened the various plasmid constructs to identify designs that produced measureable levels of fluorescence when expressed in *E. coli*. **Figure 4.3** shows that only one aptamer sequence produced strong levels of fluorescence. The design of this aptamer was based on an earlier report [29] that incorporated malachite green aptamer into a degenerate 5S rRNA scaffold. We modified the reported sequence to remove restriction endonuclease sites, and termed the resulting 126-nucleotide aptamer **MGA5S**.

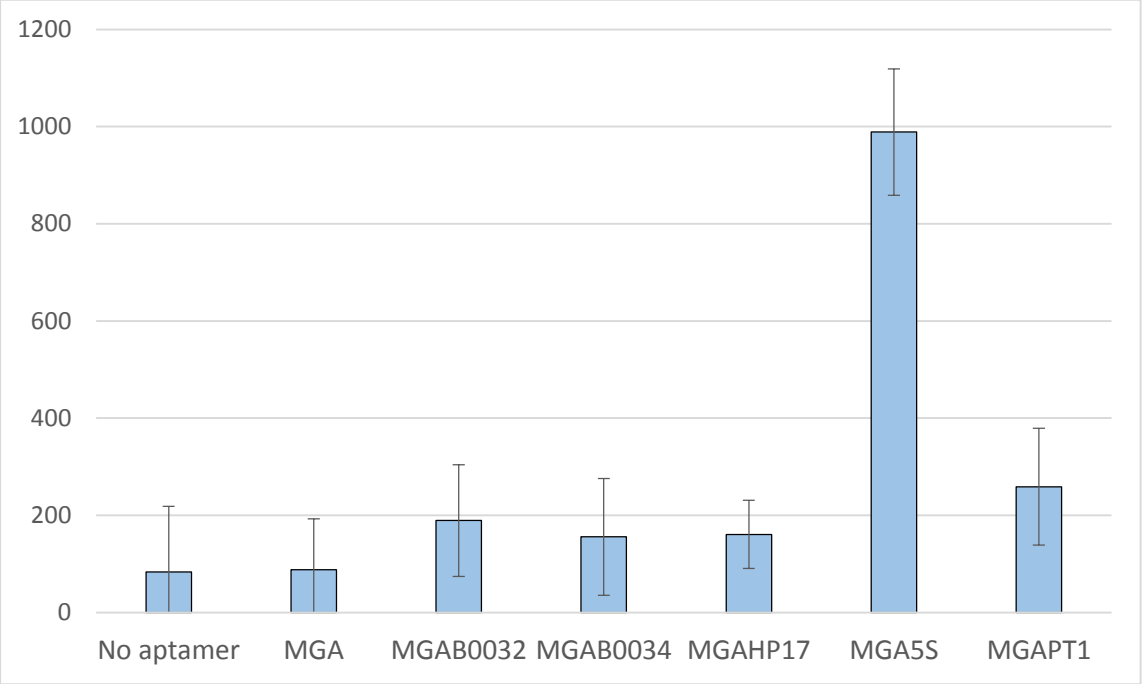


Figure 4.3. Fluorescence of malachite green aptamer expressed in conjunction with several RNA stability elements.

4.3.2 Binding specificity

It was important that fluorescence due to non-specific binding of the dye within *E. coli* was minimal. We examined binding specificity by removing the 38-nt malachite green aptamer sequence from MGA5S to create a no-aptamer control plasmid as depicted in **Figure 4.5**. Exposure

to malachite green produced a strong fluorescent signal from cells expressing MGA5S; however negligible fluorescence was detected from cells that expressed the no-aptamer control sequence. These results were observed and confirmed through fluorescence microscopy, as shown in **Figure 4.6**, as well as by population measurements made with a fluorometer, as displayed in **Figure 4.3**. These results suggest that the dye is specific to the 38-nucleotide aptamer binding sequence.

Figure 4.5. SBOL diagram of the no-aptamer control plasmid. The core, 38-nucleotide aptamer sequence was removed from MGA5S.

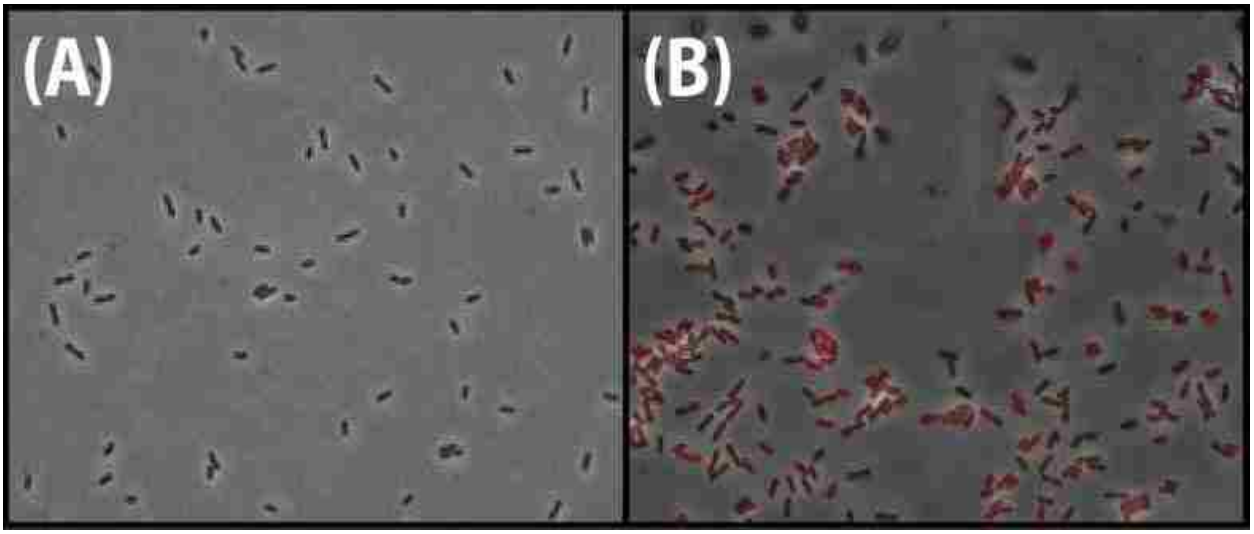


Figure 4.6. Fluorescence was detected in cells expressing MGA5S, but not in cells expressing the no-aptamer control. **(A,B)** Brightfield images of cells expressing no-aptamer control and MGA5S, respectively, are overlay with phase contrast images. For phase constant images, fluorescence intensity is false colored red.

4.3.3 Malachite green permeability

To assess cell permeability, we repeatedly pulsed 1 μM malachite green into exponentially growing cell cultures while continuously measuring fluorescence. **Figure 4.7** shows that following each pulse of malachite green, we observed a step response in fluorescence output from cell cultures. The settling time until fluorescence equilibrates was found to be approximately 2 minutes. These results suggest that malachite green quickly crosses the cell membrane in *E. coli* and rapidly forms a complex with intracellular MGA5S.

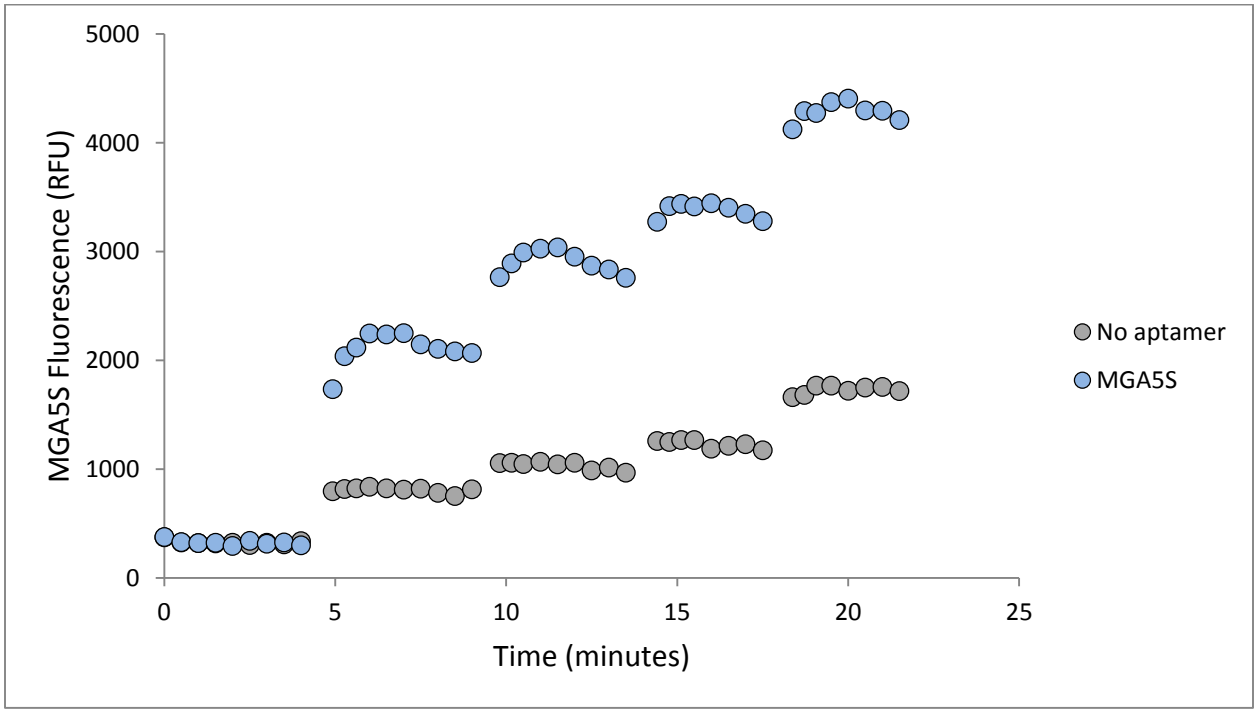


Figure 4.7. Fluorescence response of MGA5S following repeated exposure to malachite green. Cells expressing either MGA5S or the no-aptamer control were exposed to 0 μM, 1 μM, 2 μM, 3 μM, and 5 μM malachite green.

4.3.4 Malachite green toxicity

Concerns regarding cytotoxicity were investigated by culturing cells at increasing concentrations of malachite green and profiling cell growth. We measured cell growth in batch culture by recording absorbance at 600nm, and **Figure 4.9** shows the results of this experiment. Additionally, to gauge any phototoxic effects of malachite green in *E. coli*, we excited growing cultures at 620nm every 8 minutes. We calculated the maximum growth rate of all cultures. As shown in **Figure 4.10**, we found no statistically significant difference in the maximum growth rate among cultures exposed to less than 10 μM malachite green. These findings left us confident that malachite green concentrations below 10 μM would be suitable for RNA analysis.

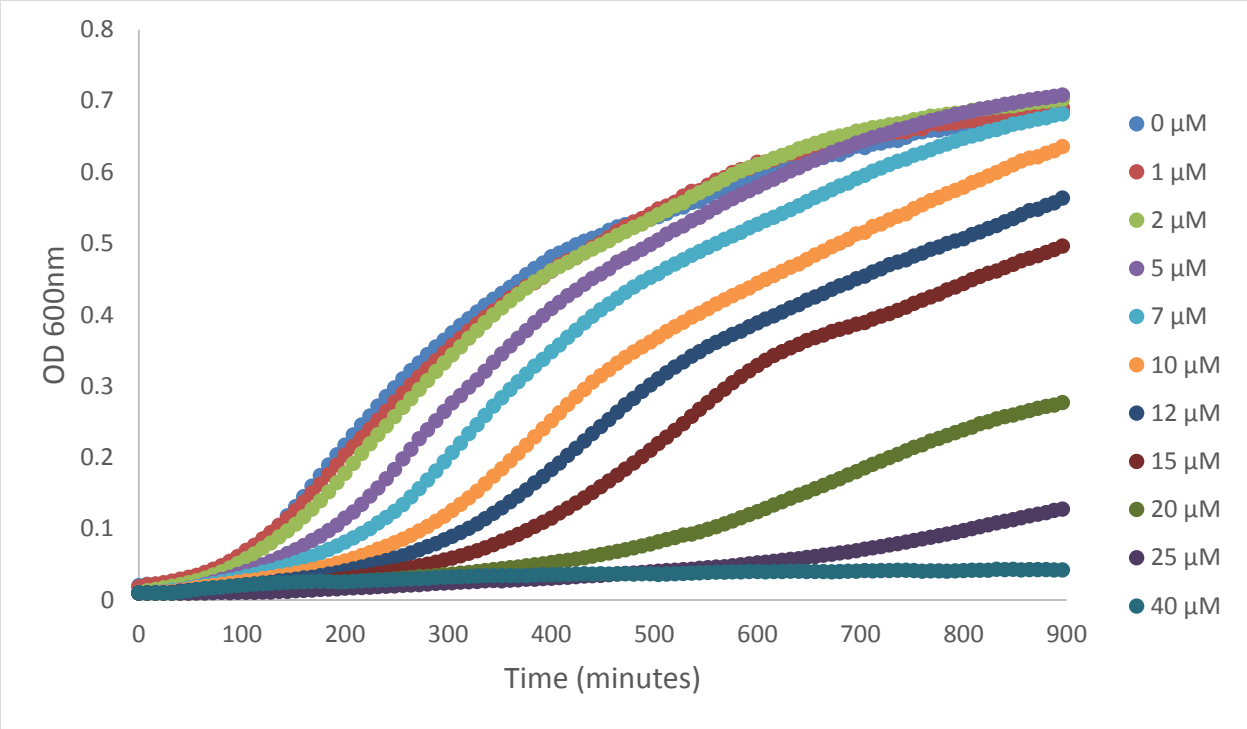


Figure 4.9. *E. coli* growth in M9CA containing increasing concentrations of malachite green. Absorbance measurements of cell cultures were subtracted from M9CA supplemented with increasing concentrations of malachite green.

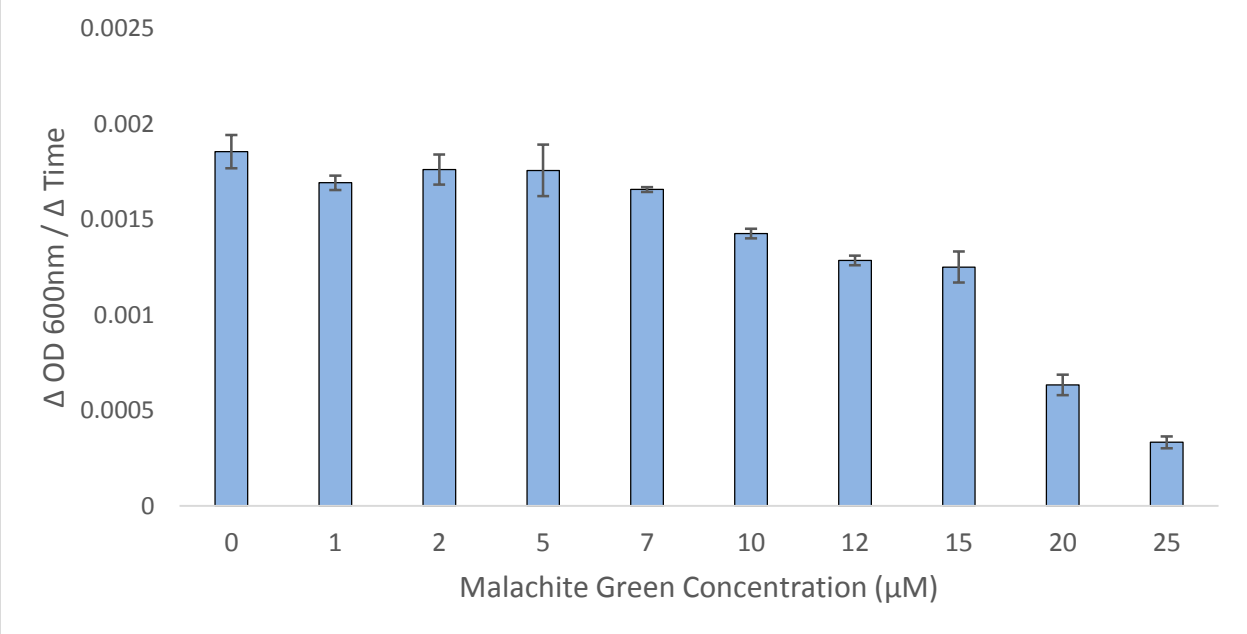


Figure 4.10. Maximum growth rates of *E. coli* in M9CA containing increasing concentrations of malachite green. Maximum growth is calculated from the slope of absorbance curves and is reported as the largest rate of increase in absorbance per unit time.

It should be noted that malachite green strongly absorbs light at 600nm. **Figure 4.11** shows that malachite green concentrations contributes strongly to optical density measurements. These data show that high concentrations of malachite green confound cell density measurements, and likely lead to an underestimation of the maximum growth rate.

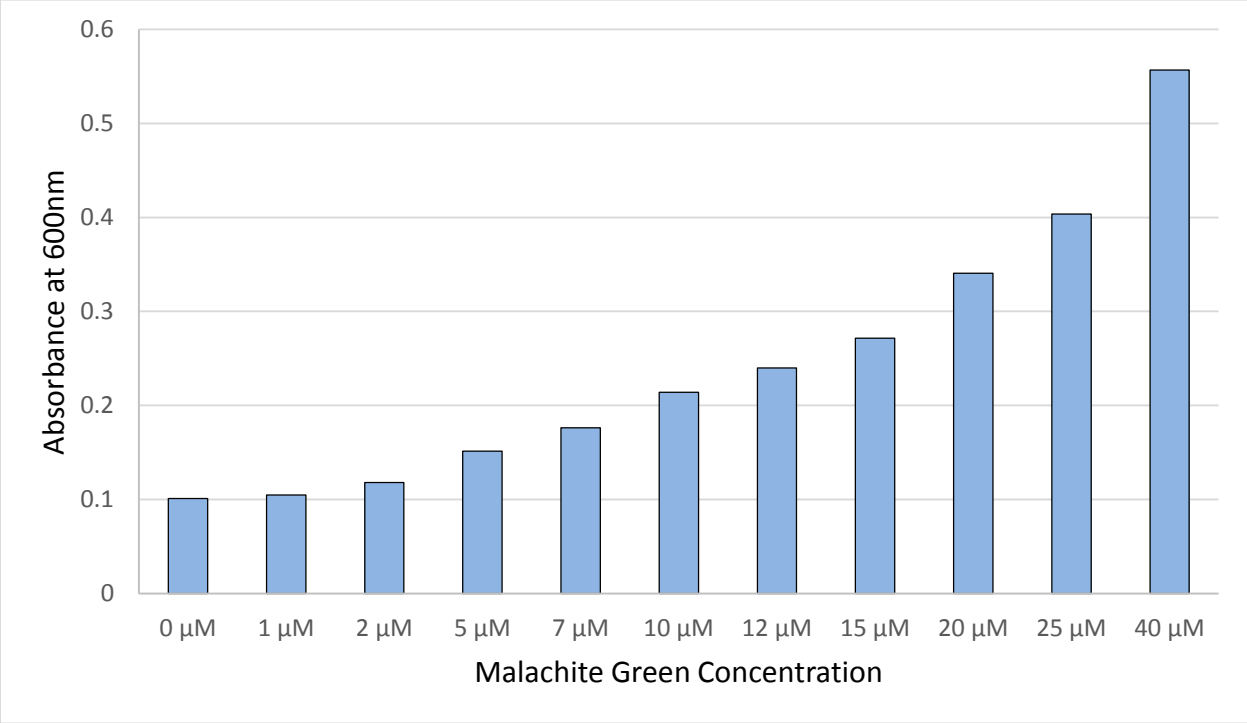


Figure 4.11. Absorbance of M9CA containing increasing concentrations of malachite green.

4.3.5 Fluorescent signal sensitivity

To investigate how sensitive the fluorescent signal is to changes in aptamer concentration, we varied intracellular MGA5S abundance and measured how well fluorescence output was able to resolve these distinct concentrations of aptamer. A plasmid was constructed to place MGA5S under the control of an IPTG-inducible promoter, BBa_R0011 (parts.igem.org) as shown in **Figure 4.13**.

Figure 4.13. SBOL diagram of plasmid used to induce MGA5S expression with IPTG. BBa_R0011 is the IPTG-inducible promoter, placi.

Cells expressing the aptamer were cultured in the presence of malachite green and varying amounts of inducer. **Figure 4.14** shows that average MGA5S fluorescence intensity increased with increasing IPTG concentrations in a sigmoidal manner. This induction response is very similar to that observed when using fluorescent proteins. Notably, the concentration range that produces a transient signal – 10 μM to 1 mM – and the saturating IPTG concentration were similar for induction of fluorescent proteins and MGA5S. We normalized fluorescence signal to that of uninduced cells and found that we could discern a 30.5-fold change in MGA5S abundance.

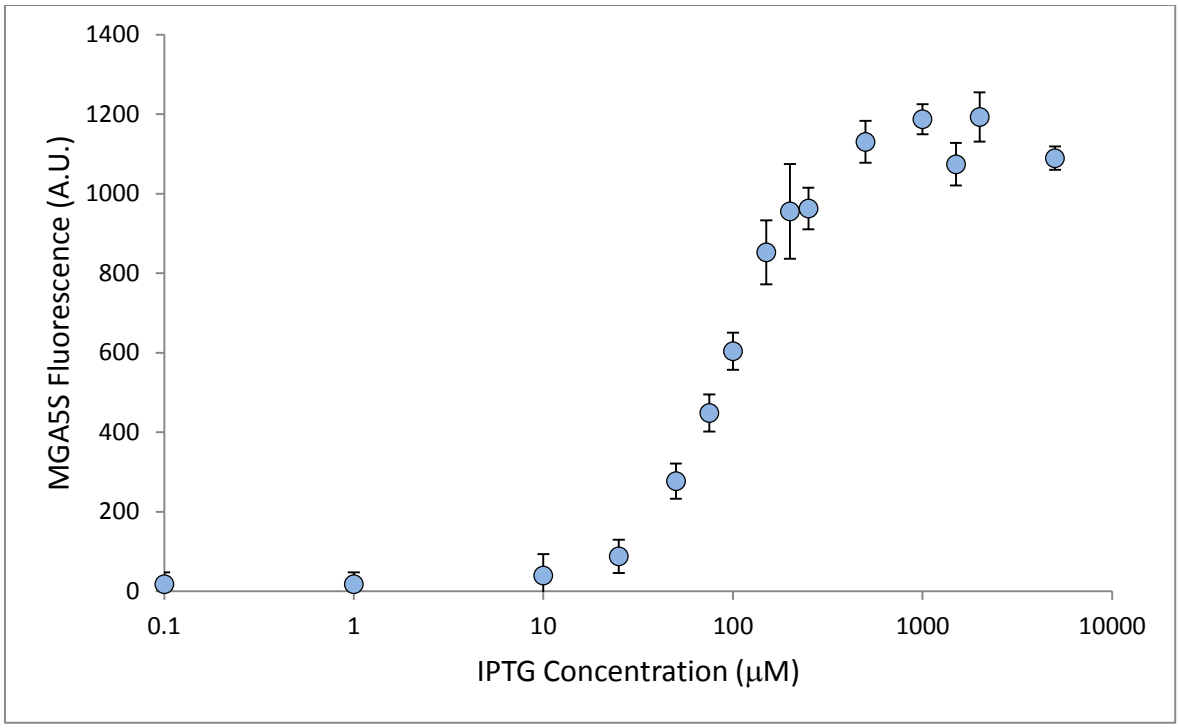


Figure 4.14. Induction of the pLac promoter with IPTG.

4.3.6 Binding stoichiometry

Binding stoichiometry of the core 38-nucleotide aptamer with malachite green was previously found to be 1:1; however, since we modified the core aptamer sequence by placing it within a 5S rRNA scaffold, it was important to ensure that stoichiometry remained unchanged. **Figure 4.15** shows the results of a continuous variation reaction between MGA5S and malachite green. In this experiment, the total concentration of malachite green and MGA5S were held constant at 10 μM; however, the mole fraction of each molecule varied. We found that fluorescence was at its highest when MGA5S and malachite green were in equimolar ratios. These data are indicative that the 1:1 binding stoichiometry between the core aptamer and malachite green is preserved in spite of the 5S rRNA scaffold.

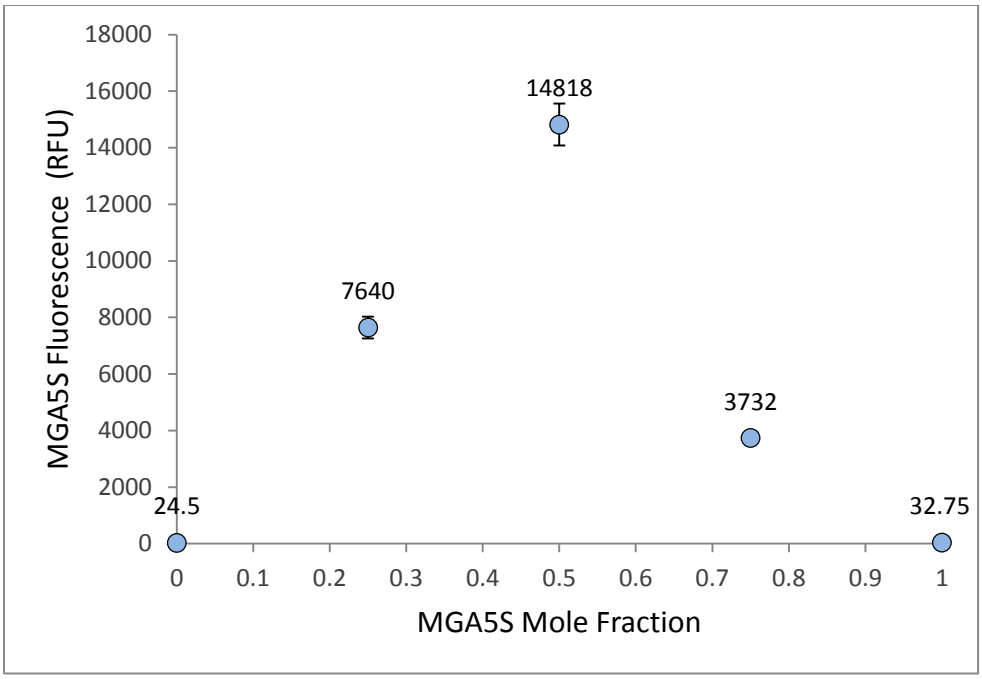


Figure 4.15. Continuous variation of MGA5S and malachite green in PBS to investigate binding affinity.

4.3.7. Dynamic measurements of RNA expression

To assess how well MGA5S reports changes in RNA concentration over time, we cultured cells expressing the aptamer while periodically measuring fluorescence. Cells expressing the no-aptamer control were also cultured and used to account for auto-fluorescence of cells in media. **Figure 4.16** and **4.17** show how expression changes over time. We observed an initial linear increase in fluorescence as cells grow; however, fluorescence levels off at high cell densities.

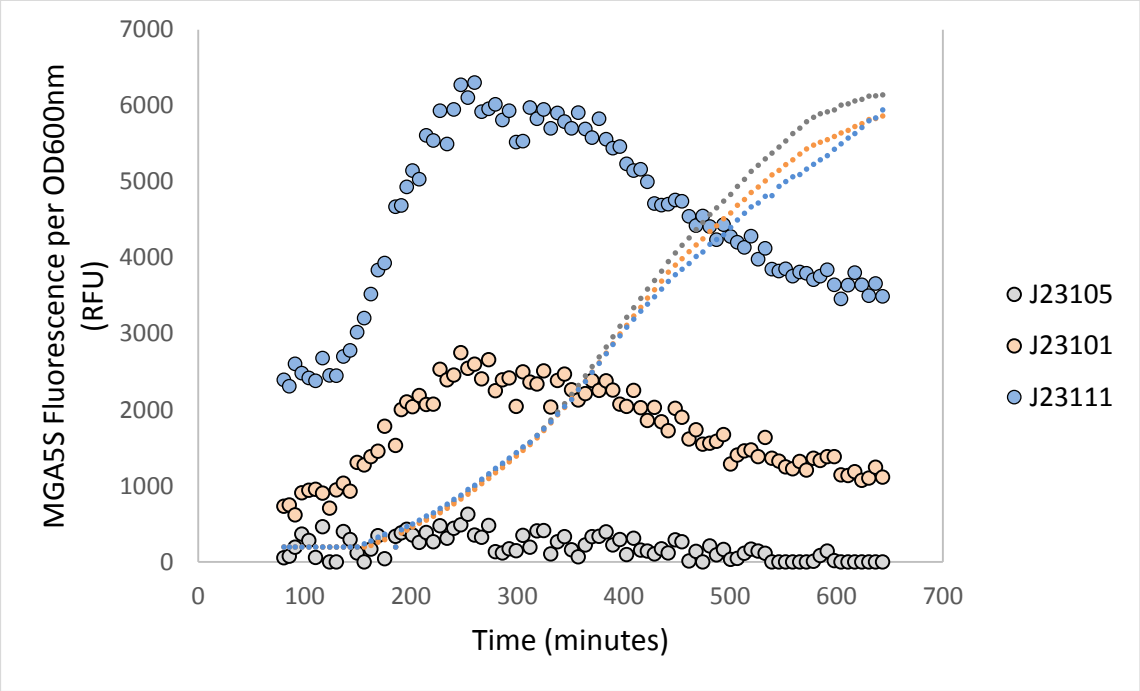


Figure 4.16. Fluorescence intensity normalized by cell density for cells growing in batch culture for 11 hours. Fluorescence measurements are overlay on absorbance measurements. As the cell population approaches stationary phase, fluorescence per absorbance decreases.

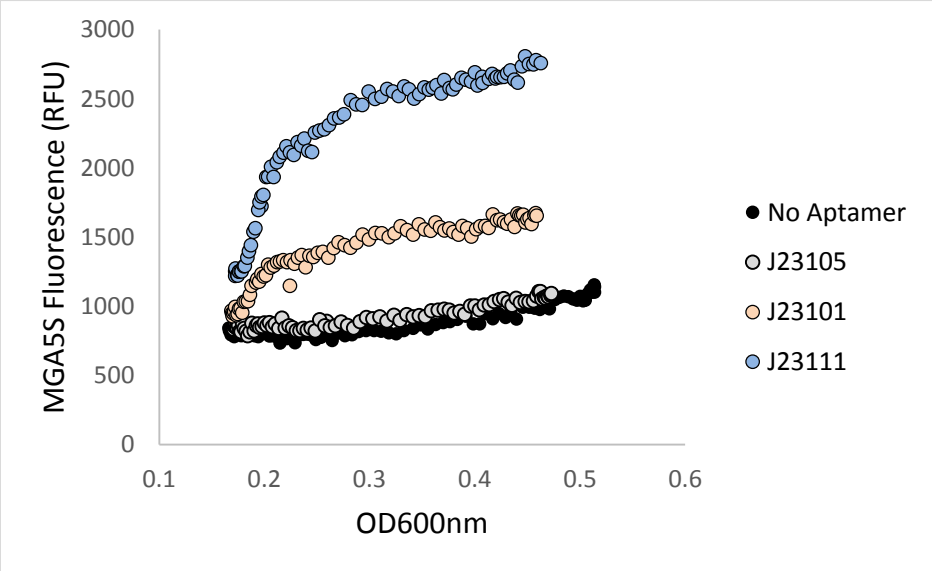


Figure 4.17. Fluorescence intensity of cultures are compared with absorbance. Initially, fluorescence increases linearly. However, fluorescence levels off at high cell densities.

Additionally, we studied dynamics by observing promoter response to an inducer. MGA5S was placed under the control of the IPTG-inducible promoter, BBa_R0011 (parts.igem.org), in *E. coli* cells expressing the Lac repressor protein. **Figures 4.18** and **4.19** show that, following induction, MGA5S fluorescence can be observed over background within 6 minutes. Furthermore, through this experiment, we observed a near constant increase in fluorescence per cell for 60 minutes following induction.

