DEVELOPMENT OF RNA-BASED TRANSLATION MODULATION TOOLS FOR SYNTHETIC BIOLOGY APPLICATIONS

RHYS HAKSTOL

Bachelor of Science, University of Lethbridge, 2014

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the Requirements for the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

DEVELOPMENT OF RNA-BASED TRANSLATION MODULATION TOOLS FOR SYNTHETIC BIOLOGY APPLICATIONS

RHYS HAKSTOL

Date of Defense: January 16, 2019

Dr. Hans-Joachim Wieden Supervisor	Professor	Ph.D.
Dr. Tony Russell Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Nehalkumar Thakor Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. Paul Hayes Chair, Thesis Examination Committee Member	Professor	Ph.D.

Dedicated to my friends and family, whose support and patience has shaped who I am today.

Abstract

The ability to reliably regulate translation of heterologous proteins is of great interest for diverse applications in the field of biological and metabolic engineering. RNA-based tools to control ribosome-dependent synthesis of proteins within *E. coli* were developed. First, adaptation of the well-studied RNA-IN and RNA-OUT system based on *E. coli* Tn*10* to reliably repress translation regardless of coding sequence context was performed. Using an invariant sequence within the RNA-IN/OUT interacting region does not interfere with the function of the RNAs involved. The same RNA-IN/RNA-OUT system was adapted to activate the translation of a specific target mRNA.

Regulation of translation extends to processing of polyproteins. In eukaryotes this can be accomplished by viral 2A peptides, avoiding reinitiating at a downstream start codon but has not been observed in bacteria. Therefore, a strategy for developing libraries of 2A peptides with potential activity across an array of organisms was developed.

Acknowledgements

I would like to thank my supervisor, Dr. Hans-Joachim Wieden, for his teaching and mentorship that nurtured my interest in synthetic biology. His passion and excitement for student inquiry led to many exciting opportunities I could not imagine.

I would like to thank other members of my supervisory committee, Dr. Tony Russell and Dr. Nehal Thakor, for their participation and guidance as well as helpful suggestions over the course of my graduate education.

I would also like to thank the Department of Chemistry and Biochemistry, especially Susan Hill, Emily Wilton and Fan Mo whose attention to detail and assistance help things in the lab and department run smoothly. I would like to acknowledge the use of the Synbridge Maker Space and the assistance and helpful suggestions of Justin Vigar, Taylor Sheahan and Luc Roberts in the execution of experiments.

My thanks to past and present members of the Wieden and Kothe Labs, including Dr. Andy Hudson and Graeme Glaister with whom I shared a lab bench, many helpful conversations and trips to and from iGEM-related gatherings. I would like to thank other members of the lab for their friendship and insight, especially Dylan Girodat, Luc Roberts, Jalyce Heller, Harland Brandon and Justin Vigar. Thank you as well to my fellow student peers across the university, Jackson Knott, Connor MacNeil, David McWatters, Nathan Kostiuk, Daniel Stuart, and especially Chelsi Harvey for their support and encouragement, and for making my graduate experience an enjoyable and memorable one. Thank you to all my friends and colleagues, your support does not go unnoticed.

My thanks also to the Natural Sciences and Engineering Research Council of Canada (NSERC) for their generous support of my education through a CGS-M award. This

support was critical to my education and success and I am deeply appreciative of the opportunities it afforded me.

I would also like to thank all those involved with Let's Talk Science at the U of L, especially Dr. Ute Wieden-Kothe. Working with Let's Talk Science was truly one of the most rewarding endeavours I have ever undertaken, and I really do appreciate it.

Finally, a sincere thank you to my family, Ken, Marisa, and Keilan who have always been there for me. You have been my mentors and role models for my entire life and I am forever grateful for your support and love.

Table of contents

List of figures

1 Introduction

In this work, a range of strategies for controlling prokaryotic gene expression are outlined as well as tools for this control. These tools are of great utility to synthetic biology and biological engineering of organisms to produce valuable or useful products and/or functions. Therefore, a background in molecular, cellular and synthetic biology will be provided as well as the rationale for the techniques and approaches used in this thesis. To begin, an introduction to prokaryotic gene expression regulation will be given, followed by the selected approach for this work, novel *trans*-acting RNAs. A more detailed background on the selected RNA-IN/RNA-OUT system will be provided. As a separate approach to controlling gene expression in both prokaryotic and eukaryotic organisms, 2A-peptides and their utility in bioengineering will be surveyed.

1.1 Prokaryotic Gene Expression

Prokaryotic genetic material in the form of DNA must go through a series of steps before a functional product, either RNA or protein can be produced (Figure 1.1). The first of these steps is transcription, in which an RNA molecule is produced by RNA polymerase in a 5′ to 3′ direction (1). This RNA may either serve a function on its own without being translated as in the case of tRNA,rRNA, sRNA (2) as well as others, or serve as messenger RNA (mRNA) and undergo translation into a protein by the 70S ribosome. The mRNA is able to engage in complementary base-pairing with the 16S rRNA at the Shine-Dalgarno region, and translation initiates followed by the step-wise decoding of the mRNA, producing a full-length polypeptide (3). At each step during gene expression,

regulation may be imposed. For example, transcription may be impacted by the binding of transcription factors to the promoter region of a gene to either upregulate or down-regulate its transcription (4). In contrast, translation may be regulated in a variety of forms including alterations to Shine-Dalgarno sequence accessibility and mRNA stability (5).

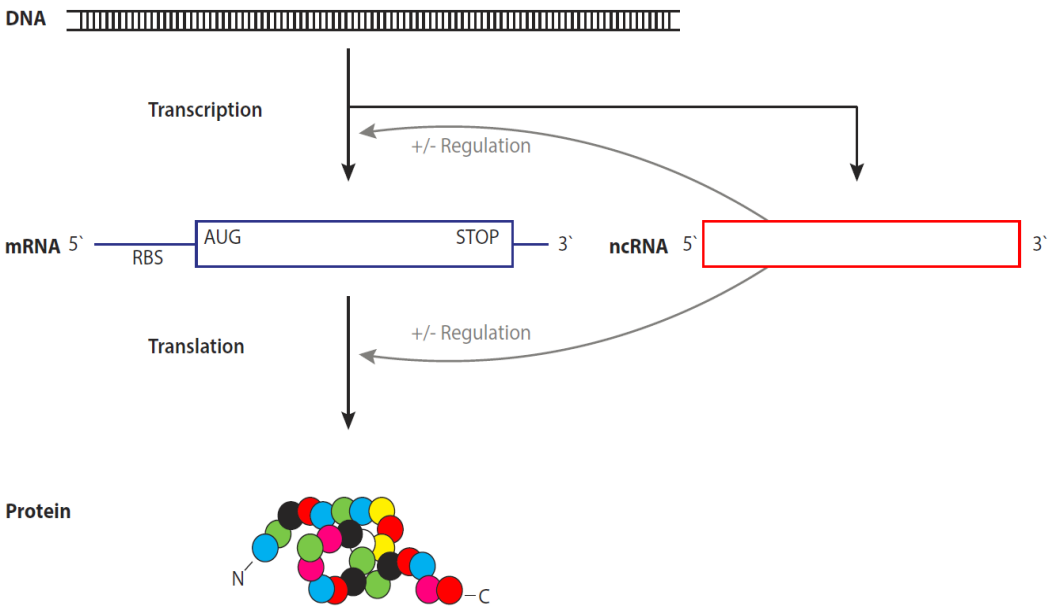


Figure 1.1. Prokaryotic gene expression. Points of control at transcription and translation are indicated for modulation of downstream product production. Non-coding RNAs (ncRNAs) can regulate both transcription and translation within the cell.

1.2 Prokaryotic Gene Expression Regulation

Gene expression may be controlled broadly at the transcriptional or translational steps. Understanding the strategies living systems employ to control gene expression can assist in the design of synthetic and reliable tools for achieving similar regulation.

1.2.1 Prokaryotic Transcription Regulation

Transcription in prokaryotes is a highly regulated step in gene expression, most often regulated through transcription factor binding to an operator region located upstream of a gene (6). These transcription factors can be involved in a variety of strategies and feedback loops to effectively modulate the RNA polymerase recruitment to a given transcription start site (7). Given the widespread use of transcription regulation strategies across prokaryotes, transcription regulation has been the target of bioengineering efforts. Namely, within prokaryotic organisms, RNA transcription attenuators have been targeted (8). These attenuators have been engineered by synthetic biologists to perform predictably, often using *E. coli* as a chassis organism in construction of complex and orthogonal genetic circuitry (9).

Transcriptional control has been widely used in *E. coli* to regulate expression of genes of interest (10). When combined, use of transcriptional regulators has enabled the construction of large predictable networks and circuitry within synthetic biological systems (11).

1.2.2 Prokaryotic Translation Regulation

Translation occurs in discrete consecutive phases known as initiation, elongation, termination and ribosome recycling (12, 13). Prokaryotic translation regulation most often

occurs at the initiation step (5, 14). With multiple strategies that can be employed in living systems for regulation of this step, it is important to understand the redundancy and overlap between these control mechanisms. The majority of the regulatory strategies for bacterial translation initiation control influence correct positioning/recognition of the mRNA on the 30S initiation complex. These control strategies can be used to either increase (15) or decrease (16) the levels of translation by impacting the accessibility of ribosomes to the initiation region of an mRNA molecule. Control of translation initiation allows for precise modulation of translation of a specific message within a bacterial cell, and optimization of stoichiometry of reaction partners, making this a critical regulatory step to consider when designing synthetic biological circuits.

One may employ a computational approach to understand the thermodynamics of translation initiation within a bacterial chassis. Computational approaches such as ribosome binding site calculators (17) have been developed to predict translation initiation strength and allow for accurate estimation of initiation strength of a given mRNA sequence, including differences in secondary structure and thermodynamics. The strength of a particular ribosome binding site is a key feature of translation initiation (18). Use of secondary structure prediction is important within mRNA, as it allows for knowledge of the ribosome binding site's accessibility, depending on a variety of conditions (17).

One such example of structurally-mediated control of translation are riboswitches. Riboswitches *in vivo* allow for cells to respond to inputs such as the presence of a small molecular ligand with an RNA-mediated response (19). Most often, riboswitches reside within the 5` UTR of a given mRNA, and respond to the presence of a ligand by binding

to it, which results in a structural change, controlling downstream gene expression. Importantly, riboswitches are able to turn translation on or off *in vivo*, which makes them attractive tools for engineering synthetic biological circuits, as they are able to respond to a variety of physical (e.g. temperature) and chemical (e.g. ligand binding) factors (19, 20).

Another RNA-based approach for regulation of translation is sRNA-mRNA interaction (21). sRNAs are usually small (50-200nt) *trans*-acting RNAs which act to influence prokaryotic gene expression. These RNAs often work by engaging in imperfect base pairing to an mRNA of interest, which thereby influences ribosome binding site accessibility as well as mRNA stability (16). sRNAs interact with mRNAs via the bacterial chaperone protein Hfq, which facilitates the interaction (22, 23). This interaction between the sRNA and mRNA can either increase (24) or decrease (25) the strength of translation initiation, which is an exploitable feature for bioengineering.

The structure of an mRNA molecule can also result in accessibility to effector proteins. The binding of a protein to a given mRNA sequence may down-regulate translation as in the case of ribosomal subunit operon synthesis (26), whereas other situations such as the bacteriophage µ *mom* gene are positively regulated by binding of a protein effector (27). Moreover, enzymatic effectors are capable of cleaving and degrading mRNA signals, effectively maintaining an optimal copy number of the respective mRNA within a prokaryotic cell (28–30). A dynamic interplay between protein and RNA is required for regulating gene expression within the complex environment of the bacterial cell.

1.3 *trans*-Acting RNA Regulation of Translation in Prokaryotes

Several strategies for regulation of gene expression employ RNA-only approaches. These strategies have been evolved in natural systems over millions of years (31) and have the advantage of being generally faster than protein regulators which must first be synthesized. They also provide a more dynamic signal, as the half-life of regulatory RNAs is rather short within the cell, and may allow for a rapid response when faced with an input (32).

1.3.1 The Tn*10*/IS*10* of *E. coli* as a Regulatory Framework

DNA transposons within bacteria have been studied for many years (33). The simplest ones encode the genes responsible for their own transposition and are termed insertion sequences (ISs). Within the IS, an inverted repeat sequence flanks a transposase gene, which once expressed recognizes a target site and inserts the entire IS into the target locus. One such IS is found in *E. coli*, and is known as IS*10* (34). IS*10* is part of a larger composite transposon known as Tn*10.* Tn*10* is composed of two IS*10* sequences flanking a tetracycline resistance gene (35) (Figure 1.2).

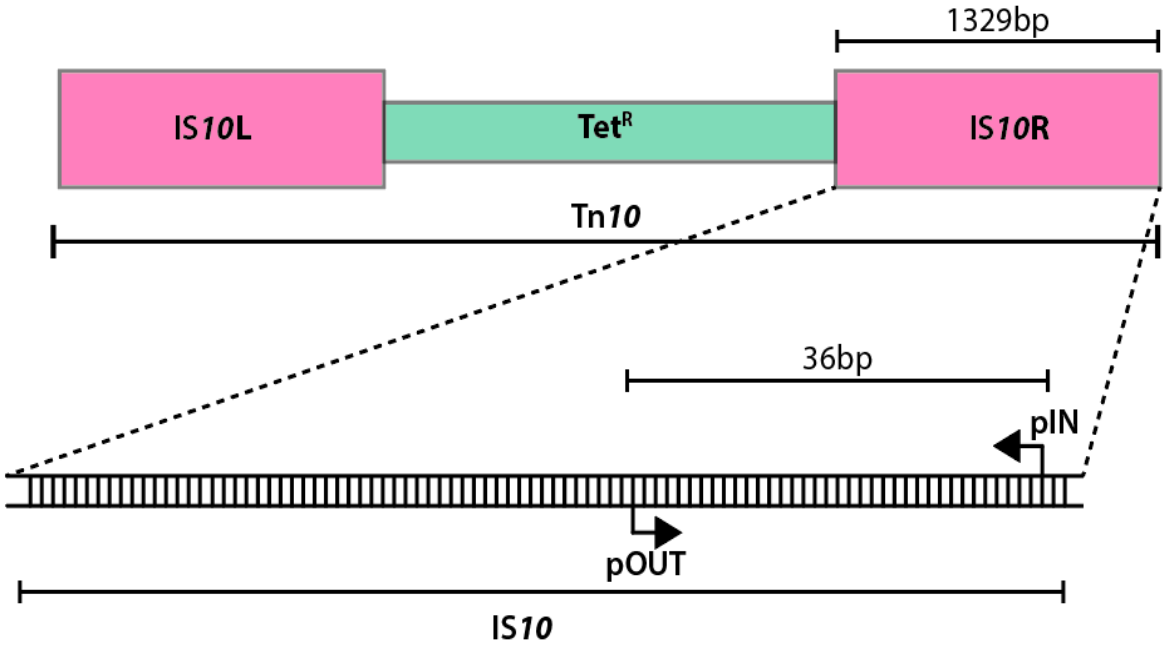


Figure 1.2. Illustration of Tn*10* of *Escherichia coli* with promoters pIN and pOUT indicated within IS*10*. These promoters give rise to two transcripts, pIN begins transcription of the transposase gene, while pOUT begins transcription of a small, *trans*-acting regulatory RNA.

IS*10* employs a *trans*-acting RNA which is encoded in *cis* to regulate the translation of the transposase by occlusion of the ribosome binding site. IS*10* contains two separate promoters, one facing outward toward the 5` end (pOUT) and one facing in toward the tet^r cassette (pIN) which give rise to transcripts known as RNA-OUT and RNA-IN, respectively (36). Importantly, the two transcripts (RNA-IN and RNA-OUT) are overlapping and complementary for approximately 36 base pairs, including a region that encodes the beginning of the transposase coding sequence (35). This allows for one RNA (RNA-OUT) to act in *trans* to base pair to the messenger RNA (RNA-IN) for the transposase gene and down-regulate its ability to be translated. This method of regulation

is reliant on the base-pairing interaction occluding the ribosome binding site, thereby disallowing translation from occurring.

This RNA-IN – RNA-OUT interaction occurs due to a base pairing event between the unstructured loop region of the short hairpin RNA-OUT molecule and a region of mRNA just 5` of the ribosome binding site (RNA-IN). This base pairing triggers a strand displacement reaction, whereby the RNA-OUT molecule is unfolded and base pairs to the mRNA of interest (Figure 1.3). The base pairing interaction creates a double-stranded RNA region which includes the ribosome binding site and start codon of the mRNA, inhibiting translation. The interaction between the unpaired bases in the RNA-OUT and RNA-IN motif is critical and is thermodynamically favoured (37). Once unwound, the RNA-IN – RNA-OUT duplex is a potential substrate for RNaseIII (38) or another analogous endoribonuclease (39). This interaction may further down regulate gene expression through destruction of the message itself, complicating the rational design of derived regulatory elements. Regardless of its mode of action, RNA-IN – RNA-OUT has indeed demonstrated its efficacy in *E. coli in vivo* in the regulation of the transposase gene and it has been conserved over evolutionary time.