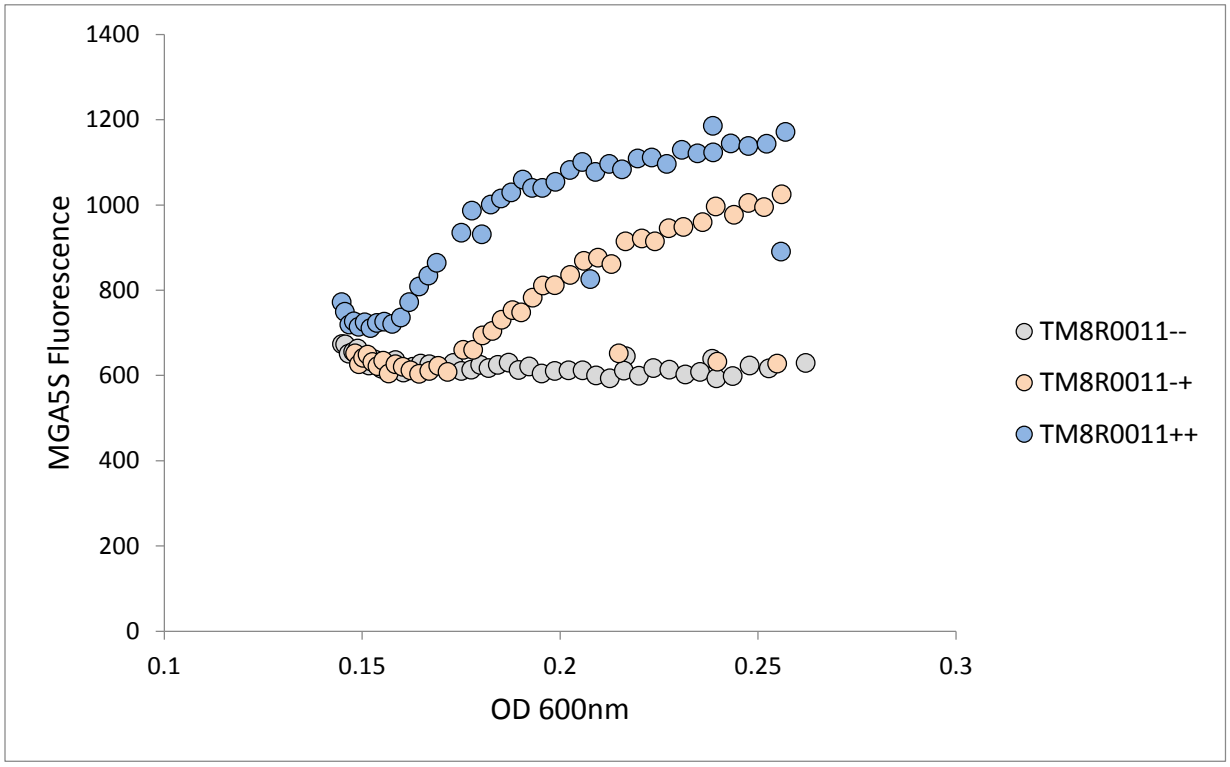


Figure 4.18. IPTG-induction of pLac leads to an increase in MGA5S expression.

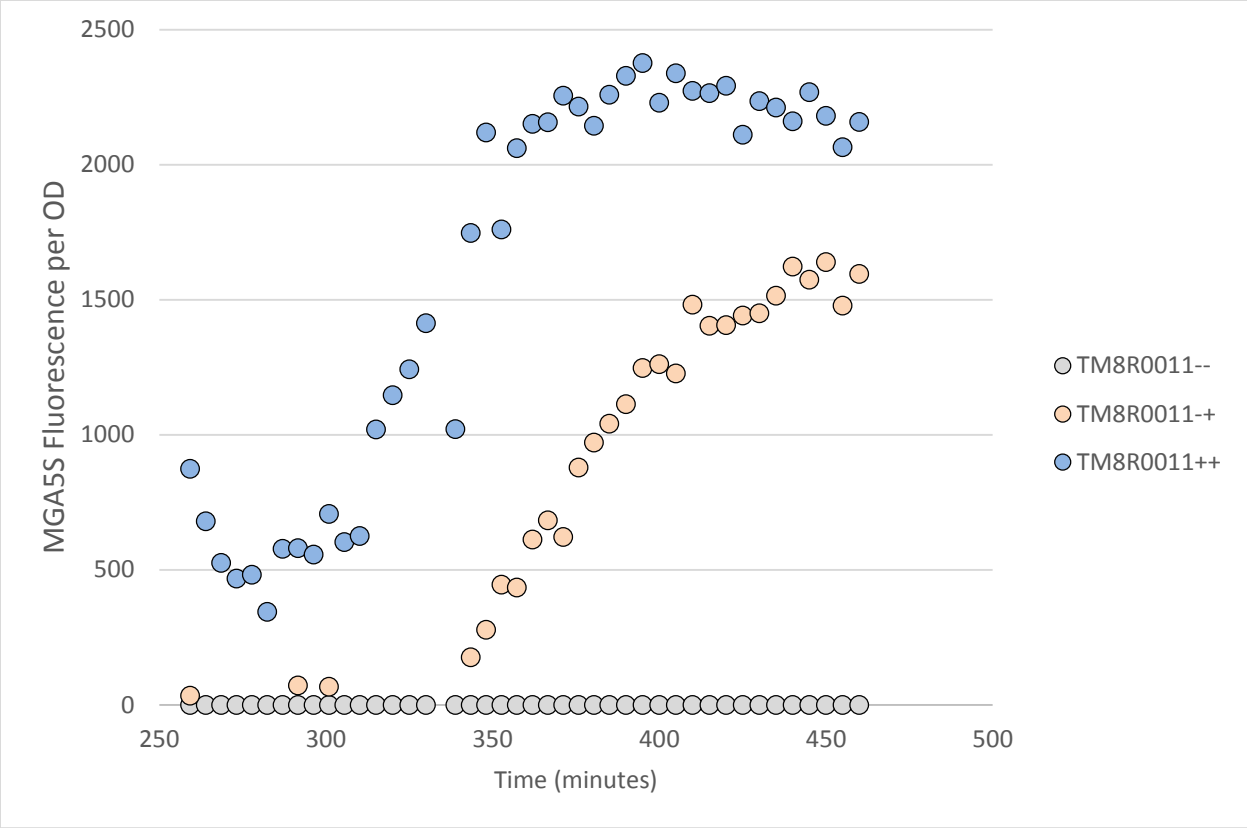


Figure 4.19. IPTG-induction of pLac leads to an increase in MGA5S expression. Cells were induced with IPTG at t=339 mins.

4.4 Discussion

Through these experiments, we demonstrate the feasibility of using the malachite green aptamer as a live-cell, RNA-based reporter for transcription. This is a meaningful achievement since the majority of current strategies that enable analysis of RNA expression require the destruction of cell samples. The work presented in this chapter thoroughly examines the strengths and limitations of malachite green aptamer as a fluorescence-based tool to measure RNA expression.

We found that we could only detect fluorescence due to aptamer-dye binding when we placed the 38-nucleotide aptamer within a scaffold that imparted intracellular stability. We

believe the scaffold provides stability in two ways. First, as other papers have demonstrated, the percent of properly folded aptamer at 37°C – the temperature at which cells are cultured - might be drastically lower than the percent of properly folder aptamer at room temperature – the temperature at which the aptamer was originally identified [25]. Second, if RNA turnover is too rapid or too few molecules exist within the cell, observing fluorescence becomes very difficult. Prokaryotes are known to experience rapid RNA degradation relative to eukaryotes [53]. We tried multiple stabilization strategies that should have ensured proper folding of malachite green aptamer; however, we only observed fluorescence when we modified the aptamer to mimic rRNA that is known to persist within E. coli. Malachite green aptamer has been found to have a dissociation constant of approximately 117 nM [25]. This means that when roughly 117 molecules of the aptamer are present in the cell, half are in the bound form. To observe a fluorescent signal, we would need to achieve relatively large copy numbers within the cell. Taken together, this leads us to believe that RNA turnover, as opposed to stable folding, is actually the more pertinent issue with regards measuring malachite green aptamer expression. However, this implies that MGA5S may not be very useful for studies that require fast degradation kinetics.

Our characterization of malachite green aptamer continued by assessing specificity, sensitivity, and response time. We found that the presence of malachite green aptamer in cells exclusively contributed to a strong fluorescence signal. We did notice that with increasing dye concentrations, the background signal increased; however, background was found to be negligible in comparison to fluorescence due to aptamer for all concentrations up to 40 μM, which is the limit of what we tested. Moreover, we found that fluorescence could discriminate varying aptamer concentrations quite well. Finally, we found that fluorescence signal responded very quickly to exposure to malachite green. This is encouraging because it shows that RNA concentration can be confidently quantified within 2 minutes of sampling. For comparison, another dye, DFHBI, was found to take about 20 minutes to reach equilibrium in when exposed to mammalian cells [2]. To our knowledge, this has been the only other characterization of a fluorescence-activating aptamer's response time to dye exposure. It would be useful to observe if the response time for DFHBI is as slow in E. coli as it was in mammalian cells.

Concerns have been raised about malachite green toxicity have been raised in previous studies. Specifically, studies have shown that the free radicals produced upon exciting the molecule can damage RNA and abolish fluorescence [54]–[56]. However, it can be argued that these studies may not be relevant to our experiment setup. For instance, phototoxicity was a concern because the researchers where exciting malachite green with a powerful laser for extended periods of time; they report that the rate of reactive oxygen species production is on the same order as that of GFP, which is often used in fluorescence-based studied. Additionally, malachite green is a known antifungal agent and may also be toxic in mammalian culture; however, we found that it was not a powerful inhibitor of *E. coli* growth. Ultimately, we found that the use of malachite green at concentrations below 10 μM were acceptable for *E. coli* culture.

Our findings allow experimentalists to make more informed decisions regarding whether the proposed expression strategy is appropriate for what they intend to measure. Concurrent to this work, a second fluorescence-activating aptamer, Spinach, has been described by another research group [27]. Fluorescence has been detected by expressing this aptamer in *E. coli* as a proof of concept; however, the fluorescent properties of this system have yet to be characterized in *E. coli* and its suitability for measuring RNA expression has not be convincingly presented. For example, to enhance the brightness of the fluorescence signal, the core aptamer sequence is placed inside a tRNA scaffold and subsequently appended to native 5S rRNA; however, the consequences of this expression strategy toward cell growth are never reported.

4.5 Methods

4.5.1 Strains and media

Turbo Competent *E. coli* (NEB) and LB media (Sigma) were used for plasmid preparation. All studies were performed using either the *E. coli* strain MG1655 (Coli Genetic Stock Center

accession number 7740) or MG1655Z1[r]. Cells used in experimental studies were cultured in either M9 Minimal media (5X Salts (Teknova; 3.0 g/L KH_2PO_4, 6.0 g/L Na_2HPO_4, 5.0 g/L NaCl, 1.0 g/L NH_4Cl), 1 mM $MgSO_4$ (Fisher), 0.1 mM $CaCl_2$ (JT Baker), 2% glucose (Fisher), and 0.5 μg/ml thiamine) or M9CA media (5X Salts; 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1% glucose, 0.5 μg/ml thiamine, and 0.1% casamino acids). All media was supplemented with antibiotics, 100 μg/ml Carbenicllin (Teknova) and 50 μg/ml Kanamycin (Teknova), and inducer, Isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma), when appropriate.

4.5.2 Plasmid construction

The malachite green aptamer sequence was found from literature [25]; promoter sequences used in the library were taken from the Registry of Standard Biology Parts (parts.igem.org), the BioFab (biofab.synberc.org), or literature [57]; and RNA stability elements were identified from literature [48], [51]. The 126-nt MGA5S sequence used in all experiments is 5′-TGCCTGGCGA CCATAGCGAT TTGGAGGATC CCGACTGGCG AGAGCCAGGT AACGAATGGA TCCAACGAAT TAGCGCCGAT GGTAGTGTGG GGTTTCCCCA TGTGAGAGTA GGACATCGCC AGGCAT-3′.

Physical DNA sequences was either obtained through the Registry of Standard Biological Parts or created through oligonucleotide synthesis (Integrated DNA Technologies). For sequences greater than 60 nucleotides, DNAWorks[r] was used to guide overlap PCR synthesis[r] and create longer linear double-stranded DNA fragments. Two plasmids were used: pSB3K3, a medium copy number plasmid with a p15A pMR101-derived origin of replication and kanamycin resistance gene; and pGA1A3, a high copy number plasmid with ampicillin resistance and a pMB1 origin of replication. Both plasmids are available through the Registry of Standard Biology Parts (parts.igem.org).

Plasmids were created by one-step isothermal DNA assembly of linearized double-standed DNA. First, DNA was linearized by PCR amplification using Phusion DNA polymerase

(NEB). Next, a one-step isothermal DNA assembly mixture that includes Phusion DNA Polymerase, T4 Ligase (NEB), and T5 Exonucleases (NEB) was used to ligate double-stranded DNA fragments together. The resulting plasmid DNA was transformed into chemically competent Turbo Compotent cells and plated on selective LB agar.

Plasmids from individual transformant colonies were prepared using QIAprep Spin Miniprep Kits (Qiagen). DNA sequences were confirmed (Genewiz) using the following primers: VF2, 5′- TGCCACCTGA CGTCTAAGAA – 3′, and VR, 5′- ATTACCGCCT TTGAGTGAGC -3′. Plasmid DNA sequences were transformed into *E. coli* MG1655 or MG1655Z1 for use in subsequent studies.

4.5.3. Cell growth in batch cultures

Cells were grown according to the following protocol unless otherwise noted. Cells were inoculated into 4 ml M9 media from -80°C freezer stocks and grown overnight at 37°C and 250 r.p.m. shaking. Overnight cultures were diluted into 1 ml of warm M9 media to an OD_{600nm} of 0.01. Every 4 hours cells were diluted 1:10 into 1 ml of warm M9 media to maintain a cell density below 10^7 cells/ml. Serial dilutions were made for 12 hours, allowing cells to experience exponential growth for roughly 9 generations.

4.5.4. Fluorescence measurement and analysis

Fluorescence of cell cultures was measured using a Tecan M200 Pro (Tecan) multimodal plate reader. Fluorescence was measured at an excitation wavelength of 620 nm and an emission wavelength of 655 nm. Gain for the instrument was set to 170 and fluorescence intensity was integrated from 40 flashes.

Cell cultures were sampled and diluted 1:100 into 1 ml of warm M9 media. 975 μl of diluted culture was added to 25 μl of a 200 μM malachite green solution in a microfuge tube. The mixture was briefly vortexed and 200 μl was added in replicate of four to a black, clear-bottom 96-well plate (Costar). Cultures were incubated within the Tecan M200 Pro at 37°C for 16 hours with 120 seconds of linear shaking at 2.5 mm every 6 minutes. Additionally, for every 6 minute interval, absorbance at 600 nm and fluorescence at 620 nm excitation and 655 nm emission were recorded for each well.

4.5.5 Fluorescence microscopy

Fluorescence microscopy images were acquired using an Eclipse Ti inverted microscope (Nikon Instruments). Brightfield and phase contrast images were collected with the assistance of NIS-Elements Imaging Software (Nikon). The setup was outfitted with a Cy5 filter cube (#41019; Chroma Technology Corporation) with the following specifications: HQ615/45x excitation, Q646lp bandpass, and HQ679/60nm emission. The entire microscope was contained inside of a temperature controlled chamber set to 37°C.

Prior to measurement, cells were suspended in warm M9 media supplemented with 5 μM malachite green. 5 μl of suspended culture was placed on 0.5% M9 agar containing 5 μM malachite green. A glass microscope slide was placed over the agar pad and sealed airtight with a 1:1:1 mixture of Vaseline, lanolin, and paraffin.

4.5.6. In vitro transcription

A linear, double-stranded DNA template suitable for *in vitro* transcription was produced by inserting the T7 promoter sequence upstream of MGA5S. The template DNA sequence was created by PCR amplification using Phusion DNA Polymerase. The forward primer for the PCR

reaction was 5′ – GAAATTAATA CGACTCACTA TAGGGAGATG CCTGGCGACC ATAGCGAT – 3′ and the reverse primer was 5′ – ATGCCTGGCG ATGTCCT – 3′.

The conversion of MGA5S DNA to RNA was performed using a HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to the manufacturers protocol for transcribing short RNA templates. Briefly, 1 μg of DNA was mixed with equal parts dNTPs and allowed to incubate at 37°C for 8 hours.

Following the *in vitro* transcription reaction, RNA was purified and template DNA was removed using a Quick RNA Miniprep Kit with on-column DNase 1 treatment (Zymo Research).

4.5.7 Malachite green permeability

E. coli growing in continuous culture were sampled, and 200 μl aliquots were placed in a 96-well tissue culture plate (Costar) in replicates of 12. Fluorescence was recorded every 30 seconds for 5 minutes. Fluorescence was recorded at an excitation wavelength of 620nm and emission wavelength of 655nm. Each addition of malachite green was performed using a multichannel pipette, where malachite green was rapidly added to replicate cultures using a volume that did not exceed 2 μl (1% of total volume). Using a 50 μl multichannel pipette, cultures were briefly resuspended and fluorescence measurement was allowed to continue.

4.5.8 Binding stoichiometry

Pure MGA5S was obtained through i*n vitro* transcription. MGA5S and malachite green were mixed in 1X phosphate-buffer saline (Sigma) in triplicate according. The total concentration of malachite green and MGA5S was held constant at 10 μM; however, the molar ratios of MGA5S to malachite green were 0:1, 0.25:0.75, 0.5:0.5 , 0.75:0.25, and 1:0, respectively.

Fluorescence was recorded at an excitation wavelength of 620 nm and an emission wavelength of 655nm.

Chapter 5

QUANTITATIVE ANALYSIS OF RNA EXPRESSION

5.1 Motivation

Can we make quantitative measurements of RNA expression in living cells? MGA5S and comparable fluorescence-activating aptamers are poised to answer that question affirmatively. To address this challenge, fluorescence intensity from MGA5S – the output, is used to accurately estimate aptamer abundance – the input.

A simple relation that scales RNA concentration and fluorescence intensity in a linear manner may be appropriate for certain situations. However, a model that considers the molecular interactions of MGA5S should prove more predictive.

The work presented in this chapter is motivated by the goal of using fluorescence output from MGA5S to quantitatively measure intracellular aptamer concentration. First, we validate that fluorescence and concentration are correlated. Next, we develop a systems model of molecular interactions. We identify unknown parameters of the model through experiments, and we explore ways to confidently estimate parameters that cannot be experimentally interrogated. Finally, we combine our refined model and experimental data to predict intracellular RNA concentrations.

5.2 Background

5.2.1. Mechanistic models in biology

Biological systems are modeled using several methods, including statistical [58], graphical [59], agent-based [60], [61], and mechanistic methods [62]. Mechanistic models aim to explain or

predict the behavior of a biological systems based on knowledge of how distinct components within the system interact [63]. Well-defined models can be incredibly informative to experimental design, and can be useful for predicting phenomena in the absence of suitable measurement techniques [64], [65]. Mechanistic models of gene expression often consider the rate at which RNA and protein molecules accumulate and diminish; and these changes are described in a time- and space-dependent manner using differential equations [65]–[67].

5.2.2. Parameter estimation

Parameter estimation is a type of inverse modeling technique that uses measured data from a physical system to infer the values of a parameterized model. Typically, this inference is the result of iterative numerical approximation. In biology, parameter estimation has been applied to problems such as identifying experimental conditions that ensure robust behavior of engineering cells [68], [69] and predicting molecular binding affinities [28].

Strategies for parameter estimation in biology traditionally fall into two categories: gradient-based and heuristic. Gradient-based methods solve a system of equations at several points in a multidimensions space and to identify the direction within the multidimension space that leads to a minimum or maximum. Examples of gradient-based methods include the Nelder-Mead method [70] and the Levenberg-Marquardt algorithm [71]. An advantage of gradient-based methods is that they can lead to very precise solutions; however, limitations are slow-convergence, inability to handle high-dimensional problems, and the propensity to identify local minima. Heuristic strategies randomly generate an initial set of parameter estimates and iteratively refine these estimates.

Heuristic methods excel at identifying parameters that reproduce empirical observations for high-dimensional problems, and they have been employed towards parameter estimation in biology for this very reason. However, it is important to remember that parameter values predicted by heuristic search methods do not necessarily represent the actual values of the model.

Instead, they represent one potential solution within a finite, or sometimes infinite, solution space.

5.2.3 Continuous culturing

Cells suspended in media are grown under two main regimes: batch culture and continuous culture. In batch culture, cells are inoculated into a fixed volume of medium in which they replicate while constantly consuming nutrients and exchanging byproducts with the media. Cell growth in batch culture can be characterized by four phases: (1) lag phase, where cells are not actively dividing; (2) exponential phase, where cells replicate vigorously; (3) stationary phase, where cell division dramatically slows as nutrients are depleted and byproducts accumulate; and (4) death phase, when the number of viable cells in suspension decreases. The simplicity of batch culturing strategies make them the primary method by which researchers grow cells in suspension. However, a major drawback of batch culture is that as the cells grow, they alter the conditions of their culturing environment and confound attempts to characterize biological responses [72].

Continuous culturing strategies strive to maintain constant environmental conditions by continuously replacing a portion of cell suspension with fresh medium. A variety of devices have been developed to enforce environmental condition, and chemostats are one notable technology developed for continuous culture [73]. Chemostats dilute cells at a fixed rate, and cell growth reaches steady-state when a limiting nutrient is completely consumed [72]. Chemostats have not experienced widespread use, in part, because they are difficult to construct and operate. However, recent efforts have modified earlier chemostat designs to simplify operation, reduce materials consumed, and increase thoroughput of continuous culture [74].

5.3 Results

5.3.1 Validating fluorescence as a reporter for RNA abundance

To accurately report intracellular RNA levels using fluorescence, we needed to understand the input-output relationship between total MGA5S concentration and fluorescence intensity for our experimental conditions. We constructed a small library of plasmids that allowed us to sample a range of RNA concentrations in *E. coli*. **Figure 5.1** provides a general diagram that illustrates the composition of plasmid constructs. In general, MGA5S expression was driven by a constitutive promoter on a medium copy plasmid.

Figure 5.1. SBOLv diagram of the general strategy for expressing MGA5S in E. coli. Various constitutive promoters of differing strength were used to control MGA5S product within cells.

Given a fixed amount of malachite green, fluorescence is expected to increase uniformly with increasing MGA5S concentration. However, when the concentration of MGA5S is in excess of the dye, fluorescence will saturate. We expressed MGA5S behind eight promoters and measured the fold-change in aptamer concentration between samples using fluorescence. To validate the use of fluorescence, we also measured fold-change in the aptamer concentration using reverse transcriptase, quantitative PCR.

Figure 5.2 demonstrates that we observed a strong linear correlation, $R^2 = 0.92$, between measurements obtained from fluorescence and reverse-transcriptase, quantitative PCR. These data indicate that our experimental conditions do not cause fluorescence saturation. We calculated error between abundance measurements made by quantitative PCR and fluorescence, and those values are provided in **Table 5.2**. Error was normalized by the average of each pair of fluorescence and reverse-transcriptase quantitative PCR measurements to gauge the extent of how close values were to one another. We found that the normalized error seldom exceeded 20% of the mean, which further indicates good agreement between the relative concentrations calculated by the two techniques. Taken together, these data show that fluorescence intensity accurately reports relative MGA5S levels in *E. coli* for our experimental conditions.

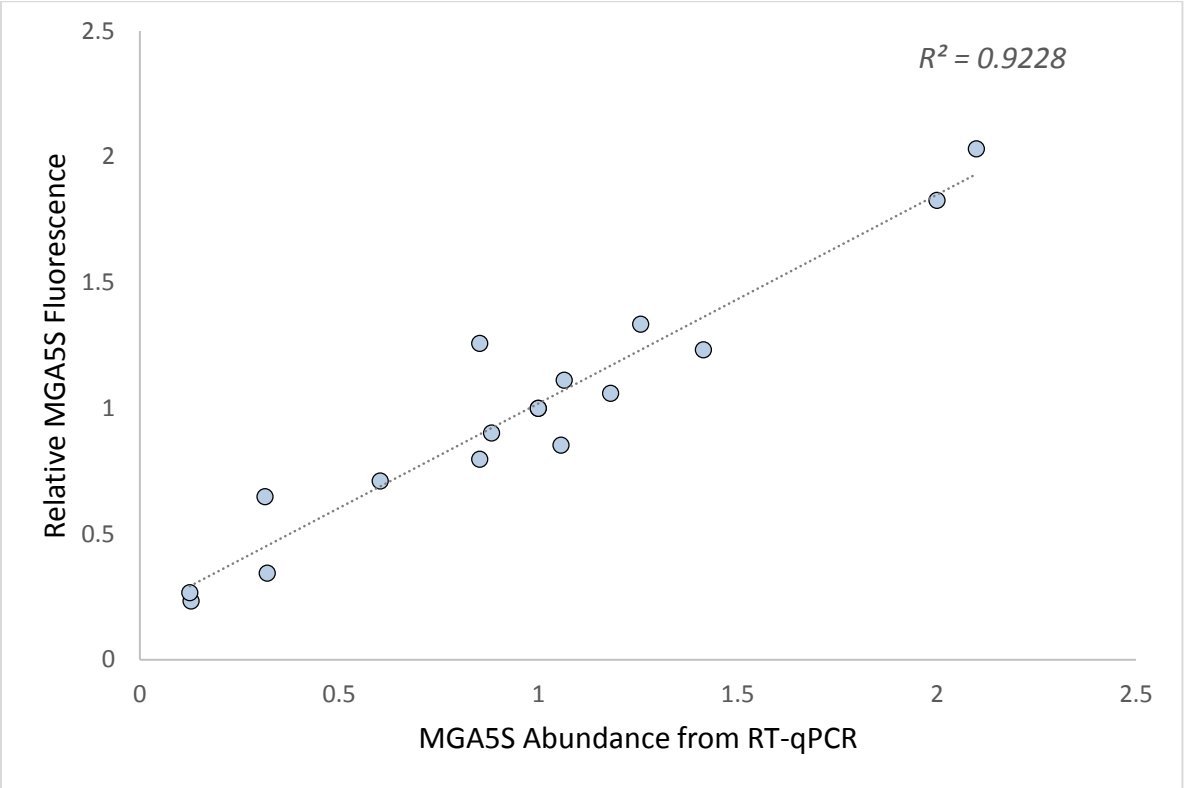


Figure 5.2. A comparison of MGA5S abundance measured by reverse-transcriptase, quantitative PCR and relative abundance measured by fluorometry.

Sample	RT-qPCR Abundance	MGA5S Fluorescence	Abs. Pct. Error
J23106	0.320	0.344	7%
J23118	0.603	0.711	16%
J23105	0.129	0.233	58%
J23101	1	1	0%
R0040	1.181	1.059	11%
FAB346	1.257	1.335	6%
J23104	2	1.826	9%
R0011	0.883	0.902	2%
R0040	1.057	0.853	21%
J23100	1.414	1.232	14%
J23101	1.000	1.000	0%
J23102	1.064	1.112	4%
J23104	2.099	2.031	3%
J23105	0.125	0.267	73%
J23106	0.314	0.648	69%
J23118	0.853	0.797	7%
FAB346	0.853	1.257	38%

Table 5.2. A comparison of MGA5S concentration predicted by two measurement strategies. Reverse-transcriptase, quantitative PCR and fluorescence show good agreement.

5.3.2. Mathematical model of molecular interactions

We developed a mechanistic model that accounts for changes in the concentrations of the aptamer and dye in order to link absolute MGA5S concentration with fluorescence intensity. **Figure 5.3** diagrams these molecular interactions.

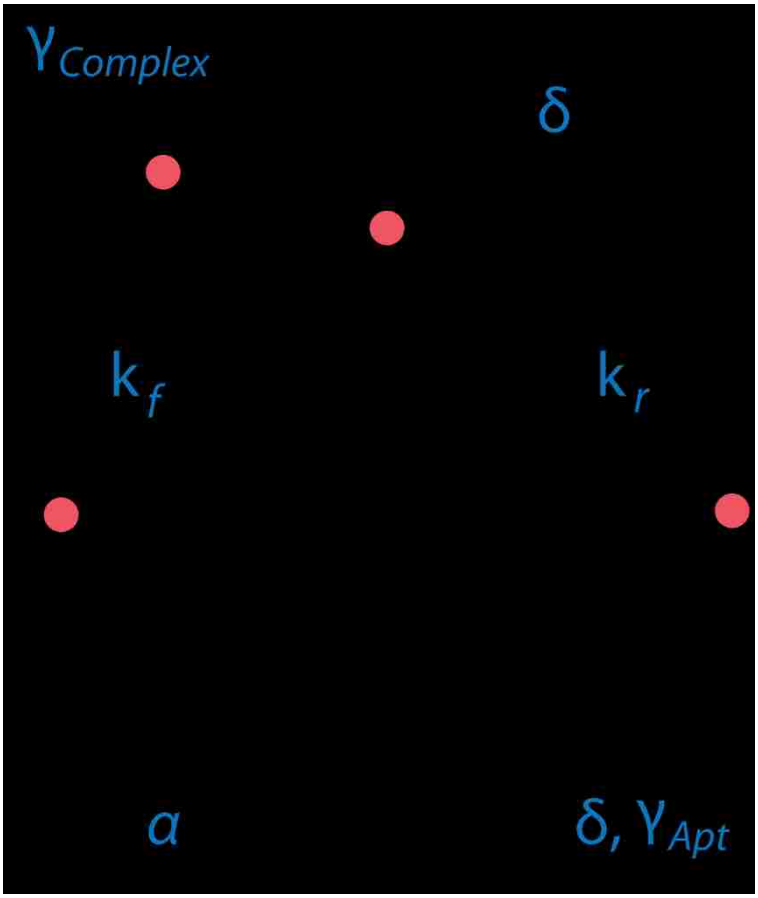


Figure 5.3. A system model for the molecular interactions of MGA5S. Aptamer concentration is affected by synthesis, degradation, dilution, binding, and unbinding events.

Time-dependent changes are described by the set of ordinary differential equations given in **Equations 5.1 – 5.3**. In this system of equations, $\boldsymbol{\alpha}$ is the rate of MGA5S synthesis, γ_{Apt} is the rate of degradation of the aptamer in its unbound form; $\gamma_{Complex}$ is the rate of degradation of the aptamer in its bound form; $\mathbf{k_f}$ and $\mathbf{k_r}$ are the rate constants describing binding and unbind of MGA5S and malachite green, respectively; $\boldsymbol{\delta}$ is the dilution rate constant due to cell growth; and **[Apt]**, **[Dye]**, and **[Complex]** are the absolute concentrations of unbound MGA5S, unbound malachite green, and MGA5S-malachite green complex, respectively.

Equation 5.1:

$$\frac{d[Apt]}{dt} = \alpha - \left(\gamma_{Apt} + \delta\right)[Apt] - k_f\,[Apt]\,[Dye] + k_r\,[Complex]$$

Equation 5.2:

$$\frac{d[Complex]}{dt} = k_f\,[Apt][Dye] - (k_r + \gamma_{Complex} + \delta)\,[Complex]$$

Equation 5.3:

$$\frac{d[Dye]}{dt} = (k_r + \gamma_{Complex})\,[Complex] - k_f[Apt][Dye]$$

Fluorescence is proportional to the concentration of MGA5S and malachite green in complex. For a fixed amount of malachite green, fluorescence is at its theoretical maximum when all malachite green is in complex with MGA5S. This allows us to propose a relationship between the absolute concentration of the aptamer-dye complex and fluorescence intensity, which is described by **Equation 5.4**.

$$FL = c_1 * [Complex]$$

and

$$FL_{MAX} = c_1 * [Dye]_{Total}$$

Equation 5.4:

$$[Complex] = [Dye]_{Total} * \frac{FL}{FL_{MAX}}$$

Finally, conservation of mass requires that the total amount of MGA5S is equal to the sum of unbound aptamer and aptamer that is in complex with malachite green. Similarly, the total amount of malachite green is equal to the sum of unbound dye and dye that is in complex with MGA5S. These equations are defined in **Equation 5.5** and **5.6**.

Equation 5.5:

$$[RNA]_{Total} = [RNA] + [Complex]$$

Equation 5.6:

$$[Dye]_{Total} = [Dye] + [Complex]$$

When RNA production is in steady-state, **Equation 5.2** can be simplified. The equation was rewritten to solve for total MGA5S concentration from steady-state. First, since $\frac{d[Complex]}{dt}$ equals 0, **Equation 5.2** was reduced to its algebraic form. Next, we substituted for [RNA] and [Dye] using the relations given in **Equation 5.5** and **5.6** to yield,

$$0 = [Complex]^2 - ([RNA]_{Total} + [Dye]_{Total} + \omega)[Complex] + [RNA]_{Total}[Dye]_{Total}$$

$$\text{with} \qquad \omega = \frac{k_r + \gamma_{Complex} + \delta}{k_f}$$

The above equation is a quadratic relation that was solved in terms of complex,

$$[Complex] = \frac{([RNA]_{Total} + [Dye]_{Total} + \omega) - \sqrt{([RNA]_{Total} + [Dye]_{Total} + \omega)^2 - 4[RNA]_{Total}[Dye]_{Total}}}{2}$$

Finally, using the relation between fluorescence intensity and the concentration of the aptamer-dye complex provided in Equation 5.4 gave,

Equation 5.7:

$$FL = \frac{FL_{MAX}}{[Dye]_{Total}} * \frac{([RNA]_{Total} + [Dye]_{Total} + \omega) - \sqrt{([RNA]_{Total} + [Dye]_{Total} + \omega)^2 - 4[RNA]_{Total}[Dye]_{Total}}}{2}$$

5.3.3 Theoretical analysis of solution to system of equations

We sought to identify the minimal amount of empirical data necessary to accurately estimate parameter values within the model by revisiting **Equation 5.7**. We were confident in our ability to estimate to degradation and to measure dilution, fluorescence, total dye concentration,

and relative aptamer abundance. To see if these data were sufficient, we explored how to apply our mathematical model to predict the remaining unknown parameter values.

To begin, we implemented a Python script to simulate experimental data by numerically integrating **Equations 5.1 – 5.3**. Parameter values used in the simulation were $\gamma_{Apt} = \gamma_{Complex}$ = 1.2e-4 / min, k_f = 1e-1, k_r = 1e-8. The rate of synthesis ranged from 1e-3 to 2e-2 μM/min, and the rate of dilution ranged from 4.17e-3 to 7.5e-3 / min. Three distinct dye concentrations, 2 μM, 5 μM, and 10 μM, were used, with respective FL_{MAX} values of 1300, 4000, and 11000.

A plot summarizing the simulated system is provided in **Figure 5.4**. As expected, we observe a linear increase in fluorescence at relatively low concentrations. The slope of the line correlating fluorescence to aptamer abundance increases with the addition of more malachite green. At the highest of MGA5S concentrations, fluorescence appears to saturate for 2 μM malachite green; however, this saturation is not observed at 5 μM and 10 μM.

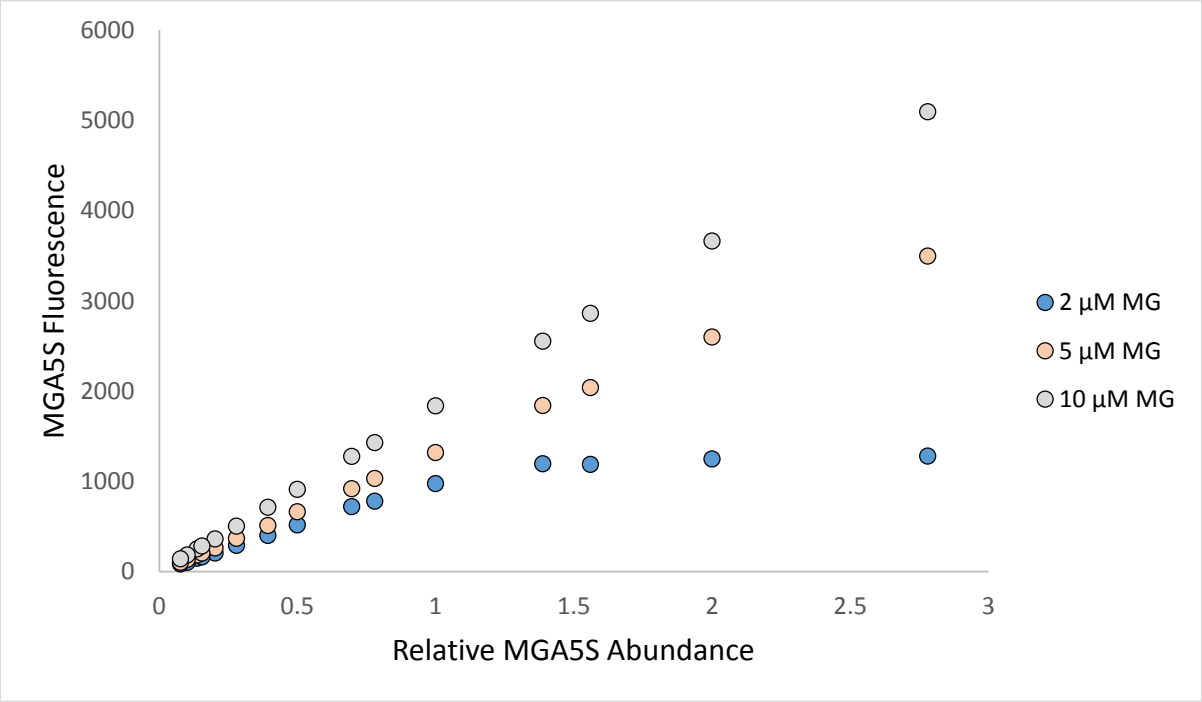


Figure 5.4. Simulated experimental data.

Next, we applied a parameter estimation strategy based on differential evolution to attempt to identify system parameters when only supplied with the following system values: (1) fluorescence, (2) dilution rate constant, (3) degradation rate constant, and (4) total dye concentration. Additionally, parameter estimation made use of knowledge about the relative concentration of RNA between samples (as this could be obtained from RTqPCR) and FL_{MAX} was shared between all samples exposed to the same dye concentration.

Figure 5.5 compares the expected total MGA5S concentration for our simulation and the predicted total MGA5S concentration following 30 optimization runs. We find that the provided amount of information is sufficient to confidently estimate total RNA concentration.

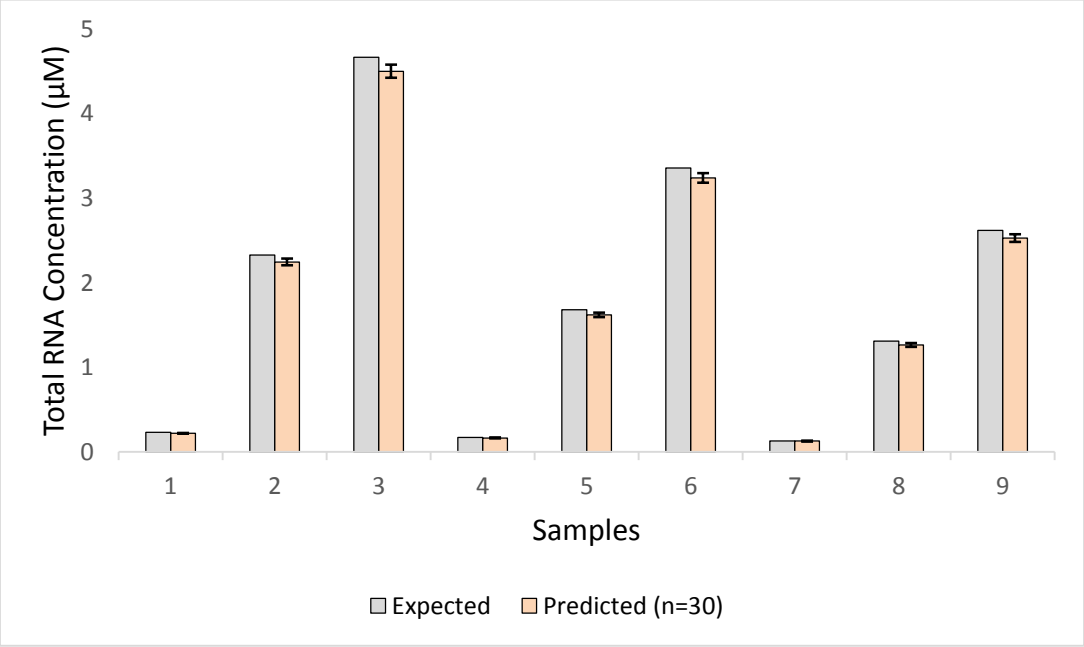


Figure 5.5. A comparison of expected and predicted total RNA concentrations from simulated data.

Upon further investigation, we found that measuring fluorescence with only a single dye concentration was sufficient for accurate approximation of absolute aptamer concentration. We

also found that a crucial requirement to obtaining a confident estimation of parameters was to obtain fluorescence data that exhibits saturation. For example, **Figure 5.6** is the same simulated data as **Figure 5.4**, except saturating fluorescence values have been removed. All attempts to accurately predict unknown parameters through differential evolution optimization were unsuccessful when provided with only linear fluorescence data.

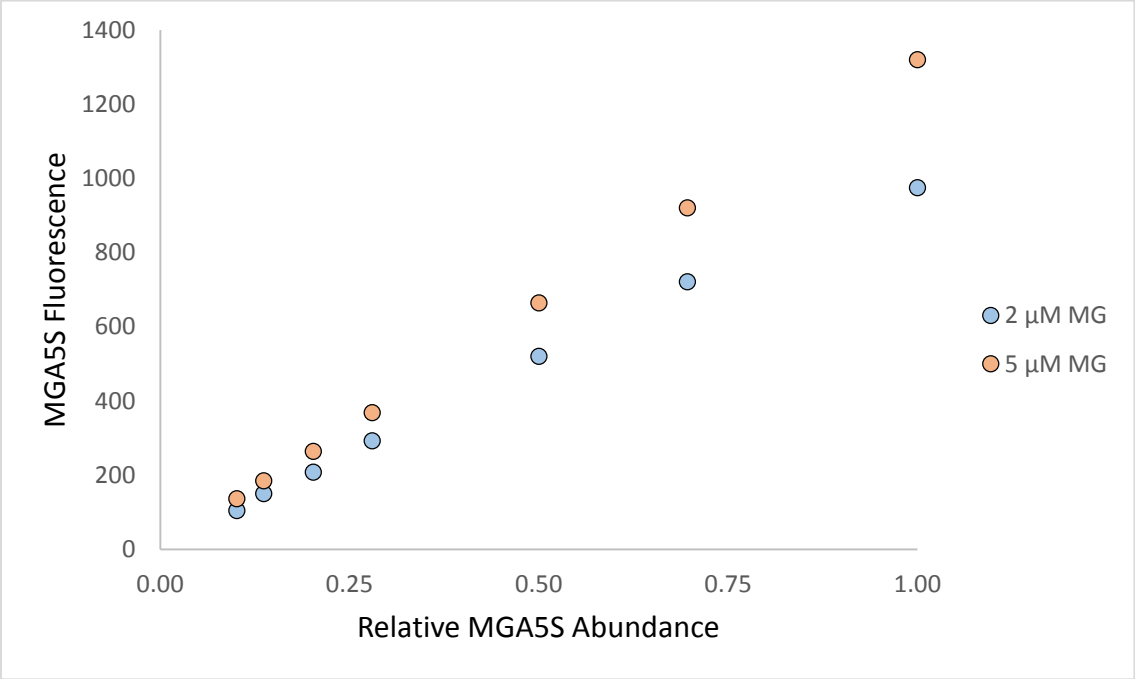


Figure 5.6. Simulated experimental data. No fluorescence saturation is observed at relatively low MGA5S abundance.

5.3.4. Controlling the rate of dilution

The dilution rate due to cell growth was explicitly controlled by culturing cells in a continuous flow bioreactor. **Equation 5.8** describes how dilution rate and cell growth are related at steady-state, where μ is the specific rate of growth for cells and δ is the dilution rate of media

in the reactor. At steady-state, the rate of growth and dilution are the same, and the cell density remains constant.

Equation 5.8:

$$\frac{d[Cells]}{dt} = (\mu - \delta) * [Cells]$$

By adjusting flow rate of media into the culturing chamber, we found that we could precisely tune the specific rate of cell growth. **Figure 5.7** shows that cell density equilibrates between 18 and 24 hours of growth in M9 medium. From this data, it was concluded that 30 hours between inoculation and measurement were the minimum required to remain confident that cells were in a physiological steady-state. A more detailed analysis of the capabilities and operation of the continuous culturing system are provided in **Appendix 1**.

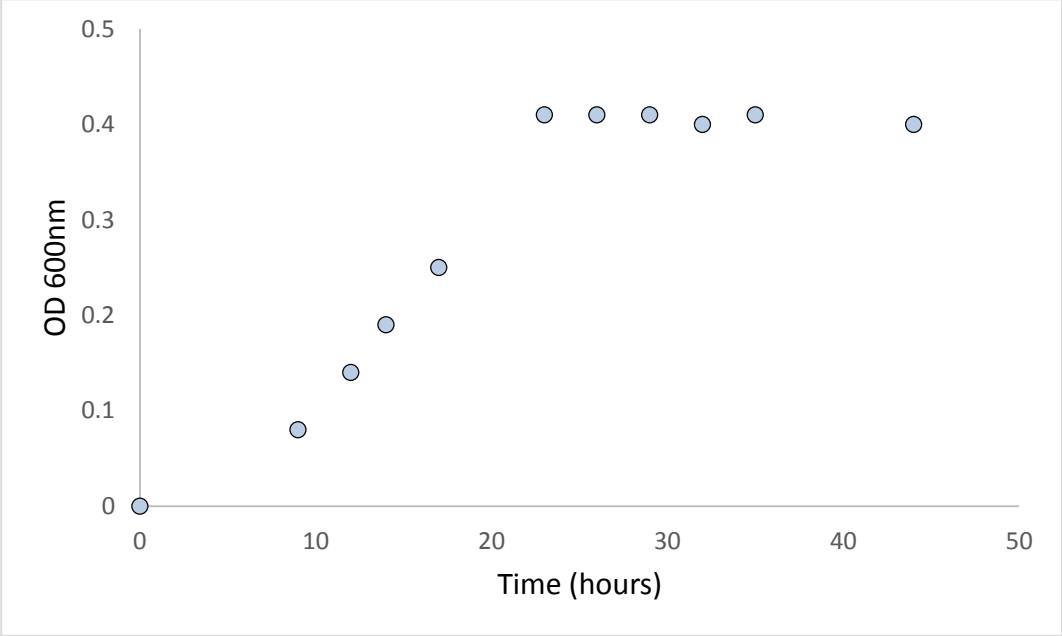


Figure 5.7. Cell density of *E. coli* growing in continuous culture over time.

5.3.5. Determining the rate of RNA degradation

In the absence of malachite green, the system of equations describing molecular interactions simplifies to a single ordinary differential equation, **Equation 5.9**, which describes the time-dependent concentration of MGA5S.

Equation 5.9:

$$\frac{d[Apt]}{dt} = \alpha - \gamma_{Apt} * [Apt]$$

When RNA synthesis (α = 0) is inhibited, the degradation rate constant of unbound MGA5S can be calculated by fitting empirical data to an exponential decay function. These conditions were realized experimentally by exposing cells expressing MGA5S to rifampicin, a molecule that inhibits bacterial RNA synthesis [75]. Cell cultures were sampled periodically following rifampicin treatment and MGA5S abundance was quantified through reverse-transcriptase quantitative PCR analysis. **Figure 5.8 and Table 5.3** show the relative change in MGA5S abundance for 40 minutes following exposure to rifampicin. The reported values for MGA5S abundance were normalized by a reference gene, 16S rRNA. Surprisingly, MGA5S showed no apparent degradation within that timeframe. These results suggest that MGA5S degrades just as slowly as 16S rRNA.

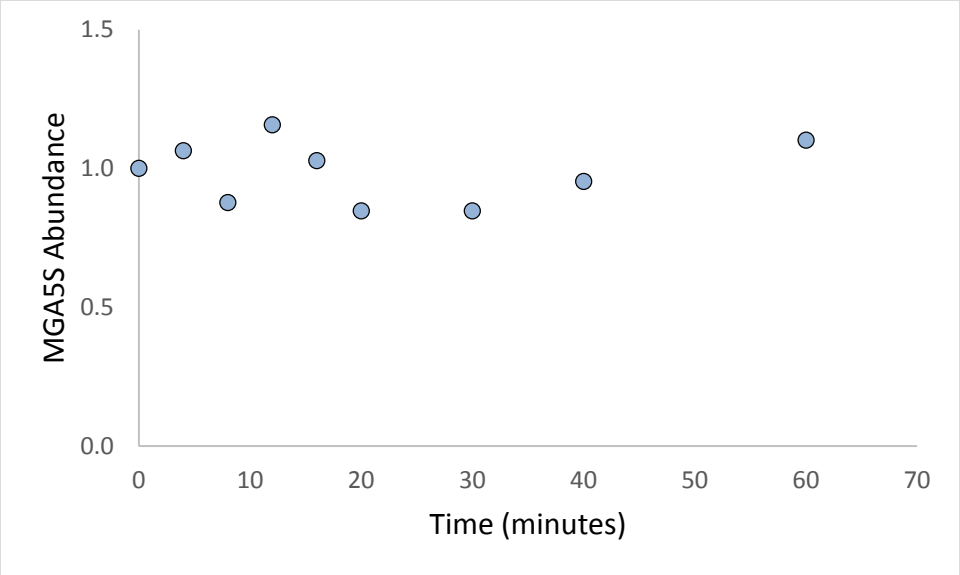


Figure 5.8. Relative change in MGA5S abundance, as measured by RTqPCR, following rifampicin exposure.

	Ct		ΔCt			
Time (minutes)	**MGA5S**	**16S**	**MGA5S**	**16S**	**ΔΔCt**	**1/2^ΔΔCt**
0	9.71	10	0.00	0.00	0.00	1.00
4	9.58	9.96	-0.13	-0.04	-0.09	1.06
8	10.03	10.13	0.32	0.13	0.19	0.88
12	10.01	10.51	0.30	0.51	-0.21	1.16
16	10.07	10.4	0.36	0.40	-0.04	1.03
20	10.18	10.23	0.47	0.23	0.24	0.85
30	10.35	10.4	0.64	0.40	0.24	0.85
40	10.43	10.65	0.72	0.65	0.07	0.95
60	10.59	11.02	0.88	1.02	-0.14	1.10

Table 5.3. Relative change in MGA5S abundance following rifampicin exposure.

Alternatively, we attempted to estimate the degradation rate constant from fluorescence. This was based on that assumption that binding to malachite green does not significantly alter degradation of the aptamer. Cells expressing malachite green were grown in batch culture until MGA5S expression per cell reached steady-state. Once MGA5S was shown to be produced at a constant rate, cells were exposed to rifampicin. Fluorescence was measured throughout the entirety of this experiment, and the results are displayed in **Figures 5.9 – 5.11**. As expected, cultures that were not exposed to rifampicin continued to produce MGA5S at a constant rate until they reached high cell densities. However, we observed that both the fluorescence and absorbance of cultures exposed to malachite green remained largely unchanged following rifampicin exposure. It should be noted that rifampicin alters the optical properties of the media and resulted in a shift in both fluorescence and absorbance.

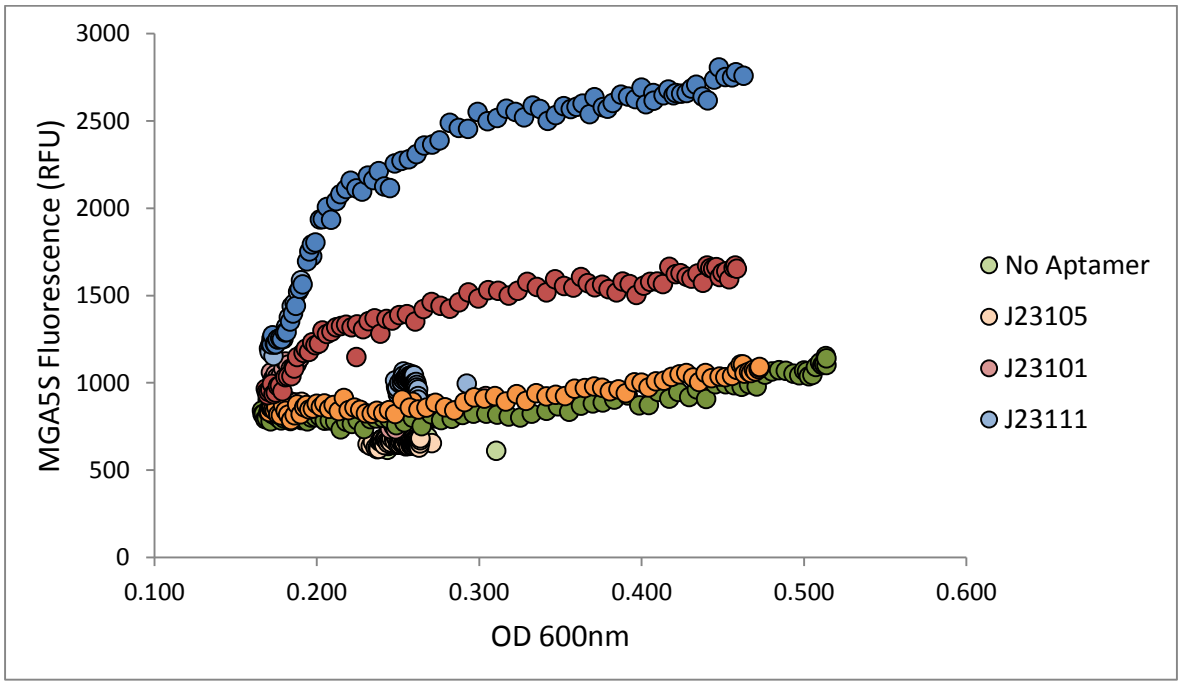


Figure 5.9. Fluorescence trajectory of cells with and without rifampicin. Darker filled circles represent cells that are not exposed to rifampicin. Lighter filled circles represents cells exposed to rifampicin.

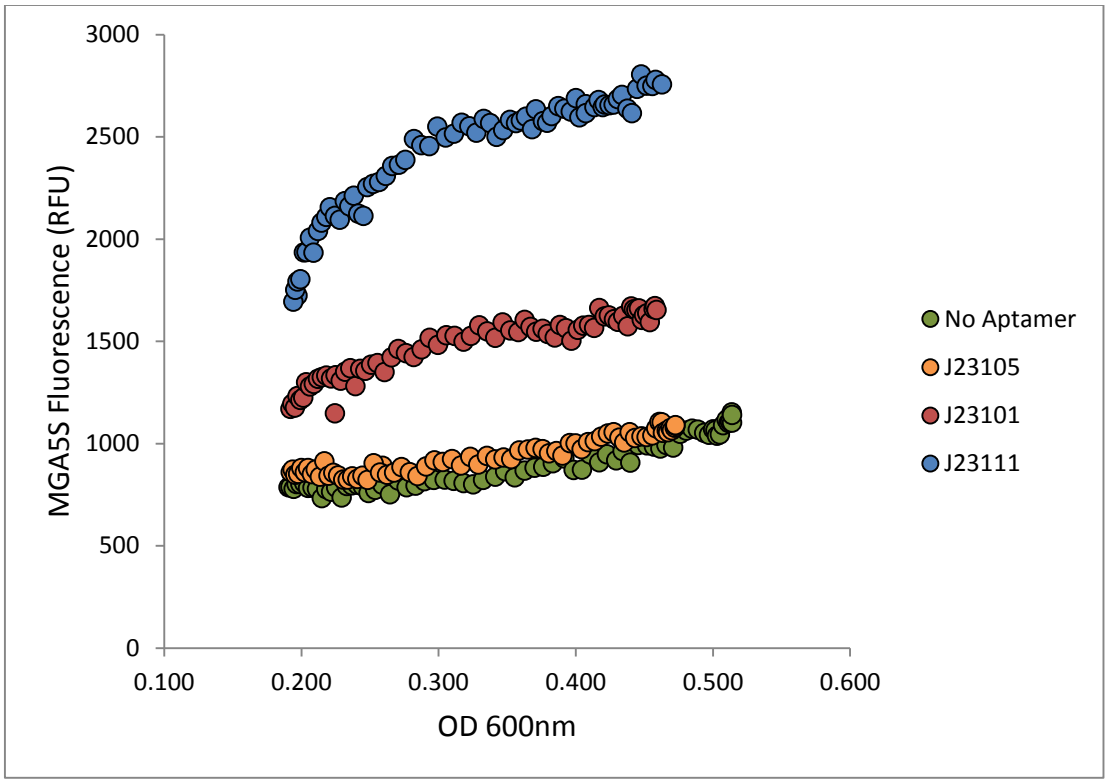


Figure 5.10. Fluorescence trajectory of cells not exposed to rifampicin.

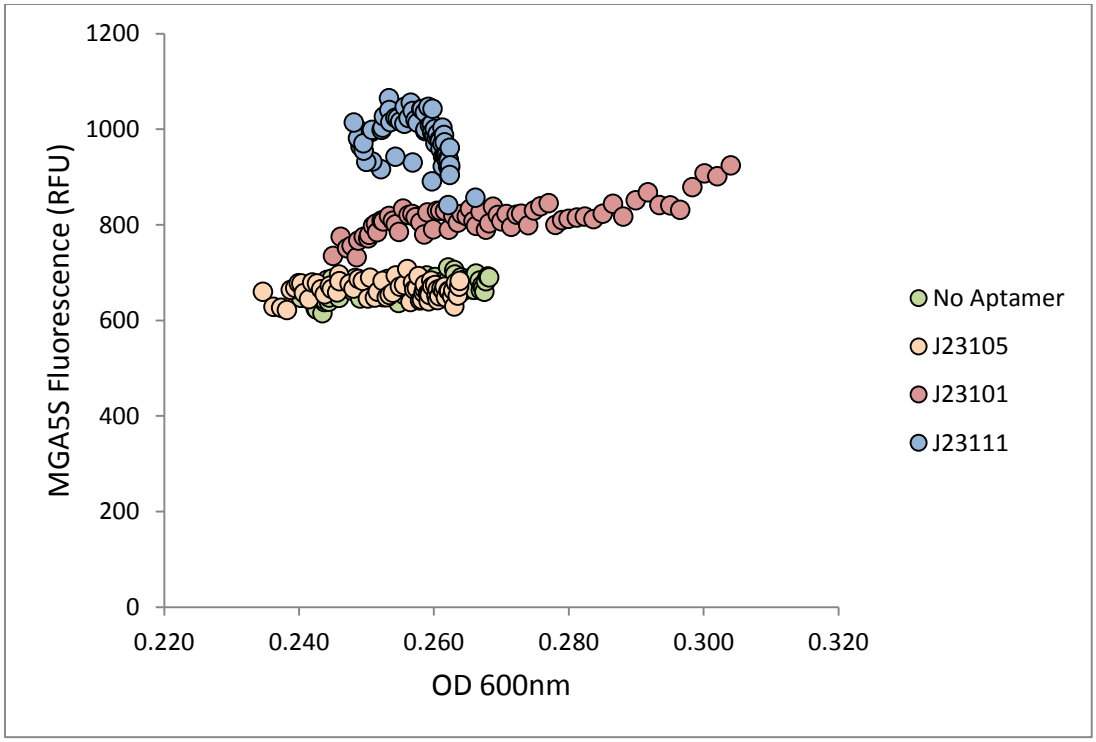


Figure 5.11. Fluorescence trajectory of cells exposed to rifampicin.

The results of both RT-qPCR and fluorometry experiments suggest that any decrease in MGA5S in *E. coli* is predominantly due to dilution that is the result of cell growth. As a consequence, the degradation rate constant for both the unbound and bound aptamer (γ_{Apt} and $\gamma_{Complex}$, respectively) are set to 0 in the systems model.

5.3.6. Estimating absolute RNA concentration

Using experimentally determined parameter values from the previous sections, we applied differential evolution to **Equation 5.7** in order to predict absolute RNA concentrations in *E. coli*. The optimization routine was supplied the following information: (1) fluorescence values measured by fluorometry (2) $\gamma_{Complex} = 0$, (3) total dye concentration = 5 μM, (4) dilution rates measured from continuous culture, (5) relative MGA5S abundance as measured by RTqPCR, and (6) kr = kf * 116e-9.

Figure 5.13 and **Table 5.4** summarize the results of eight distinct optimizations. Samples were predicted to have RNA abundances that ranged from as little as 390 ± 60 copies per cell to as many as 8,350 ± 1,320 copies per cell. As the predicted concentration of MGA5S increases, so does the error surrounding the prediction. As a result, attempting to estimate the absolute concentration of the largest MGA5S concentrations results in a large standard deviation.

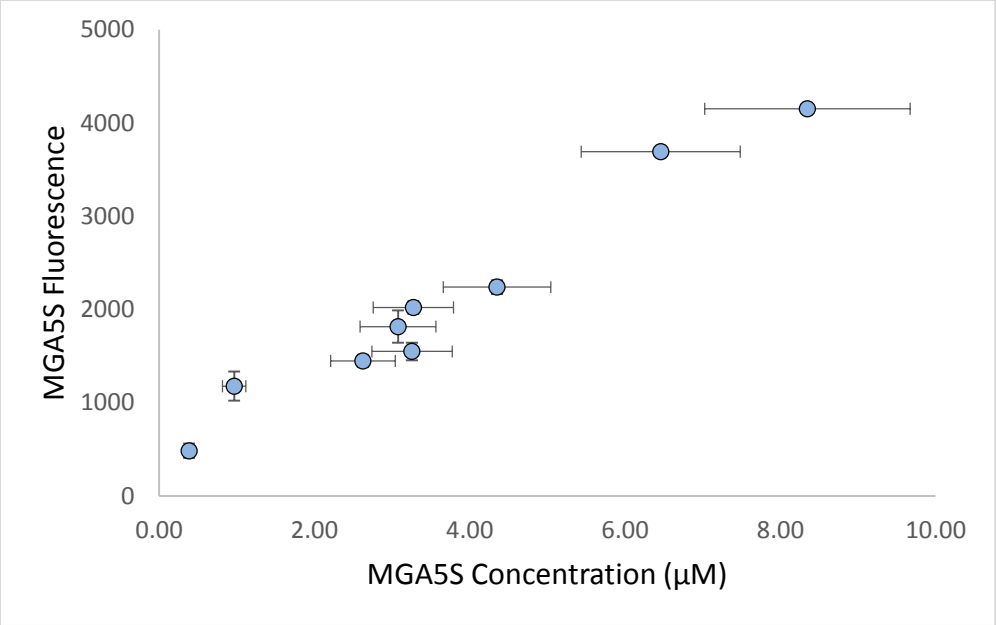


Figure 5.12. Absolute MGA5S concentration in *E. coli* is predicted from experimental fluorescence data.

	Concentration		**Fluorescence**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev**
J23105	0.39	0.06	486	77
J23106	0.97	0.15	1178	156
J23118	2.62	0.42	1449	65
R0040	3.26	0.52	1550	95
J23101	3.08	0.49	1817	172
J23102	3.27	0.52	2021	68
J23100	4.35	0.69	2239	75
J23104	6.46	1.03	3691	18
J23111	8.35	1.32	4152	47

Table 5.4. Absolute MGA5S concentration in E. coli is predicted from experiment fluorescence data.

5.3.7. Considering alternative models

An assumption of the model is that the change in total dye concentration within cells reaches steady-state. This assumption is supported by earlier findings such as those presented in **Figure 4.7**, where fluorescence equilibrates for a fixed total MGA5S and total dye concentration. A mechanism that would explain steady-state total dye concentrations is that the dye rapidly diffuses across the cell membrane. After some initial influx of dye into cells following exposure to malachite green, we would expect the total dye concentration to remain largely unchanged.

However, we investigated the effect of active transport of dye into cells. An equation that explains the change in total dye concentration is given by **Equation 5.10**.

Equation 5.10:

$$\frac{d[Dye]_{Total}}{dt} = \frac{d[Dye]}{dt} + \frac{d[Complex]}{dt} = j - \delta\,[Complex]$$

In this equation, j represents the net flux of dye into the cell. If j is negative, this suggests that malachite green is being actively transported out of the cell. When j is 0, the net flux of malachite green into and out of *E. coli* is unchanging. Another way of stating this second case is to say that the flux of malachite green into cells is at steady-state; however, the total concentration of dye within cells is expected to decrease due to an additional loss of dye due to cell growth. A third, and final case worth considering is when $j = \delta$ [Complex]. Under this regime, the net rate of malachite green import is equal to the rate of malachite green loss due to dilution. In this final case, the change in total dye is at steady-state.

The concentration of aptamer-dye complex and unbound malachite green were simulated under each of these regimes, and the results are shown in **Figure 5.13 – 5.15**. Assuming active transport of malachite green into cells at a rate greater than δ * [Complex], **Figure 5.13** shows that

dye concentration is expected to continually increase. This would mean that fluorescence would continue to rise until all aptamer molecules within cells are bound. Moreover, when the fluorescence finally saturates, adding more dye should not cause a further increase in fluorescence. Contrary to this hypothetical scenario, experimental data has demonstrated that for a given dye and fluorescence equilibrate, following addition of more malachite green, fluorescence signal once again rises (see **Figure 4.7**).