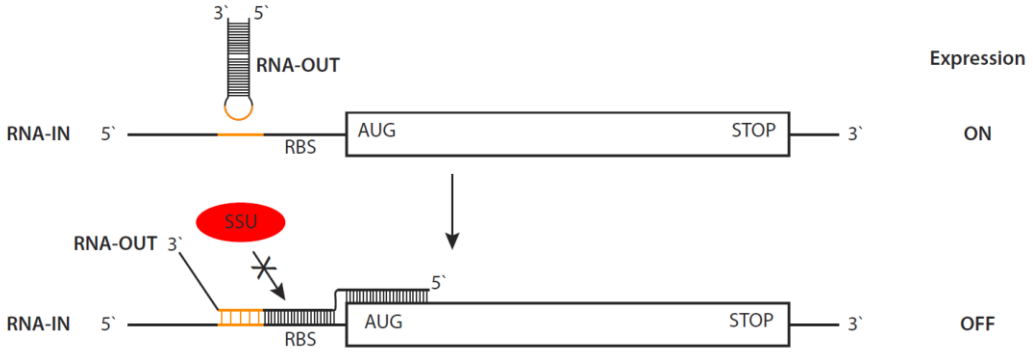


Figure 1.3. Mode of action of naturally-occurring RNA-IN/RNA-OUT pairing to down-regulate translation within *E. coli.* Ribosomal small subunit (SSU) in red.

1.3.2 RNA-IN and RNA-OUT Applications in Synthetic Biology

Based on its modularity, the RNA-IN and RNA-OUT system was recognized as a potentially useful tool from a bioengineering standpoint as well (40). In synthetic biology applications, it is often necessary to express multiple genes at controlled levels which integrate multiple internal and external signals. The RNA-IN – RNA-OUT regulatory system has successfully been utilized as one such tool for use in metabolic engineering networks due to its ability to be regulated with high specificity and efficiency (40). In their work, Mutalik and colleagues generated a library of RNA-IN and RNA-OUT variants by randomizing the 5nt loop sequence of the RNA-OUT and the cognate sequence of the RNA-IN in order to identify orthogonal pairs. They were able to generate pairs of RNA-IN and RNA-OUT which do not cross-react with other RNA species and provide predictable and high levels of repression. Mutalik and colleagues also generated a predictive model for the performance of other variants and experimentally validated this for two of their variants (40).

Additionally, an analysis of the regions of complementarity after extensive mutations was performed (41). These *in vivo* mutations were selected to either strengthen or weaken the interaction between the RNA-IN and RNA-OUT molecules and thereby modulate the level of repression. Their analysis revealed that the IS*10* antisense RNA system has been optimized over millions of years of evolution, and changes to this interaction platform are generally disfavoured. However, compensatory mutations made within the interacting region are tolerated and do not decrease the repression of downstream translation (41).

Once utilized in a bioengineering context, the mechanism of the RNA-IN – RNA-OUT interaction was investigated (42, 43). A reasonable conjecture was that Hfq, a protein

responsible to facilitating base-pairing for many cellular mRNAs and small RNAs, might function as a chaperone for this interaction. Utilizing EMSA analysis, Ross and colleagues were able to determine that both RNA-IN and RNA-OUT do indeed interact with Hfq with nanomolar affinity (42, 43). They further confirmed this via structural probing using hydroxyl radical, RNase A, T1 or V1 cleavage to analyze the areas protected by Hfq (43). This work illustrated the elements of the RNA-IN/OUT interaction that are stabilized by Hfq to accelerate the base-pairing process, as well as structural rearrangements made by Hfq *in vitro*. Thus, a more complete understanding of the RNA-IN/OUT system and its mechanism of action has been gained since the work performed by Mutalik and colleagues (40).

1.4 2A Peptide Regulation of Gene Expression

Viruses have evolved to utilize host gene expression machinery to proliferate and persist (44). The family *Picornaviridae* contains important mammalian pathogens (45) of which several members have devised a strategy to translate their positive-sense RNA genome without multiple translation initiation regions (46). This strategy employs the use of a short peptide linker between the 2A protein and 2B protein within the non-structural region of the viral polyprotein (47). The linker results in ribosome skipping and re-initiation on the 2B protein without any other translational signal such as a stop codon and subsequent start codon being present (Figure 1.4).

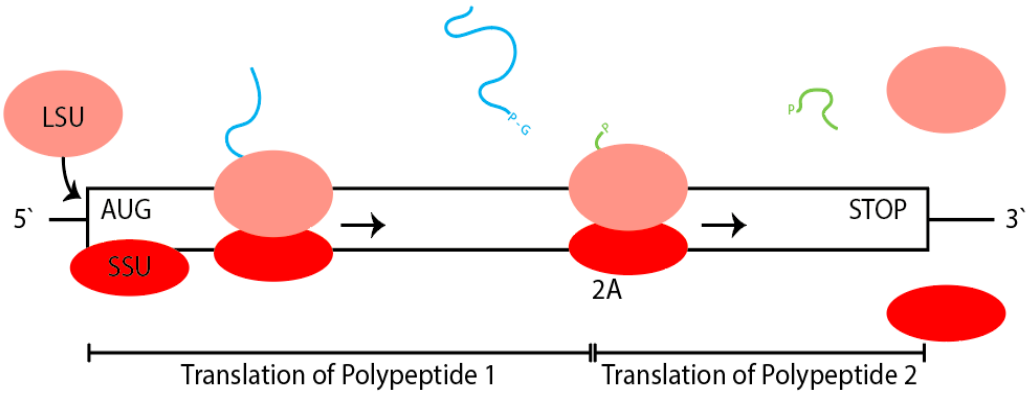


Figure 1.4. Schematic of 2A peptide activity within a eukaryotic organism. The 2A peptide results in a ribosomal skipping event, and re-initiation on a second open reading frame without standard initiation taking place. Ribosomal small subunit (SSU) and large subunit (LSU) are indicated in red and pink, respectively.

1.4.1 2A Peptides in Eukaryotic Synthetic Biology

As the picornaviruses natively infect eukaryotic hosts, mechanistic information regarding 2A peptide activity is mainly available from work performed in eukaryotic cells (48, 49). 2A peptides have clear uses in expression of multiple different proteins from the same mRNA, without using multiple ribosome binding signals (50). Additionally, several members of the family *Picornaviridae* possess such a strategy, with differing ribosome-skipping efficiency depending upon the host (51). Each of the 2A peptide sequences also differ depending on the virus (52). Therefore, sequence and host organism play important roles in the efficiency of 2A peptide cleavage (51).

Within a synthetic biology context, 2A peptides may be used for production of multiple peptides of interest in equimolar amounts, while decreasing the number of total ribosomes required to generate these proteins (53). Often, when synthetic biologists wish to express multiple genes within cells, multicistronic vectors are employed. In the case of eukaryotic cells, an internal ribosome entry site (IRES) may be engineered and used in between open

reading frames to optimize expression of multiple proteins *in vivo* (54). However, IRES sequences have different expression levels, often lower than canonical initiation regions and are generally large (over 500nt) (55). In contrast, the 2A peptide sequence required for ribosomal skipping is between 54 and 66nt and each has a known skipping efficiency. This means that an experimenter can reliably predict the relative amount of gene products made if it is preceded by a given 2A peptide sequence.

In eukaryotic chassis organisms, 2A peptides have been used in many synthetic biology designs in order to optimize the expression of multiple genes within the same construct (51). However, the individual sequences of the *Picornaviridae* 2A peptides have not been demonstrated to operate within bacterial chassis (56).

1.5 Research Objectives

1.5.1 Novel RNA-OUTs Capable of Repressing Translation in a Sequence-Independent Manner

Mutalik and colleagues (40) engineered the area of interaction to include a portion of the coding sequence for sfGFP. However, to repress production of a gene with a different sequence, this interaction area will not be complementary and may result in leaky translation. To circumvent this, an approach was designed so that a N-terminal fusion tag would be generated (e.g. 6x-His, FLAG) encoded by the mRNA containing the RNA-IN element. In this way, any mRNA containing this N-terminal fusion coding sequence may be regulated by the same RNA-OUT species, regardless of downstream coding sequence.

1.5.2 Generation of Translation-Activating RNA-IN/RNA-OUT Pairs

While in the natural Tn*10* and synthetically designed RNA-IN/OUT systems the result is negative regulation of translation, through careful design it may be possible to adapt the RNA-IN/RNA-OUT base pairing interaction into one which activates translation (Figure 3.1D).

1.5.3 Development of 2A Peptide Libraries

2A peptides have the advantage of producing a predictable ratio of proteins due to their ribosome-skipping activity. However, 2A peptides have not been shown to function in bacteria (56).

Therefore, this objective concerns the design and construction of a large 2A peptide library and analyzing its activity *in vivo* within *E. coli.*

References

1. McAdams, H. H., and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 814–819

2. Waters, L. S., and Storz, G. (2009) Regulatory RNAs in Bacteria. *Cell*. **136**, 615–628

3. Shine, J., and Dalgarno, L. (1974) The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 1342–1346

4. Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods*. **11**, 508–520

5. Vigar, J. R. J., and Wieden, H. J. (2017) Engineering bacterial translation initiation — Do we have all the tools we need? *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 3060–3069

6. Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juárez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J., and Morett, E. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PLoS One*. 10.1371/journal.pone.0007526

7. Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**, 64–68

8. Buskirk, A. R., Kehayova, P. D., Landrigan, A., and Liu, D. R. (2003) In vivo evolution of an RNA-based transcriptional activator. *Chem. Biol.* **10**, 533–540

9. Takahashi, M. K., and Lucks, J. B. (2013) A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Res.* **41**, 7577–7588

10. Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210

11. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science*. **297**, 1183–1186

12. Hershey, J. W. B., Sonenberg, N., and Mathews, M. B. (2012) Principles of Translational Control: An Overview. *Cold Spring Harb. Perspect. Biol.* **4**, a011528–a011528

13. Sonenberg, N., and Hinnebusch, A. G. (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*. **136**, 731–745

14. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods*. **10**, 354–360

15. Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W.L., and Breaker, R.R. (2014) A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression *Science*. **306**, 275–279

16. Caron, M. P., Lafontaine, D. A., and Massé, E. (2010) Small RNA-mediated regulation at the level of transcript stability. *RNA Biol.* **7**, 140–144

17. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950

18. Carrier, T. A., and Keasling, J. D. (1999) Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. *Biotechnol. Prog.* **15**, 58–64

19. Breaker, R. R. (2012) Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, 1–15

20. Mellin, J. R., and Cossart, P. (2015) Unexpected versatility in bacterial riboswitches. *Trends Genet.* **31**, 150–156

21. Lavi-Itzkovitz, A., Peterman, N., Jost, D., and Levine, E. (2014) Quantitative effect of target translation on small RNA efficacy reveals a novel mode of interaction. *Nucleic Acids Res.* **42**, 12200–12211

22. Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004) The bacterial Sm-like protein Hfq: A key player in RNA transactions. *Mol. Microbiol.* **51**, 1525–1533

23. Urban, J. H., and Vogel, J. (2007) Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res.* **35**, 1018–1037

24. Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y. (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.* **31**, 170–174

25. Lahiry, A., Stimple, S. D., Wood, D. W., and Lease, R. A. (2017) Retargeting a Dual-Acting sRNA for Multiple mRNA Transcript Regulation. *ACS Synth. Biol.* **6**, 648–658

26. Petersen, C. (1989) Long-range translational coupling in the rplJL-rpoBC operon of Escherichia coli. *J. Mol. Biol.* **206**, 323–332

27. Wulczyn, F. G., Bölker, M., and Kahmann, R. (1989) Translation of the bacteriophage Mu mom gene is positively regulated by the phage com gene

product. *Cell*. **57**, 1201–1210

28. Blaszczyk, J., Tropea, J. E., Bubunenko, M., Routzahn, K. M., Waugh, D. S., Court, D. L., and Ji, X. (2001) Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure*. **9**, 1225–1236

29. Maisonneuve, E., Shakespeare, L. J., Jorgensen, M. G., and Gerdes, K. (2011) Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13206–13211

30. Neubauer, C., Gao, Y. G., Andersen, K. R., Dunham, C. M., Kelley, A. C., Hentschel, J., Gerdes, K., Ramakrishnan, V., and Brodersen, D. E. (2009) The Structural Basis for mRNA Recognition and Cleavage by the Ribosome-Dependent Endonuclease RelE. *Cell*. **139**, 1084–1095

31. Simons, R. W. (1988) Naturally occurring antisense RNA control--a brief review. *Gene*. **72**, 35–44

32. Raser, J. M., and O'Shea, E. K. (2005) Molecular biology - Noise in gene expression: Origins, consequences, and control. *Science.* **309**, 2010–2013

33. Hickey, D. (1992) Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. *Genetica*

34. Kleckner, N., Barker, D. F., Ross, D. G., and Botstein, D. (1978) Properties of the translocatable tetracycline resistance element Tn10 in Escherichia coli and bacteriophage lambda. *Genetics*. **90**, 427–461

35. Simons, R. W., Hoopes, B. C., Mcclure, W. F. L., and Kleckner, N. (1983) Three Promoters near the Termini of IS10: pIN , pOUT , and pIII. *Cell*. **34**, 673–682

36. Simons, R. W., and Kleckner, N. (1983) Translational control of IS10 transposition. *Cell*. **34**, 683–691

37. Kittle, J. D., Simons, R. W., Lee, J., and Kleckner, N. (1989) Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5′ end of the target RNA and a loop in the anti-sense RNA. *J. Mol. Biol.* **210**, 561–572

38. Case, C. C., Roels, S. M., Jensen, P. D., Lee, J., Kleckner, N., and Simons, R. W. (1989) The unusual stability of the IS10 anti-sense RNA is critical for its function and is determined by the structure of its stem-domain. *EMBO J*. **8**, 4297–4305

39. Case, C. C., Simons, E. L., and Simons, R. W. (1990) The IS10 transposase mRNA is destabilized during antisense RNA control. *EMBO J.* **9**, 1259–1266

40. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454

41. Jain, C. (1995) IS10 antisense control in vivo is affected by mutations throughout the region of complementarity between the interacting RNAs. *J. Mol. Biol.* **246**, 585–594

42. Ross, J. A., Wardle, S. J., and Haniford, D. B. (2010) Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Mol. Microbiol.* **78**, 607–621

43. Ross, J. A., Ellis, M. J., Hossain, S., and Haniford, D. B. (2013) Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA*. **19**, 670–684

44. Ghedin, E., Sengamalay, N. A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., Spiro, D. J., Sitz, J., Koo, H., Bolotov, P., Dernovoy, D., Tatusova, T., Bao, Y., St George, K., Taylor, J., Lipman, D. J., Fraser, C. M., Taubenberger, J. K., and Salzberg, S. L. (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature*. **437**, 1162–1166

45. Stanway, G. (1990) Structure , function and evolution of picornaviruses. *J. Gen. Virol.* **71**, 2483–2501

46. Forss, S., Strebel, K., Beck, E., and Schaller, H. (1984) Nucleotide sequence and genome organization of foot-and-mouth disease virus. *Nucleic Acids Res.* **12**, 6587–6601

47. Hughes, P. J., and Stanway, G. (2000) The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation. *J. Gen. Virol.* **81**, 201–207

48. Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., and Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal "skip." *J. Gen. Virol.* **82**, 1013–1025

49. Yang, X., Cheng, A., Wang, M., Jia, R., Sun, K., Pan, K., Yang, Q., Wu, Y., Zhu, D., Chen, S., Liu, M., Zhao, X. X., and Chen, X. (2017) Structures and corresponding functions of five types of picornaviral 2A proteins. *Front. Microbiol.* **8**, 1–14

50. Szymczak, A. L., and Vignali, D. A. (2005) Development of 2A peptide-based strategies in the design of multicistronic vectors. *Expert Opin. Biol. Ther.* **5**, 627–638

51. Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A., and Choi, S. Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*. **6**, 1–8

52. Luke, G. A., de Felipe, P., Lukashev, A., Kallioinen, S. E., Bruno, E. A., and Ryan, M. D. (2008) Occurrence, function and evolutionary origins of "2A-like" sequences in virus genomes. *J. Gen. Virol.* **89**, 1036–1042

53. De Felipe, P., Hughes, L. E., Ryan, M. D., and Brown, J. D. (2003) Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth disease virus 2A peptide. *J. Biol. Chem.* **278**, 11441–11448

54. Pfleger, B. F., Pitera, D. J., Smolke, C. D., and Keasling, J. D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24**, 1027–1032

55. Chan, H. Y., Sivakamasundari, V., Xing, X., Kraus, P., Yap, S. P., Ng, P., Lim, S. L., and Lufkin, T. (2011) Comparison of IRES and F2A-based locus-specific multicistronic expression in stable mouse lines. *PLoS One*. 10.1371/journal.pone.0028885

56. Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M. D. (2001) The 'cleavage'activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like'sequences. *J. Gen. Virol.* **82**, 1027–1041

2 *Trans*-Acting RNA Down-Regulation of Translation

2.1 Introduction

2.1.1 Background

Biological and metabolic engineering depend on the ability to reliably regulate translation of heterologous proteins (1). Antisense RNA offers an attractive method for regulation of gene expression, and has been used extensively (2). However, the specificity of the interaction between antisense and sense RNA is of concern, as base-pairing interactions can occur between regions lacking perfect complementarity (i.e. bulged nucleotides, wobble base pairs) (2–6). Adaptation of the well-studied RNA-IN and RNA-OUT system based upon *E. coli* Tn*10* to behave orthogonally and reliably repress translation for use in synthetic biology was accomplished in 2012 (7). This strategy was used to generate a library of repressing *trans*-acting RNA molecules with limited cross-talk and reliable repression profiles (7). Several cognate RNA-IN and RNA-OUT pairs from within this library were selected for *in vivo* experimentation and validation. These pairs differed in their respective sense/antisense, or seed/anti-seed regions responsible for the initial base-pairing interaction prior to strand displacement and eventual down-regulation of translation.

Alternative approaches to genetic regulation utilizing *trans*-acting antisense RNA molecules have been employed previously (8–10). These interactions may indeed be

engineered to be orthogonal (8) or may be modifications of an existing natural system to be highly specific (7). The interaction between sense-antisense RNAs which impact translation can be thermodynamically governed, with highly stable duplex formation occurring predictably (8, 11). However, more recent work illustrates that with respect to sense-antisense regulation, this may not always be the case (12, 13) indicating that a complex interplay of factors is at work in the regulatory pathway. Within the RNA-IN/RNA-OUT framework the interacting sequences are within the unpaired region at the apex of the RNA-OUT stem-loop structure and a complementary motif just upstream of the ribosome binding site of the respective RNA-IN molecule (Figure 2.1). This region is critical for the recognition and specificity of this system, as it is the site of strand displacement and duplex formation which drives the eventual down-regulation of translation (11, 14). The seed/anti-seed regions have been the subject of extensive study (7, 14) with mutational analysis revealing that sequence alterations within these regions significantly decreases down-regulation of translation. Another important aspect of the design of these molecules is the interaction between RNA-IN and RNA-OUT. Once duplex formation occurs, it often extends into the coding sequence (Figure 2.1 A) of the downstream open reading frame. From a synthetic biology perspective this is a disadvantage as it necessitates the re-design of a *trans*-acting RNA-OUT molecule for each and every new coding sequence one wishes to control, limiting the interchangeability of the reported regulatory elements (7). Additionally, the design of Mutalik and colleagues only used a single fluorescent reporter (sfGFP), which does not allow for calibration to another gene or control for plasmid copy number within each individual cell. Therefore, an appropriate adaptation to the outlined design is to include an invariant sequence just downstream of the start codon to isolate the sequence of the gene

of interest from the regulatory module (sequence). The use of such a spacer sequence will facilitate a modular approach in which translation of an mRNA can be regulated irrespective of the gene of interest and enables the inclusion of well characterized N-terminal tag sequences (e.g. 6x His, FLAG) which interact with the *trans*-acting regulatory RNA. Moreover, the coding sequence is interchangeable, allowing for an array of mRNAs to be regulated in response to a specific *trans*-acting regulatory RNA species. A second adaptation is the inclusion of an additional fluorescent reporter acting as a noise module within the cell. Because this regulatory system is reliant upon RNA-RNA interactions, it is interesting to investigate the impact the interaction has upon transcript level. Therefore, a third consideration is to monitor the transcript levels during the experiments and to determine what effect the RNA-IN/OUT duplex formation has on mRNA levels.

In order to account for these criteria for improvement, a novel design was employed to assess and develop universally active RNA-IN/OUT pairs (Figure 2.1 B, C). A dual-fluorescent reporter system was used to measure the baseline level for transcriptional and translational variability within the cell as a control. By comparing the baseline level of one fluorescent reporter to the one being regulated, a relative measurement of regulation by the *trans*-acting RNA can be made. An invariant sequence within the mRNA was inserted to ensure that base pairing could be extended to other mRNAs regardless of coding sequence as long as the invariant sequence is present. Finally, mRNA levels of the transcripts were monitored to examine the effect that duplex formation might have on the lifetime of the RNA species in question.

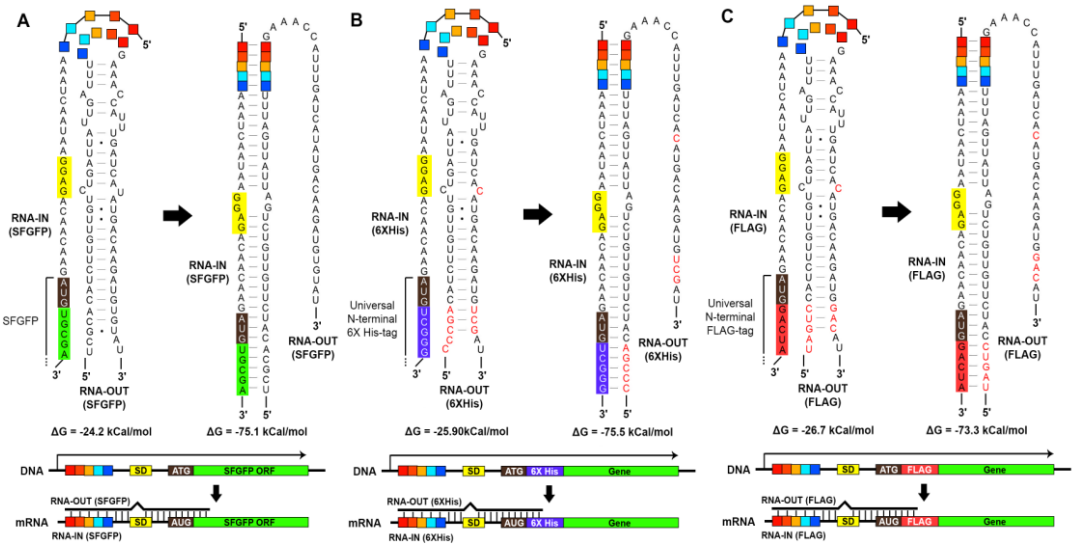


Figure 2.1. Illustration of RNA-IN and RNA-OUT base-pairing interactions with ΔG values computed using mfold software for **A)** a fluorescent protein coding sequence, **B)** a coding sequence with an N-terminal 6x His fusion, and **C)** a coding sequence with an N-terminal FLAG tag sequence.

2.1.2 Design of Genetic Constructs

With respect to the reporter plasmid, two fluorescent proteins were selected to be expressed from a single transcript. In this way, promoter noise and transcript number should be similar across experiments. The fluorescent proteins themselves were selected to have similar excitation wavelengths (allowing for excitation with a single wavelength) and high brightness for ease of measurement. Specifically, the fluorescent proteins eGFP and mBeRFP (15) have unique fluorescent properties which make them suitable for the analysis pipeline carried out: eGFP has an excitation maximum at 488nm and an emission maximum at 510nm, while mBeRFP has an excitation maximum at 450nm and an emission maximum at 610nm. Given their spectral properties (Figure 2.2) excitation at 488nm will excite both proteins to a degree sufficient to read-out fluorescence intensity. The ability to detect both fluorescent proteins when excited at 488nm was validated prior to proceeding. As expected, the mBeRFP reached approximately half its maximal

brightness as it was being excited at approximately half its maximal absorption wavelength as can be seen in Figure 2.7.

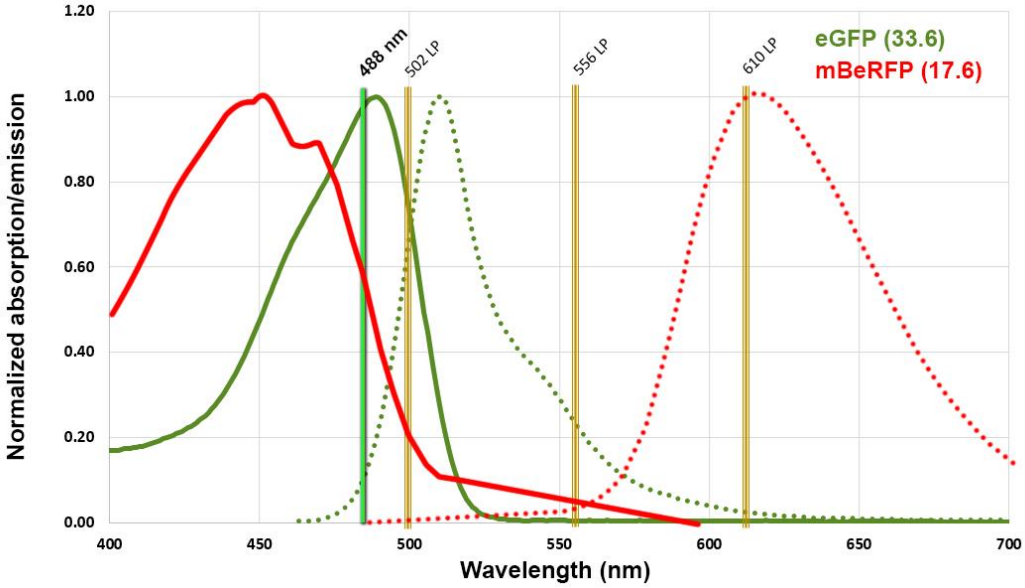


Figure 2.2. Spectral properties of both reporter fluorescent proteins selected for study (15, 20, 31). Excitation spectra are represented by solid lines, while emission spectra are represented by dotted lines. Additionally, the brightness of each fluorescent protein once matured is noted in parentheses.

The ability to induce *trans*-acting RNA expression in response to IPTG was an additional consideration. As this approach was designed to have applications in biological and metabolic engineering, the importance of having a reliable method of turning on expression to modulate response is important. For this, a well-characterized LacO-1(16, 17) promoter was used in the plasmid pRNA-OUT in order to allow expression of the corresponding RNA-OUT to be activated. Expression of the single mRNA encoding both fluorescent proteins was driven by a well characterized (17), constitutive promoter from the iGEM Registry of Standard Biological Parts (BBa_J23119). In this way, a predictable, well established and characterized number of transcripts for regulation are available within the cell (17).

Additionally, the RNA-IN and RNA-OUT motifs selected (A4/S4 and A5/S5) were specifically chosen for their previously described orthogonality (7). Each RNA-OUT should be acting on only its cognate RNA-IN sequence, and thereby repressing translation of only one type of open reading frame within the cell (Figure 2.3). To expand the utility of a specific RNA-IN/OUT pair, the interaction was engineered to included tag sequences at the 5` end of the open reading frame for each respective fluorescent protein. A corresponding region of complementarity was engineered into the *trans*-acting RNA-OUT to allow for universal repression (Figure 2.1 B, C). The 6x His and FLAG tag sequences were selected to enable purification or tracking of the downstream protein products. In essence, by creating N-terminal fusion proteins, one may be able to reliably repress translation of several open reading frames with a single species of *trans*-acting RNA-OUT. This is an important design criterion not previously addressed, improving the utility of this approach.

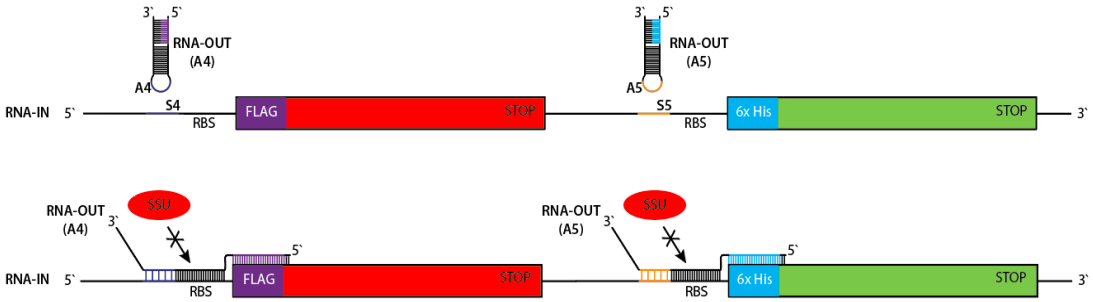


Figure 2.3. Bicistron regulation by separate RNA-OUT *trans*-acting regulatory RNAs. The A4 motif interacts with the S4 motif and subsequent unwinding of the RNA-OUT down-regulates production of mBeRFP while the A5 and S5 motifs interact which results in down-regulation of eGFP. Ribosomal small subunit (SSU) is indicated in red.

2.2 Methods

2.2.1 Construction of Genetic Elements

Plasmids were designed to facilitate measurement of translation down-regulation by a synthetic, inducible *trans*-acting RNA molecule. To accomplish this, and mimic the natural Tn*10* system, two plasmids were engineered. The first, was a high-copy number plasmid harbouring an RNA-OUT sequence of interest under control of the LacO-1 inducible promoter. The plasmid primarily encoding the *trans*-acting RNA (pRNA-OUT) was derived from pBbE6a-RFP (Addgene plasmid #35826). Construction of pRNA-OUT was accomplished by restriction digestion of pBbE6a-RFP with EcoRI and XhoI (NEB) (1µg of plasmid DNA was digested in a 50µL reaction with 10 units of each restriction enzyme) and subsequent removal of the RFP coding sequence by agarose gel extraction (1% agarose gel, 10V/cm for 30 minutes). Oligonucleotides (Appendix E) were obtained (IDT) specific to each unique desired RNA-OUT molecule. By oligonucleotide extension and subsequent 20µL PCR (using Phusion high-fidelity DNA polymerase HF mastermix, Thermo Fisher; 0.2µM primers; 30 extension cycles of 15 seconds), a full-length RNA-OUT cassette was generated. In a similar fashion to the backbone plasmids, the full-length RNA-OUT cassette was digested with EcoRI and XhoI (NEB). The digested RNA-OUT cassette was then ligated into the backbone vector in a 10µL ligation using T4 DNA ligase (100ng of digested plasmid DNA along with 300ng of insert DNA were combined in a 10µL reaction with 400 units of T4 DNA ligase overnight at 16°C) (NEB) (Figure 2.4).

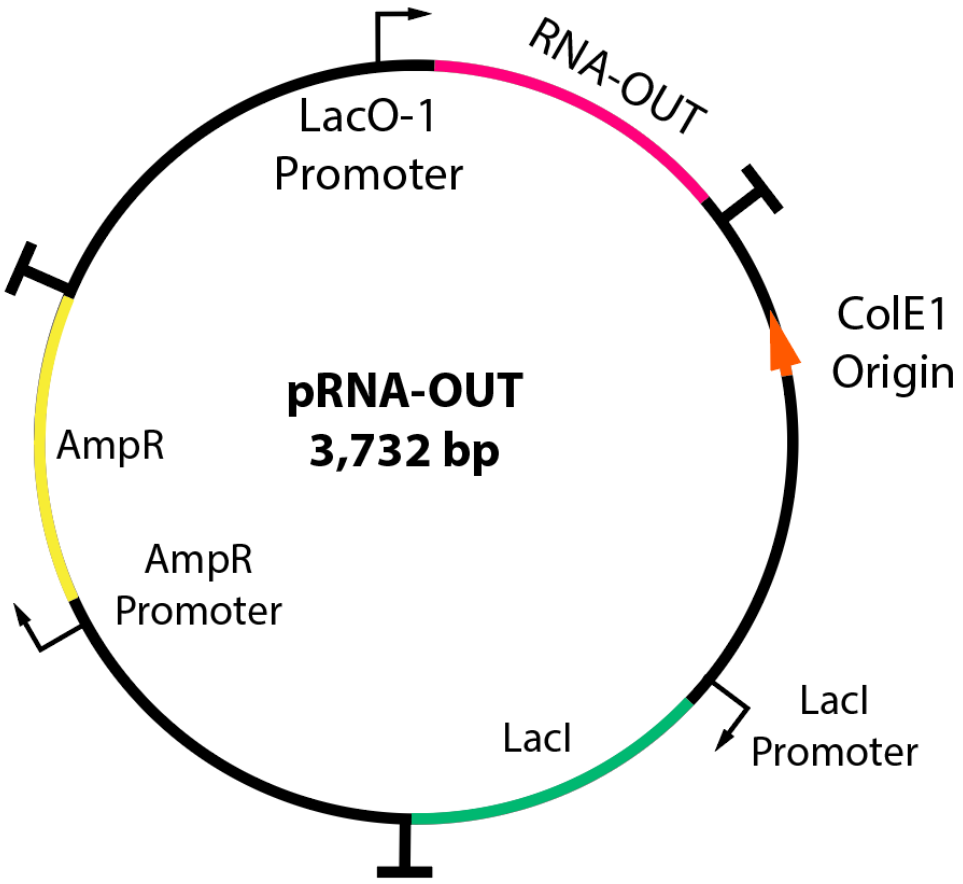


Figure 2.4. Schematic diagram of a high-copy number plasmid encoding an inducible *trans*-acting RNA capable of repressing a specific RNA-IN. Production of the RNA-OUT species of interest is activated by induction with IPTG. Right-angled arrows indicate promoter regions while T-shaped segments indicate terminator regions.