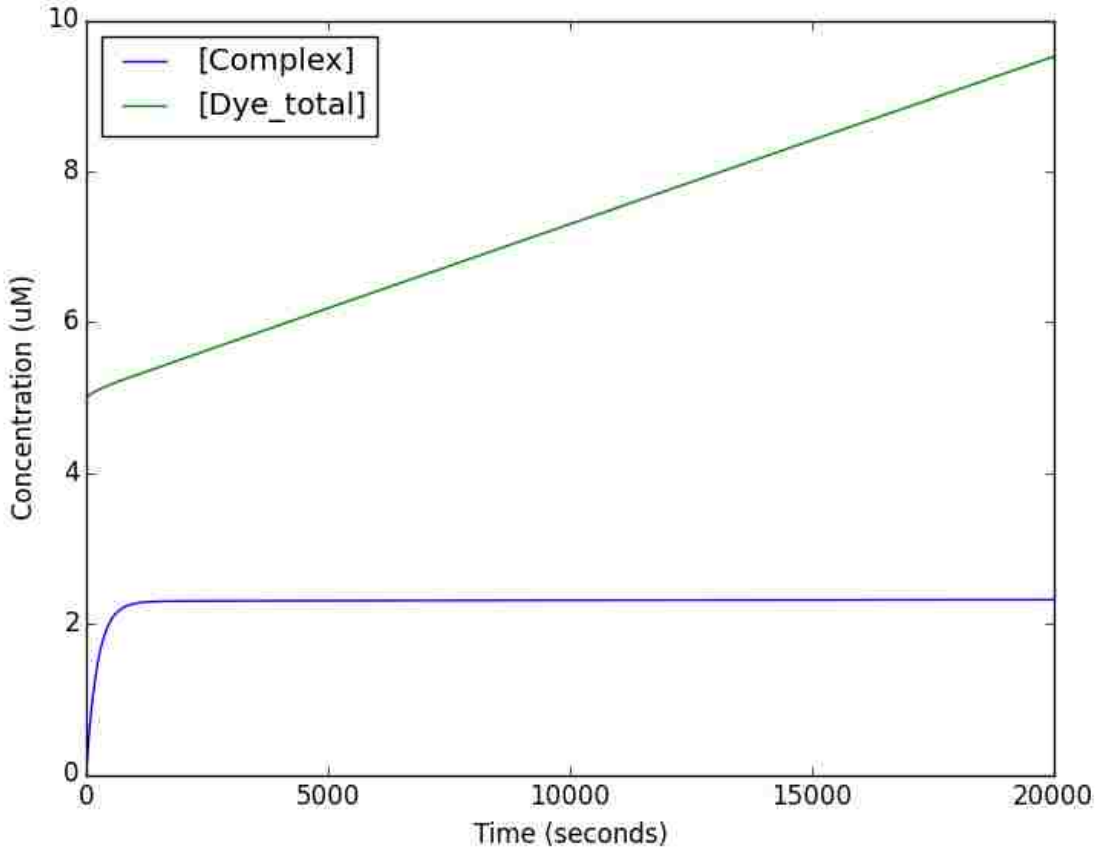


Figure 5.13. Active transport of malachite green into cells at a rate greater than the rate of malachite green loss due to dilution results in a constant increase in intracellular total malachite green, and consequently, fluorescence.

Figure 5.14 shows that if the rate of influx of malachite green into cells is less than the rate of malachite green loss due to cell growth, then eventually the intracellular total dye concentration will reach 0 μM. This would result in a situation where no fluorescence is observed from any cells given enough time. This second scenario is incongruent with experimental observations.

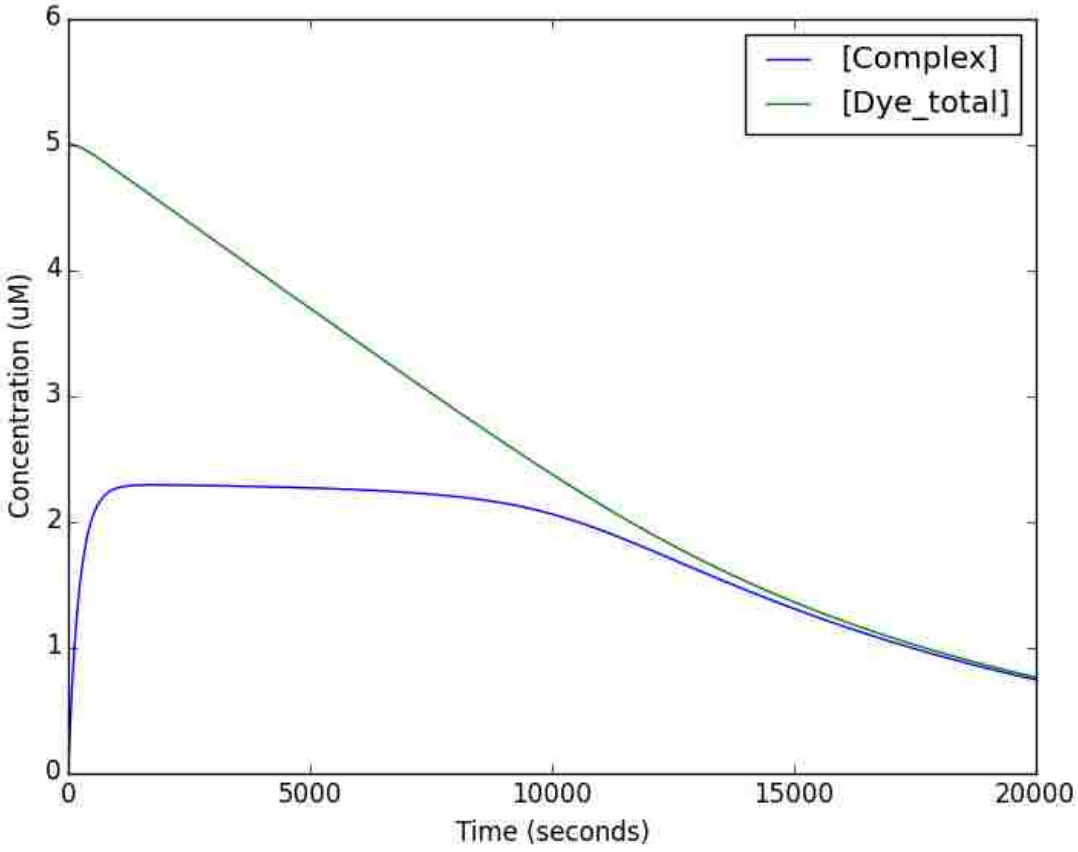


Figure 5.14. When malachite green within cells is loss due to cell growth at a rate greater than its influx into cells, the concentration of the aptamer-dye complex as well as the total dye concentration decrease over time. This leads to a decreasing fluorescence signal that eventually disappears.

The results of the final case are presented in **Figure 5.15**. The assumption of this scenario is that the rate of influx of malachite green into cells is equal to the rate of loss of malachite green because of dilution due to cell growth at all times that are sufficiently later than the time of initial exposure of malachite green to cells. In this scenario, we observe an initial rise in the concentration of the MGA5S-malachite green complex and an eventual plateau in its concentration. This corresponds to the experimentally observed rise in fluorescence followed by an equilibration in fluorescence.

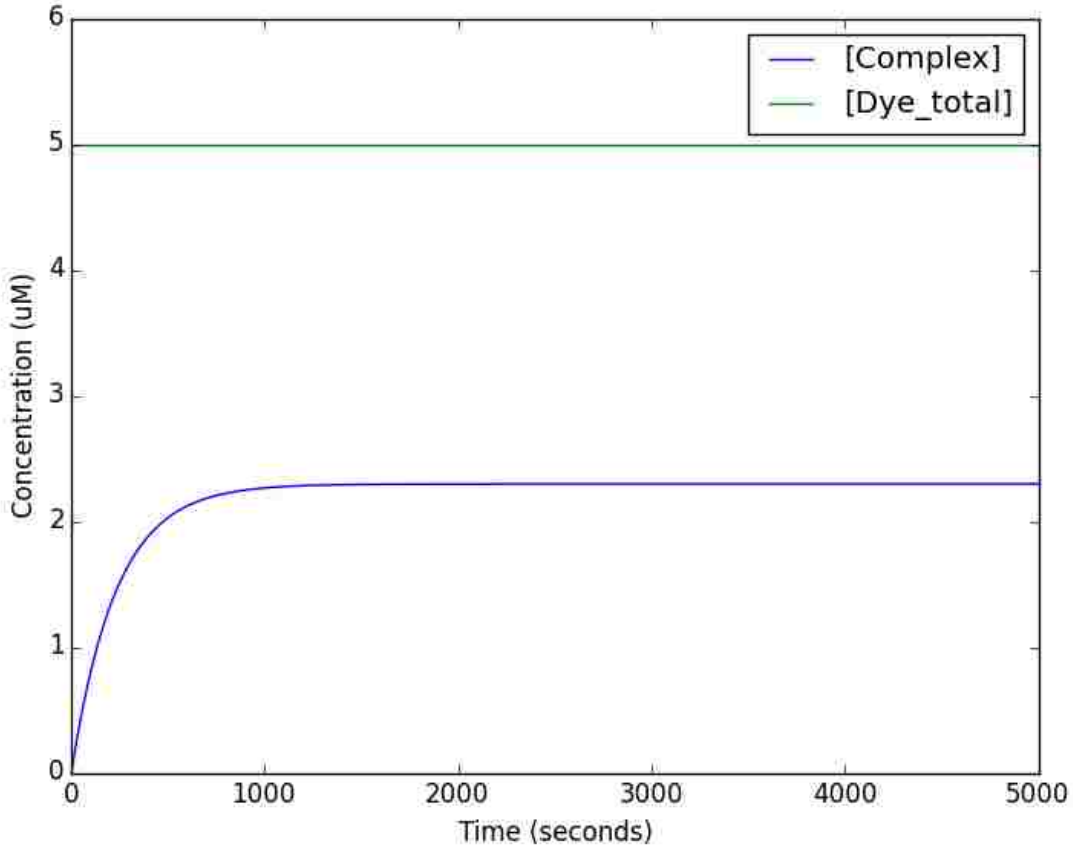


Figure 5.15. A third mechanism assumes that the change in total intracellular dye reaches steady-state. This requires that the rate of import and the rate of loss due to cell growth are equal. As a result, fluorescence initially increases and eventually equilibrates.

5.4 Discussion

In this chapter, we focus on using MGA5S as a quantitative reporter for RNA abundance. To realize the quantitative capability of MGA5S, we relied heavily on a combination of experimental methods and mathematical modeling. Ultimately, we predict intracellular MGA5S abundance from fluorescence intensity. This is a meaningful achievement because alternative, non-destructive methods for RNA expression analysis are either qualitative or, at best, semi-quantitative.

We began by validating that fluorescence intensity corresponds to *in vivo* aptamer concentration. Demonstrating this concept was a prerequisite for any further analysis. Next, we were able to show that there was a strong correlation between fluorescence intensity and intracellular concentration by comparing MGA5S abundance measurements made by fluorometry and by using reverse-transcriptase, quantitative PCR.

In theory, a model of molecular interactions should allow us to accurately relate absolute MGA5S concentration to fluorescence intensity. However, the predictive accuracy of such a model would be strongly influenced by our ability to confidently parameterize the model. Consequently, we devised experiments to measure several unknown parameters and we developed strategies to directly control other parameters.

We found that MGA5S degraded very slowly. This should not be surprising since its structure mimics rRNA – a molecule that is known to persist in *E. coli* for generations [76]. We measured RNA stability using reverse-transcriptase, quantitative PCR. Due to its enhanced stability, we were not able to compare the degradation of MGA5S with a reference sequence, like 16S rRNA, as has been done in other studies [5], [53], [77]. To further increase confidence in our assumption that the rate of MGA5S degradation was negligible compared to the rate of dilution due to cell growth, we observed degradation of MGA5S fluorescence following transcriptional

arrest. After observing no significant decrease in fluorescence per cell over the course of 10 hours, we were confident in setting the rate of MGA5S degradation in our model to 0.

Dilution rate can be estimated by intermittently measuring the absorbance for cells growing in batch culture. By carefully determining the relationship between absorbance and cell density, an accurate estimation of dilution can be obtained. However, continuous culturing strategies provide a more robust strategy for estimating dilution since the rate of dilution can be precisely controlled, and we used this method to influence the rate of dilution.

With a more refined model, we identified what empirical data would be necessary to predict MGA5S concentration with high confidence. The parameter estimation routine predicted RNA abundances that ranged from 390 ± 60 to 8350 ± 1320 transcripts per cell. The large majority of RNA in cells exists as ribosomal RNA [78], [79]. In *E. coli* 16S rRNA abundance has been suggested to reach levels as high as 10^4 to 10^5 molecules per cell [80], [81] through fluorescence *in situ* hybridization studies. We compared the RT-qPCR cycle times of MGA5S and 16S rRNA, our reference gene, to consider whether the *in vivo* MGA5S concentrations predicted by our model were reasonable. The strain with the lowest MGA5S abundance had a cycle time of 12.67 ± 0.37 and its respective 16S cycle time was 8.76 ± 0.16 - approximately 4 cycle less. Based on these results, if 16S rRNA has an abundance of 10^4, then MGA5S for this strain would be roughly 330 copies per cell; the model predicts an abundance of 390 ± 60 per cell. The strain with the highest MGA5S concentration exhibited an MGA5S cycle time of 8.25 ± 0.34 and its respective 16S cycle time was 8.78 ± 0.09. Assuming 16S rRNA is 10^4 copies per cell, we would expect MGA5S concentration to be roughly 14,000 copies per cell; the model predicts an abundance of 8,350 ± 1,320 copies per cell. It has been noted that estimating very high concentrations from fluorescence is difficult due to high variability.

Taken together, the work in this chapter shows that MGA5S can be used to quantify absolute RNA concentration in living cells. At present, no other method can claim to quantify RNA abundance in a non-destructive manner. Moreover, a general framework for quantifying any fluorescence-activating aptamer is outlined so that other aptamers can be rapidly deployed for use in quantitative applications.

5.5 Methods

5.5.1 Continuous culture

E. coli were inoculated into 4 ml M9 media containing the appropriate antibiotic and grown overnight at 37°C and 250 rpm. Cells were diluted into 1 ml of fresh M9 media to a final OD_{600nm} of 0.1. The diluted sample was used to inoculate a chemostat reactor containing greater than 10 ml of fresh M9 media with the appropriate antibiotic. Cells were grown in the reactor at 37°C and a specified dilution rate for at least 40 hours prior to sampling and measurement.

5.5.2 Single time point fluorescence measurement

E. coli were grown in continuous culture before being sampled and diluted in fresh M9 media to a final OD_{600nm} of 0.2. 975 µl of diluted culture was added to 25 µl of 200 µM malachite green solution in a microfuge tube and briefly vortexed. 200 µl of culture was added in replicate of four to a black, clear-bottom 96-well plater (Costar). Culture samples were measured at a fluorescence excitation wavelength of 620 nm and fluorescence emission wavelength of 655 nm. The fluorescence of all samples were subtracted by the fluorescence of a no-aptamer control prior to reporting.

Simultaneous to exposing diluted cell cultures to malachite green, 400 µl of diluted culture was added to 800 µl Bacterial RNA Protect Reagent (Qiagen). Cells in RNA Protect Reagent were briefly vortexed and then centrifuged at 14,000xg for 10 minutes. Supernatant was carefully decanted and the cell pellet was immediately frozen at -80°C.

5.5.3 Dynamic fluorescence measurement

E. coli were grown in continuous culture before being sampled and diluted to a final OD_{600nm} of 0.1 in 1 ml of warm M9 media. 975 μl of diluted culture was added to 25 μl of a 200 μM malachite green solution in a microfuge tube. The mixture was briefly vortexed and 200 μl was added in replicate of four to a black, clear-bottom 96-well plate (Costar). The multiwall plate was covered with a Breathe-easy gas-permeable membrane (Sigma) to prevent evaporation of fluid volume. Cultures were incubated within a Tecan M200 pro at 37°C for 16 hours with 120 seconds of linear shaking at 2.5 mm every 6 minutes. Additionally, every 6 minutes, absorbance at 600 nm and fluorescence at a 620 nm excitation wavelength and 655 nm wavelength were recorded for each well. Fluorescence of all samples were substracted by the fluorescence of a no-aptamer control prior to reporting.

5.5.4 Promoter induction

E. coli expressing MGA5S under the control of an IPTG-inducible promoter were grown in the presence and absence of inducer in continuous culture. A final IPTG concentration of 1 mM was used for induction. Cultures were sampled and incubated in a 96-well plate in a manner identical to that described in **Section 5.5.3**. When MGA5S fluorescence per unit absorbance at OD_{600nm} remained constant for 30 minutes, a portion of uninduced cells were exposed to inducer and fluorescence and absorbance measurements were allowed to continue.

5.5.5 Measuring RNA degradation

RNA degradation was measured using both reverse-transcriptase quantitative PCR and using MGA5S fluorescence.

E. coli expressing MGA5S were grown in continuous culture. 5 ml of culture was drawn and added to 50 μl of 50 mg/ml rifampicin in DMSO. The sample was briefly vortexed and subsequently incubated at 37°C and 250 rpm. 400 μl aliquots of cells in rifampicin were drawn and added to 800 μl Bacterial RNA Protect Reagent (Qiagen) at the following times following rifampicin exposure: 4, 8, 12, 16, 20, 30, 40, and 60 mins. Reverse-transcriptase, quantitative PCR proceeded as described in **Section 5.5.7**. Following experimental measurement, degradation was estimated by comparing the abundance of MGA5S and 16S rRNA prior to rifampicin exposure to samples obtained following rifampicin exposure.

Degradation was also estimated by MGA5S fluorescence. Cells were cultured, sampled, and placed in replicate of 12 in a 96-well plate as described in **Section 5.5.3**. When fluorescence per unit absorbance at OD_{600nm} remained constant for 30 minutes, half of the replicates were exposed to rifampicin to produce a final concentration of 500 μg/ml rifampicin. Simultaneously, an *E. coli* strain expressing the no-aptamer control was exposed to rifampicin to account for changes in fluorescence intensity due to a change in the optical properties of the medium following rifampicin exposure.

5.5.6 Parmeter estimation by differential evolution

The unknown parameters of our systems model, k_f, k_r, FL_{MAX}, and $[RNA]_{Total}$, were estimated through differential evolution, where optimization was directed by the rand/1/bin strategy. Control parameters for the routine were NP=30, CR=0.8, and F=0.8. Negative parameters estimates were not allowed to avoid unsolvable equations. Differential evolution is a heuristic search method, and parameter estimation was performed 100 times to obtain error estimates. Additionally, experimental fluorescence measurements for each biological sample were recorded using four technical replicates. The parameter estimation routine randomly sampled, with replacement, the measured value from a single replicate for each biological sample during each

round of parameter estimation. Python code that implements the described routine is supplied in Supplementary information.

5.5.7 Reverse-transcriptase quantitative PCR

Continuously growing cell cultures were sampled and diluted into warm M9 media to an OD_{600nm} of 0.2. 400 μl of cells were sampled and immediately exposed to Bacterial RNA Protect Reagent (Qiagen). Cells were briefly vortex and then centrifuged at 10,000xg for 10 minutes. Supernatant was decanted and cells were either immediately processed for RNA extraction or frozen at -80°C for later processing.

Cell lysis proceeded by exposing thawed cells to 5 mg/ml lysozyme in 1X TE (Sigma) for 10 minutes at room temperature and constant shaking. RNA extraction and purification was performed using Quick-RNA Mini Prep Kits (Zymo Research) according to protocol. Following extraction, total RNA was quantified on a Nanodrop 1000 using a mass extinction constant of 32.43 ng·cm/μl.

One-step RT-qPCR was performed using iTaq Universal SYBR Green One-Step Kit (BioRad) and a Rotor-Gene Q (Qiagen) according to protocol. 10 ng of total RNA from each sample, 300 nM forward primer, and 300 nM reverse primer were added to each reaction mixture. The forward and reverse primers for MGA5S are 5′ – GAGAGCCAGG TAACGAATGG – 3′ and 5′ – TGGCGATGTC CTACTCTCAC – 3′, respectively. Primers were designed and synthesis by Integrated DNA Technologies. Primers for a reference gene, 16S, were from literature and are 5′-CCGGATTGGA GTCTGCAACT – 3′ and 5′ – GTGGCATTCT GATCCACGAT TAC – 3′, for the forward and reverse, respectively. Fold-change in RNA abundance is reported after normalizing samples 16S and to a reference, TM8J23101, using the ΔΔCt method.

Chapter 6

CHARACTERIZING THE TRANSCRIPTIONAL STRENGTH OF PROMOTERS

6.1 Motivation

Work in previous chapters demonstrates that MGA5S can be used as a quantitative, live-cell measurement tool. Yet, a pressing question arises from that work: "How can MGA5S be applied to advance biological understanding and engineering?"

Characterizing the function of synthetic gene network components is critical to the realization of complex biological engineering applications [6]. Well-characterized genetic elements promise to improve *a priori* modeling of engineered biological systems [65], [82]–[84] and to facilitate the development of standard descriptions for the function of gene network components [85], [86]. However, a lack of tools to probe the internal state of cellular systems limits researchers' ability to characterize many components of synthetic biological systems.

The work in this chapter is motivated by a desire to provide a use-case for MGA5S as a tool that can address a challenging biological problem. We apply the aptamer to characterize the function of DNA regulatory elements in *E. coli*. Since MGA5S is a direct reporter of transcription, we use the aptamer to precisely characterize relative promoter activity in a much more direct manner than has been previously achieved.

6.2 Background

6.2.1 Current methods for promoter characterization

Gene expression in prokaryotes is, in part, controlled by the ability of promoter sequences to recruit RNA polymerase and initiate transcription [87], [88]. As a consequence, efforts to rationally modify gene expression in bacteria often involve altering these DNA regulatory sequences [89], [90]. In synthetic biology, the frequency with which promoter sequences are manipulated and the desire to achieve precise expression levels have led researchers to attempt to meaningfully describe the function of various promoter sequences.

In general, promoter characterization is performed by expressing a fluorescent protein under near identical conditions – only the promoter used to drive protein expression is changed. These techniques use a protein to report on RNA synthesis; therefore, many assumptions are made regarding transcription and translation processes. Previously, this strategy has been used to estimate the activity of DNA regulatory elements, and important findings have resulted from those studies. One report proposed a criteria for information that should be included in a "part characterization," and it described a unit of measure for promoter activity, polymerase per second (PoPS) [91]. A second paper highlighted the importance of minimizing interactions between the promoter and its surrounding DNA sequence to increase the robustness of measurement to changing genetic context [92]. A final work linked protein fluorescence to a mathematical model to estimate rates of synthesis [64]. Additionally, it showed that reporting relative activity, as opposed to absolute activity, facilitated consistent measurements between laboratories.

6.3 Results

6.3.1. Measuring relative promoter activity

Precisely characterizing the function of DNA regulatory elements improves the predictive power of computational models of synthetic gene networks. To demonstrate the utility of fluorescence-activating aptamers as a tool for better characterization of DNA components, we used MGA5S to measure the relative strength of promoter sequences in a small library.

We expanded our library of plasmids from Chapter 5 to estimate the strength of many promoters. As before, constructs were built by placing MGA5S under the control of varying promoters on, otherwise, genetically identical p15A plasmids. **Figure 6.1** is a plasmid map of the general composition of plasmids that were included in the library.

Promoters used to create the library were primarily derived from a set of constitutive promoters developed by the Anderson laboratory (parts.igem.org). However, promoters from the BioFAB (biofab.synberc.org) and Registry of Standard Biological Parts (parts.igem.org) were also included in the library to assess the function of DNA regulatory elements that are frequently used in synthetic biology applications. Ultimately, the library contained synthetic constructs that utilized an assortment of promoters that varied in total length (35 to 55 base pairs), in 5′ UTR length (0 or 1 bp), and in the nature of their regulation. A list of promoter sequences as well as their attributes are provided in **Table 6.1**.

Figure 6.1. SBOLv diagram of the architecture of plasmids used to characterize promoter strength.

Name	Sequence	Length	5' UTR	Regulation
J23100	TTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGC	35	0	Constitutive
J23101	TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGC	35	0	Constitutive
J23102	TTGACAGCTAGCTCAGTCCTAGGTACTGTGCTAGC	35	0	Constitutive
J23104	TTGACAGCTAGCTCAGTCCTAGGTATTGTGCTAGC	35	0	Constitutive
J23105	TTTACGGCTAGCTCAGTCCTAGGTACTATGCTAGC	35	0	Constitutive
J23106	TTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGC	35	0	Constitutive
J23107	TTTACGGCTAGCTCAGCCCTAGGTATTATGCTAGC	35	0	Constitutive
J23108	CTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC	35	0	Constitutive
J23110	TTTACGGCTAGCTCAGTCCTAGGTACAATGCTAGC	35	0	Constitutive
J23111	TTGACGGCTAGCTCAGTCCTAGGTATAGTGCTAGC	35	0	Constitutive
J23114	TTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGC	35	0	Constitutive
J23116	TTGACAGCTAGCTCAGTCCTAGGGACTATGCTAGC	35	0	Constitutive
J23117	TTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGC	35	0	Constitutive
J23118	TTGACGGCTAGCTCAGTCCTAGGTATTGTGCTAGC	35	0	Constitutive
R0040	TCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCAC	54	0	Inducible, aTc
R0011	AATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACA	55	1	Inducible, IPTG
FAB295	TTGCCTCTTAATCATCGGCTCGTATAATGTGTGGA	35	0	Constitutive
FAB346	TTGACAATTAATCATCCGGCTCGTAATGTTTGTGGA	36	1	Constitutive

Table 6.1. The sequence and attributes of promoters included in our library.

Growth conditions for cell cultures were strictly controlled prior to sampling and measurement to ensure consistency between measurements. Cells were grown in continuous culture with a media dilution rate of 0.347 hr^{-1}, which corresponds to a 2 hrs doubling time. Cells were grown under these conditions for at least 20 generations in order to allow them to reach physiological steady-state [74]. We assessed how well cultures were able to maintain a steady-state level of MGA5S expression by periodically sampling cultures, exposing cells to malachite green, and measuring fluorescence. **Figure 6.2** shows that MGA5S fluorescence remains stable following 20 generations of growth.

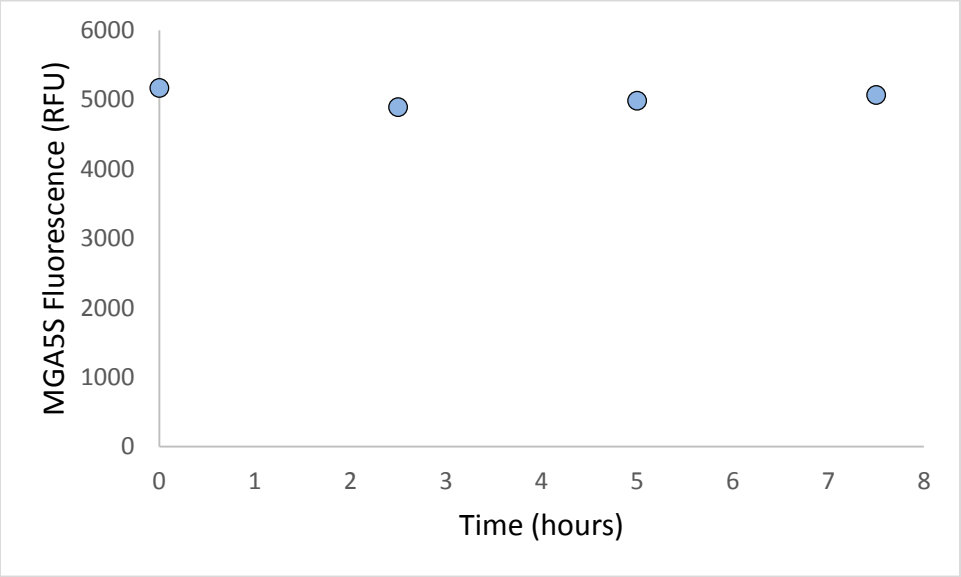


Figure 6.2. Reactors were sampled for 7.5 hours after cells were believed to be in steady-state growth. MGA5S fluorescence remains constant suggesting that expression is in steady-state.

Additionally, culture samples were diluted to the same cell density prior to exposing cells to malachite green and recording fluorescence. Since all strains experience identical experimental conditions, changes in MGA5S concentration, and relative fluorescence, can be attributed to differences in the rate of aptamer synthesis due to a particular promoter sequence.

Similar to previous studies [64], we report relative promoter units (RPUs) as the ratio of the rate of synthesis of each promoter relative to that of a reference promoter, BBa_J23101. **Figure 6.3** and **Table 6.2** show that our library is composed of promoters that exhibit relative promoter activities spanning a 47-fold range, from 0.05 ± 0.04 to 2.36 ± 0.18.

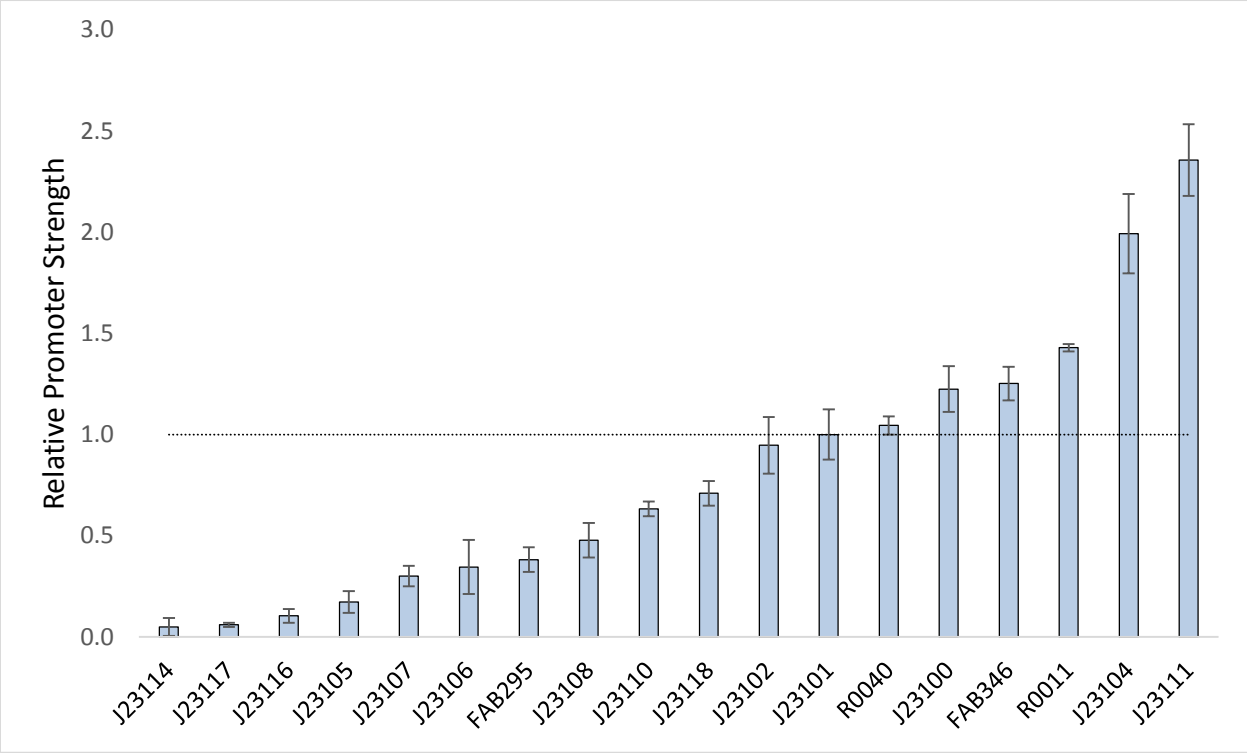


Figure 6.3. Relative promoter activity characterized by MGA5S.

	Relative Fluorescence	Std. Dev.
J23114	0.05	0.04
J23117	0.06	0.01
J23116	0.11	0.03
J23105	0.17	0.05
J23107	0.30	0.05
J23106	0.35	0.13
FAB295	0.38	0.06
J23108	0.48	0.09
J23110	0.63	0.04
J23118	0.71	0.06
J23102	0.95	0.14
J23101	1.00	0.12
R0040	1.05	0.05
J23100	1.23	0.11
FAB346	1.25	0.08
R0011	1.43	0.02
J23104	1.99	0.20
J23111	2.36	0.18

Table 6.2. Relative promoter activity characterized by MGA5S.

Prior to this work, promoter strength has only been inferred by using fluorescent proteins. We compared promoter strength measured by MGA5S with that measured by others using fluorescent proteins. The Registry of Standard Biological Parts reports fluorescence due to expression of RFP behind many of the promoter used in our library. The reported values were

normalized to the J23101 promoter, and compared with relative promoter activity measured using MGA5S.

Overall, we observed qualitative agreement between the predicted synthesis rates when using either MGA5S or RFP. However, **Figure 6.4** shows that we did observe substantial differences in the predicted strengths for three promoters: J23110, J23104, and J23111.