The second plasmid designed was a low-copy number plasmid harbouring two fluorescent reporter genes under control of a constitutive, strong promoter with distinct, specific RNA-IN motifs upstream of their respective ribosome binding sites. This plasmid was derived from pBBRBB-eGFP (Addgene plasmid #32549). Specific RNA-IN constructs with distinct 5` tag sequences (S4-FLAG, S5-6xHis) were designed and obtained as oligonucleotides (IDT) and subsequently underwent oligonucleotide extension and 20µL PCR (using Phusion high-fidelity DNA polymerase HF mastermix, Thermo Fisher; 0.2µM primers; 30 extension cycles of 15 seconds) to generate a full-length RNA-IN motif. A second section of the reporter plasmid was ordered as a gBlock

(IDT) (Appendix E) and contained the coding sequence for mBeRFP as well as a different RNA-IN motif. This segment was digested with EcoRI and BamHI (1µg of DNA was digested in a 50µL reaction with 10 units of each restriction enzyme) (NEB), while the previously mentioned RNA-IN construct was cut with BamHI and NdeI for insertion just upstream of the coding sequence for eGFP. pBBRBB-eGFP was digested with NdeI and EcoRI as described above. All segments were gel-purified from a 1% agarose gel run at 10V/cm for 30 minutes. A 10µL, three-part ligation was performed on the resultant fragments using T4 DNA ligase (100ng of digested plasmid DNA along with 300ng of each insert DNA were combined in a 10µL reaction with 400 units of T4 DNA ligase overnight at 16°C) (NEB) to generate a full plasmid containing both reporter genes with their own ribosome binding sites as well as two distinct RNA-IN motifs (Figure 2.5). In the case of a single fluorescent reporter gene (eGFP), both pBBRBB-eGFP and a PCR product of the construct containing the desired RNA-IN motif, invariant tag and BBa_J23119 strong constitutive promoter were digested with NdeI and BamHI (1µg of DNA was digested in a 50µL reaction with 10 units of each restriction enzyme) (NEB). Thereafter, they were gel purified and ligated together (Figure 2.6). The use of a high-

copy regulatory plasmid and a low-copy reporter plasmid should reflect the natural stoichiometry of RNA-OUT to RNA-IN of approximately 20:1 (18).

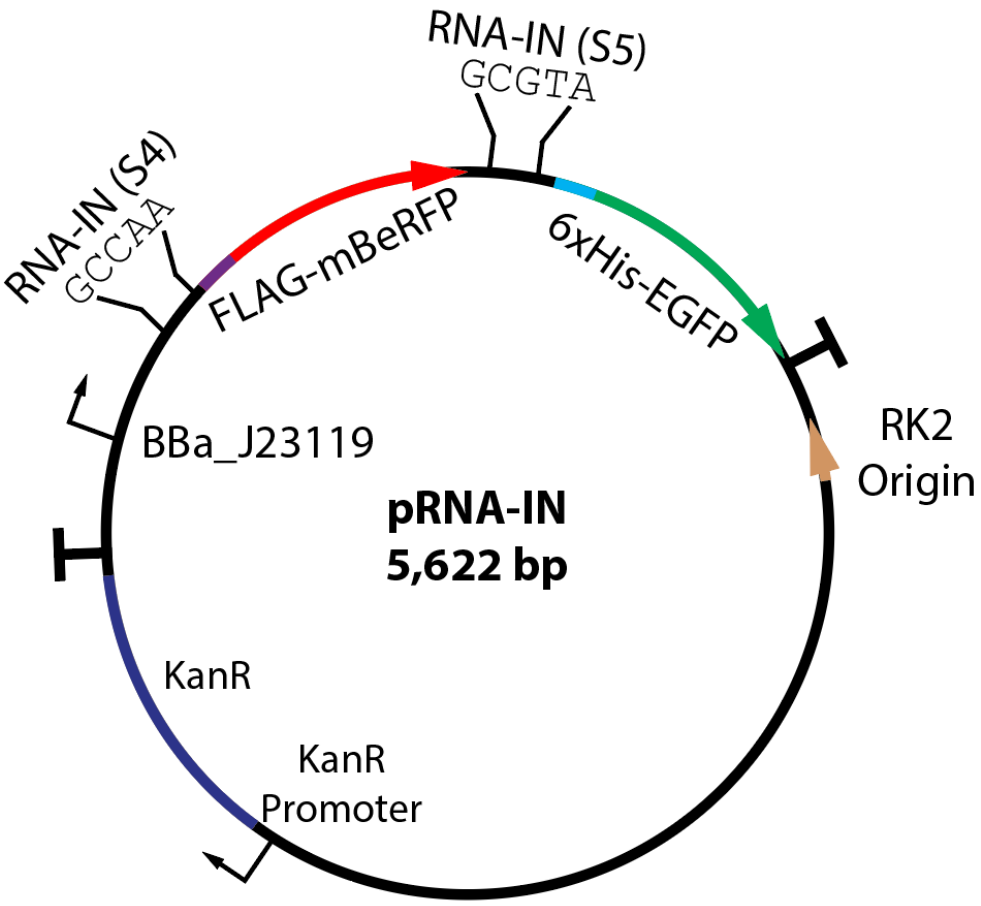


Figure 2.5. Schematic diagram of low-copy number reporter plasmid used in RNA-IN/RNA-OUT examination. A strong, constitutive promoter drives production of a bicistronic transcript encoding both a red (red) and green fluorescent protein (green) under control of separate unique RNA-OUT molecules. Right-angled arrows indicate promoter regions while T-shaped segments indicate terminator regions. Purple and cyan are indicative of FLAG and 6x His tags, respectively.

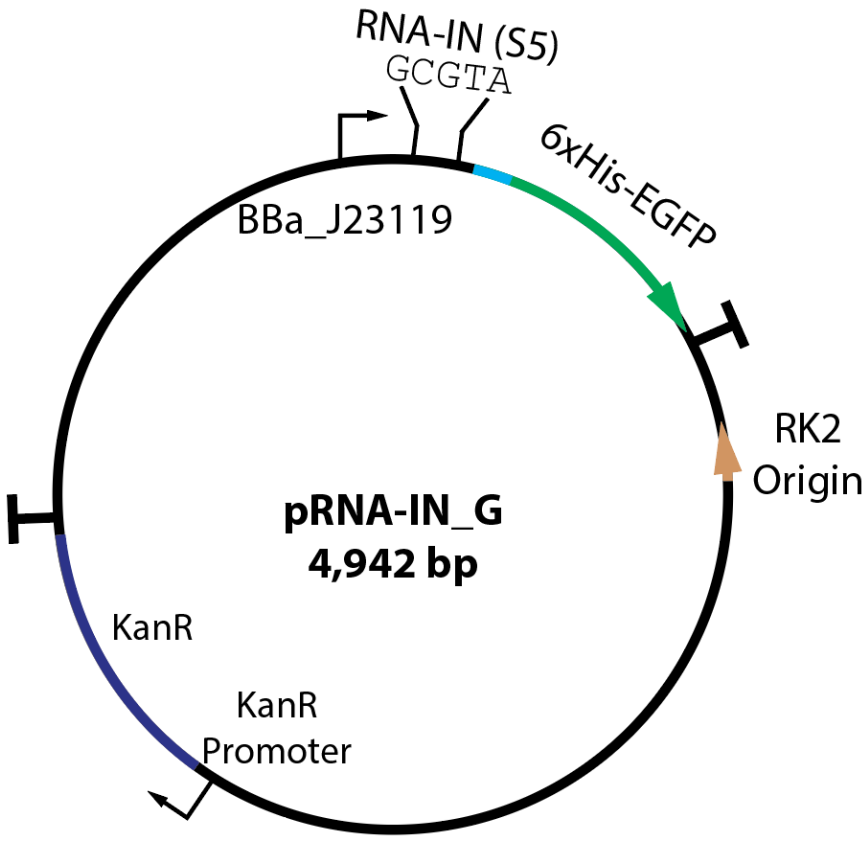


Figure 2.6. Schematic diagram of low-copy number single reporter plasmid used in RNA-IN/RNA-OUT examination. A strong, constitutive promoter drives production of a single cistron transcript encoding a green fluorescent protein (green) with a 6x His tag (cyan) under control of an engineered, unique RNA-OUT molecule. Right-angled arrows indicate promoter regions while T-shaped segments indicate terminator regions.

2.2.2 Cell Growth and Fluorescence Analysis

2.2.2.1 Electroporation

50 mL LB media (10g/L NaCl, 10g/L tryptone, 5g/L yeast extract) was inoculated with an overnight culture (1:100 dilution) and incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000) for approximately 2 hours, until an OD_{600nm} of 1.0 was reached. Cells were centrifuged at 4°C for 10 minutes at 3000 x g using a Sorvall Legend RT. Exponentially growing *Escherichia coli* DH5α cells (Thermo) were washed twice with ice-cold sterile H_2O. The pellet resulting from centrifugation of 50 mL of the

harvested culture suspension was resuspended in 1.5 mL ice-cold sterile H_2O. 100 μL of the cell suspension (i.e. washed DH5α cells in cold, sterile water) was mixed with 1 μL of pRNA-IN or pRNA-OUT ligation. This mixture was transferred to a 0.2 cm electroporation cuvette (Fisher) and 2.5 kV (12.5 kV/cm) was applied using a BioRad Gene Pulser Xcell resulting in a time constant of ~4.5 msec. Cells were diluted immediately with 800 μL room temperature SOC media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose). The cell suspension was incubated for 60 min at 37 °C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). 250 μL of cell suspension was plated on LB agar plates with 50μg/mL kanamycin (for pRNA-IN transformants) or 100μg/mL ampicillin (for pRNA-OUT transformants) and incubated overnight at 37°C. Colonies were selected, and plasmids were isolated using a BioBasic EZ-10 Spin Column DNA Miniprep Kit and sent for Sanger sequencing (Genewiz). Successfully ligated plasmids were co-transformed (one of each in a cognate pair of pRNA-IN and pRNA-OUT) into *Escherichia coli* BL21-Gold(DE3) (Agilent) cells as described above.

2.2.2.2 Flow Cytometry

5mL overnight cultures of co-transformants were diluted to an OD_{600nm} of 0.05 in 50 mL fresh LB media with 50μg/mL kanamycin and 100μg/mL ampicillin. Cultures were incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). IPTG was added to a final concentration of 1mM to one culture at an OD_{600nm} of 0.3, while one culture remained uninduced, both were grown until reaching an OD_{600nm} of 1.0. Cells were then harvested by centrifugation at 5000 x g for 2 min using a Sorvall Legend Micro 17 (Thermo). The cell pellet was washed in 1 mL ice-cold sterile 1 X Phosphate Buffered

Saline (PBS: 0.137M NaCl, 0.00270M KCl, 0.0100M Na_2HPO_4, 0.0180M KH_2PO_4) before resuspension in 500 µL sterile 1 X PBS. The resuspended cells were allowed to incubate for 90 min at 20°C to facilitate the maturation of fluorescent proteins and subsequently diluted 10X in 1X PBS and red and green fluorescence was analyzed upon excitation at 488nm. Using a BD FACS Aria Fusion, eGFP fluorescence was measured with a 502LP mirror and bandpass filter of 530/30nm, mBeRFP fluorescence was measured using a 600LP mirror and a bandpass filter of 610/20nm. A scatter gate was set using the forward and side scatter area of *E. coli* cells harbouring no plasmid ($1 X 10^5$ cells from each culture were analyzed). Gates were set, including no less than 50,000 cells and analysis performed in Flowing Software v2.5.1 (Flowing Software, Perttu Terho).

2.2.2.3 Analysis of Fluorescent Protein Stability *in vivo*

To assess the turnover and degradation of fluorescent proteins *in vivo*, overnight cultures of *E.coli* BL21-Gold (DE3) (Agilent) co-transformants were diluted to an OD_{600nm} 0.05 in 50 mL fresh LB media with 50µg/mL kanamycin and 100µg/mL ampicillin. Cultures were incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). IPTG was added to a final concentration of 1mM at an OD_{600nm} 0.3, while one culture remained uninduced, both were grown until reaching an OD_{600nm} 1.0. Cells were then harvested by centrifugation at 5000 x g for 10 min (Sorvall Legend RT). The cell pellet was washed with ABT minimal media (19, 20) containing 15.1mM $(NH_4)_2SO_4$, 42.2mM Na_2HPO_4, 22.0mM KH_2PO_4, 51.4mM NaCl, 0.8mM Na_2SO_4, 2.1mM $MgCl_2$, 90.1µM $CaCl_2$, 1.7µM $FeCl_3 \cdot 7H_2O$, and 9.4µM thiamine. The pellet was then resuspended in 50mL of ABT minimal media and incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ

4000). Samples were obtained at 30 min intervals for 24 hours and the OD_{600nm} was measured, an additional sample was saved for flow cytometric analysis. Each sample was washed in 1 mL ice-cold sterile 1 X PBS before resuspending in 500 μL sterile 1 X PBS. The resulting suspension was then 10-fold diluted 1X PBS and analyzed. Using a BD FACS Aria Fusion, exciting eGFP at 488nm eGFP fluorescence was measured with a 502LP mirror and bandpass filter of 530/30nm. A similar procedure as described above was used to gate and analyze samples.

2.2.2.4 Fluorescence Spectroscopy

In an approach similar to that outlined above, overnight cultures of co-transformants were diluted to an OD_{600nm} = 0.05 in 5mL fresh LB media with 50μg/mL kanamycin and 100μg/mL ampicillin. Cultures were incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). IPTG was added to a final concentration of 1mM at an OD_{600nm} = 0.3, while one culture remained uninduced, both were grown until reaching an OD_{600nm} = 1.0. Cells were then harvested by centrifugation at 5000 x g for 2 min. The cell pellet was washed in 1 mL ice-cold sterile 1 X PBS before resuspending in 500 μL sterile 1 X PBS. Again, the cells were allowed to incubate for 90 min at 20°C to facilitate the maturation of fluorescent proteins and 200 μL of the resulting cell suspension was placed in a 6mm x 6mm x 50mm quartz cuvette (Starna Cells) and cuvette holder (Varian). The fluorescence of the cell suspension was analyzed using an emission scan on a Cary Eclipse Fluorescence Spectrophotometer (Varian, Inc. Agilent Technologies). Upon baseline setup of the spectrophotometer with PBS, relative fluorescence was obtained for each sample. Excitation of both fluorophores at 488nm and emission read from 500nm-700nm (1nm step size) allowed for relative determination of fluorescent

protein abundance in both the absence and presence of a specific *trans*-acting regulatory RNA. The peak fluorescence in both the induced and uninduced treatments was used for comparison and to calculate a percent repression value.

$$Percent\ Repression = 1 - \left(\frac{Maximum\ Fluorescence_{\ Induced\ Sample}}{Maximum\ Fluorescence_{\ Uninduced\ Sample}} \times 100\%\right)$$

(Equation 2.1)

2.2.3 Transcript Analysis by RT-qPCR

2.2.3.1 Primer Design

Reference gene primer sequences were chosen based upon previously published experimental data (21) to amplify an approximately 100bp region of several described reference genes (*hcaT, rrsA, cysG*). In a similar fashion, primers were designed to amplify an approximately 100bp region of the transcript encoded on pRNA-IN corresponding to either mBeRFP or eGFP. The products generated by the primers after reverse transcription and subsequent PCR amplification were verified by gel electrophoresis (Appendix A). The primer sequences are summarized in Table 2.1.

Table 2.1. Primers used in quantitative RT-PCR. These primer sequences were used to specifically amplify regions of mBeRFP, eGFP, cysG, hcaT and rrsA for transcript stability analysis.

Primer Name	**Primer Sequence (5`- 3`)**
mBeRFP CDS FWD	CTACATGGCCCTGAAGCTC
mBeRFP CDS REV	TCGGCCTCCTTGATTCTTTC
eGFP CDS FWD	GGCATCAAGGTGAACTTCAAGA
eGFP CDS REV	TTTGCTCAGGGCGGACT
cysG FWD	TTGTCGGCGGTGGTGATGTC
cysG REV	ATGCGGTGAACTGTGGAATAAACG

hcaT FWD	GCTGCTCGGCTTTCTCATCC
hcaT REV	CCAACCACGCTGACCAACC
rrsA FWD	CTCTTGCCATCGGATGTGCCCA
rrsA REV	CCAGTGTGGCTGGTCATCCTCTCA

2.2.3.2 RNA Isolation and Quantitative Real-Time PCR

E. coli BL21-Gold (DE3) cells co-transformed with pRNA-IN containing RNA-IN motifs S4 and S5, and pRNA-OUT containing the RNA-OUT motif A5 were grown overnight in 5mL of LB media at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). Overnight cultures were then diluted to an OD_{600nm} 0.05 in 5mL fresh LB media with 50μg/mL kanamycin and 100μg/mL ampicillin. One culture had IPTG added to a final concentration of 1mM when at an OD_{600nm} 0.3, while the other culture remained uninduced. Both cultures were then grown until reaching an OD_{600nm} of 1.0.

Total RNA from *E. coli* was extracted using the Bio Basic EZ-10 Spin Column Total RNA Miniprep Kit according to the manufacturer's instructions. Total RNA was collected for each sample. RNA concentration was quantified using a BioDrop μLITE spectrophotometer, and the 260nm/280nm and 260nm/230nm absorbance ratios were examined for protein and solvent contamination. The integrity of all RNA samples was confirmed by a 1% denaturing formaldehyde agarose gel (10V/cm for 30 minutes) (Appendix B). 800ng of total RNA was reverse transcribed in a total volume of 20μL containing qScript Reverse Transcriptase (Quanta Biosciences) for 60 min at 42°C according to the manufacturer's instructions. The reaction was terminated by incubation at 85°C for 5 min.

The cDNA levels were then quantified using an Applied Biosystems StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) with SYBR Green detection. Samples were measured in triplicate in 8 tube strips (Thermo Fisher Scientific) containing a 50µL reaction mixture with 25µL PerfeCta SYBR Green SuperMix, ROX (2X) (Quanta Biosciences), 100nM forward primer. 100nM reverse primer, 10µL of cDNA template and remaining volume nuclease-free water. Real-time PCR was performed using an initial denaturation of 10 min at 95°C, followed by 40 cycles of 15 sec at 95°C, 60 sec at 52°C, and 30 sec at 72°C. Serial dilutions were performed for each reference gene to generate a concentration range sufficient to establish a standard curve. Confirmation of a single product being amplified was performed by fluorescent detection at the annealing phase and during subsequent dissociation curve analysis (15 sec at 95°C, one minute at 60°C-ramping to 95°C at an increase 0.3°C per cycle). The threshold cycles (Ct) and other data analysis were performed using StepOne Software v2.3 (Thermo Fisher Scientific). Real-time PCR quantifications were all done with no-template and no-reverse transcriptase controls in tandem (Appendix C). The ratio between a target and reference gene was calculated using the following equation:

$$Ratio = \frac{(E_{Target})^{\Delta C_t\ target}}{(E_{Reference})^{\Delta C_t\ reference}} \qquad \text{(Equation 2.2) (22)}$$

Where E is the efficiency of amplification, which can be determined as the slope of a plot of threshold cycle for a target as a function of $-\log_{10}$ concentration of the number of target molecules (Figure 2.16). ΔCt for both the target and reference case are calculated as the Ct of induced sample subtracted from the Ct of uninduced sample for target and reference genes, respectively. A ratio between a target and reference gene can then be calculated by

comparing the obtained ΔCt values. Standard error is calculated using Equation 2.3 for the measurement of ΔCt as each measurement is performed in triplicate:

$$SEM = \frac{\sqrt{\frac{\sum|x-\mu|^2}{n}}}{\sqrt{n}} \qquad \text{(Equation 2.3)}$$

Where x is any value within the data set, μ is the mean of the data set and n is the number of data points within the data set.

Additionally, a single mean t-test was performed using Equation 2.4, and a corresponding assignment of a p-value.

$$t = \frac{\bar{x}-\mu}{SEM} \qquad \text{(Equation 2.4)}$$

Where $\bar{x}$ is the sample mean and μ is the defined population mean and SEM is the standard error measurement.

2.3 Results

2.3.1 Engineered RNA-OUT Molecules Are Able to Regulate RNA-INs in a Bicistron *in vivo*

E. coli BL21-Gold (DE3) co-transformed with a pRNA-IN and pRNA-OUT variant were either treated with 1mM IPTG (to induce production of a cognate RNA-OUT) or not (for comparison). The expression of mBeRFP and eGFP in these cells was first assessed by fluorescence spectroscopy in ensemble measurements. Upon induction of expression of RNA-OUT-A5, designed to down-regulate eGFP translation when interacting with the S5 RNA-IN motif, a corresponding decrease in eGFP emission signal at 510nm is observed (Figure 2.7). However, a similar decrease in the mBeRFP emission signal at 610nm is

also observed. The corresponding 23.4% decrease in eGFP signal is accompanied by a 36.5% decrease mBeRFP signal.

In a similar fashion, a co-transformant containing pRNA-OUT-A4 and pRNA-IN with both S4 and S5 motifs was utilized in a bulk fluorescence experiment. Upon induction of

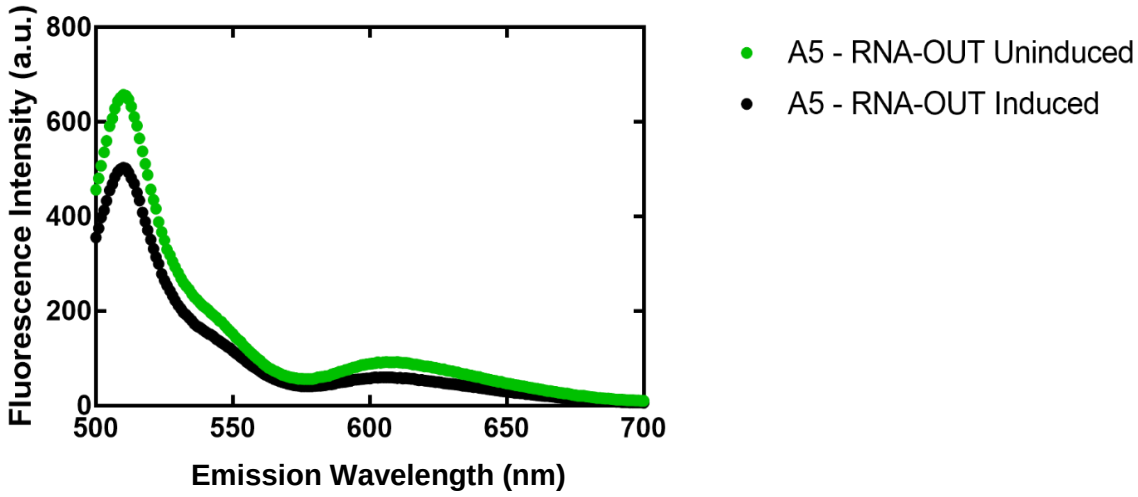


Figure 2.7. Ensemble fluorescence emission scan upon excitation at 488nm comparing fluorescence of cells with induced RNA-OUT (A5) to those in which RNA-OUT (A5) has not been induced. Emission peaks at 510nm and 610nm are indicative of relative abundance of eGFP and mBeRFP respectively.

an RNA-OUT-A4 molecule designed to down-regulate mBeRFP translation upon interaction with the S4 RNA-IN motif, a decrease in mBeRFP emission signal at 610nm is observed (Figure 2.8). Likewise, a decrease in eGFP emission signal at 510nm is observed. A 14.8% decrease in peak eGFP signal is mirrored by a 16.3% decrease in peak mBeRFP signal (Equation 2.1).

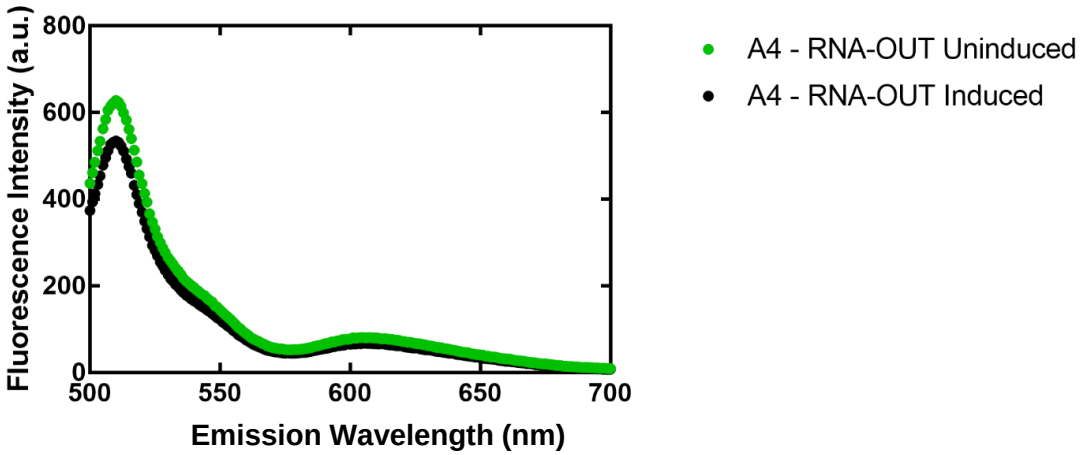


Figure 2.8. Ensemble fluorescence emission scan upon excitation at 488nm comparing fluorescence of cells with induced RNA-OUT (A4) to those in which RNA-OUT (A4) has not been induced. Emission peaks at 510nm and 610nm are indicative of relative abundance of eGFP and mBeRFP respectively.

Ensemble fluorescent measurements yield only an average snapshot of the fluorescence of a population of cells at a given point in time. To obtain the fluorescence properties of individual cells flow cytometry was used. The ability to measure single cells allows one to assess if subpopulations exist within the ensemble fluorescence measurement. Flow cytometry gates were selected to isolate only singlet cells from the 100,000 cells analyzed (Figure 2.9).

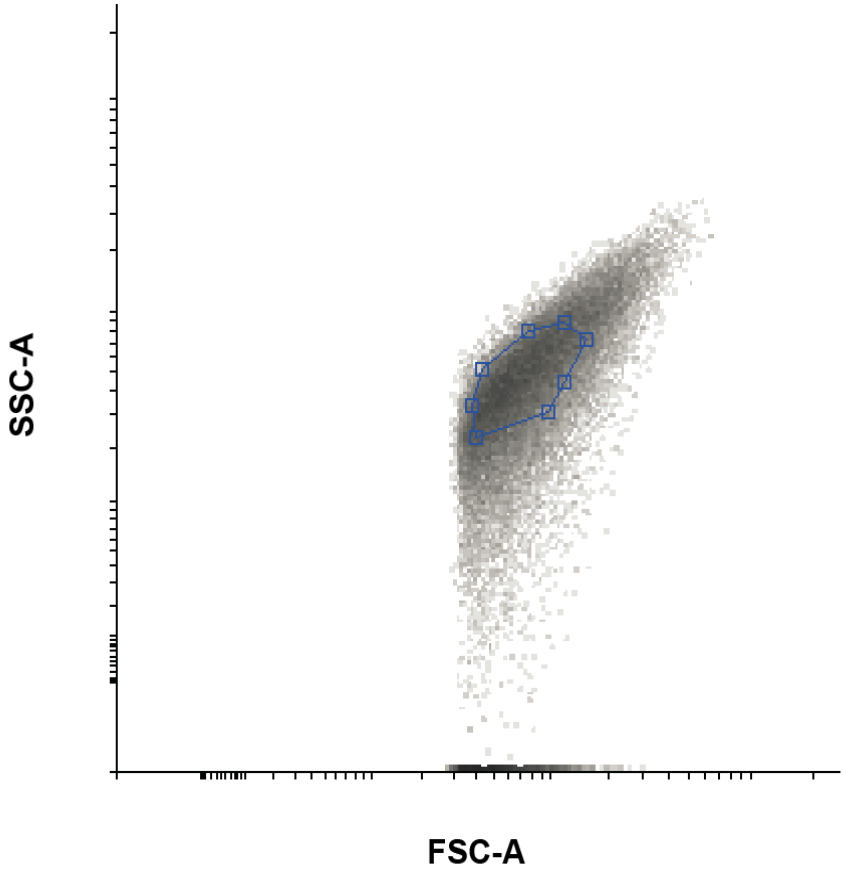


Figure 2.9. Flow cytometry density comparison of side scatter and forward scatter in BL21(DE3) *E. coli* harbouring both pRNA-IN and pRNA-OUT, without induction of the trans-acting regulatory RNA-OUT (A4) molecule. Area within the blue selection indicates a gate drawn to analyze single cells only.

Having gated for single cells, green and red fluorescence within the single cells were compared in the absence (Figure 2.10) and presence (Figure 2.11) of a specific induced *trans*-acting regulatory RNA-OUT (A5). A change in both green and red fluorescence can be observed which corresponds to a 58.2% decrease in mean eGFP fluorescence, and a 70.5% reduction in mean mBeRFP fluorescence.

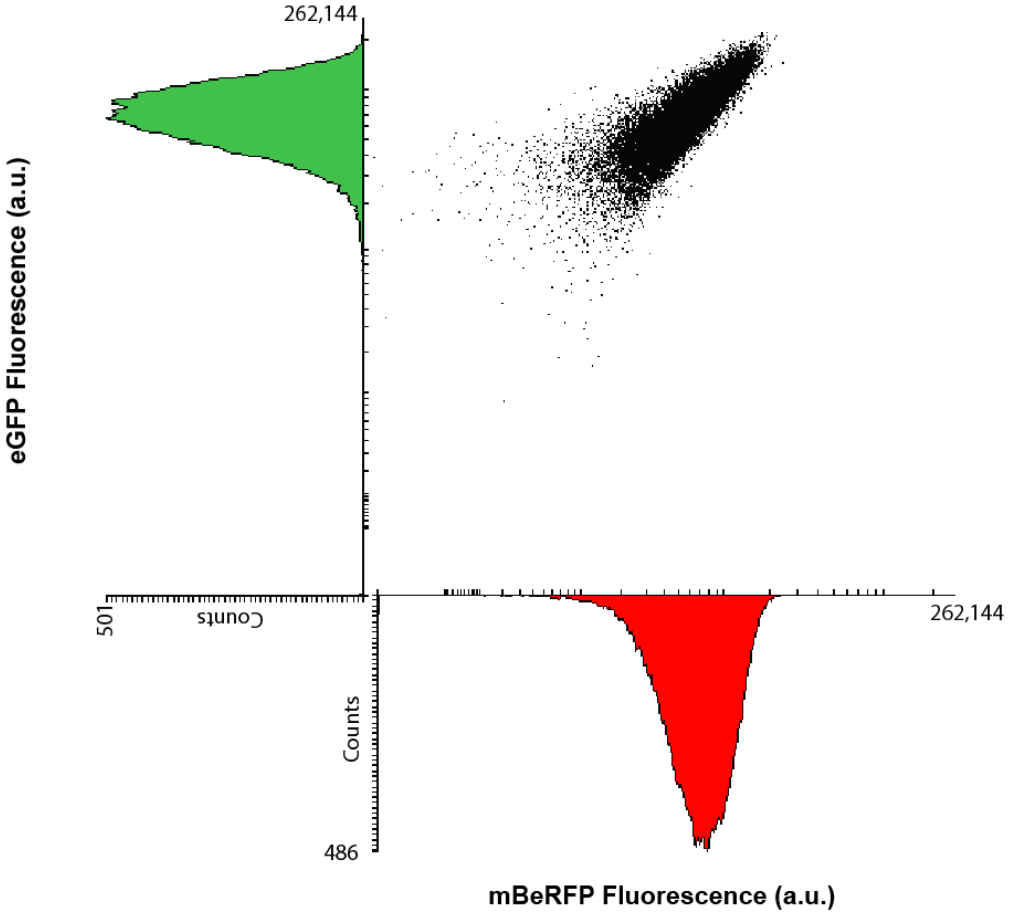


Figure 2.10. Flow cytometry comparison of eGFP and mBeRFP fluorescence in gated single BL21(DE3) *E. coli* cells harbouring both pRNA-IN and pRNA-OUT, however, without induction of the trans-acting regulatory RNA-OUT (A5) molecule. Histograms on each axis indicate the respective distribution of red and green fluorescence.

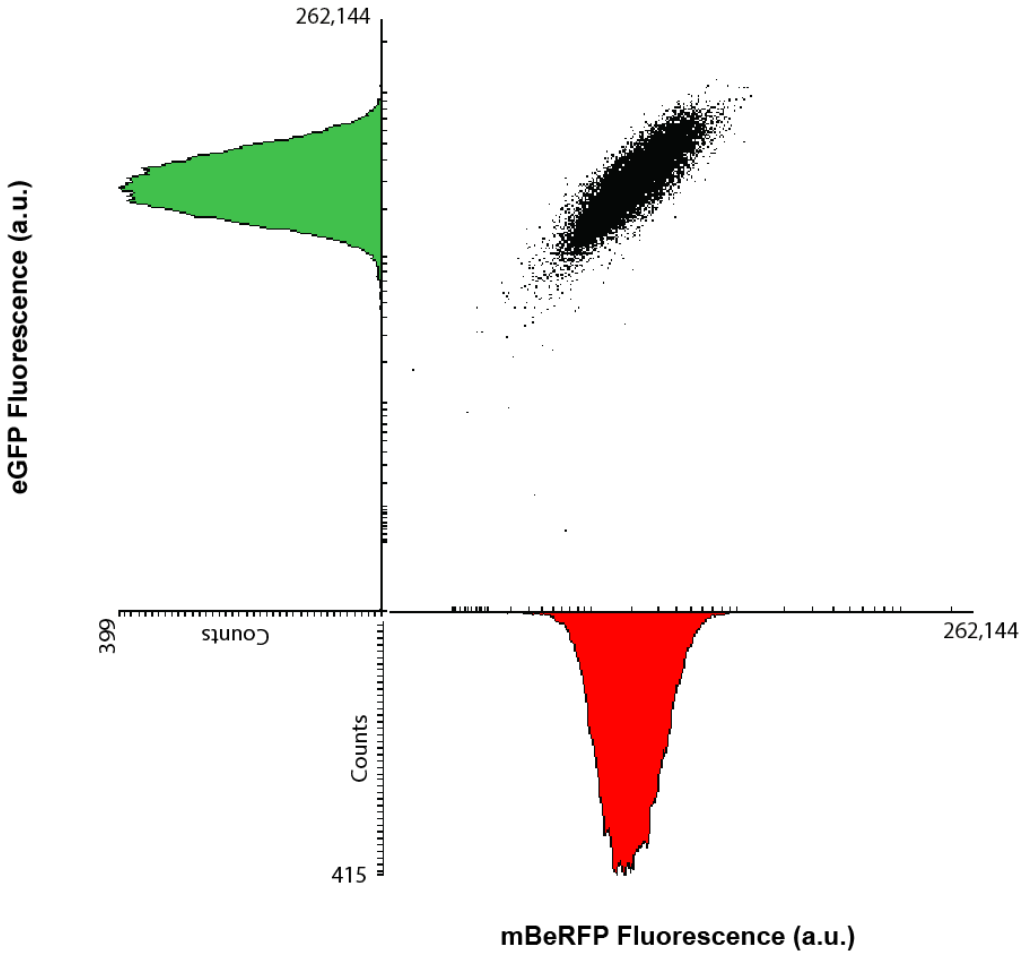


Figure 2.11. Flow cytometry comparison of eGFP and mBeRFP fluorescence in gated single BL21(DE3) *E. coli* cells harbouring both pRNA-IN and pRNA-OUT, upon induction of the trans-acting regulatory RNA-OUT (A5) molecule. Histograms on each axis indicate the respective distribution of red and green fluorescence.