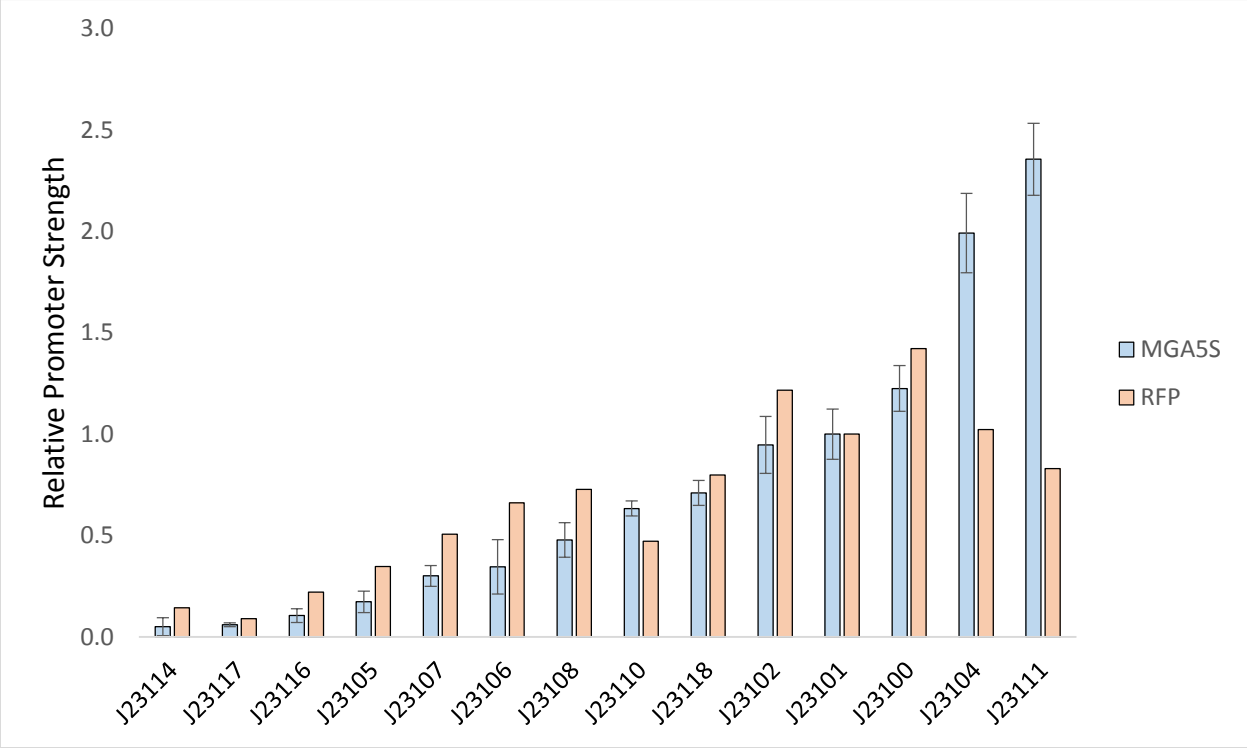


Figure 6.4. A comparison of relative promoter strength predicted by MGA5S and RFP fluorescence.

6.3.2. Changing experimental conditions

Recent studies have highlighted the importance of considering context when reporting quantitative measurements of biological systems [93]. We examined the robustness of our promoter strength assay to different culturing strategies.

Recognizing that many laboratories do not have access to continuous culturing equipment, we also characterized promoter strength from batch culture and observed whether there were significant differences between measurements made from continuous culture. Cells were grown in minimal media supplemented with malachite green and both fluorescence and absorbance at 600nm were periodically recorded. Fluorescence during exponential growth was normalized by cell density and used to describe relative promoter activity. **Figure 6.5** and **Table 6.3** show the predicted promoter strengths from cells measured in batch culture.

Overall, we found good agreement between the two culturing strategies. However, there were a few noteworthy differences. First, absolute fluorescence values from cells grown in batch culture were lower than the fluorescence produced from their counterparts grown in continuous culture. Second, we found that there was noticeably more noise between replicate samples when grown in batch culture. Third, continuous culture and batch culture produced qualitatively different results for the promoters J23104, J23108, and J23110. Finally, it was more difficult to resolve fluorescence of the weakest promoters in our library when compared to the no-aptamer control.

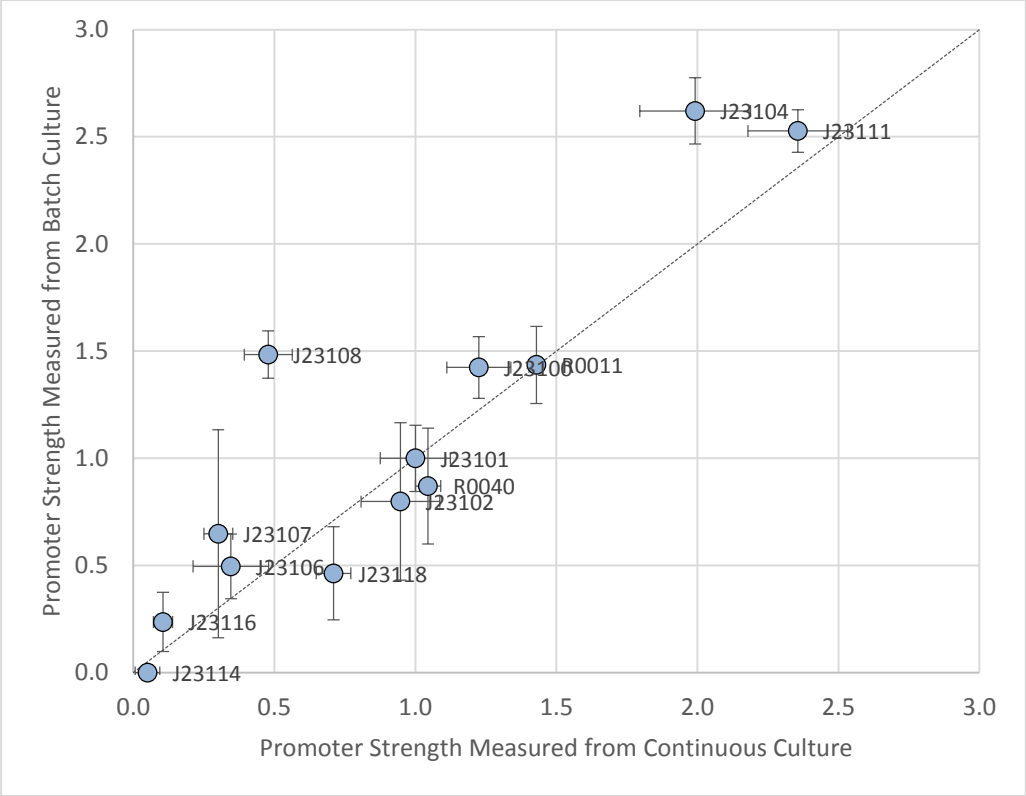


Figure 6.5. Comparison of promoter strength characterized from cells grown in continuous culture as opposed to those grown in batch culture.

	Blank Reduced Fluorescence		Relative Fluorescence	
Sample	**Mean**	**Std Dev**	**Mean**	**CV**
J23110	0	392	0.000	0.000
J23114	0	76	0.000	0.000
J23105	188	69	0.171	0.368
J23116	260	36	0.236	0.138
J23118	509	110	0.463	0.217
J23106	545	82	0.496	0.151

J23107	712	346	0.648	0.485
J23102	878	322	0.798	0.367
R0040	957	259	0.870	0.270
J23101	1100	170	1.000	0.154
J23100	1566	226	1.424	0.144
R0011	1579	283	1.436	0.179
J23108	1632	181	1.484	0.111
J23111	2779	277	2.527	0.100
J23104	2882	445	2.620	0.155

Table 6.3. Promoter strengths characterized from cells grown in batch culture.

Another aspect of context that we studied was cell growth. An *E. coli* strain that expressed MGA5S under the control of BBa_J23101 was grown in continuous culture at different dilution rates. **Figure 6.8** shows that, as expected, fluorescence decreases as the rate of cell growth increases.

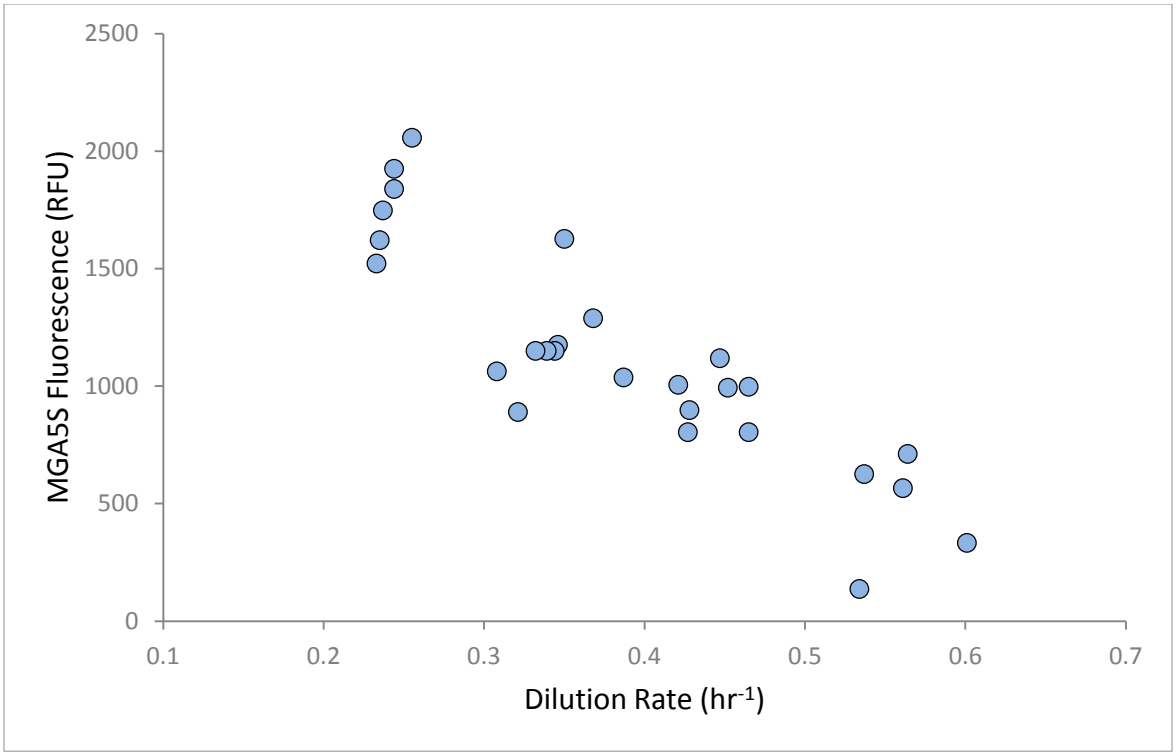


Figure 6.8. MGA5S abundance decreases linearly with dilution rate.

We observed how that cell growth was not consistent among all strains grown in batch culture. Notably, as shown in **Figure 6.9**, cells expressing MGA5S behind the promoter J23108 grew significantly slower than all other cultures. This slower growth rate leads to an overestimate in promoter activity when characterizing fluorescence once a specific cell density is achieved.

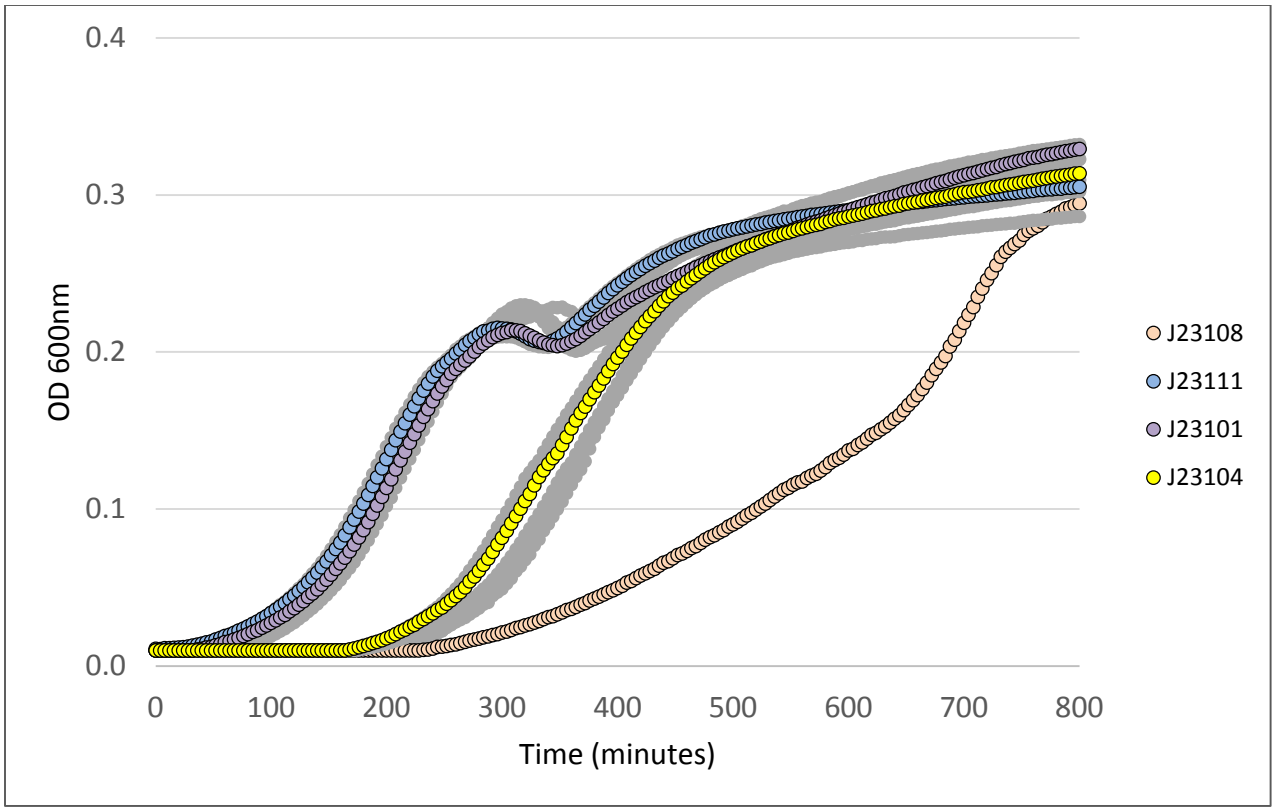


Figure 6.9. Cell growth, as measured by OD at 600nm, of *E. coli* strains expressing MGA5S behind various promoters. Each strain experiences different growth rates.

6.4 Discussion

Experimental results presented in this chapter present an application for MGA5S. Synthetic biology is a discipline that will strongly benefit from an improved ability to characterize cellular events such as the rate of RNA synthesis. In this chapter, we use our quantitative report of RNA expression to investigate the degree to which several DNA regulatory sequences promote gene expression.

In contrast to previous efforts that use fluorescent proteins as an indirect reporter of RNA synthesis, our strategy uses MGA5S, a direct product of RNA synthesis. An RNA-based reporter offers several advantages over an indirect reporter such as GFP. First, RNA experiences rapid

changes in synthesis and degradation under certain circumstances [94], [95]. These fast events are not observed when using protein reporters because a significant time delay exists between when mRNA levels fluctuate and when protein levels reflect that change. This delay is the consequence of additional gene expression steps such as post-transcriptional processing, translation, and protein maturation. Second, attempts to characterize promoter activity through the use of fluorescent proteins model a more expansive biological system. Doing so requires estimates for many more molecular components and processes: a challenge that is often circumvented by simply using parameter estimates gleaned from literature. This is not ideal since the experimental conditions that produced to parameter estimates described in literature may be entirely different than the ones used to characterize promoter strength. When using MGA5S, all model parameters were determined under the same conditions that promoter strength was characterized.

We observed qualitative agreement between promoter activity measurements made using MGA5S and those made by other groups using fluorescent proteins. However, we did identify a handful of promoter sequences for which MGA5S and fluorescent proteins measurements led to incoherent predictions. While these discrepancies were not investigated in depth, we hypothesize that they are the result of inadvertent interactions between the promoter sequence and 5' UTR [82].

It is important to consider the context in which biological systems are characterized. This concept has received increasing attention in recent years, as studies have demonstrated that DNA sequences can have unintentional interactions [92], [96] and identical synthetic gene networks can perform dramatically differently in different strains [97]. We considered various ways in which individuals might apply MGA5S to characterize promoter strength. Batch culturing strategies are used much more often than continuous culturing methods; therefore, it is likely that many attempts to investigate promoter activity will be performed in batch culture. As mentioned in earlier chapters, the extracellular environment is constantly changing in batch culture. As a consequence, it is important that researchers identify ways to make consistent measurements between samples. We attempted to maintain consistency by measuring the fluorescence of cell samples during log phase growth as they cross a predefined absorbance threshold – a technique

commonly used in batch culture measurements [98]. We find that batch culture and continuous culture produce results that generally agree; however, there were a few instances where measurements did not coincide.

Disagreement between batch and continuous culture measurements for the promoters J23104 and J23108 may be explained by their rates of growth. We initially inoculated media in each well using the same cell density. However, once inoculated growth is no longer strictly regulated and each strain reproduces at different rates. From absorbance measurements, we found that J23104 and J23108 experienced the slowest growth rates among all strains. As we show in **Figure 6.8**, reducing the rate of MGA5S dilution leads to an increase in fluorescence signal. Assuming that growth is the same between the reference strain, J23101, and slower growing strains like J23104 and J23108 should lead to an overestimation of promoter strength for the slow growers. Unsurprisingly, this was what we observed.

An explanation for the discrepancy between batch and continuous culture measurements for the promoter J23110 is not obvious. The promoter was characterized as a very weak promoter (~ 0 RPUs) in batch culture but was found to be approximately 0.6 RPUs in continuous culture. Reverse-transcriptase, quantitative PCR confirmed MGA5S abundance to be approximately that predicted by MGA5S fluorescence in continuous culture, and J23110 promoter strength was characterized in continuous culture numerous times (n = 6).

A final way in which we examined the effect of context on promoter activity measurements was in our choice of promoters. We use promoters with varying total lengths, 5′ UTR lengths, and the nature of regulation.

Taken together, we measure promoter activity with MGA5S and identify important considerations for improving characterization efforts. Contrasting predictions for promoter strength when using MGA5S as opposed to fluorescent proteins led us to speculate that certain promoters were not functioning independently of their neighboring genetic elements. Specifically, we speculate that the 5′ UTR was affecting the promoter's function. Additionally, we highlight the importance of comparing measurements between various strains when they are

experiencing similar contexts. Traditionally, gene expression studies choose an arbitrary cell density to make fluorescence measurements. However, we imagine that more robust normalization procedures, such as comparing fluorescence at an identical rate of growth, may lead to more accurate measurements of biological processes.

6.5 Methods

6.5.1. Measuring relative promoter activity

E. coli expressing MGA5S under the control of several promoters were grown in either continuous or batch culture according to the protocols outlined in **Sections 5.5.1** and **5.5.3**. Cell samples drawn from continuous culture were diluted to an OD_{600nm} of 0.2 prior to fluorescence measurement. The fluorescence of cells growing in batch culture were recorded once OD600nm became greater than 0.15. For each strain, fluorescence was measured at an excitation wavelength of 620 nm and emission wavelength of 655 nm and subtracted from the fluorescence measured from a no-aptamer control strain. Relative fluorescence was determined by dividing the fluorescence of each strain by the fluorescence from a reference strain, TM8J23101.

6.5.2. Growth-rate dependence of fluorescence

To examine the effect of cell growth rate on MGA5S fluorescence, cells were continuously cultured similar to the methods described in **Section 5.5.1**. Cells were grown in M9 media with the appropriate antibiotic at dilution rates ranging from 0.2 hr^{-1} to 0.6 hr^{-1}. Cell cultures were diluted to an OD_{600nm} of 0.15 prior to fluorescence measurement. Single-time point fluorescence measurements proceeded according to the methods outlined in **Section 5.5.2**.

6.5.3. Promoter library construction

A set of plasmids were constructed according to the methods outlined in Section 4.5.2. Each plasmid contained a distinct promoter that was obtained through parts.igem.org or synthesized from a sequence provided in literature.

Chapter 7

FLUORESCENCE DETECTION OF DIVERSE TRANSCRIPTS

7.1 Motivation

Fluorescent proteins have revolutionized the study of protein expression and dynamics in cells [99], [100]. Similarly, fluorescence-activating aptamers have the potential to transform the way RNA expression and dynamics are studied [101]. It can be argued that the utility of MGA5S and other fluorescence-activating aptamers as a tool for biomedical researcher hinges on its ability to enable the investigation of diverse mRNA sequences.

One particular strategy to broadly interrogate RNA expression in prokaryotes involves integrating these aptamers within operons: a concept analogous to tagging fluorescence proteins to endogenous proteins. However, it has been shown that placing a fluorescence-activating aptamer in either the 5'- or 3' untranslated region (UTR) of a gene of interest decreases fluorescent output of the aptamer [30]. Undesirable consequences of diminished fluorescence include a smaller range of detection and a reduced ability to resolve minor fluctuations in intracellular transcript concentrations.

It has been hypothesized that the aptamer's propensity to misfold into a structure incapable of producing fluorescence is increased through interactions with proximal nucleotides. Attempts to strengthen the fluorescence signal of aptamers linked to gene-coding sequences have involved reengineering the aptamer's sequence [2] or have focused on identifying optimal placement of the aptamer with respect to a particular gene [30]. Unfortunately, these strategies cannot guarantee that the aptamer will produce acceptable levels of fluorescence in diverse contexts, and repeatedly identifying new strategies for specific, non-ideal cases can be both time intensive and costly.

Therefore, the primary motivation for the work presented in this chapter is to develop and demonstrate a robust and generally-applicable method to tag diverse transcripts with fluorescence activating aptamers to monitor gene expression of transcripts in *E. coli*.

7.2 Background

7.2.1 RNA cleavage

Bacteria possess a variety of mechanisms to cleavage of RNA sequences following transcription [102]–[104]. From these naturally observed RNA cleavage systems, researchers have developed ways to apply *trans*-acting ribonucleases [105] and *cis*-acting ribozymes [106] towards cell engineering efforts in *E. coli*.

One ribonuclease, Csy4, is used throughout the work presented in this chapter. Csy4 is a part of prokaryotic CRISPR systems and cleaves RNA sequences upon recognition of the 28 nucleotide motif, 5′-GUUCACUGCC GUAUAGGCAG CUAAGAAA-3′ [107]. Following RNA cleavage, the first 20 nucleotides of the Csy4 recognition sequence remain with the upstream portion of the separated sequence, and the last 8 nucleotides remain with the downstream portion as illustrated in **Figure 7.1**. We refer to these residual nucleotides as the cleavage remnant. Additionally, Csy4 cleavage modifies the end chemistry of the 5′-end of the downstream sequence from 5′-PPP to 5′-OH [5]. This has been shown to increase the intracellular stability of the transcript due to a decrease in RppH catalyzed degradation [103].

Figure 7.1. Csy4 cleavage is initiated by a specific sequence.

Ribozymes are a special class of RNA molecules that exhibit catalytic activity, and numerous ribozymes have been discovered and characterized. In this chapter, we make use of two autocatalytic ribozymes, RiboJ [92], [106], [108] and LtsvJ [109]. RiboJ is a 75-nucleotide ribozyme that cleaves itself to produce a 7-nucleotide and a 68-nucleotide transcript as shown in **Figure 7.2**. LtsvJ is an 80-nucleotide ribozyme that cleaves itself to produce an 8-nucleotide and a 72-nucleotide sequence as shown in **Figure 7.3**. These shorter sequences remain linked to any nucleotides upstream and downstream of the cleavage site, and we refer to them throughout this chapter as cleavage remnants. Similar to Csy4 cleavage, the RNA end chemistry at the site of cleavage is changed from 5′-PPP to 5′-OH.

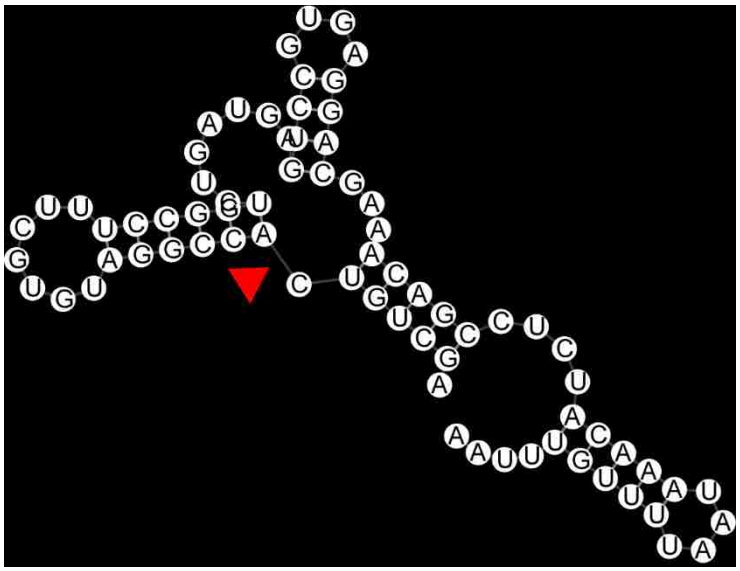


Figure 7.2. RiboJ cleavage occurs at a conserved catalytic site (red). Following cleavage, 7 nucleotides remain with the sequence directly upstream of the ribozyme.

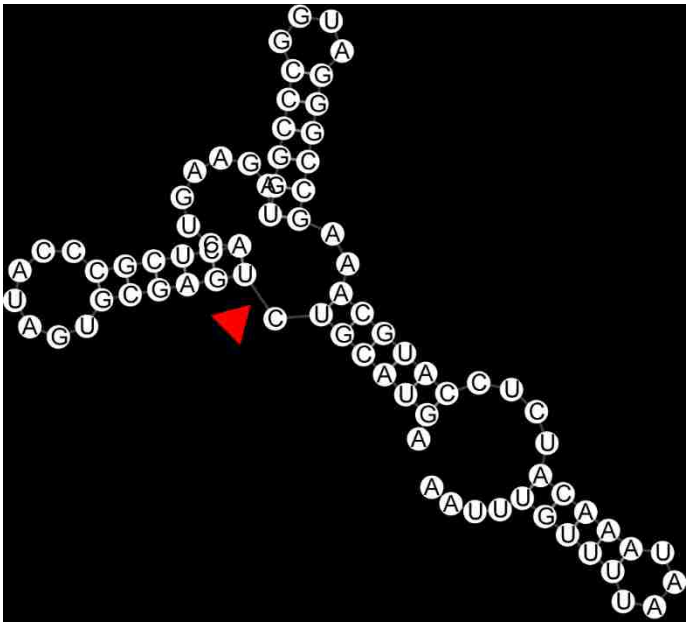


Figure 7.3. LtsvJ cleavage occurs at a conserved catalytic site (red). Following cleavage, 8 nucleotides remain with the sequence directly upstream of the ribozyme.

7.2.2 RNA sequence-structure-function relationship

RNA function is a consequence of its structure; the structure is a direct product of its sequence. This relationship is central to engineering strategies presented in the latter half of this chapter.

The strong dependence of function on structure and sequence is observed throughout nature and has been demonstrated with rationally designed biomolecules such as riboswitches [110] and aptazymes [111]. In recent years, engineering efforts have leveraged this sequence-function relationship to create novel RNA molecules that exhibit desirable functions [112], [113]; however, d*e novo* design of RNA sequences that enable a specific function is still very challenging. Subtle changes to a sequence – sometimes even a single nucleotide – can drastically alter the molecule's structure. Fortunately, computational tools and algorithms that excel at predicting RNA structure have been developed to facilitate *de novo* design [114]–[116].

7.3 Results

7.3.1 Dual-domain RNA

We began with a simple solution that tagged MGA5S to mRNA: MGA5S was placed either upstream or downstream of the GFP protein-coding sequence. Ideally, the MGA5S portion of the RNA sequence would not affect the protein coding region following transcription, and vice versa. We hypothesized that if there were undesirable effects on protein expression or a reduction in aptamer signal strength, base-pairing interactions between nucleotides encoding MGA5S and GFP would be the likely culprit. Accordingly, short sequences of nucleotides were used to buffer interactions between the two domains. Sequences expected to form flexible linkers as well as ones expected to form rigid hairpins were placed between the two domains to encourage functional isolation of MGA5S and GFP, as illustrated in **Figure 7.4**.

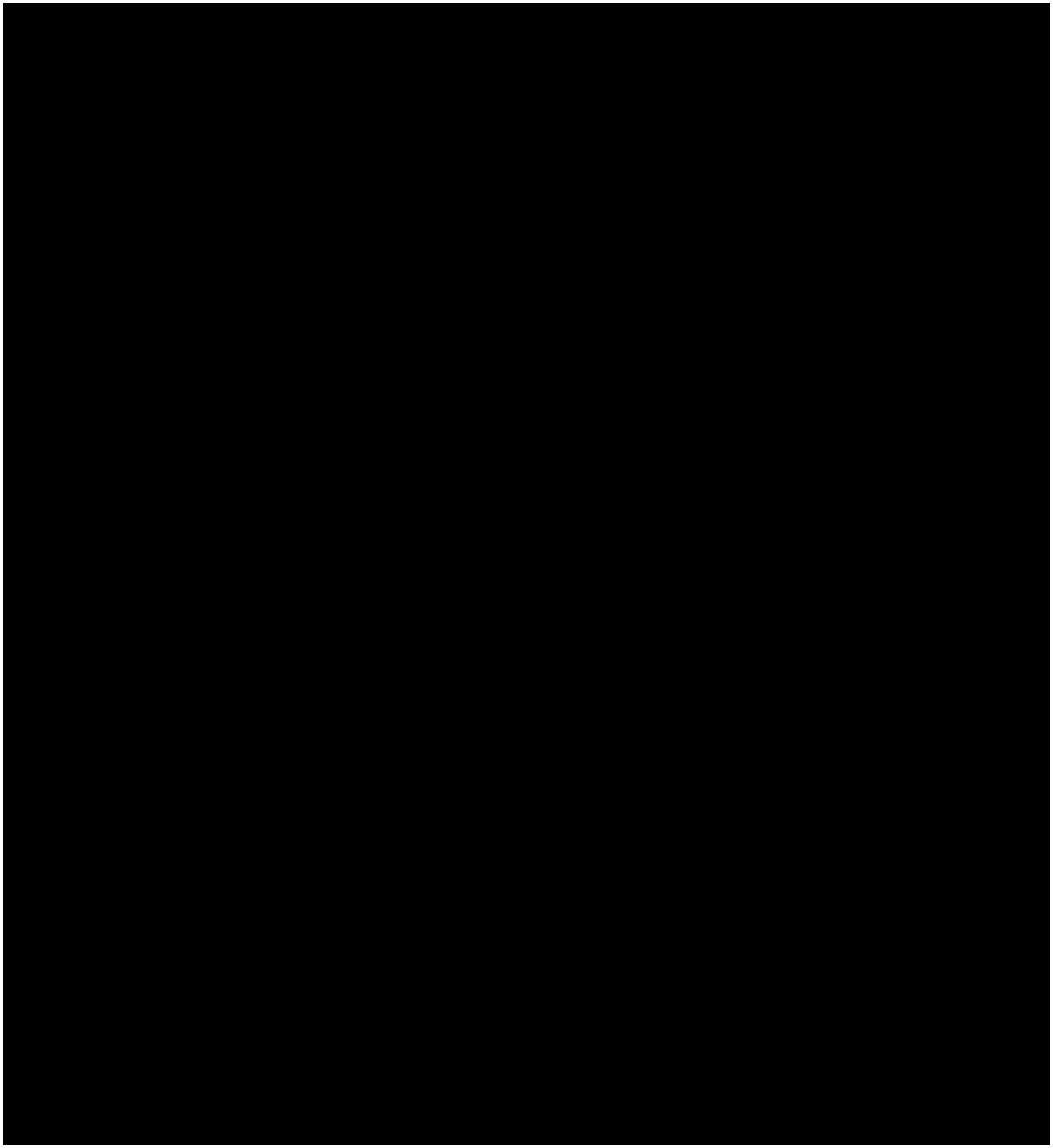


Figure 7.4. Sequence and predicted secondary structure of nucleotide linkers that were intended to buffer interactions between MGA5S and the GFP-coding sequence. (**A, B**) Flexible linker placed with MGA5S placed upstream and downstream of GFP. (**C, D)** *De novo* designed hairpin forming sequence. **(E, F**) Hairpin forming sequence from literature [48].

To evaluate the effectiveness of the various gene network designs, fluorescence was measured at the emission wavelengths for both MGA5S and GFP in cells expressing the dual-domain RNA constructs. Fluorescence output was compared to the fluorescence output of cells expressing only MGA5S or only GFP to gauge changes in signal intensity. **Figure 7.5** shows that when compared to cells expressing only one fluorophore, all but one dual-domain RNA designs resulted in a relatively weak fluorescence signal for both MGA5S and GFP. These results suggest that tagging a protein coding sequence with MGA5S negatively impacts not only aptamer fluorescence, but it can also undermine protein expression.