In the interests of testing the effects of different pairs of RNA-IN/OUT with single-cell resolution, a second pair of sense/antisense motifs was interrogated for its impact on translation regulation. Specifically, to examine if a different *trans*-acting regulatory RNA would exert a similar effect on the translation of both fluorescent proteins, a second regulatory RNA-OUT motif (A4) designed to down-regulate mBeRFP translation was examined by flow cytometry. Cells from two different treatments, one not producing the *trans*-acting regulatory RNA species (Figure 2.12) and one in which the production of

RNA-OUT (A4) was induced (Figure 2.13). A difference in fluorescence can be observed upon induction of RNA-OUT (A4), corresponding to a 4.2% increase in mean eGFP fluorescence and a 16.5% increase in mean mBeRFP fluorescence.

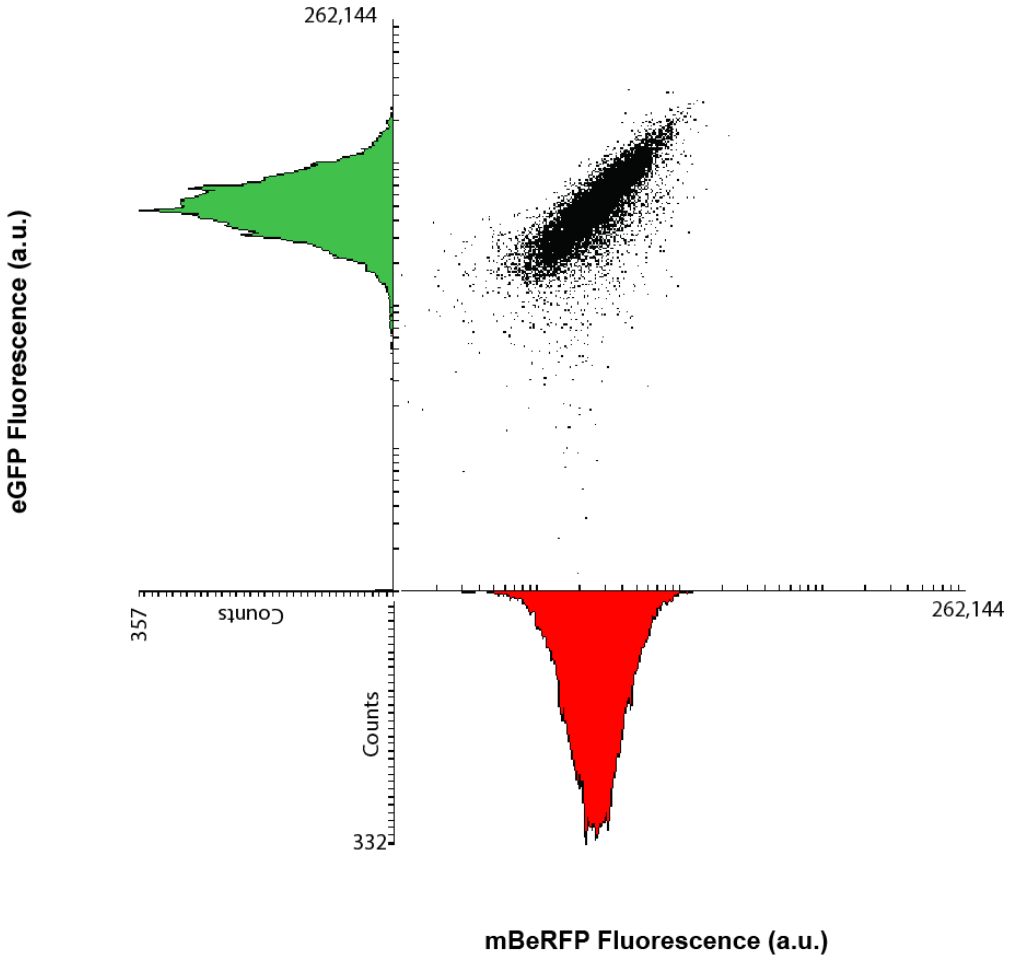


Figure 2.12. Flow cytometry comparison of eGFP and mBeRFP fluorescence in gated single BL21(DE3) *E. coli* cells harbouring both pRNA-IN and pRNA-OUT, however, without induction of the trans-acting regulatory RNA-OUT (A4) molecule. Histograms on each axis indicate the respective distribution of red and green fluorescence.

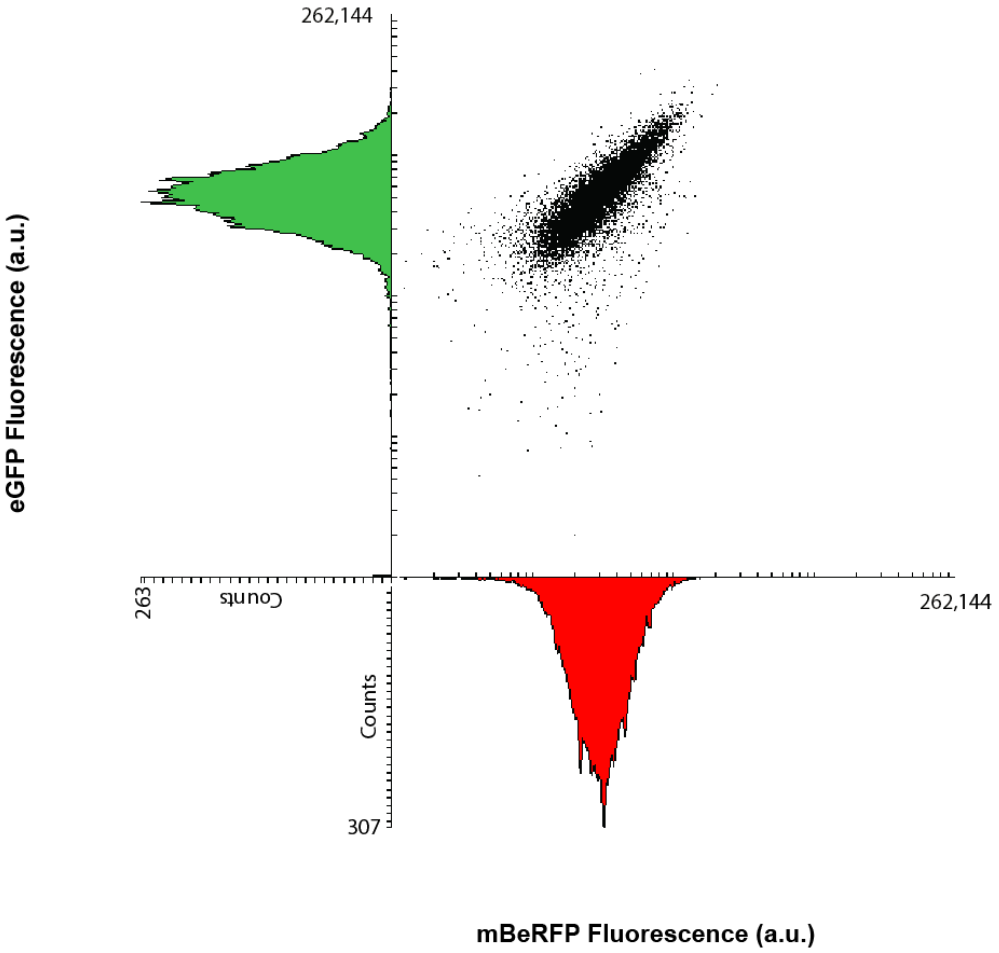


Figure 2.13. Flow cytometry comparison of eGFP and mBeRFP fluorescence in gated single BL21(DE3) *E. coli* cells harbouring both pRNA-IN and pRNA-OUT, upon induction of the trans-acting regulatory RNA-OUT (A4) molecule. Histograms on each axis indicate the respective distribution of red and green fluorescence.

2.3.2 Engineered RNA-OUT is Able to Regulate a Single Cistron Expressing a Fluorescent Protein *in vivo*

The apparent down-regulation of both fluorescent proteins as opposed to the desired single, specific repression suggests that the RNA-OUT/RNA-IN interaction may be impacting the stability of the entire cistron. Therefore, an attempt to modulate translation of a single cistron, to recapitulate the approach utilized by Mutalik *et al.*, 2012 (7) was constructed as described above. *E. coli* BL21-Gold (DE3) were co-transformed with a pRNA-IN_G, as it only contained the green fluorescent reporter, and a cognate pRNA-OUT. Upon induction of an RNA-OUT specifically designed to down regulate translation of the green fluorescent reporter, a corresponding decrease in eGFP emission at 510nm is observed (Figure 2.14). The decrease in eGFP peak fluorescence is observed to be 83.3%.

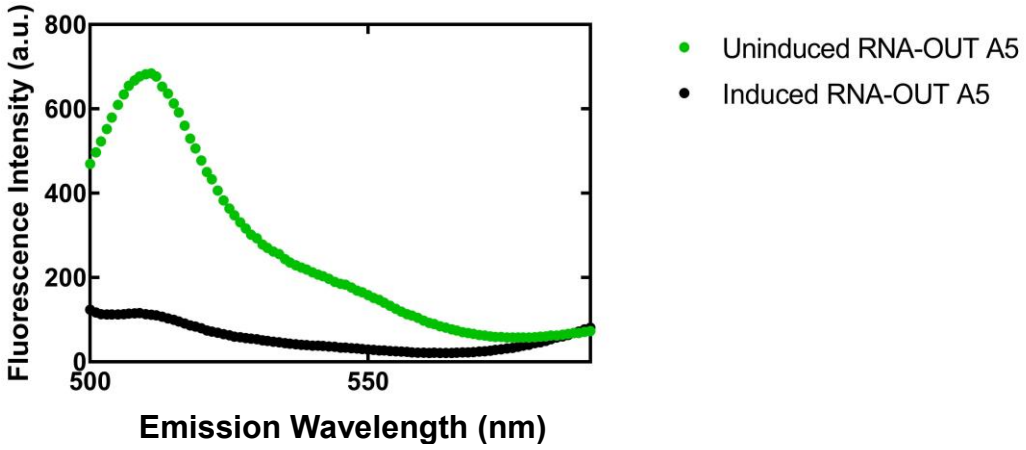


Figure 2.14. Ensemble fluorescence emission scan upon excitation of 488nm comparing fluorescence of cells with only one reporter construct (eGFP) upon induction of a cognate RNA-OUT (A5) compared to those in which RNA-OUT (A5) has not been induced. Emission peak at 510nm is indicative of relative abundance of eGFP.

2.3.3 eGFP Turnover Analysis

To assess the ability of our constitutively expressed reporter system to respond to changes caused by the expression of repressor RNA, it is important to determine the half-life of already expressed reporter proteins. In order to determine the turnover of eGFP, cells were maintained in ABT minimal media for 21 hours after reaching an OD_{600nm} of 1.0 and green fluorescence was measured on 100,000 cells by flow cytometry (Figure 2.15). In this way, the degradation of eGFP can be measured as a decrease in eGFP fluorescence assuming that under these conditions no new eGFP will be produced and essentially no cell growth that would dilute the eGFP fluorescence occurs. Cells in which a regulatory RNA, RNA-OUT A5 was not induced have higher fluorescence throughout the entire eight hours of measurement. However, green fluorescence increases, and an opposite trend is seen in the remaining two timepoints, approximately one day later, but overall, eGFP fluorescence levels remain stable over this time.

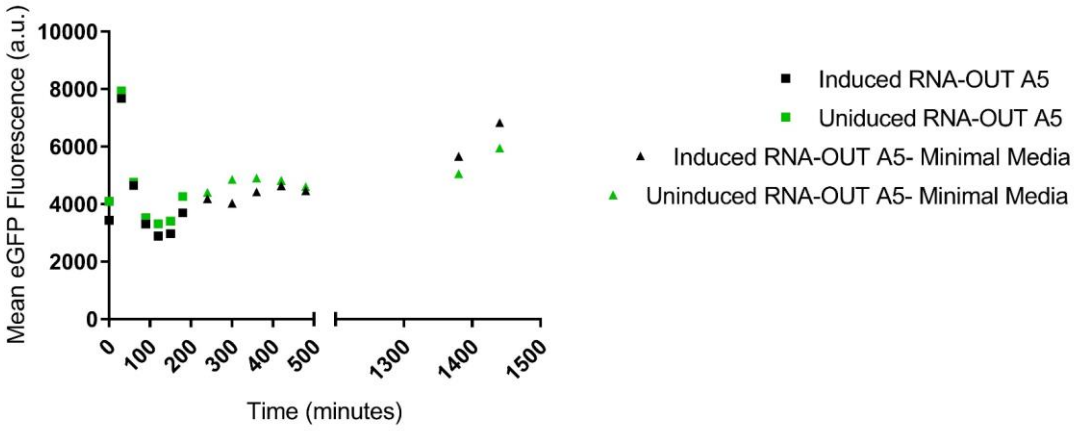


Figure 2.15. Mean green (eGFP) fluorescence of BL21-Gold (DE3) *E. coli* containing both pRNA-IN with two fluorescent reporters and pRNA-OUT (A5) measured by flow cytometry. Cells were grown to an OD600 of 1.0 and transferred to ABT minimal media where fluorescence was monitored for the next 21 hours.

2.3.4 Reporter Transcript Level Determination by Quantitative Reverse-Transcriptase PCR

Down regulation of the entire dual fluorescent reporter system by a single *trans*-acting regulatory RNA suggests that regulation may be happening at the level of transcript stability. It is possible that the entire transcript is degraded prior to being translated, ultimately reducing the number of mRNA copies available (transcript level) for translation. Therefore, to investigate transcript levels, quantitative RT-PCR was conducted. Following the above protocol, total RNA was extracted from *E. coli* BL21-Gold (DE3) which was co-transformed with a dual fluorescent reporter pRNA-IN and a cognate pRNA-OUT designed to regulate only the eGFP signal. Dilutions of reference genes allowed establishment of a standard curve and determining the PCR efficiency (Figure 2.16). From the standard curve, a relative amount of each transcript can be determined.

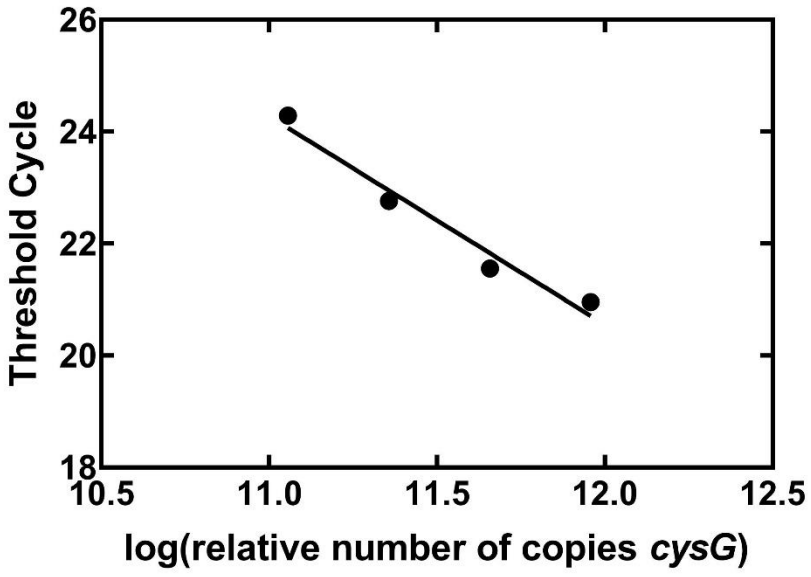


Figure 2.16. qPCR standard curve generated across different concentrations of initial total RNA. The *cysG* transcript exhibited high steady-state stability throughout experiments and was selected as a reference.

A ratio between the transcript levels in an induced sample and an uninduced sample of identical co-transformants was determined and fold change was calculated relative to the *cysG* reference gene sample (Figure 2.17).

The eGFP mRNA signal upon induction of the A5 RNA-OUT is decreased to a level of 0.85 compared to its level in the uninduced treatment. The mBeRFP mRNA signal is also decreased to a level of 0.85 once the *trans*-acting RNA OUT A5 is induced. These decreases are not statistically significant by a student's t-test (p=0.13 for eGFP and p=0.19 for mBeRFP).