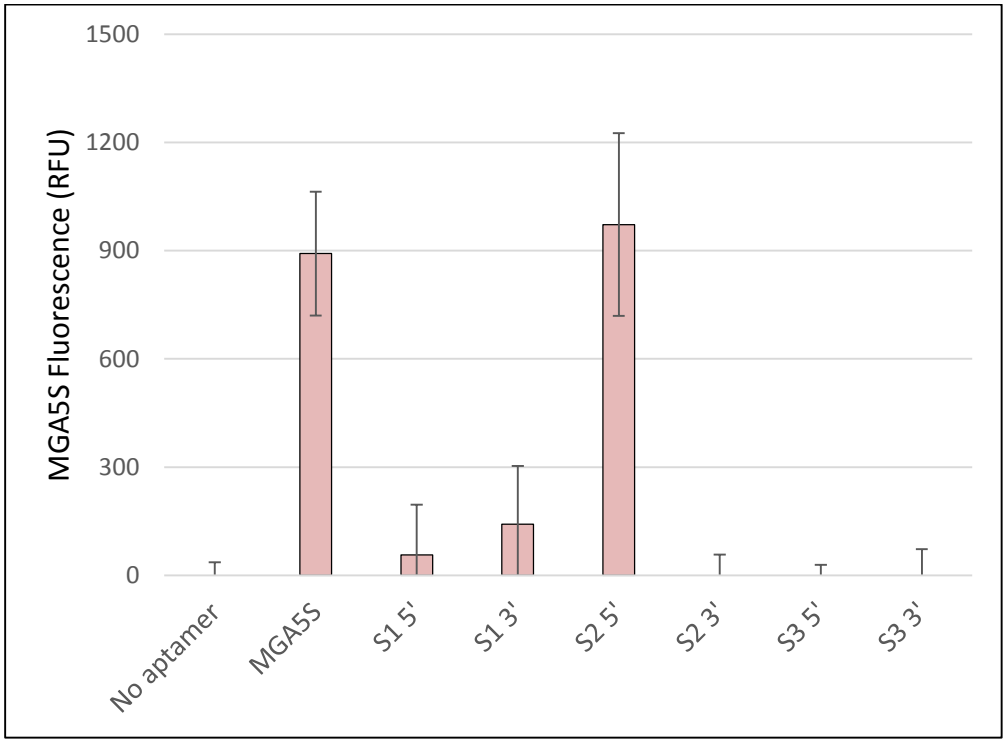


Figure 7.5: Inserting MGA5S into a GFP-coding operon eliminates fluorescence. The fluorescent output from cells expressing dual-domain transcripts was compared to that from cells expressing only MGA5S (MGA5S+) and only GFP (GFP+). For almost all dual-domain transcripts, both MGA5S and GFP fluorescence signals were markedly reduced.

Importantly, we observed an unfavorable attribute of the lone rigid linker that was able to preserve both GFP and MGA5S fluorescence: it caused a dramatic increase in GFP fluorescence, presumably due to a lack of transcript degradation. We reasoned that as long as MGA5S remained physically connected to a protein-coding sequence, it would affect its intracellular stability and obstructive it native function.

7.3.2 Trans-cleaved RNA

Given on our prior observations, we turned to strategies that would physically separate the two domains following transcription. An initial plasmid was assembled in which GFP- and MGA5S-coding domains were separated by a 28-nucleotide sequence recognized by the Csy4 ribonuclease. A plasmid map of this construct is provided in **Figure 7.6**. To assess the impact of post-transcriptional cleavage, fluorescence was observed in both the presence and absence of Csy4. Once again, we compared the fluorescence from cells that simultaneously expressed GFP and MGA5S to the fluorescence of cells expressing either only MGA5S or only GFP. **Figure 7.7** shows that in the absence of Csy4, both MGA5S and GFP fluorescence were negligible. However, in the case of Csy4 expression, GFP fluorescence returned to a level comparable to cells expressing only GFP. Unfortunately, MGA5S fluorescence remained negligible.

Figure 7.6. SBOL diagram of Csy4-cleaved RNA gene network design. MGA5S was placed downstream of a GFP coding sequence. Additionally, the GFP and MGA5S domains were separated by the 28 nucleotide Csy4 recognition sequence.

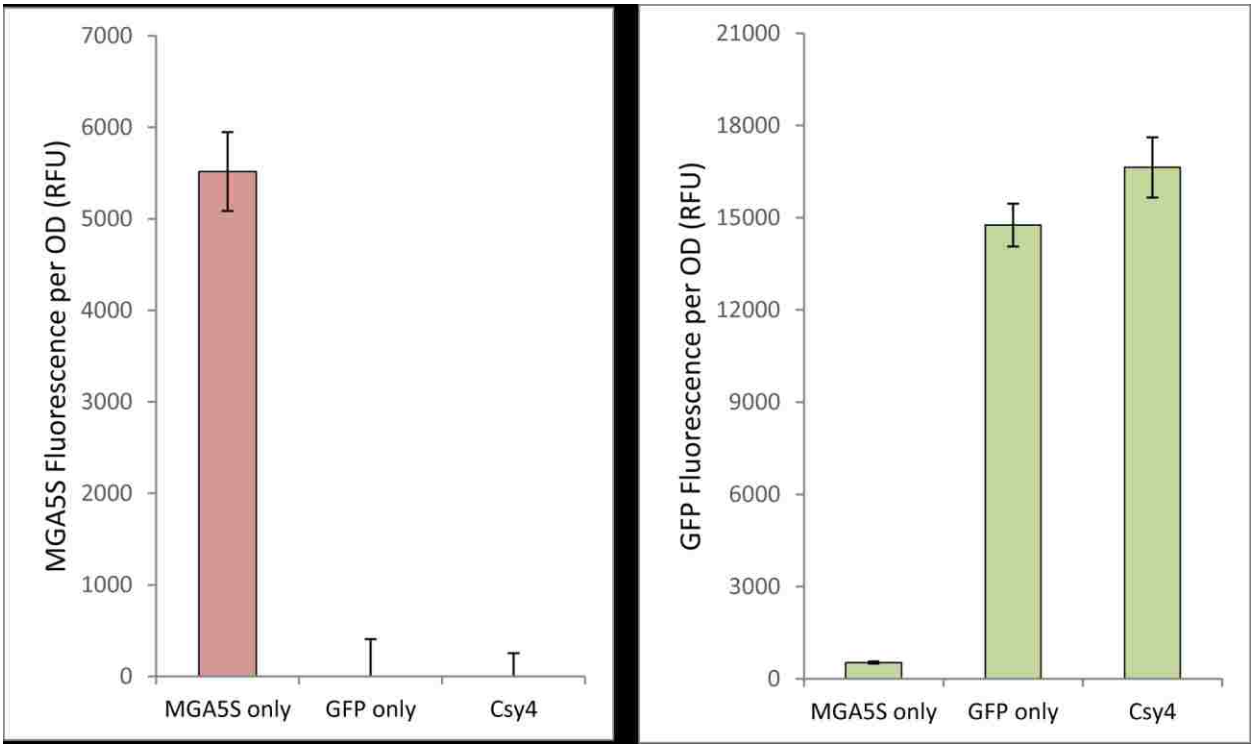


Figure 7.7: Csy4 cleavage enhances GFP signal, but not MGA5S signal. The fluorescent output from cells expressing Csy4-cleaved transcripts was compared to that from cells expressing only MGA5S (MGA5S+) and only GFP (GFP+). In combination with Csy4 expression, GFP signal is recovered.

The Vienna RNA secondary structure prediction package [117] was used to examine the structure of the MGA5S domain after separation by Csy4. Figure 7.8 shows the predicted secondary structure of the MGA5S domain following Csy4 cleavage.

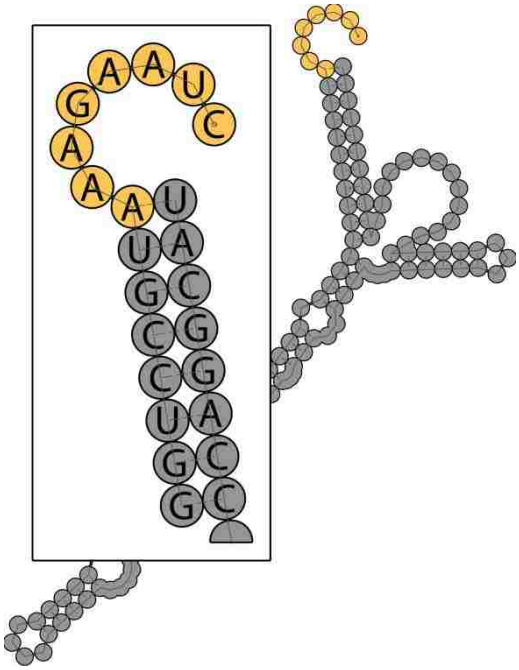


Figure 7.8. Diagram of predicted secondary structure of MGA5S domain following Csy4 cleavage. MGA5S (grey) is expected to retain its native conformation. Residual nucleotides from the Csy4 recognition sequence remain with MGA5S after cleavage (yellow). All but one of these nucleotides are predicted to remain unpaired.

The residual nucleotides following cleavage appear to have a strong tendency to remain unpaired. We surmised that this group of unpaired nucleotides might decrease intracellular stability of MGA5S by increasing the rate of transcript degradation or by increasing its tendency to misfold.

To test the effects of these residual nucleotides on aptamer fluorescence, a synthetic gene network was created where the residual nucleotides were placed at the 5′ end of MGA5S. Importantly, this construct contained neither a GFP coding region nor the Csy4 recognition site to remove potentially cofounding interactions. The measurements summarized in **Figure 7.9** demonstrate that the residual nucleotides imparted on MGA5S following Csy4 cleavage strongly reduce fluorescence output. These results reinforced our hypothesis that the cleavage remnant interferes with the aptamer's ability to fluorescence.

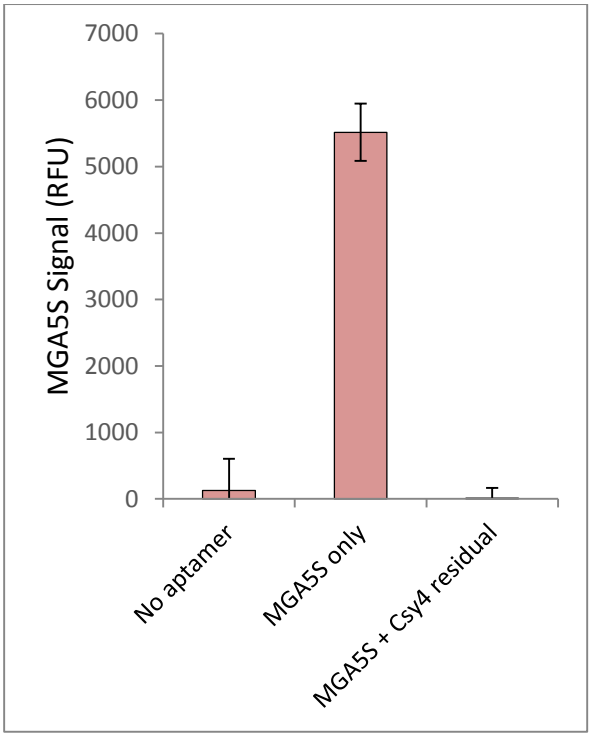


Figure 7.9. Adding the residual nucleotides upstream of MGA5S diminish fluorescence.

Based on the observation that the cleavage remnant practically eliminates MGA5S fluorescence, we inquired further as to why. Presumably, the unpaired nature of the residual nucleotides was significant to this finding. Towards the goal of promoting base-pairing with the nucleotides belonging to the cleavage remnant, a set of gene circuits were constructed that added nucleotides to encourage base-pairing interactions with the residual nucleotides following cleavage.

The sequence and predicted secondary structures for these design are show in **Figure 7.10**. The added sequences were complementary to the residual nucleotides, and were placed directly upstream or downstream of MGA5S. We refer to the additional nucleotides that base-pair with residual nucleotides as the stem-promoting sequence.

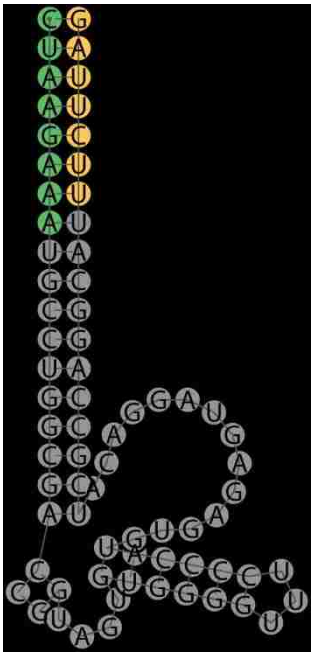


Figure 7.10. Diagram of predicted secondary structures of various modified MGA5S domains following Csy4 cleavage.

Figure 7.11 shows that all gene circuit designs that included a stem-promoting sequence restored MGA5S fluorescence output to levels comparable to native MGA5S. This finding further supported our notion that the residual, unpaired nucleotides following cleavage were the underlying cause of an observed decrease in MGA5S fluorescence.

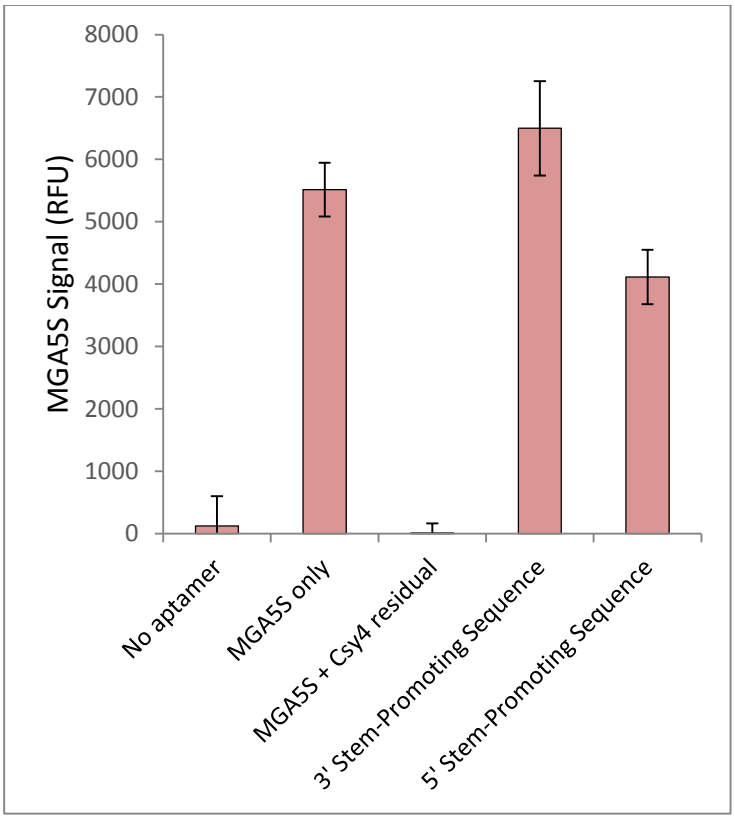


Figure 7.11. Incorporating nucleotides that base-pair with the cleavage remnant into gene network designs restores MGA5S fluorescence.

Finally, a synthetic gene network encoding GFP, the Csy4 recognition site, MGA5S, and the stem promoting sequence was assembled. An SBOL diagram of this circuit is shown in **Figure 7.12**. Significant levels of fluorescence at both GFP and MGA5S emission wavelengths were detected as shown in the results of **Figure 7.13**. Notably, very high levels of Csy4 expression resulted in fluorescence that was significantly stronger than the fluorescence measured from positive controls. This intense signal is likely due to a decrease in the rate of RNA degradation following ribonuclease cleavage [5], and similar findings have been reported in other works that have employed Csy4 cleavage [105].

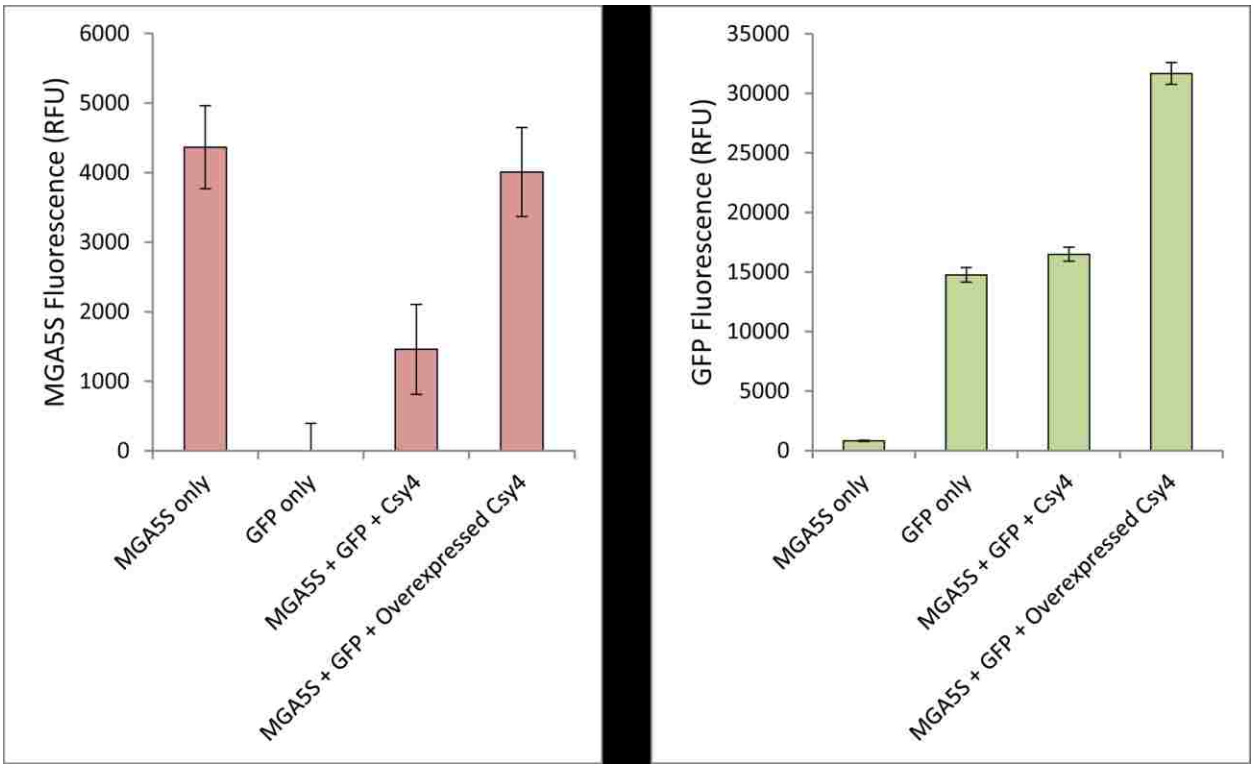


Figure 7.13. Simultaneous observation of MGA5S and Csy4 fluorescence.

7.3.3 Cis-cleavage strategies

In addition to *trans*-cleavage strategies, we explored separation of transcripts containing MGA5S and GFP domains through the use of the self-cleaving ribozymes RiboJ and LtsvJ [106]. **Figure 7.14** depicts the general composition of synthetic gene circuits expected to separate MGA5S and GFP domains via *cis*-cleavage. In contrast to the relative orientations of MGA5S and GFP on plasmids encoding Csy4-cleavage, MGA5S was placed upstream, and GFP downstream, of the ribozyme sequences. This was done so that fewer residual nucleotides would remain with MGA5S, as opposed to with GFP, following cleavage. Furthermore, the unpaired nucleotides that remain with MGA5S following cleavage were accounted for during plasmid design by placing nucleotides upstream of MGA5S that were complementary to the expected remnant.

Figure 7.14. SBOL diagram of synthetic gene networks that utilize *cis*-cleavage to separate aptamer and protein-coding domains.

Figure 7.15 shows fluorescence measurements of *cis*-cleaved transcripts. Surprisingly, no GFP was observed from the synthetic gene network that used RiboJ and MGA5S fluorescence was not detected from cultures that used either RiboJ or LtsvJ.

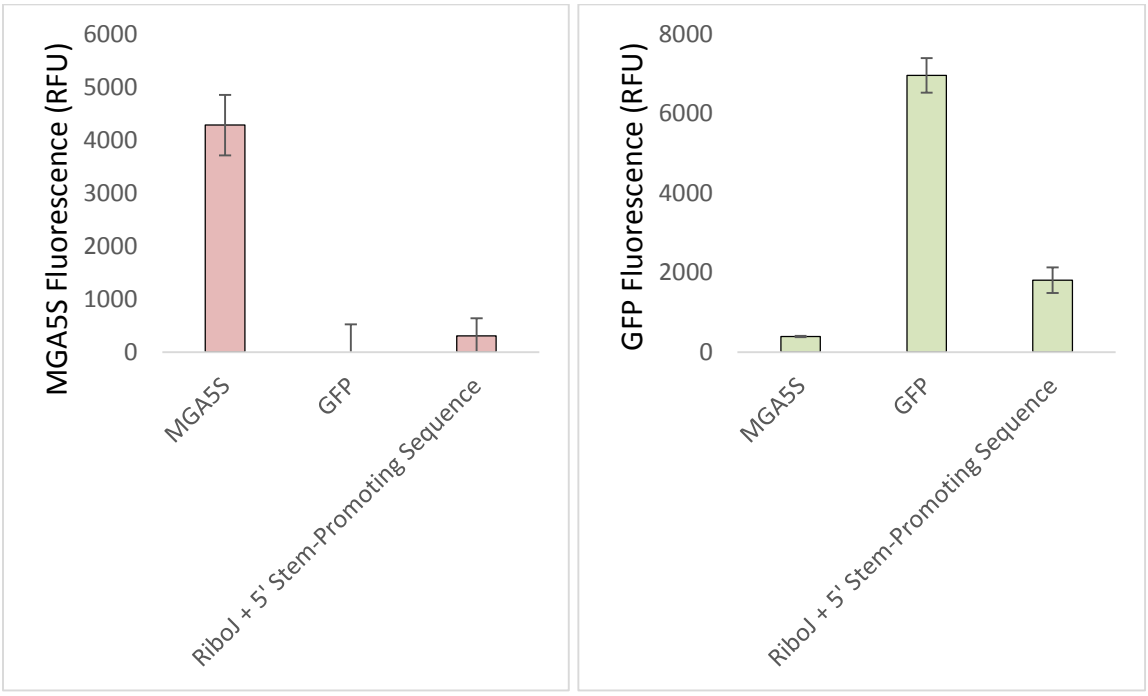


Figure 7.15. Adding a stem-promoting sequence to hybridize with residual nucleotides following RiboJ cleavage ablates fluorescence.

The computationally predicted secondary structures of the transcripts prior to cis-cleavage are displayed in **Figure 7.16**. Close examination of these structures revealed that the stem-promoting sequence likely interferes with the ability of both ribozymes to properly fold into a conformation capable of autocatalysis. According to the folding algorithm, the nucleotides that were intended to base-pair with residual nucleotides following cleavage form a stem with part of the ribozyme sequence and hinder the formation of the conserved catalytic core [106]. Therefore, we sought to identify a short sequence of nucleotides that would preferentially base-pair with residual nucleotides after cleavage occurs, but remain unpaired prior to ribozyme cleavage.

An experimental screen was devised in which eight to ten random nucleotides were placed upstream of MGA5S as shown in **Figure 7.16**. Cells expressing our randomized gene networks were screened, first, for GFP and, subsequently, for MGA5S fluorescence. After three high-throughput screens of a total of 285 transformant colonies, no cells were identified that produced significant levels of both MGA5S and GFP fluorescence. The results of these screens are aggregated in **Figure 7.17**.

Figure 7.16. SBOL diagram of plasmids that contain random stem-promoting sequences.

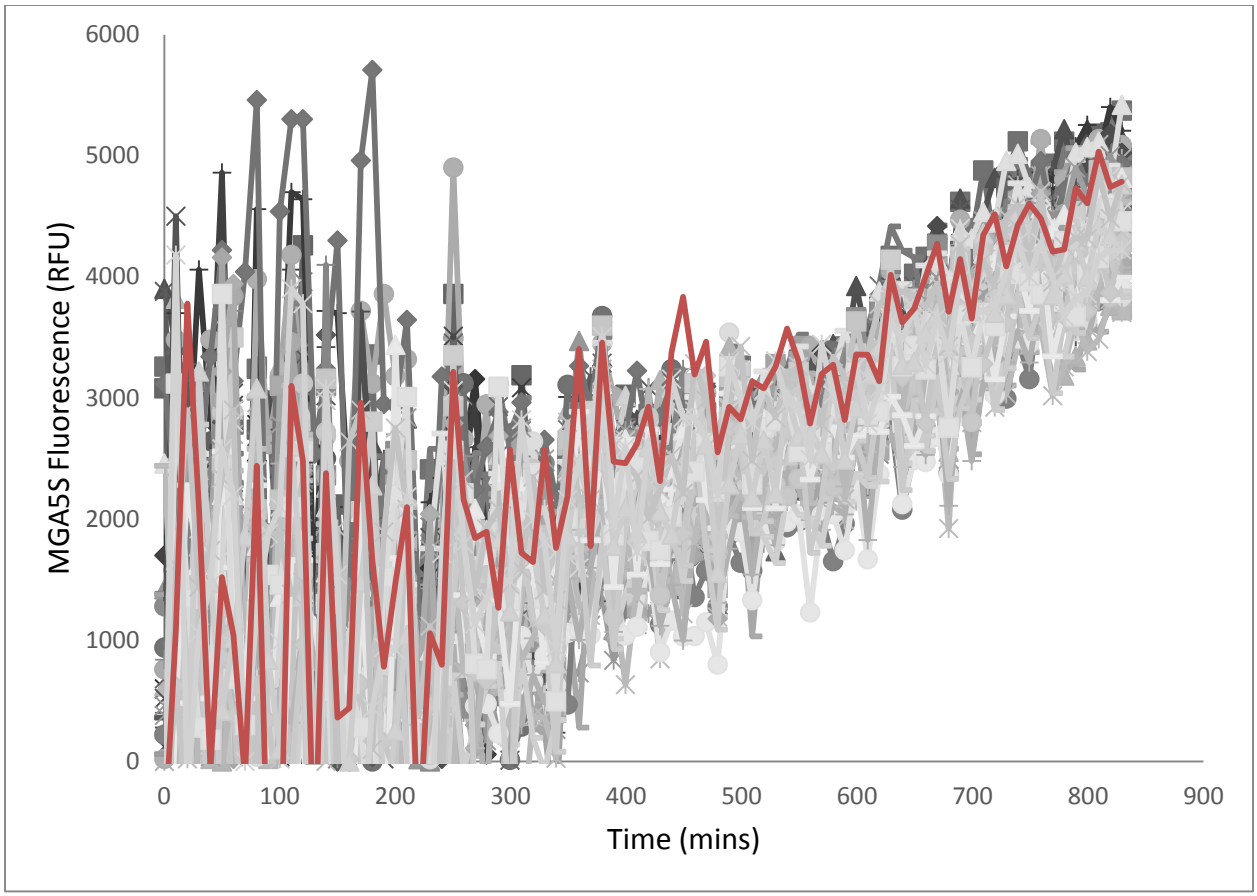


Figure 7.17. Multiple experimental screens of random sequences failed to produce a viable stem-promoting sequence.

In theory, there are 1.4 million unique combinations for an RNA sequence that is 10 nucleotides in length. Given the large number of potential sequences that would need to be experimentally screened and our relatively low-throughput ability to screen each sequence, we turned to computational methods. A custom Python script that implemented the Vienna RNAfold API [117] was used to screen all possible sequence variations. The code for this script is supplied in **Appendix D**. The program considered the predicted secondary structure of both the pre-cleaved and post-cleaved sequences for each variant. From the pool of sequences, ones that were predicted to allow ribozymes to properly fold prior to cleavage but sequester remnant nucleotides following cleavage were recorded and ranked. Without any precedent on which to

base our scoring strategy, sequences were rated according to numerous criteria and the top candidates within each grouping were selected.

By optimizing the code to make use of multiple CPUs, compute times were drastically reduced. Additionally, this increase in computational efficiency allowed us to screen many more designs. For example, we assessed variable sequences that were entirely upstream of MGA5S, entirely downstream of MGA5S, or both upstream and downstream of MGA5S as shown in **Figure 7.18**.

Figure 7.18. The orientation of random sequences, with respect to MGA5S and ribozymes, were varied during the computational screen.

In total, over 9.8 million sequences were screened and the highest ranking candidates were used to guide experimental design during a second iteration of plasmid construction. From the 42 plasmids that were selected for experimental evaluation, two were found to exhibit significant fluorescence at both MGA5S and GFP emission wavelengths as shown in **Figure 7.19**. **Table 7.1** shows that both of these sequences ranked highly when scores accounted for base pairing probabilities of the entire pre-cleavage and post-cleavage structures as well as the base pairing probabilities of the remnant nucleotides. Additionally, both of the successful designs placed a 10-nucleotide stem promoting sequence at the 3′ end of MGA5S.

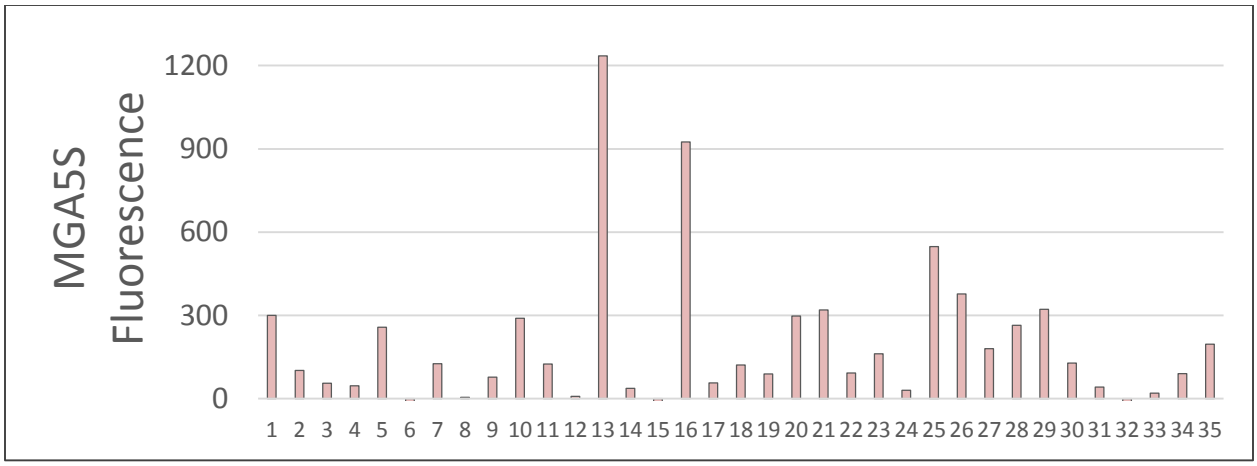


Figure 7.19. Two strains (#13 and #16) that utilized RiboJ cleavage to link MGA5S and a GFP-coding sequence were found to have significant levels of fluorescence.

Sequence	Position	Pre-Cleavage	Post-Cleavage	Remnant	Composite
{MGA5S}-AUAGUUAAGG	3′	0.492	0.577	0.741	1.673
{MGA5S}-GCACGCUAGG	3′	0.495	0.574	0.675	1.486
{MGA5S}-CGAUAGAAG	3′	0.499	0.577	0.659	1.450
{MGA5S}-AAUAGCAG	3′	0.502	0.579	0.643	1.414
{MGA5S}-AUAGCAG	3′	0.519	0.583	0.513	1.136
AUAGCA-*{MGA5S}*-G	Mixed	0.494	0.558	0.496	1.054
AAUGGUA-*{MGA5S}*-G	Mixed	0.493	0.554	0.489	1.043
ACAUUG-*{MGA5S}*-C	Mixed	0.493	0.553	0.478	1.031
ACAUUUAUC-*{MGA5S}*-U	Mixed	0.498	0.557	0.472	1.029
CCUAAUGGU-*{MGA5S}*	5′	0.484	0.550	0.472	1.022
CCUAGUUGGA-*{MGA5S}*	5′	0.493	0.549	0.467	1..016

Table 7.1. Nucleotide sequence and scores for computationally determined stem-promoting sequence candidates.

The predicted secondary structure of MGA5S and the ribozyme prior to and following cis-cleavage is shown in **Figure 7.20** and **7.21**.