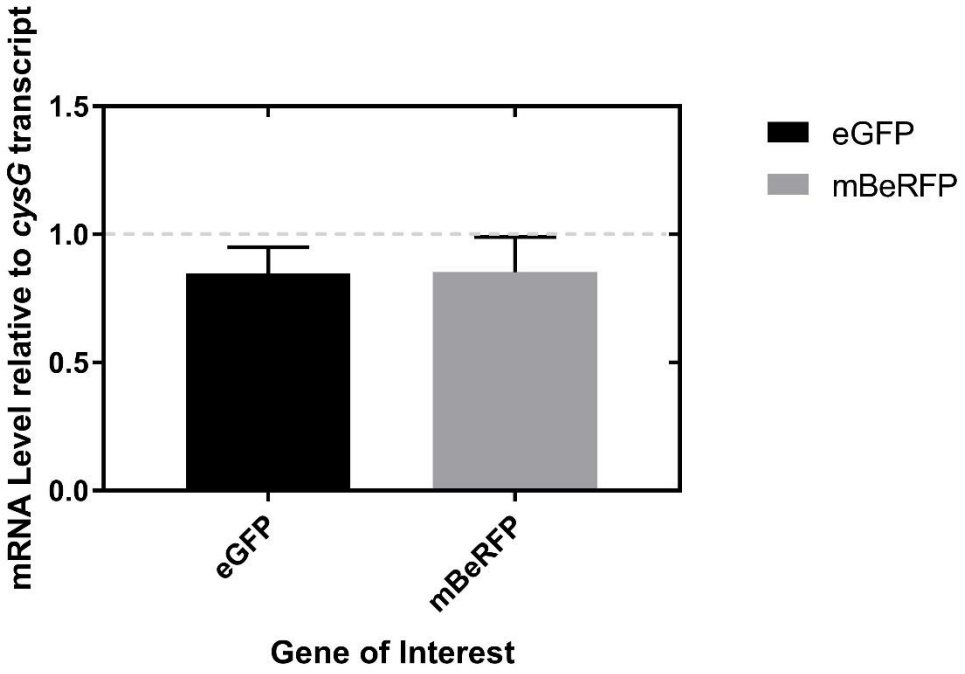


Figure 2.17. mRNA fold change of eGFP (black) and mBeRFP (grey) transcripts measured by qPCR upon induction of RNA-OUT (A5) trans-acting RNA. Error bars represent standard error.

2.4 Discussion

2.4.1 Regulation of Reporter Translation by Specific *trans*-Acting RNA-OUT Molecules

In a bicistronic reporter system harbouring two fluorescent reporters with separate N-terminal fusion regions, specific interactions between *trans*-acting RNA-OUT molecules and messenger RNA just upstream of the ribosome binding site down-regulate translation of the mRNA. Surprisingly, the interaction between an RNA-OUT designed to regulate the downstream fluorescent reporter has an impact on the expression of both fluorescent reporter proteins (Figure 2.11). The level of down-regulation of the two fluorescent proteins in response to the induction of regulatory RNA is similar. This suggests that the specific RNA-IN/RNA-OUT interaction is likely having an impact on the overall stability of the reporter transcript. Even when the *trans*-acting RNA targeted to the upstream reporter is induced, the signal for both reporters is decreased to a similar extent. This can be explained by degradation of the transcript as a whole following the formation of the RNA-IN/OUT duplex, likely by RNase III. Furthermore, it is consistent with previous observations identifying the degradation of the native transcript as an important factor *in vivo* (23). Given that the RNase III recognition motif (24, 25) is double-stranded RNA of at least 22bp in length, the 33bp of interaction (although interrupted by a region of discontinuous base pairing) between the RNA-IN and RNA-OUT molecules may provide a suitable substrate for degradation (26). This suggests that the *trans*-acting RNA likely affects translation of reporter genes, although not at the highly-specific level that was hypothesized. In addition, the reported experiments confirm that the inclusion of a 5` tag sequence on the fluorescent reporter gene and a corresponding complementary sequence

within the regulatory RNA sequence does not adversely impact the interaction between the RNA species.

The observation that the signal from both fluorescent reporters decreases in intensity in ensemble measurement first led to the consideration that perhaps subpopulations of cells existed in which the regulatory RNA was specifically only regulating one portion of the mRNA and down-regulating translation. To examine this, flow cytometry was performed on identical co-transformant treatments (i.e. identical clones with RNA-OUT either induced or not induced). In examining the flow cytometry measurements, a large decrease in the whole analysed population was observed for both red and green fluorescence upon induction of the RNA-OUT A5 molecule. The mean fluorescence change is noticeable in both fluorescent proteins and is stronger than the peak fluorescence decrease when measured by ensemble fluorescence spectroscopy upon induction of RNA-OUT A5. However, the effect is far less pronounced, and in fact reversed in the situation in which RNA-OUT A4 is induced (Figure 2.12, 2.13). This RNA-OUT variant was designed to specifically regulate the FLAG tagged mBeRFP portion of the dual fluorescent reporter, but again, its impact appears to be exerted across the entire transcript (Figure 2.3). While Mutalik and colleagues (7) also used a flow cytometry approach to validate the orthogonality of their approach, their repression is reported mainly with one fluorescent reporter (however two fluorescent reporters are briefly discussed). The repression value reported by Mutalik and colleagues (7) from flow cytometry in the case of A5/S5 RNA-OUT/RNA-IN interaction is 86.2%, while in the case of A4/S4 RNA-OUT/RNA-IN interaction it is 85.3%. In contrast, the data gathered here show a 58.2% decrease in green fluorescence in the case of A5/S5 RNA-OUT/RNA-IN interaction. The repression is

indeed not as drastic as shown in the previous report, and down regulation of both fluorescent reporters is observed. This result, similar to the ensemble fluorescence measurements, also suggests that perhaps RNase III cleavage and degradation of the entire message likely contributes to the observed effect. This does not exclude a contribution of binding site accessibility to regulation, however, the fact that both fluorescent reporters were down-regulated to a similar extent suggests RNA degradation facilitated by the duplex formation between the regulatory RNA to reporter mRNA. A model summarizing the RNase III degradation of mRNA molecules mediated by the *trans*-acting RNA molecules is summarized in Figure 2.18.

To investigate the role of RNA turnover in the regulation of fluorescent reporter translation, quantitative reverse transcriptase PCR was performed on regions of both the mBeRFP and eGFP coding sequences. A similar approach has been used in the cyanobacterium *Synechococcus* to investigate the impact of RNA-OUT on RNA-IN transcript levels , which showed a 96% reduction in mRNA transcript level in response to RNA-OUT (27). However, this type of investigation has not been performed in *E. coli* in the context of dual fluorescent reporters.. Therefore, qPCR experiments using total RNA extracted from co-transformed *E. coli* in which the cognate regulatory RNA had, or had not been induced. The observed decrease in transcript levels clearly supports the hypothesis, that down-regulation and degradation of the entire transcript is part of the mechanism responsible for the observed decrease in reporter protein expression. While this degradation is not responsible for a massive decrease in transcript number, it is possible that the decrease in the mBeRFP (14.7%) and eGFP (15.3%) transcripts may still exert a large influence on translation of fluorescent reporters (28).

In order to determine if the RNA-OUT molecules were indeed capable of specifically repressing the translation of a single message, a second reporter plasmid was tested. In this reporter plasmid design, a single fluorescent protein product was subject to regulation (6xHis-eGFP) by a single induced regulatory RNA (RNA-OUT A5). In this case, when fluorescence spectroscopy bulk measurements were performed relative to an uninduced control, the decrease in eGFP fluorescence was observed to be 83.3% ± 6.1% which is consistent with the 86.2% repression observed by Mutalik and colleagues (7). Therefore, the inclusion of the universal tag sequence in both the RNA-IN and RNA-OUT molecules did not hamper down-regulation, suggesting that this approach is indeed a useful tool for future engineering applications.

While a portion of the down-regulation observed may have indeed been due to ribosome binding site occlusion, it seems clear from RT-qPCR analysis that this fluorescence protein translation decrease is likely due to degradation of the entire transcript. However, further work should be focused on testing more interaction motifs and comparing the transcript levels in the single fluorescent reporter case. Additionally, use of these RNA-IN/OUT molecules could be adapted for engineering of biological circuitry in eukaryotic organisms such as yeast, which has not been examined to date, but would likely need to be amended due to differences in translation initiation (29, 30). Nonetheless, RNA-IN and RNA-OUT motifs provide a robust method of controlling translation in prokaryotes, and could also be explored in more prokaryotic species. Due to the high specificity of RNA-RNA interactions, the applications of the RNA-IN/OUT system, especially in the modified universal case, are widespread. For example, to regulate a suite of genes in a metabolic pathway to specifically tailor reactant stoichiometry, a set of RNA-OUT

molecules could potentially be designed to regulate any number of complementary engineered mRNAs to a defined, predictable extent. Additionally, in a situation in which gene products are toxic and the levels must be modulated to maintain cell viability, RNA-OUT molecules may be used to specifically down-regulate translation of toxic products to a sustainable level. Overall, this repression strategy can be employed in future studies in which the goal is dynamic and quick response to a given stimulus or experimental desire.

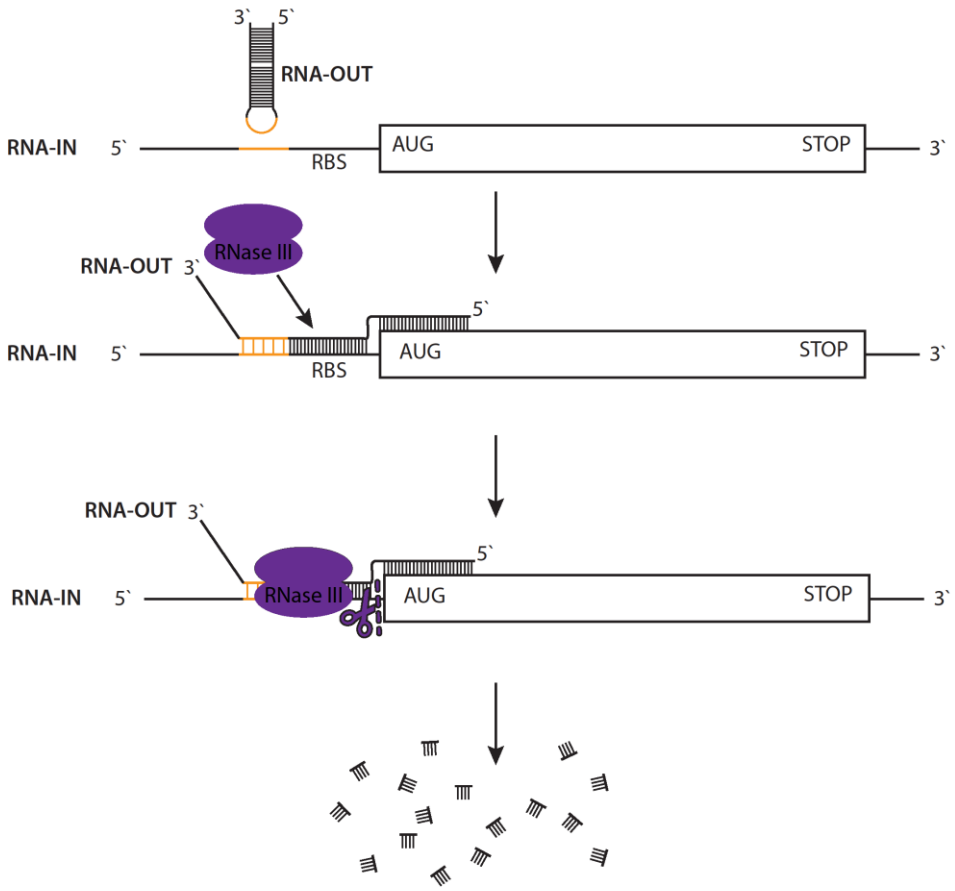


Figure 2.18. Model of RNase III-mediated degradation model of translation down-regulation in response to duplex formation.

References

1. Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods*. **11**, 508–520

2. Brophy, J. A., and Voigt, C. A. (2016) Antisense transcription as a tool to tune gene expression. *Mol. Syst. Biol.* 10.15252/msb.20156540

3. Bradley, R. W., Buck, M., and Wang, B. (2016) Tools and Principles for Microbial Gene Circuit Engineering. *J. Mol. Biol.* **428**, 862–888

4. Wagner, G. H., and Simons, R. W. (1994) Antisense RNA Control in Bacteria, Phages, and Plasmids. *Annu. Rev. Microbial.* **48**, 713–742

5. Franch, T., Petersen, M., Wagner, E. G. H., Jacobsen, J. P., and Gerdes, K. (1999) Antisense RNA regulation in prokaryotes: Rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.* **294**, 1115–1125

6. Waters, L. S., and Storz, G. (2009) Regulatory RNAs in Bacteria. *Cell*. **136**, 615–628

7. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454

8. Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841–847

9. Man, S., Cheng, R., Miao, C., Gong, Q., Gu, Y., Lu, X., Han, F., and Yu, W. (2011) Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Res.* 10.1093/nar/gkr034

10. Engdahl, H. M., Hjalt, T. A. H., and Wagner, E. G. H. (1997) A two unit antisense RNA cassette test system for silencing of target genes. *Nucleic Acids Res.* **25**, 3218–3227

11. Jain, C. (1997) Models for pairing of IS10 encoded antisense RNAs in vivo. *J. Theor. Biol.* **186**, 431–439

12. Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D., and Arkin, A. P. (2011) Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8617–8622

13. Ross, J. A., Ellis, M. J., Hossain, S., and Haniford, D. B. (2013) Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA*. **19**, 670–684

14. Jain, C. (1995) IS10 antisense control in vivo is affected by mutations throughout the region of complementarity between the interacting RNAs. *J. Mol. Biol.* **246**, 585–594

15. Yang, J., Wang, L., Yang, F., Luo, H., Xu, L., Lu, J., Zeng, S., and Zhang, Z. (2013) mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *PLoS One*. **8**, 6–11

16. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014) Linear DNA for rapid prototyping of synthetic biological circuits in an escherichia coli based TX-TL cell-free system. *ACS Synth. Biol.* **3**, 387–397

17. Registry of Standard Biological Parts

18. Kleckner, N. (1990) Regulating tn10 and is10 transposition. *Genetics*. **124**, 449–454

19. Clark, D. J., and Maaloe, O. (1967) DNA Replication and the Division Cycle in Escherichia coli. *J. Mol. Biol.* **23**, 99–112

20. Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjørn, S. P., Givskov, M., and Molin, S. (1998) New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246

21. Zhou, K., Zhou, L., Lim, Q. 'En, Zou, R., Stephanopoulos, G., and Too, H.-P. (2011) Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. *BMC Mol. Biol.* **12**, 18

22. Pfaffl, M. W. (2006) Relative quantification. *Real-time PCR*. 10.1186/1756-6614-3-5

23. Case, C. C., Simons, E. L., and Simons, R. W. (1990) The IS10 transposase mRNA is destabilized during antisense RNA control. *EMBO J.* **9**, 1259–1266

24. Blaszczyk, J., Tropea, J. E., Bubunenko, M., Routzahn, K. M., Waugh, D. S., Court, D. L., and Ji, X. (2001) Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure*. **9**, 1225–1236

25. Schweisguth, D. C., Chelladurai, B. S., Nicholson, A. W., and Moore, P. B. (1994) Structural characterization of a ribonuclease III processing signal. *Nucleic Acids Res.* **22**, 604–612

26. Gordon, G. C., Cameron, J. C., and Pfleger, F. (2017) RNA Sequencing Identifies New RNase III Cleavage Sites in Escherichia coli and Reveals Increased Regulation of mRNA. *mBio.* **8**, 1–18

27. Zess, E. K., Begemann, M. B., and Pfleger, B. F. (2016) Construction of new synthetic biology tools for the control of gene expression in the cyanobacterium Synechococcus sp. strain PCC 7002. *Biotechnol. Bioeng.* **113**, 424–432

28. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124

29. Sonenberg, N., and Hinnebusch, A. G. (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*. **136**, 731–745

30. Vigar, J. R. J., and Wieden, H. J. (2017) Engineering bacterial translation initiation — Do we have all the tools we need? *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 3060–3069

31. Iizuka, R., Yamagishi-Shirasaki, M., and Funatsu, T. (2011) Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem.* **414**, 173–178

3 Translation-Activating RNA-IN/OUT Pairs

3.1 Introduction

3.1.1 Activation of Bacterial Translation

Similar to translation down-regulation, activating translation in bacteria can be achieved by several different methods (1). Often this activation occurs through modulation of the 5`-UTR structure. Given that the interaction between RNA-IN and RNA-OUT occurs within the 5`-UTR of an mRNA, this is an appropriate target for potential activation design. This will provide a novel tool for bioengineers in addition to existing strategies such as riboswitches, toehold switches and small RNAs, extending the utility of the RNA-IN/RNA-OUT interaction.

Riboswitches - Engineered riboswitches are capable of activating translation (2). This sensitive response to a ligand and subsequent unfolding of secondary structure containing the ribosome binding site facilitates translation of specific mRNAs (Figure 3.1A). Importantly, riboswitches are not limited to one single ligand, as new aptamer domains may be evolved synthetically via SELEX for response to multiple different ligands (3–5) . Riboswitches are useful tools for bioengineers as they allow for specific and predictable modulation of translation in response to ligand concentration (6). Their flexibility and specificity, in addition to the rapid response, make riboswitches amenable to engineering and rational design within biological systems. An additional riboswitch design consideration is the inclusion of the expression platform into the ribosome binding site, allowing ribosome binding to be occluded unless a specific ligand is present (7).