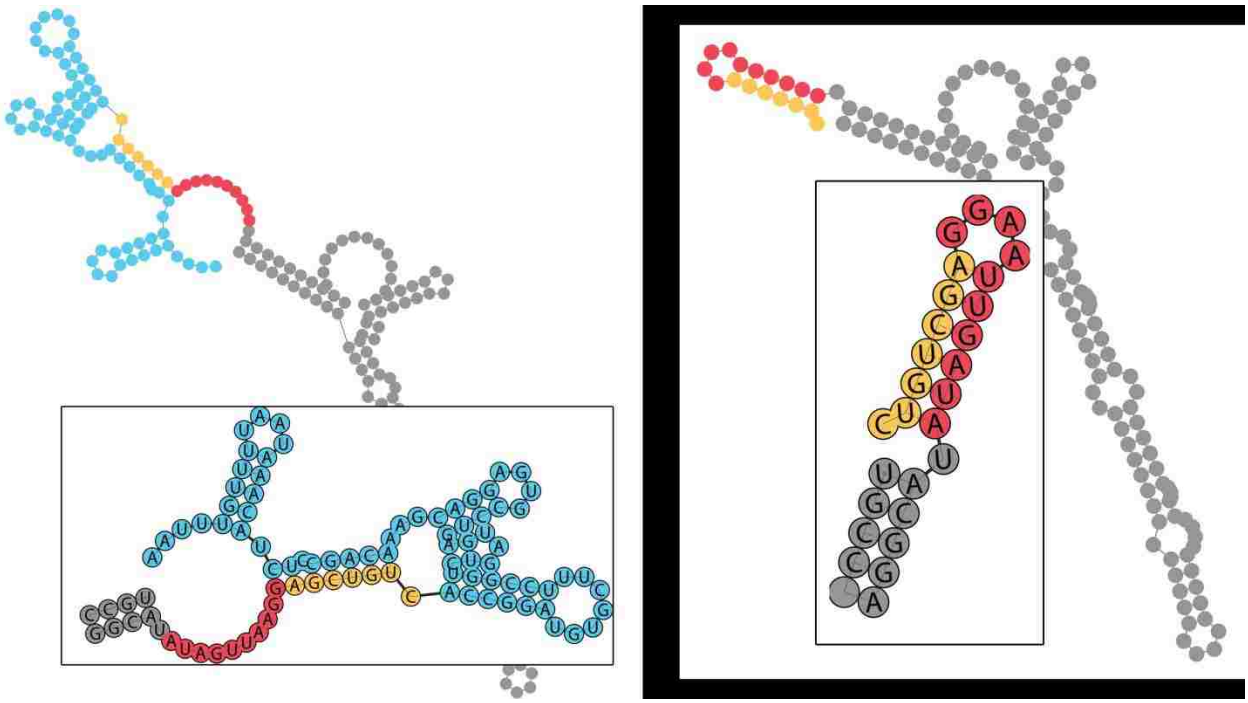


Figure 7.20 and 7.21. Following cleavage, 7 residual nucleotides of RiboJ (yellow) remain with MGA5S and the rest (blue) remain with GFP. A stem-promoter sequence (red) is placed between MGA5S and RiboJ to encourage base-pairing interactions with residual nucleotides following cleavage.

7.4 Discussion

Recent efforts have used fluorescence-activating aptamers to interrogate mRNA levels during gene expression [30]. These RNA-based fluorophores offer plenty of advantages over more traditional transcription measurement methods like quantitative PCR and MS2-GFP. A few noteworthy benefits are that the aptamers can be directly encoded into the RNA sequence, and that they are non-destructive and non-obstructive. However, others have reported fluorescence signal attenuation when coupling the aptamer with other sequences [2]. This attenuation has been, in part, attributed to aptamer misfolding, and these results should not be surprising given detailed studies on the ability of an RNA sequence to maintain a target structure when surrounded by additional nucleotides. In certain cases, *ad hoc* revisions to aptamer placement were made to ensure that the aptamers produced a reliable fluorescence signal. However, we

applied the reported strategy [30] to a previously untested fluorescence-activating aptamer, MGA5S, with little success.

We found that by using the strategy outlined, MGA5S fluorescence was not just diminished, but ablated. Moreover, we found that in certain instances, placing MGA5S upstream of a coding sequence not only affected aptamer fluorescence but also affected protein expression; we observed no GFP fluorescence. This suggests that tagging a protein coding sequence with MGA5S not only negatively impacts aptamer fluorescence, but it also has the potential to weaken protein expression. While we did eventually identify a short nucleotide sequence that preserves fluorescence of both MGA5S and GFP, there is no guarantee that our strategy will work with novel aptamers in the future.

These findings highlights the need to move away from *ad hoc* tagging strategies in favor of more robust and broadly-applicable methods to link diverse transcripts with fluorescence-activating aptamers. Consequently, the majority of the data presented in this chapter chronicles our attempts to create a robust strategy to link RNA sequences with MGA5S.

We imagined that since slight modifications to an RNA sequence can yield drastic changes to its structure, it would be nearly impossible to find a sequence to link MGA5S to every possible protein coding sequence in a way that discourages interactions between the two domains. Therefore, the focus for engineering efforts quickly turned towards integrating post-transcriptional cleavage mechanisms into the design. However, it became readily apparent that the sequence-structure relationship of post-transcriptional separated RNA molecules would need to be considered.

We used both *trans*-acting ribonucleases and *cis*-cleaving ribozymes, and identified several design considerations while tailoring these cleavage systems for aptamer tagging. Foremost, separation of the transcript leaves residual nucleotides on both the aptamer and protein-coding domains; these residual nucleotides have a propensity to remain unpaired. While these residual nucleotides do not necessarily affect protein expression, we found both strategies require that additional nucleotides are placed near the aptamer to encourage base-pairing with

the remaining nucleotides. Furthermore, we found that for cleavage catalyzed by either Csy4 or ribozymes, many unique sequences could accomplish this task.

Overall, simultaneous observation of GFP and MGA5S by way of Csy4 cleavage was easier to achieve. A major advantage of this strategy is that the sequence used to connect the aptamer and mRNA is quite small when compared to the length of ribozyme sequences. Additionaly, Csy4 seems to cleave MGA5S-tagged sequences with greater efficiency, as we ultimately observed a fluorescence signal just as strong as a positive control that only expressed MGA5S. However, it is noted that post-transcriptional cleavage modifies the end chemistry at the site of cleavage, which in turn affects degradation. Although a change in aptamer stability may be acceptable, an ideal strategy would not drastically affect the degradation rate of the transcript being detected. Future research might focus strategies to preserve the native degradation rate of an mRNA sequence.

Cis-cleavage presents a more elegant solution for tagging mRNA sequences. In contrast to Csy4 cleavage, no additional proteins are required and researchers interested in modeling the mechanism are required to account for fewer molecular interactions. But, we found it very difficult to identify a *cis*-cleaving ribozyme that produced reliable fluorescence at both GFP and MGA5S emission wavelengths. This challenge arises from strict sequence and structure requirements for the stem-promoting sequence; the sequence must be neither too willing to base-pair prior to cleavage nor too unwilling to base-pair following cleavage.

A computational approach was used to screen stem promoting sequences, and the top candidates were selected for experimental testing. Despite many encouraging computational predictions, very few stem-promoting sequences actually produced a detectable fluorescence signal for MGA5S. It should be noted, however, that all predictions did produce GFP fluorescence. This indicates that, at the very least, ribozyme was restored due to all candidates identified with our computational approach. It is possible that improved techniques that experimentally investigate the structure of RNA sequences [118]–[120] could assist future efforts to separate transcripts with self-cleaving ribozymes, and this could be particularly useful as more fluorescence-activating aptamers are developed.

An alternative cleavage strategy that was not explored in this chapter involves the use of *trans*-cleaving ribozymes [121], [122]. Similar to *cis*-acting ribozymes, they would still be synthesized and degraded on a time scale similar to other RNA molecules, which is not true for protein ribonucleases such as Csy4, and could still allow for purely RNA-based synthetic gene network designs. However, unlike *cis*-cleavage strategies, considerably less effort may be required to identify a stem promoting sequence.

7.5 Methods

7.5.1. Plasmid construction

The sequence for Csy4, the Csy4 recognition site, RiboJ, and LtsvJ were described in literature. Physical DNA for Csy4 was obtained through PCR amplification of the gene sequence from *Pseudomonas aeruginosa* genomic DNA. Physical DNA for the Csy4 recognition sequence, RiboJ, and LtsvJ were obtained through overlap PCR synthesis. Integration of DNA components into a pSB3K3 vector proceeded analogous to the methods described in **Section 4.5.2**.

7.5.2 Fluorescence measurement

Fluorescence of cell cultures was measured using a Tecan M200 Pro following the protocols outlined in **Sections 5.5.2** and **5.5.3**. However, in addition to measuring the fluorescence of MGA5S at an excitation wavelength of 620 nm and emission wavelength of 655 nm, GFP fluorescence was measured at an excitation wavelength of 485 nm and emission wavelength of 520 nm.

7.5.3. RNA secondary structure prediction

RNA secondary structure was determined using the Vienna RNA secondary structure package [115], [117] web servers and standalone packages. Secondary structures were predicted at 37°C using default options and parameter values, unless otherwise stated. The predicted centroid secondary structure was assumed to better reflect the physical secondary structure in comparison to the predicted minimum free energy structure, and was used to inform rational modifications.

7.5.4. Experimentally screening linker sequences.

Linker sequences were randomized through incorporation of random oligonucleotides (Integrated DNA Technology) into plasmids. Specifically, 8 to 10 were placed within a DNA sequence that is complimentary to the DNA sequence of synthetic gene networks. Plasmid construction proceeded in a manner as previously described in **Section 4.5.2**.

Transformant colonies were screened first for GFP fluorescence using a DR-88M blue light transilluminator (Clare Chemical Research). Transformant colonies with green fluorescence were inoculated into 1 ml of M9 media with appropriate antibiotics in a deep, 96-well plate and incubated overnight at 37°C and 250 rpm. Cells were diluted to an OD_{600nm} of 0.1 in 1 ml of fresh M9 media supplemented with 5 μM malachite green. 200 μl aliquots of diluted culture were placed in a 96-well plate and incubated in at Tecan M200 Pro at 37°C with 2.5mm linear shaking. MGA5S fluorescence was recorded at an excitation wavelength of 620 nm and emission wavelength of 655 nm. GFP fluorescence was recorded at an excitation wavelength of 485 nm and emission wavelength of 520 nm. Absorbance was recorded at 600 nm. Fluorescence of transformant colonies was compared to the fluorescence of the no aptamer control to assess fluorescence enhancement.

7.5.5. Computationally screening linker sequences.

The Vienna RNA Secondary Structure Package was built and compiled in a LINUX environment with a Python wrapper as described in the package's documentation (http://www.tbi.univie.ac.at/RNA/). A Python routine was written to predict the secondary structure of MGA5S containing various surrounding sequences. The surrounding sequences included the 28-nt Csy4 recognition sequence, RiboJ sequence, LtsvJ sequence. Additionally, the Python routine placed a variable 8 to 10 nucleotide sequence either upstream or downstream of the cleavage site to promote base-pairing following cleavage. All possible combinations of nucleotides for the 8 to 10 nucleotide sequences were considered. The code for this routine is described in **Appendix E**.

The calculated free energy of the predicted secondary structures, and the number of residual nucleotides that were predicted to form base pairs were used as primary criteria for ranking candidate sequences. The top 40 designs were selected and examined experimentally in a manner similar to the methods described in **Section 7.5.4**.

Chapter 8

CONCLUSION

8.1 Summary

RNA dynamics are key to cell function. Synthesis, degradation, localization, and binding are examples of RNA actions that impact phenotype in both prokaryotes and eukaryotes. In recent years, several methods have been developed to improve our ability to study the native function of RNA molecules. Certain strategies excel at quantifying hundreds to thousands of RNA sequences simultaneously and are amenable to high-throughput studies. Some boast excellent sensitivity: they can detect the presence of a specific sequence with single-molecule resolution. Others excel at measuring RNA dynamics with outstanding spatial and temporal resolution. Each technique has distinct advantages over alternatives, and these advantages stipulate its use. However, no existing method adequately addresses the challenge of quantitatively measuring RNA dynamics in living cells.

This dissertation presents a novel solution that employs fluorescence-activating aptamers to overcome this technical challenge. Fluorescence-activating aptamers are genetically-encoded molecules that produce a distinct light signal when exposed to a specific dye. Researchers do not need to destroy cells to detect the presence of fluorescence-activating aptamers, and the strength of their signal is proportional to the amount of molecules present.

We thoroughly investigate the fluorescence properties of one particular aptamer within *E. coli*, and we present several conclusions:

(1) **Heterologous expression and fluorescence detection of malachite green aptamer in *E. coli* is feasible.** Following engineering efforts to stabilize its expression, malachite green aptamer produces a strong, specific, and sensitive signal indicating its presence.

(2) **Fluorescence intensity can be used to quantitatively predict *in vivo* malachite green aptamer concentration.** By integrating mathematical models of transcription and experimental data, we find that it is possible to estimate absolute aptamer concentration within reason.

(3) **Malachite green aptamer can be applied towards biological investigation and engineering.** Latter chapters demonstrate this concept through two proof-of-concept applications: (a) the aptamer is used to characterize DNA regulatory elements in synthetic gene networks, and (b) malachite green aptamer is appended to a protein-coding sequence to simultaneously measure RNA and protein expression.

8.2 Future directions

The work described in this thesis can be expanded upon in many ways. First, the fluorescence-activating can be applied in diverse research settings that were not the primary concern of this work. Second, we identify shortcomings of malachite green aptamer which reveal remaining challenges in understanding how to use fluorescence-activating aptamers as effective tools for RNA expression analysis.

8.2.1 Improving malachite green aptamer

We achieve stable malachite green aptamer expression by incorporating it within an rRNA scaffold. This scaffold imparts stability and protects malachite green aptamer from rapid turnover in *E. coli*. In Chapters 4 and 5, we show that MGA5S persists for generations and the best way to reduce aptamer levels is through dilution due to cell growth. While we found the rRNA scaffold necessary for fluorescence measurement, we recognize that it prevents malachite

green aptamer from being used in RNA degradation studies. It is not hard to imagine that future work could explore alternative stabilization strategies for the malachite green aptamer. We tried a few alternative strategies with little success; however, others strategies exist in literature [48], [51]. Ideally, we envision a collection of stability elements will be discovered than allow researchers to precisely tune the degradation rate of malachite green aptamer. Studies related to topics like transcriptional bursting [94] and noise control [123], [124] would benefit from aptamer degradation rates that are only slow enough to ensure fluorescence can be detected. Other studies may want the rate of dilution and degradation to be equivalent. Finally, it is worth mentioning that RNA is typically more stable in eukaryotes; therefore, engineering strategies involving scaffolds that mimic rRNA may not be necessary when expressing malachite green aptamer in that context.

Another way to potentially improve malachite green aptamer would be to attempt to increase its brightness, and we see two major strategies that can accomplish this goal. First, these aptamer were identified by *in vitro* selection [125], [126] from a pool of many candidate sequences. It is possible that by performing an entirely new selection procedure, one might identify an RNA sequence that results in a stronger fluorescence signal when the aptamer and dye form a complex. Second, the sequence can be refined through mutagenesis studies to investigate and improve thermal stability in a manner analogous to improvements to the Spinach aptamer [2]. Since the aptamer was original selected using room temperature experiments, we would expect that thermal stability for cellular environments would benefit from mutational analyses. However, we recognize that a more stable sequence at higher temperatures does not guarantee an increase in brightness.

8.2.2 Improving malachite green

Malachite green is not suitable for many biological studies. For example, malachite green is an antifungal agent and should not be used for expression studies in yeast. Additionally,

malachite green does not readily cross mammalian cell members. Issues such as cell permeability have been addressed by modifying the dye [38]. Interestingly, these modifications also reduced toxicity in yeast and enabled fluorescence measurement without inhibiting cell growth. These findings are encouraging because they suggest that not only can biologists improve properties of the malachite green aptamer at the genetic level, but chemists can improve properties of the fluorescence-activating aptamer by altering the molecular structure of its cognate dye. This was done for the Spinach aptamer [27] as well as other dye-aptamer combinations [127], [128].

8.2.3. Novel aptamers and dyes

Green fluorescence protein was the first fluorescent protein to be discovered and cloned [129]–[131]. Since then, GFP has been refined for enhanced brightness [132] and faster maturation [133]. Even more important, a plethora of fluorescent proteins spanning the visible light spectrum have been discovered or engineered [134]. We anticipate a similar expansion in the number of fluorescence-activating aptameres.

A major result of this dissertation is that it establishes a framework through which any fluorescence-activating aptamer can be used in a quantitative manner. Aside from malachite green, Spinach aptamer [27] and RNA Mango [26] have been developed and used within cells. In particular, Mango has been shown to strongly bind to its cognate dye with a dissociation constant of 3.2 nM its cognate dye. For comparison, malachite green aptamer binds with 117 nM affinity and Spinach binds with 300 – 540 nM affinity. This is significant because Mango and future aptamers with a similar binding affinity might enable more sensitivity detection of transcripts with fluorescence – even to very low (< 10) copies per cells. For malachite green aptamer, the fluorescence detection limit is likely in the range of 10^2 copies per cell due to its relatively high dissociation constant.

In addition to novel aptamers and dyes with enhanced fluorescence properties, we envision that an expansive set of fluorescence-activating aptamers will allow for multiplex

studies. Assuming that their fluorescence spectra do not overlap, multiple aptamers can be simultaneously employed to observe the expression dynamics of multiple

8.2.4. RNA tags

In Chapter 7 we explore various strategies for tagging diverse transcripts with malachite green aptamer. We identified and overcame a variety of design challenges; yet, we did not identify an ideal strategy for tagging mRNA sequences. Certain methods resulted in direct interactions between the aptamer and protein-coding domains. Other methods insulated the two domains but increased stability of the protein-coding transcript leading to increased protein production. In other words, all attempts were obstructive to native function. Future work might search for alternative ways to tag mRNA sequences with fluorescence-activating aptamers.

We hypothesize that physical separation of the aptamer and protein-coding domains still represents a viable strategy in bacteria. However, it will like require heterologous expression of proteins that either (1) restore 5′- and 3′-end-chemistry or (2) specifically target cleaved transcripts for degradation at a rate similar to native RNases.

For eukaryotic organisms, even more post-transcriptional modification techniques exist. For example, the aptamer can be embedded into an intron of an RNA sequence. In this scenario, it may be determined that engineering strategies to insulate interactions between the two domains is not necessary. Additionally, if the aptamer imparts stability on the intron sequence, it may not necessarily negatively affect the stability of the protein-coding sequence.

8.2.5. Applications

There have been several reports of RNA-based synthetic gene networks [83], [135]. We anticipate that fluorescence-activating aptamers will eventually be used as quantitative and semi-

quantitative readouts for these types of synthetic systems. As proof of principle, we explore the use of malachite green aptamer and Spinach aptamer as fluorescent reporters in an RNA-based toggle switch (*data not show in this dissertation*). The development of the system was based on research that developed RNA regulators of transcription [83], [136]. We combined the results of that work with our Csy4-based mRNA tagging strategy to try to visualize bistability. We believe that this endeavor represents a feasible proof-of-concept synthetic gene network design in which fluorescence-activating aptamers can be immediately applied in a quantitative manner to predict network behavior prior to construction.

Another application for malachite green could involve quantitative studies of RNA expression in situations where existing techniques like RNA-Seq and RT-qPCR cannot be easily used. Situations that involve single-cell measurements such as fluorescence microscopy, flow cytometry, and microfluidic technologies. In these cases, fluorescence intensity can still be used as a quantitative readout of RNA abundance once the equipment has been calibrated to determine the theoretical maximum fluorescence. Dynamic and quantitative single-cell measurements of RNA expression allow for investigation of many interesting questions pertaining to how noise and variability in gene expression affects phenotype.

8.2.6. Standardization of biological parts

As a synthetic biologist, I was motivated to pursue this line of research, in large part, because the field of synthetic biology lacks precise ways to characterize synthetic gene network elements. Developing standard descriptions of biological components is a tenet of synthetic biology. Great progress has been made towards that goal in the last decade with respect to describing the structure and composition of biological networks [85]. However, there is a dearth in standard ways to describe the function of components in biological networks.

As others before me have argued, advancements in synthetic biology will require scientists and engineers to develop better tools to characterize biological networks [6], [137], [138];

and, I agree. Without strong tools to interrogate the operation of synthetic gene networks and precisely characterize their function, there is little hope of developing useful standards.

In this work, we show how malachite green aptamer can be used to characterize DNA promoter sequences in a more accurate and precise manner than existing techniques. Furthermore, we report promoter activity in terms of relative units because it has been shown to provide a more robust measurement when comparing activity across different equipment or between laboratories. However, the quantitative nature of malachite green aptamer could easily allow researchers to translate promoter activity into a more intuitive metric like absolute rates of synthesis that can be described using SI units that describe amount or concentration per unit time (ie. μM/s or ng/min). In addition to characterizing promoter activity, fluorescence-activating aptamers can be used to characterize properties like terminator efficiency, ribosomal binding site strength, and how gene length affects transcription rates.

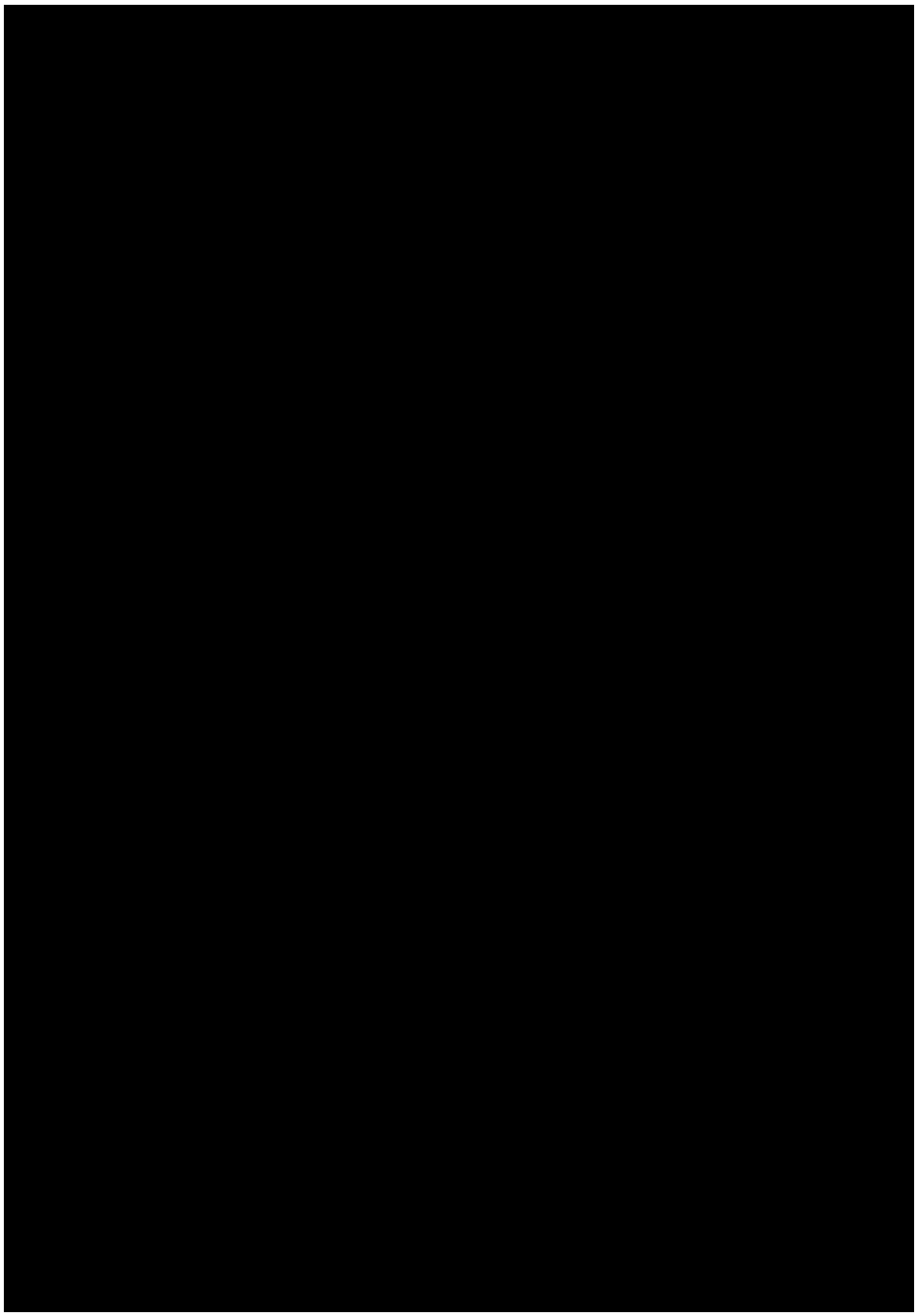
Appendix A

LIST OF ABBREIVIATED TERMS

Amp	Ampicillin
Cb	Carbenicillin
cDNA	Complementary DNA
dsDNA	Double-stranded DNA
GFP	Green fluorescence protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
LB	Luria-Bertani medium
M9CA	Minimal M9 salt medium supplemented with 0.2% casamino acids
MGA5S	Malachite green aptamer integrated into a degenerate 5S rRNA sequence
MG1655	*E. coli* K-12 strain
RBS	Ribosome binding site
RFU	Relative fluorescence unit
RPU	Relative promoter unit
RTqPCR	Reverse-transcriptase, quantitative polymerase chain reaction

Appendix B

LIST OF SBOL VISUAL ICONS

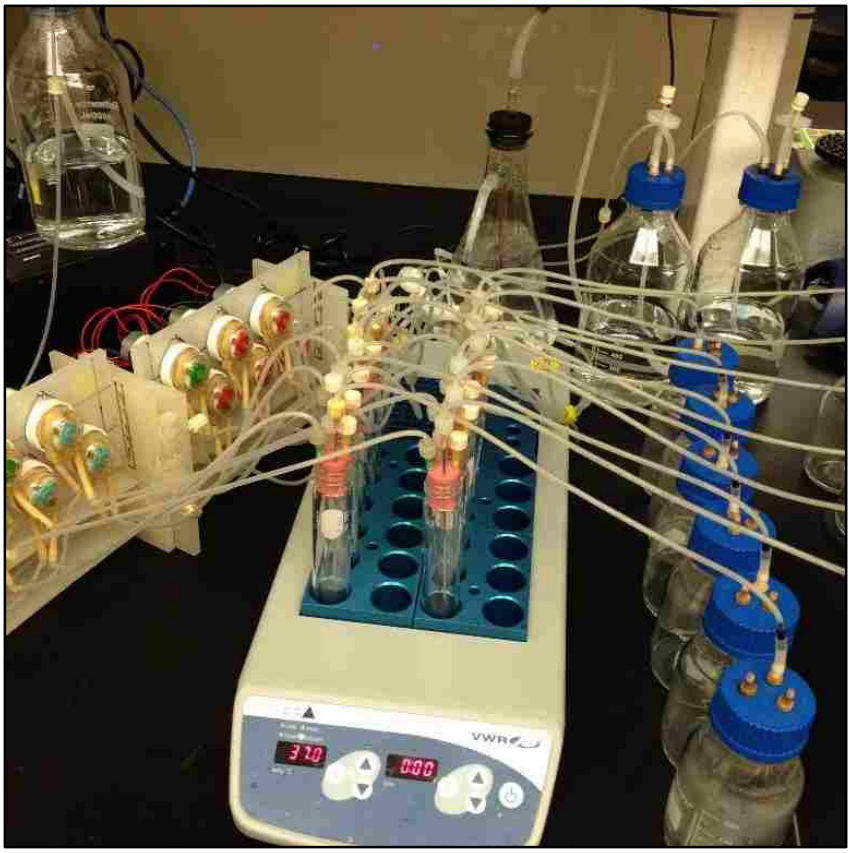


Appendix C

CHEMOSTAT ASSEMBLY AND OPERATION

C.1 Overview

All electronic circuit diagrams, dimensional drawings, and code is freely available at www.github.com/uwsysbio under the "Chemoduino" folder. This appendix provides a bill of materials, describes how to assemble the chemostat array that was used for continuous culturing throughout this dissertation, and provides an example of how to use the software to operate the chemostat system.



C.2 Assembly instructions

C.2.1. Building media reservoirs.

Using a ¼" drill bit, create 3 holes in the caps for the Kimax 1 L glass storage bottles. Partially tap holes using a ¼"-28 tap making sure not to tap the entire way through the cap. Place the ¼" panel mount connectors (PMS230-1) into the partially tapped holes and, using a wrench, screw the plastic connector completely through. This should create a very tight seal. Use Tygon tubing, straight connectors, and reduction couplers as necessary to achieve desired routing. One connector should be connected to a 0.4 μm PTFE filter and used to allow sterile air to enter the reservoir. A second connector should be capped with a male lock ring plug (LP4-1) and can serve as an injection port. The third connector should be used to carry media out of the sterile reservoir.

C.2.2. Building effluent chambers.

Using a ¼" drill bit, create 2 holes in the caps for the Kimax 250 ml glass storage bottles. Tap the caps in a manner similar to that described in **Section C.2.1**. One panel mount connector should be connected to the effluent line of a chemostat culturing chamber. Use Tygon tubing, straight connectors, and reduction couplers as necessary to achieve desired routing.

C.2.3. Building equipment stand

Mill a ¼" sheet of polypropylene according to the dimensions specified in the dimensional drawings provided on GitHub (https://github.com/uwsysbio). Completely tap ¼" holes using a ¼"-28 tap and insert female panel mount to barb connectors (FTLLB210-J1A) within the appropriate holes. Completely tap 10-32" holes and screw milled polypropylene sheets together

as described on the dimension drawings. Finally, Mill 1-1/8" holes into a flat polypropylene sheet that will be used to mount the peristaltic pumps. A dimensional drawing is provided on GitHub.

C.2.4. Culturing chamber

The culturing chamber uses four needles to deliver and remove air and fluid from the culturing chamber. The air agitator needle, the needle connected to the effluent line, and the needle connected to the media inlet are identical to those described in previous works [74]. A fourth needle is used to allow for repeated inoculation or reagent delivery. This line should be capped with a male luer plug.

C.3 Bill of materials

Description	**Vendor**	**Item #**	**Quantity**
1/16" ID x 1/8" OD x 1/32" Wall Tygon® Sanitary Silicone Tubing	US Plastics	57287	100 ft
1/8" ID x 1/4" OD x 1/16" Wall PharMed® Tubing	US Plastics	57318	10 ft
Peristaltic Pump	Welco	WPM1-P3BB-BP	12
1/4" Polypropylene Sheet	TAP Plastics	591	6 sq ft.
Female Luer Thread Style Panel Mount 1/4-28 UNF to 200 Series Barb, 1/16" (1.6 mm) ID Tubing, Natural Kynar PVDF	Value Plastics	FTLLB210-J1A	100
Elbow Reduction Tube Fitting with 200 Series Barbs, 1/8" (3.2 mm) and 1/16" (1.6 mm) ID Tubing, Natural Kynar PVDF	Value Plastics	L230/210-J1A	100

Male Luer Integral Lock Ring to 200 Series Barb, 1/16" (1.6 mm) ID Tubing, Natural Kynar PVDF	Value Plastics	MTLL210-J1A	100
Straight Through Reduction Tube Fitting with 200 Series Barbs, 1/8" (2.4 mm) and 1/16" (1.6 mm) ID Tubing	Value Plastics	N230/210-J1A	100
Straight Through Reduction Tube Fitting with 200 Series Barbs, 3/32" (2.4 mm) and 1/16" (1.6 mm) ID Tubing	Value Plastics	N220/210-J1A	100
Y Tube Fitting with 200 Series Barbs, 1/16" (1.6 mm) ID Tubing	Value Plastics	Y210-J1A	100
Male Luer Integral Lock Ring Plug	Value Plastics	LP4-1	100
Panel Mount Connector 1/4-28 UNF to 200 Series Barbs, 1/8" (3.2 mm) ID Tubing, White Nylon	Value Plastics	PMS230-1	100
Kimble 250 ml glass media storage bottle	Cole Parmer	EW-34523-02	12
Kimble 1 L glass media storage bottle	Cole Parmer	EW-34523-06	4
18 gauge x 1.5" hypodermic needles	Air Tite		12
20 gauge x 1.5" hypodermic needles	Air Tite		12
8-40 Screws	Home Depot		50
8-40 hex nuts	Home Depot		50
10-32 Screws	Home Depot		50
10-32 hex nuts	Home Depot		50
Silent Air Pumps *	Aquarium Guys.com	212422	2

PTFE filters, 0.45 u, for air filtration*	Cole Parmer	HV-02915-22	24
PYREX® 55mL Screw Cap Culture Tubes*	Corning Life Sciences	9825-25	12
medium tubing clamps*	VWR	63022-405	1 pack
Air-tite Regular hypodermic white hub needle, 16G, 5 in. length for effluent line*	Fisher	14-817-105	1 pack
Long Length Spinal Needles, 18G 6" Pink hub for air line *	VWR	BD408360	1 pack

* These items were taken directly from the Bill of Materials provided by Aaron Miller of the Dunham lab (http://dunham.gs.washington.edu/AppendixA.xls).

C.4 Software documentation

All software is made freely available through GitHub (https://github.com/uwsysbio). In this section, common commands for chemostat operation will be described.

Connecting to the chemoduino:

Ensure that both the Raspberry Pi and all Arduinos are connected to (1) a 3.3V or 5V power source, (2) to each other through USB serial, (3) to the internet (Raspberry Pi only), and (4) to the motor controller board (Arduinos only). Also, be sure that the motor controlled board is connected to a 12V power supply. Finally, load the entire chemoduino folder from GitHub onto the Raspberry Pi, and flash the Arduino memory with the files located in "…/Arduino/v2.0."

Using PuTTY, or any other SSH client, connect to the Raspberry Pi through its IP address. Navigate to the folder `/chemoduino/Python`. If using Python 2.7.X, type: `python PyArdComm_27.py` to run the chemostat controller script. If using Python 3.X, type: `python PyArdComm.py`. Finally, you will be prompted about which serial port you would like to connect to.

Starting and stopping a pump:

Motors can be run in three operational modes: continuous, pulsatile, and automated. Continuous operation makes the motor pump fluid at a constant, continuous rate until it is stopped. Pulsatile mode causes the motor to pump fluid at a specific frequency and duration. Automated mode uses a timer to turn pumps on and off at specified frequencies without end. Pulsatile operation stops automatically after a specific number of cycles; however continuous and automated operation must be stopped manually.

To start the pump in continuous mode, enter the following command:

```
-start # -cont
```

To start the pump in pulsatile mode, enter the following command:

```
-start # -pulse
```

To start the pump in automated mode, enter the following command:

```
-start # -auto
```

To stop a pump, enter the following command:

```
-stop #
```

The # symbol represents the pump id, a number between 0 and 5.

Setting and getting pump attributes:

The write pin connects a pump id (ie. 0 to 5) with the corresponding PWM output pin of the Arduino microcontroller. Setting and getting which pump corresponds to a particular PWM output pin on the Arduino can be done with the following command:

```
-set <id> –wpin <pin>
-get <id> -wpin
```

In the above command, <id> is the pump id (0 to 5) and <pin> is the corresponding Arduino pin. If using the printed circuit board layout provided on GitHub, available Arduino pin options are 13, 11, 10, 9, 6, and 5.

Pulse width is used to control how long a motor remains on in its automated or pulsatile operational mode. By default, this value is set to 50 milliseconds, and we would not encourage going much lower than this value without evaluating the repeatability of the volume of fluid that is dispensed. Pulse width can be set or get with the following command:

-set <id> -pfreq <pin>

-get <id> -wpin

Pulse frequency is used to control how often a motor is activated while operating in automated mode. For low-geared 12VDC motors that use ¼" PharMed tubing, we recommend using an initial pulse frequency of 99950 milliseconds and refining that value using effluent volume measurements. For high-geared 12VDC motors, we recommend using an initial pulse frequency of 14950 milliseconds.

Appendix D

PARAMETER ESTIMATION OF THE TRANSCRIPTION MODEL

D.1 Overview

All Python code is freely available at www.github.com/uwsysbio under the "AbsApt" folder. The code was developed to facilitate researchers attempting to predict absolute aptamer concentration from fluorescence intensity. This appendix provides an example of how to use the script to predict RNA abundance from fluorescence.

Appendix E

RNA STRUCTURE SCREENING

E.1 Overview

All Python code is freely available at www.github.com/uwsysbio under the "AptCleavage" folder. The code implements the ViennaRNA Package. Prior to running the Python script that we provide, users must build the ViennaRNA library with the Python interface enabled. Instructions for building the library are found at http://www.tbi.univie.ac.at/RNA/. We can confirm that the library can be built using Ubuntu; however, we did not explicitly test whether this package can be built under Windows.

BIBLIOGRAPHY

[1] X. Zhuang, "SINGLE-MOLECULE RNA SCIENCE," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, no. 1, pp. 399–414, Jun. 2005.

[2] R. L. Strack, M. D. Disney, and S. R. Jaffrey, "A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA," *Nat. Methods*, vol. 10, no. 12, pp. 1219–1224, Oct. 2013.

[3] M.-C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang, "Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes," *Science*, vol. 329, no. 5992, pp. 689–693, Aug. 2010.

[4] B. F. Pfleger, D. J. Pitera, C. D. Smolke, and J. D. Keasling, "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes," *Nat. Biotechnol.*, vol. 24, no. 8, pp. 1027–1032, Aug. 2006.

[5] J. M. Carothers, J. A. Goler, D. Juminaga, and J. D. Keasling, "Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression," *Science*, vol. 334, no. 6063, pp. 1716–1719, Dec. 2011.

[6] T. K. Lu, A. S. Khalil, and J. J. Collins, "Next-generation synthetic gene networks," *Nat. Biotechnol.*, vol. 27, no. 12, pp. 1139–1150, Dec. 2009.

[7] T. A. Cooper, L. Wan, and G. Dreyfuss, "RNA and Disease," *Cell*, vol. 136, no. 4, pp. 777–793, Feb. 2009.

[8] R. Morin, M. Bainbridge, A. Fejes, M. Hirst, M. Krzywinski, T. Pugh, H. McDonald, R. Varhol, S. Jones, and M. Marra, "Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing," *BioTechniques*, vol. 45, no. 1, pp. 81–94, Jul. 2008.

[9] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, Jan. 2009.

[10] G. K. Geiss, R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, and L. Hood, "Direct multiplexed measurement of gene expression with color-coded probe pairs," *Nat. Biotechnol.*, vol. 26, no. 3, pp. 317–325, Mar. 2008.

[11] M. Morrow and J. Donaldson, "Nanostring's nCounter--A True Digital Target Profiling Technology," *PDA J. Pharm. Sci. Technol.*, vol. 65, no. 6, pp. 692–692, Nov. 2011.

[12] U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder, "The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing," *Science*, vol. 320, no. 5881, pp. 1344–1349, Jun. 2008.

[13] C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter, "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nat. Biotechnol.*, vol. 28, no. 5, pp. 511–515, May 2010.

[14] V. A. Malkov, K. A. Serikawa, N. Balantac, J. Watters, G. Geiss, A. Mashadi-Hossein, and T. Fare, "Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter™ Assay System," *BMC Res. Notes*, vol. 2, no. 1, p. 80, 2009.

[15] E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, "Localization of ASH1 mRNA Particles in Living Yeast," *Mol. Cell*, vol. 2, no. 4, pp. 437–445, Oct. 1998.

[16] A. M. Femino, "Visualization of Single RNA Transcripts in Situ," *Science*, vol. 280, no. 5363, pp. 585–590, Apr. 1998.

[17] A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi, "Imaging individual mRNA molecules using multiple singly labeled probes," *Nat. Methods*, vol. 5, no. 10, pp. 877–879, Oct. 2008.

[18] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells," *Science*, vol. 329, no. 5991, pp. 533–538, Jul. 2010.

[19] I. Golding and E. C. Cox, "RNA dynamics in live Escherichia coli cells," *Proc. Natl. Acad. Sci.*, vol. 101, no. 31, pp. 11310–11315, Aug. 2004.

[20] N. Daigle and J. Ellenberg, "λN-GFP: an RNA reporter system for live-cell imaging," *Nat. Methods*, vol. 4, no. 8, pp. 633–636, Aug. 2007.

[21] A. Calapez, "The intranuclear mobility of messenger RNA binding proteins is ATP dependent and temperature sensitive," *J. Cell Biol.*, vol. 159, no. 5, pp. 795–805, Dec. 2002.

[22] Y. Shav-Tal, "Dynamics of Single mRNPs in Nuclei of Living Cells," *Science*, vol. 304, no. 5678, pp. 1797–1800, Jun. 2004.

[23] V. L. Zimyanin, K. Belaya, J. Pecreaux, M. J. Gilchrist, A. Clark, I. Davis, and D. St Johnston, "In Vivo Imaging of oskar mRNA Transport Reveals the Mechanism of Posterior Localization," *Cell*, vol. 134, no. 5, pp. 843–853, Sep. 2008.

[24] K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder, "Translation-Independent Localization of mRNA in E. coli," *Science*, vol. 331, no. 6020, pp. 1081–1084, Feb. 2011.

[25] J. R. Babendure, S. R. Adams, and R. Y. Tsien, "Aptamers Switch on Fluorescence of Triphenylmethane Dyes," *J. Am. Chem. Soc.*, vol. 125, no. 48, pp. 14716–14717, Dec. 2003.

[26] E. V. Dolgosheina, S. C. Y. Jeng, S. S. S. Panchapakesan, R. Cojocaru, P. S. K. Chen, P. D. Wilson, N. Hawkins, P. A. Wiggins, and P. J. Unrau, "RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking," *ACS Chem. Biol.*, vol. 9, no. 10, pp. 2412–2420, Oct. 2014.

[27] J. S. Paige, K. Y. Wu, and S. R. Jaffrey, "RNA Mimics of Green Fluorescent Protein," *Science*, vol. 333, no. 6042, pp. 642–646, Jul. 2011.

[28] T. P. Constantin, G. L. Silva, K. L. Robertson, T. P. Hamilton, K. Fague, A. S. Waggoner, and B. A. Armitage, "Synthesis of New Fluorogenic Cyanine Dyes and Incorporation into RNA Fluoromodules," *Org. Lett.*, vol. 10, no. 8, pp. 1561–1564, Apr. 2008.

[29] X. Zhang, A. S. R. Potty, G. W. Jackson, V. Stepanov, A. Tang, Y. Liu, K. Kourentzi, U. Strych, G. E. Fox, and R. C. Willson, "Engineered 5S ribosomal RNAs displaying aptamers recognizing vascular endothelial growth factor and malachite green," *J. Mol. Recognit.*, vol. 22, no. 2, pp. 154–161, Mar. 2009.

[30] G. Pothoulakis, F. Ceroni, B. Reeve, and T. Ellis, "The Spinach RNA Aptamer as a Characterization Tool for Synthetic Biology," *ACS Synth. Biol.*, vol. 3, no. 3, pp. 182–187, Mar. 2014.

[31] R. L. Strack, W. Song, and S. R. Jaffrey, "Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria," *Nat. Protoc.*, vol. 9, no. 1, pp. 146–155, Jan. 2014.

[32] J. S. Paige, T. Nguyen-Duc, W. Song, and S. R. Jaffrey, "Fluorescence Imaging of Cellular Metabolites with RNA," *Science*, vol. 335, no. 6073, pp. 1194–1194, Mar. 2012.

[33] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox," *Nat. Protoc.*, vol. 2, no. 3, pp. 727–738, Mar. 2007.

[34] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert, "A Whole-Cell Computational Model Predicts Phenotype from Genotype," *Cell*, vol. 150, no. 2, pp. 389–401, Jul. 2012.

[35] S. Dmitrieff, M. Rao, and P. Sens, "Quantitative analysis of intra-Golgi transport shows intercisternal exchange for all cargo," *Proc. Natl. Acad. Sci.*, vol. 110, no. 39, pp. 15692–15697, Sep. 2013.

[36] D. Chandran, W. B. Copeland, S. C. Sleight, and H. M. Sauro, "Mathematical modeling and synthetic biology," *Drug Discov. Today Dis. Models*, vol. 5, no. 4, pp. 299–309, Dec. 2008.

[37] D. Grate and C. Wilson, "Laser-mediated, site-specific inactivation of RNA transcripts," *Proc. Natl. Acad. Sci.*, vol. 96, no. 11, pp. 6131–6136, May 1999.

[38] C. Szent-Gyorgyi, B. A. Schmidt, Y. Creeger, G. W. Fisher, K. L. Zakel, S. Adler, J. A. J. Fitzpatrick, C. A. Woolford, Q. Yan, K. V. Vasilev, P. B. Berget, M. P. Bruchez, J. W. Jarvik, and A. Waggoner, "Fluorogen-activating single-chain antibodies for imaging cell surface proteins," *Nat. Biotechnol.*, vol. 26, no. 2, pp. 235–240, Feb. 2008.

[39] S. Srivastava, R. Sinha, and D. Roy, "Toxicological effects of malachite green," *Aquat. Toxicol.*, vol. 66, no. 3, pp. 319–329, Feb. 2004.

[40] G. Brawerman, "Determinants of messenger RNA stability," *Cell*, vol. 48, no. 1, pp. 5–6, Jan. 1987.

[41] G. Brawerman, "mRNA decay: Finding the right targets," *Cell*, vol. 57, no. 1, pp. 9–10, Apr. 1989.

[42] D. W. Selinger, "Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation," *Genome Res.*, vol. 13, no. 2, pp. 216–223, Feb. 2003.

[43] R. Rauhut and G. Klug, "mRNA degradation in bacteria," *FEMS Microbiol. Rev.*, vol. 23, no. 3, pp. 353–370, Jun. 1999.

[44] M. L. Brock and D. J. Shapiro, "Estrogen stabilizes vitellogenin mRNA against cytoplasmic degradation," *Cell*, vol. 34, no. 1, pp. 207–214, Aug. 1983.

[45] M. Grunberg-Manago, "M ESSENGER RNA S TABILITY AND I TS R OLE IN C ONTROL OF G ENE E XPRESSION IN B ACTERIA AND P HAGES," *Annu. Rev. Genet.*, vol. 33, no. 1, pp. 193–227, Dec. 1999.

[46] M. P. Deutscher, "Degradation of RNA in bacteria: comparison of mRNA and stable RNA," *Nucleic Acids Res.*, vol. 34, no. 2, pp. 659–666, Jan. 2006.

[47] S. C. Makrides, "Strategies for achieving high-level expression of genes in Escherichia coli," *Microbiol. Rev.*, vol. 60, no. 3, pp. 512–538, Sep. 1996.
[48] T. A. Carrier and J. D. Keasling, "Controlling Messenger RNA Stability in Bacteria: Strategies for Engineering Gene Expression," *Biotechnol. Prog.*, vol. 13, no. 6, pp. 699–708, Dec. 1997.
[49] P. Alifano, C. B. Bruni, and M. S. Carlomagno, "Control of mRNA processing and decay in prokaryotes," *Genetica*, vol. 94, no. 2–3, pp. 157–172, Jun. 1994.
[50] N. Mertens, E. Remaut, and W. Fiers, "Increased stability of phage T7g10 mRNA is mediated by either a 5′- or a 3′-terminal stem-loop structure," *Biol. Chem.*, vol. 377, no. 12, pp. 811–817, Dec. 1996.
[51] N. Nakashima, T. Tamura, and L. Good, "Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli," *Nucleic Acids Res.*, vol. 34, no. 20, pp. e138–e138, Oct. 2006.
[52] F. Akter and Y. Yokobayashi, "RNA Signal Amplifier Circuit with Integrated Fluorescence Output," *ACS Synth. Biol.*, p. 141107155723000, Nov. 2014.
[53] J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, and S. N. Cohen, "Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays," *Proc. Natl. Acad. Sci.*, vol. 99, no. 15, pp. 9697–9702, Jul. 2002.
[54] O. Tour, R. M. Meijer, D. A. Zacharias, S. R. Adams, and R. Y. Tsien, "Genetically targeted chromophore-assisted light inactivation," *Nat. Biotechnol.*, vol. 21, no. 12, pp. 1505–1508, Dec. 2003.
[55] D. Hoffman-Kim, T. J. Diefenbach, B. K. Eustace, and D. G. Jay, "Chromophore-Assisted Laser Inactivation," in *Methods in Cell Biology*, vol. 82, Elsevier, 2007, pp. 335–354.
[56] J. C. Liao, J. Roider, and D. G. Jay, "Chromophore-assisted laser inactivation of proteins is mediated by the photogeneration of free radicals," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 7, pp. 2659–2663, Mar. 1994.
[57] R. Lutz, "Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements," *Nucleic Acids Res.*, vol. 25, no. 6, pp. 1203–1210, Mar. 1997.
[58] W. Young, A. E. Raftery, and K. Yeung, "Fast Bayesian inference for gene regulatory networks using ScanBMA," *BMC Syst. Biol.*, vol. 8, no. 1, p. 47, 2014.
[59] N. Friedman, "Inferring Cellular Networks Using Probabilistic Graphical Models," *Science*, vol. 303, no. 5659, pp. 799–805, Feb. 2004.
[60] G. An, Q. Mi, J. Dutta-Moscato, and Y. Vodovotz, "Agent-based models in translational systems biology," *Wiley Interdiscip. Rev. Syst. Biol. Med.*, vol. 1, no. 2, pp. 159–171, Sep. 2009.
[61] S. D. Hester, J. M. Belmonte, J. S. Gens, S. G. Clendenon, and J. A. Glazier, "A Multi-cell, Multi-scale Model of Vertebrate Segmentation and Somite Formation," *PLoS Comput. Biol.*, vol. 7, no. 10, p. e1002155, Oct. 2011.
[62] C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J.-B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young, "Transcriptional regulatory code of a eukaryotic genome," *Nature*, vol. 431, no. 7004, pp. 99–104, Sep. 2004.

[63] C. F. Craver, "When mechanistic models explain," *Synthese*, vol. 153, no. 3, pp. 355–376, Nov. 2006.

[64] J. R. Kelly, A. J. Rubin, J. H. Davis, C. M. Ajo-Franklin, J. Cumbers, M. J. Czar, K. de Mora, A. L. Glieberman, D. D. Monie, and D. Endy, "Measuring the activity of BioBrick promoters using an in vivo reference standard," *J. Biol. Eng.*, vol. 3, no. 1, p. 4, 2009.

[65] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," *Nature*, vol. 456, no. 7221, pp. 516–519, Nov. 2008.

[66] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, "Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing," *Proc. Natl. Acad. Sci.*, vol. 101, no. 30, pp. 10955–10960, Jul. 2004.

[67] H. Kobayashi, M. Kaern, M. Araki, K. Chung, T. S. Gardner, C. R. Cantor, and J. J. Collins, "Programmable cells: Interfacing natural and engineered gene networks," *Proc. Natl. Acad. Sci.*, vol. 101, no. 22, pp. 8414–8419, Jun. 2004.

[68] G. Lillacci and M. Khammash, "Parameter Estimation and Model Selection in Computational Biology," *PLoS Comput. Biol.*, vol. 6, no. 3, p. e1000696, Mar. 2010.

[69] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom, "Systems biology: parameter estimation for biochemical models: Parameter estimation in systems biology," *FEBS J.*, vol. 276, no. 4, pp. 886–902, Feb. 2009.

[70] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, Jan. 1965.

[71] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters.," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.

[72] C. N. Takahashi, A. W. Miller, F. Ekness, M. J. Dunham, and E. Klavins, "A Low Cost, Customizable Turbidostat for Use in Synthetic Circuit Characterization," *ACS Synth. Biol.*, p. 140801103853005, Aug. 2014.

[73] A. Novick and L. Szilard, "Description of the Chemostat," *Science*, vol. 112, no. 2920, pp. 715–716, Dec. 1950.

[74] A. W. Miller, C. Befort, E. O. Kerr, and M. J. Dunham, "Design and Use of Multiplexed Chemostat Arrays," *J. Vis. Exp.*, no. 72, Feb. 2013.

[75] E. A. Campbell, N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb, and S. A. Darst, "Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase," *Cell*, vol. 104, no. 6, pp. 901–912, Mar. 2001.

[76] G. A. Mackie, "RNase E: at the interface of bacterial RNA processing and decay," *Nat. Rev. Microbiol.*, vol. 11, no. 1, pp. 45–57, Dec. 2012.

[77] C. D. Smolke, T. A. Carrier, and J. D. Keasling, "Coordinated, Differential Expression of Two Genes through Directed mRNA Cleavage and Stabilization by Secondary Structures," *Appl. Environ. Microbiol.*, vol. 66, no. 12, pp. 5399–5405, Dec. 2000.

[78] S. Ide, T. Miyazaki, H. Maki, and T. Kobayashi, "Abundance of Ribosomal RNA Gene Copies Maintains Genome Integrity," *Science*, vol. 327, no. 5966, pp. 693–696, Feb. 2010.

[79] G. N. Basturea, D. R. Dague, M. P. Deutscher, and K. E. Rudd, "YhiQ Is RsmJ, the Methyltransferase Responsible for Methylation of G1516 in 16S rRNA of E. coli," *J. Mol. Biol.*, vol. 415, no. 1, pp. 16–21, Jan. 2012.

[80] K. Zwirglmaier, W. Ludwig, and K.-H. Schleifer, "Recognition of individual genes in a single bacterial cell by fluorescence in situ hybridization - RING-FISH: RING-FISH," *Mol. Microbiol.*, vol. 51, no. 1, pp. 89–96, Nov. 2003.

[81] J. Kim and N. Wang, "Characterization of copy numbers of 16S rDNA and 16S rRNA of Candidatus Liberibacter asiaticus and the implication in detection in planta using quantitative PCR," *BMC Res. Notes*, vol. 2, no. 1, p. 37, 2009.

[82] V. K. Mutalik, J. C. Guimaraes, G. Cambray, C. Lam, M. J. Christoffersen, Q.-A. Mai, A. B. Tran, M. Paull, J. D. Keasling, A. P. Arkin, and D. Endy, "Precise and reliable gene expression via standard transcription and translation initiation elements," *Nat. Methods*, vol. 10, no. 4, pp. 354–360, Mar. 2013.

[83] J. B. Lucks, L. Qi, V. K. Mutalik, D. Wang, and A. P. Arkin, "Versatile RNA-sensing transcriptional regulators for engineering genetic networks," *Proc. Natl. Acad. Sci.*, vol. 108, no. 21, pp. 8617–8622, May 2011.

[84] V. K. Mutalik, L. Qi, J. C. Guimaraes, J. B. Lucks, and A. P. Arkin, "Rationally designed families of orthogonal RNA regulators of translation," *Nat. Chem. Biol.*, vol. 8, no. 5, pp. 447–454, Mar. 2012.

[85] M. Galdzicki, K. P. Clancy, E. Oberortner, M. Pocock, J. Y. Quinn, C. A. Rodriguez, N. Roehner, M. L. Wilson, L. Adam, J. C. Anderson, B. A. Bartley, J. Beal, D. Chandran, J. Chen, D. Densmore, D. Endy, R. Grünberg, J. Hallinan, N. J. Hillson, J. D. Johnson, A. Kuchinsky, M. Lux, G. Misirli, J. Peccoud, H. A. Plahar, E. Sirin, G.-B. Stan, A. Villalobos, A. Wipat, J. H. Gennari, C. J. Myers, and H. M. Sauro, "The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology," *Nat. Biotechnol.*, vol. 32, no. 6, pp. 545–550, Jun. 2014.

[86] L. J. Kahl and D. Endy, "A survey of enabling technologies in synthetic biology," *J. Biol. Eng.*, vol. 7, no. 1, p. 13, 2013.

[87] R. M. Saecker, M. T. Record, and P. L. deHaseth, "Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis," *J. Mol. Biol.*, vol. 412, no. 5, pp. 754–771, Oct. 2011.

[88] D. J. Lee, S. D. Minchin, and S. J. W. Busby, "Activating Transcription in Bacteria," *Annu. Rev. Microbiol.*, vol. 66, no. 1, pp. 125–152, Oct. 2012.

[89] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, "Genetic programs constructed from layered logic gates in single cells," *Nature*, vol. 491, no. 7423, pp. 249–253, Oct. 2012.

[90] B. C. Stanton, A. A. K. Nielsen, A. Tamsir, K. Clancy, T. Peterson, and C. A. Voigt, "Genomic mining of prokaryotic repressors for orthogonal logic gates," *Nat. Chem. Biol.*, vol. 10, no. 2, pp. 99–105, Dec. 2013.

[91] B. Canton, A. Labno, and D. Endy, "Refinement and standardization of synthetic biological parts and devices," *Nat. Biotechnol.*, vol. 26, no. 7, pp. 787–793, Jul. 2008.

[92] J. H. Davis, A. J. Rubin, and R. T. Sauer, "Design, construction and characterization of a set of insulated bacterial promoters," *Nucleic Acids Res.*, vol. 39, no. 3, pp. 1131–1141, Feb. 2011.

[93] V. K. Mutalik, J. C. Guimaraes, G. Cambray, Q.-A. Mai, M. J. Christoffersen, L. Martin, A. Yu, C. Lam, C. Rodriguez, G. Bennett, J. D. Keasling, D. Endy, and A. P. Arkin, "Quantitative

estimation of activity and quality for collections of functional genetic elements," *Nat. Methods*, vol. 10, no. 4, pp. 347–353, Mar. 2013.

[94] H. H. McAdams and A. Arkin, "It's a noisy business! Genetic regulation at the nanomolar scale," *Trends Genet.*, vol. 15, no. 2, pp. 65–69, Feb. 1999.

[95] L. S. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schaffer, "Stochastic Gene Expression in a Lentiviral Positive-Feedback Loop: HIV-1 Tat Fluctuations Drive Phenotypic Diversity," *Cell*, vol. 122, no. 2, pp. 169–182, Jul. 2005.

[96] V. A. Rhodius, V. K. Mutalik, and C. A. Gross, "Predicting the strength of UP-elements and full-length E. coli E promoters," *Nucleic Acids Res.*, vol. 40, no. 7, pp. 2907–2924, Apr. 2012.

[97] R. G. Egbert and E. Klavins, "Fine-tuning gene networks using simple sequence repeats," *Proc. Natl. Acad. Sci.*, vol. 109, no. 42, pp. 16817–16822, Oct. 2012.

[98] S. Yang, S. C. Sleight, and H. M. Sauro, "Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits," *Nucleic Acids Res.*, vol. 41, no. 1, pp. e33–e33, Jan. 2013.

[99] R. Y. Tsien, "THE GREEN FLUORESCENT PROTEIN," *Annu. Rev. Biochem.*, vol. 67, no. 1, pp. 509–544, Jun. 1998.

[100] B. N. G. Giepmans, "The Fluorescent Toolbox for Assessing Protein Location and Function," *Science*, vol. 312, no. 5771, pp. 217–224, Apr. 2006.

[101] S. Tyagi, "Imaging intracellular RNA distribution and dynamics in living cells," *Nat. Methods*, vol. 6, no. 5, pp. 331–338, May 2009.

[102] A. J. Carpousis, "The RNA Degradosome of *Escherichia coli*: An mRNA-Degrading Machine Assembled on RNase E," *Annu. Rev. Microbiol.*, vol. 61, no. 1, pp. 71–87, Oct. 2007.

[103] A. Deana, H. Celesnik, and J. G. Belasco, "The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal," *Nature*, vol. 451, no. 7176, pp. 355–358, Jan. 2008.

[104] S. Gottesman, "THE SMALL RNA REGULATORS OF *ESCHERICHIA COLI*: Roles and Mechanisms*," *Annu. Rev. Microbiol.*, vol. 58, no. 1, pp. 303–328, Oct. 2004.

[105] L. Qi, R. E. Haurwitz, W. Shao, J. A. Doudna, and A. P. Arkin, "RNA processing enables predictable programming of gene expression," *Nat. Biotechnol.*, vol. 30, no. 10, pp. 1002–1006, Sep. 2012.

[106] C. Lou, B. Stanton, Y.-J. Chen, B. Munsky, and C. A. Voigt, "Ribozyme-based insulator parts buffer synthetic circuits from genetic context," *Nat. Biotechnol.*, vol. 30, no. 11, pp. 1137–1142, Oct. 2012.

[107] R. E. Haurwitz, M. Jinek, B. Wiedenheft, K. Zhou, and J. A. Doudna, "Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease," *Science*, vol. 329, no. 5997, pp. 1355–1358, Sep. 2010.

[108] J. M. Buzayan, W. L. Gerlach, and G. Bruening, "Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing," vol. 83, no. 23, pp. 8859–8862, 1986.

[109] A. C. Forster and R. H. Symons, "Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site," *Cell*, vol. 50, no. 1, pp. 9–16, Jul. 1987.

[110] G. Rodrigo, T. E. Landrain, and A. Jaramillo, "De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells," *Proc. Natl. Acad. Sci.*, vol. 109, no. 38, pp. 15271–15276, Sep. 2012.

[111] J. T. Stevens and J. M. Carothers, "Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways," *ACS Synth. Biol.*, p. 141027155148007, Oct. 2014.

[112] M. N. Win and C. D. Smolke, "Higher-Order Cellular Information Processing with Synthetic RNA Devices," *Science*, vol. 322, no. 5900, pp. 456–460, Oct. 2008.

[113] M. N. Win and C. D. Smolke, "A modular and extensible RNA-based gene-regulatory platform for engineering cellular function," *Proc. Natl. Acad. Sci.*, vol. 104, no. 36, pp. 14283–14288, Sep. 2007.

[114] I. L. Hofacker, "Vienna RNA secondary structure server," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3429–3431, Jul. 2003.

[115] A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neubock, and I. L. Hofacker, "The Vienna RNA Websuite," *Nucleic Acids Res.*, vol. 36, no. Web Server, pp. W70–W74, May 2008.

[116] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce, "NUPACK: Analysis and design of nucleic acid systems," *J. Comput. Chem.*, vol. 32, no. 1, pp. 170–173, Jan. 2011.

[117] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, "ViennaRNA Package 2.0," *Algorithms Mol. Biol.*, vol. 6, no. 1, p. 26, 2011.

[118] C. E. Hajdin, S. Bellaousov, W. Huggins, C. W. Leonard, D. H. Mathews, and K. M. Weeks, "Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots," *Proc. Natl. Acad. Sci.*, vol. 110, no. 14, pp. 5498–5503, Apr. 2013.

[119] Y. Wan, M. Kertesz, R. C. Spitale, E. Segal, and H. Y. Chang, "Understanding the transcriptome through RNA structure," *Nat. Rev. Genet.*, vol. 12, no. 9, pp. 641–655, Aug. 2011.

[120] E. Rivas and S. R. Eddy, "A dynamic programming algorithm for RNA structure prediction including pseudoknots," *J. Mol. Biol.*, vol. 285, no. 5, pp. 2053–2068, Feb. 1999.

[121] B. A. Sullenger and T. R. Cech, "Ribozyme-mediated repair of defective mRNA by targeted trans-splicing," *Nature*, vol. 371, no. 6498, pp. 619–622, Oct. 1994.

[122] J. Kim, S. Jeong, A. Kertsburg, G. A. Soukup, and S.-W. Lee, "Conditional and Target-Specific Transgene Induction through RNA Replacement Using an Allosteric Trans-Splicing Ribozyme," *ACS Chem. Biol.*, vol. 9, no. 11, pp. 2491–2495, Nov. 2014.

[123] K. H. Kim and H. M. Sauro, "Measuring Retroactivity from Noise in Gene Regulatory Networks," *Biophys. J.*, vol. 100, no. 5, pp. 1167–1177, Mar. 2011.

[124] K. H. Kim and H. M. Sauro, "Adjusting Phenotypes by Noise Control," *PLoS Comput. Biol.*, vol. 8, no. 1, p. e1002344, Jan. 2012.

[125] A. D. Ellington and J. W. Szostak, "In vitro selection of RNA molecules that bind specific ligands," *Nature*, vol. 346, no. 6287, pp. 818–822, Aug. 1990.

[126] C. Tuerk and L. Gold, "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase," *Science*, vol. 249, no. 4968, pp. 505–510, Aug. 1990.

[127] S. Sando, A. Narita, and Y. Aoyama, "Light-Up Hoechst–DNA Aptamer Pair: Generation of an Aptamer-Selective Fluorophore from a Conventional DNA-Staining Dye," *ChemBioChem*, vol. 8, no. 15, pp. 1795–1803, Oct. 2007.

[128] S. Sando, A. Narita, M. Hayami, and Y. Aoyama, "Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: a blue fluorescent RNA," *Chem. Commun.*, no. 33, p. 3858, 2008.

[129] O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea," *J. Cell. Comp. Physiol.*, vol. 59, no. 3, pp. 223–239, Jun. 1962.

[130] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier, "Primary structure of the Aequorea victoria green-fluorescent protein," *Gene*, vol. 111, no. 2, pp. 229–233, Feb. 1992.

[131] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, and D. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, no. 5148, pp. 802–805, Feb. 1994.

[132] G. Zhang, V. Gurtu, and S. R. Kain, "An Enhanced Green Fluorescent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells," *Biochem. Biophys. Res. Commun.*, vol. 227, no. 3, pp. 707–711, Oct. 1996.

[133] J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, and G. S. Waldo, "Engineering and characterization of a superfolder green fluorescent protein," *Nat. Biotechnol.*, vol. 24, no. 1, pp. 79–88, Jan. 2006.

[134] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, "A guide to choosing fluorescent proteins," *Nat. Methods*, vol. 2, no. 12, pp. 905–909, Dec. 2005.

[135] Y. Nomura, D. Kumar, and Y. Yokobayashi, "Synthetic mammalian riboswitches based on guanine aptazyme," *Chem. Commun.*, vol. 48, no. 57, p. 7215, 2012.

[136] M. K. Takahashi and J. B. Lucks, "A modular strategy for engineering orthogonal chimeric RNA transcription regulators," *Nucleic Acids Res.*, vol. 41, no. 15, pp. 7577–7588, Aug. 2013.

[137] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 6, pp. 410–422, Jun. 2009.

[138] M. Heinemann and S. Panke, "Synthetic biology--putting engineering into biology," *Bioinformatics*, vol. 22, no. 22, pp. 2790–2799, Nov. 2006.