However, identifying riboswitch-ligand pairs *in vivo* is limited due predominantly to issues of ligand solubility and toxicity. For this reason, computational methods have been designed to generate likely riboswitch structures *in silico* which can then be tested experimentally (8).

Toehold switches - A second strategy for activation of translation in bacterial (9) and in cell-free systems (10) is the use of toehold switches. Toehold switches are purely RNA-based, and consist of a trigger RNA that binds to a transducer RNA which contains the message (Figure 3.1B). The dynamic range of activation achieved by these switches is greater than 400-fold (most other riboregulators attain a maximal range of around 55-fold) (11) and they can be utilized to construct complex logic gates and circuits (9, 11). Toehold switches are recognized for their orthogonality, with negligible cross-reactivity (12) which allows them to be powerful diagnostic tools for healthcare applications (10). Additionally, because of their independence from protein cofactors, they can be far more portable to remote areas with less required preparation, increasing their utility as diagnostic tools.

Small RNAs - A third class of translation activation approaches is the use of sRNAs (Figure 3.1C). Usually, the sRNAs are *trans*-acting in nature, similar to the RNA-IN/RNA-OUT system, and their action is facilitated by the chaperone Hfq (13–15). These share the advantages of being able to rapidly respond to changes within the bacterial environment, and predictable alteration of translation strength. Due to their small size, rapid turnover, and high specificity synthetic sRNAs have been used to modulate expression of value-added products in the bioengineering field such as 1,3-diaminopropane, which is used in the manufacturing of plastics (15, 16). The utility of

small RNAs is known to be wide-ranging, from down-regulating gene expression in a sense-antisense manner to modulating protein activity within the cell (17). Engineering of sRNA molecules has led to several new applications, including for the RNA-IN/RNA-OUT system. The Tn*10* system was previously reengineered in tandem with an aptamer domain to operate as a transcription modulator (18). However, in the classical approach, both natural and synthetic circuits within *E. coli* containing the Tn*10* system only use RNA-OUT as a down-regulator of translation (19, 20). The tight interaction and orthogonality of designed RNA-IN/RNA-OUT motif pairs affords a design opportunity with respect to adaptability for translation activation. These studies build upon previous reports and employ a strategy similar to synthetic sRNA design to allow for activation of translation to be observed.

To investigate the ability of the RNA-IN/RNA-OUT system to behave as a translation activator, the mRNA with its ribosome binding site constrained in secondary structure was designed to have the structure rearranged upon binding of a cognate *trans*-acting RNA molecule harbouring an appropriate RNA-IN motif to the RNA-OUT motif upstream of the ribosome binding site of the mRNA (Figure 3.1D). This approach is designed to allow for inducible and selective activation of translation of a specific target gene with minimal off-target effects and a dynamic range similar to other sRNA systems.

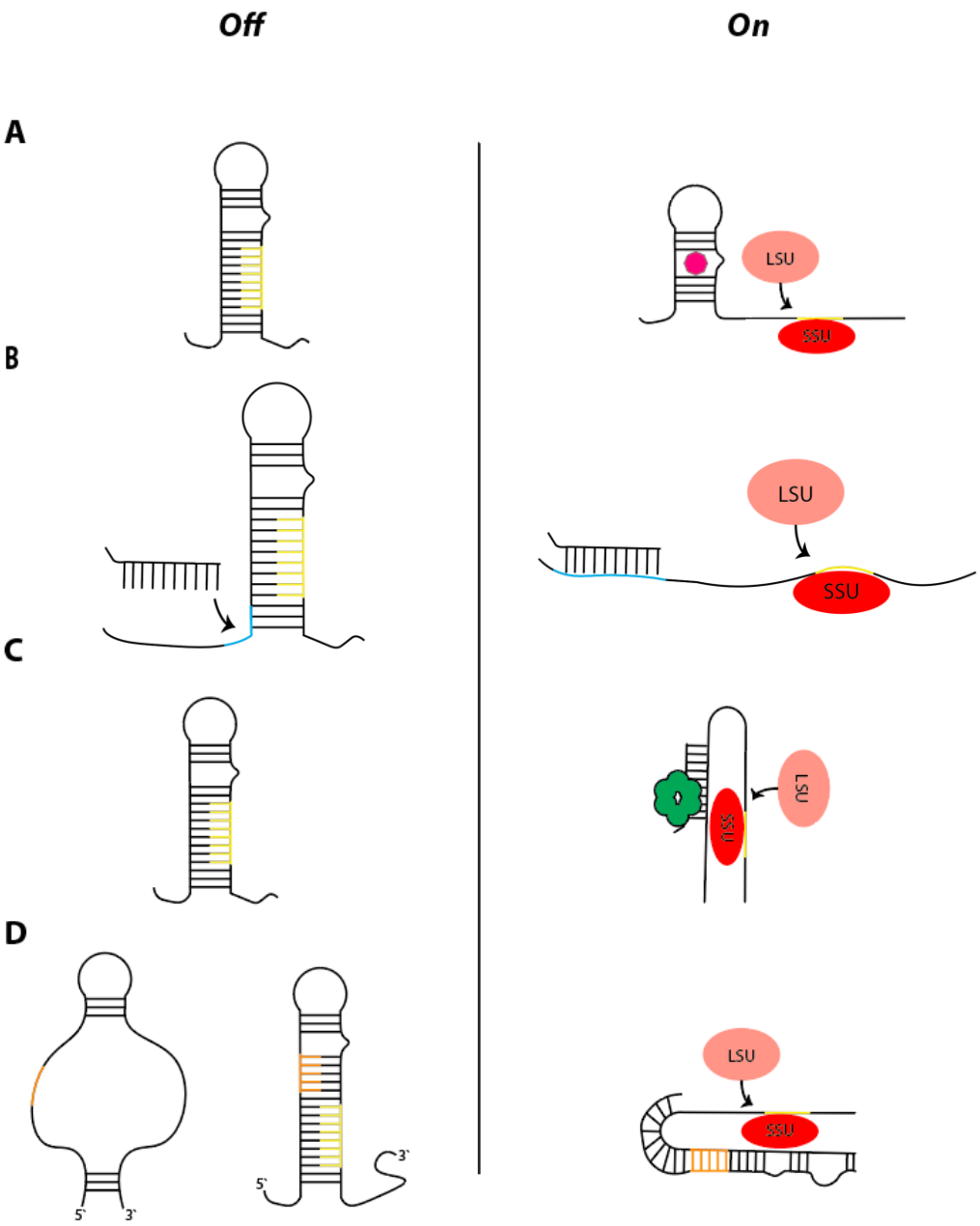


Figure 3.1. Strategies for upregulating translation in *E. coli*. Riboswitches turn on translation upon binding of a cognate ligand to relieve structure surrounding the ribosome binding site (yellow) (a). Toehold switches operate in the absence of protein cofactors to relieve structure surrounding the ribosome binding site upon binding of a cognate RNA molecule to the region indicated in cyan (b). Small RNA binding facilitated by Hfq results in ribosome binding site accessibility (c). A structured RNA molecule harbouring an RNA-OUT motif (orange) engages in specific base-pairing with an unstructured *trans*acting RNA with an RNA-IN motif to relieve structure and facilitate translation (d). Ribosomal small subunit (SSU) and large subunit (LSU) are indicated in red and pink, respectively.

3.1.2 Design of Genetic Constructs

Plasmids were designed to facilitate measurement of translation up-regulation by a synthetic, inducible *trans*-acting RNA molecule. This RNA molecule would generally lack large regions of secondary structure and harbour a motif corresponding to a unique RNA-IN sequence. In designing the interacting RNA molecules, the reporter mRNA molecule had to be carefully designed as well. The reporter mRNA for both 6xHis-eGFP and FLAG-mBeRFP was constructed to be involved in tight secondary structure, with the ribosome binding site occluded due to base pairing. Upstream of the ribosome binding site, a motif corresponding to a unique RNA-OUT sequence is present. The RNA-IN and RNA-OUT sequences provide specificity for the interaction between the two molecules, with the interaction altering the secondary structure of the mRNA, facilitating translation of the fluorescent protein.

While this approach was intended to be an inversion of the standard RNA-IN/RNA-OUT Tn*10* system, the design of the interacting RNA species was no less important. In this design the *trans*-acting RNA is no longer occupied in secondary structure as in the natural system. In order to construct this RNA, mutations were introduced along the 5` segment of the RNA specifically to disrupt the intramolecular Watson-Crick pairing (i.e. disfavouring helical formation) while leaving the 3`segment available to base pair with the messenger RNA. At the same time, the messenger RNA containing the RNA-OUT motif was engineered to be highly structured, and for this structure to be relieved upon base pairing with the *trans*-acting regulatory RNA molecule. The base pairing of the RNA-IN and RNA-OUT motifs confers specificity of the interaction and facilitates the strand displacement and eventual up-regulation of translation. The RNA-IN and RNA-

OUT motifs selected were previously shown to be orthogonal and not to cross-react (19). Again, the use of a high-copy regulatory plasmid and a low-copy reporter plasmid allowed for mimicking of the natural stoichiometry of RNA-OUT to RNA-IN of approximately 20:1 (21). Similar to the prior design, the coding sequence included an invariant spacer sequence, making this approach a universal, RNA-only translation activating strategy.

3.2 Methods

3.2.1 Construction of Genetic Elements

In designing the reporter plasmid, a previously constructed low-copy number pRNA-IN plasmid was adapted. A digest was performed on pRNA-IN with BamHI and NdeI (NEB) (1µg of plasmid DNA was digested in a 50µL reaction with 10 units of each restriction enzyme) and an identical digest was performed on a PCR product containing the highly structured interacting motif and 6xHis tag. The digestion products were subsequently gel purified from a 1% agarose gel run at 10V/cm for 30 minutes and the resulting purified fragments were then ligated in a 10µL reaction using T4 DNA ligase (NEB) (100ng of digested plasmid DNA along with 300ng of insert DNA were combined in a 10µL reaction with 400 units of T4 DNA ligase overnight at 16°C) to generate a reporter plasmid now harbouring a highly structured interaction motif. Presence of the insert was confirmed by Sanger sequencing (Genewiz).

In the design of the plasmid harbouring an inducible *trans*-acting RNA molecule for activation of translation, a previously constructed pRNA-OUT plasmid was used as a

starting point. The backbone plasmid was digested with EcoRI and XhoI (NEB) to remove the RNA-OUT cassette previously contained within. An identical EcoRI and XhoI digest was performed on a PCR product of the engineered *trans*-acting RNA coding sequence, which was gel purified and ligated into the backbone as described above. Presence of the insert was confirmed by Sanger sequencing (Genewiz). This plasmid allows for inducible expression of the unstructured *trans*-acting RNA in an attempt to activate translation.

3.2.2 Cell Growth and Fluorescence Spectroscopy Analysis

Escherichia coli BL21-Gold(DE3) (Agilent) were co-transformed by electroporation with purified 100ng of pRNA-OUT and 100ng of pRNA-IN containing an inducible trans-acting RNA and dual fluorescent reporter cassette, respectively. Overnight cultures of co-transformants were diluted to an OD_{600nm} = 0.05 in 5mL fresh LB media with 50μg/mL kanamycin and 100μg/mL ampicillin. Cultures were incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). IPTG was added to a final concentration of 1mM in one culture at an OD_{600nm} 0.3, while the other culture remained uninduced. Both cultures were grown until reaching an OD_{600nm} = 1.0. Cells were then harvested by centrifugation at 5000 x g for 2 min. The cell pellet was washed in 1 mL ice-cold sterile 1 X PBS before resuspension in 500 μL sterile 1 X PBS. The resulting suspension was incubated for 90 min at 20°C to facilitate the maturation of the produced fluorescent proteins and subsequently analyzed using a Cary Eclipse Fluorescence Spectrophotometer (Varian, Inc. Agilent Technologies) and its emission scan program within the Cary Eclipse software suite. Upon establishing a reference fluorescence baseline with PBS, relative determinations of fluorescence were obtained. 200μL of each washed cell

suspension was loaded into a 6mm x 6mm x 50mm quartz cuvette (Starna Cells) and then placed within the fluorescence spectrometer. Both fluorophores were excited at 488nm and fluorescence from 500nm-700nm was recorded, allowing for relative determination of fluorescent protein abundance in both the absence and presence of a specific *trans*-acting activating RNA species (i.e. in an induced and an uninduced sample). Peak fluorescence at 510nm is indicative of eGFP levels, while peak fluorescence at 610nm is indicative of mBeRFP levels. The peak fluorescence in both the induced and uninduced treatments was used for comparison and calculation of percent activation value using Equation 3.1.

$$Percent\ Activation = \left(\frac{Maximum\ Fluorescence_{\ Induced\ Sample} - Maximum\ Fluorescence_{\ Uninduced\ Sample}}{Maximum\ Fluorescence_{\ Uninduced\ Sample}} \times 100\%\right)$$

(Equation 3.1)

3.2.3 Flow Cytometry

Overnight cultures of co-transformants were diluted to an OD_{600nm} = 0.05 in 50 mL fresh LB media with 50μg/mL kanamycin and 100μg/mL ampicillin. Cultures were incubated at 37°C while shaking at 220 rpm (Thermo Fisher MaxQ 4000). IPTG was added at a final concentration of 1mM to one culture at an OD_{600nm} = 0.3, while the other culture remained uninduced. Both cultures were grown until reaching an OD_{600nm} of 1.0. Cells were subsequently harvested by centrifugation at 5000 x g for 2 min. The resulting cell pellet was then washed in 1 mL ice-cold sterile 1 X PBS before resuspending in 500 μL sterile 1 X PBS. The resuspension was allowed to incubate at 20°C for 90 min based on the reported maturation times of the fluorophores (22, 23), to facilitate the maturation of

fluorescent proteins. The suspension was further diluted 10X in 1X PBS and subjected to flow cytometry. Given the use of identical fluorophores (mBeRFP and eGFP) to those used in prior RNA-IN/RNA-OUT flow cytometry experiments, the same protocol was used as described previously to obtain relative fluorescence characteristics of single cells. Using a BD FACS Aria Fusion, both fluorophores were excited at 488nm and eGFP fluorescence was measured with a 502LP mirror and bandpass filter of 530/30nm. mBeRFP fluorescence was measured using a 600LP mirror and a bandpass filter of 610/20nm. A scatter gate was set using the forward and side scatter area of *E. coli* cells containing no plasmid. 1 X 10^5 cells from each culture were analyzed. Gates were set, including no less than 50,000 cells and analyses were performed in Flowing Software v2.5.1 (Flowing Software, Perttu Terho).

3.3 Results

3.3.1 Engineered RNA-OUT Molecules Up-Regulate Translation of a Bicistron *in vivo*

In order to determine the extent to which the engineered RNA-OUT RNA is able to activate translation, fluorescence measurements were conducted. *E. coli* BL21-Gold (DE3) co-transformed with a pRNA-IN and activating pRNA-OUT variant were either treated with 1mM IPTG (to induce production of a cognate RNA-OUT) or not (for comparison). Upon induction of the RNA-OUT_Act1H RNA transcription, which is designed to up-regulate eGFP translation upon interaction with the specifically designed RNA-IN_Act1H motif (Figure 3.1D). This induction resulted in a 20.4% increase in peak eGFP fluorescence (Figure 3.2). The increase in mean fluorescence across triplicate trials is less pronounced at an increase of 4.1% ± 12.4% (Appendix F)

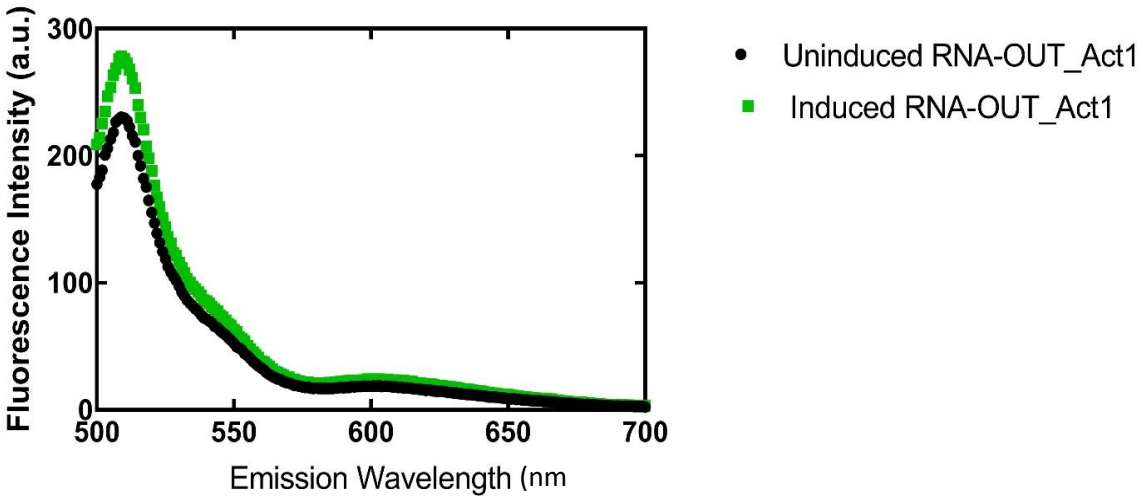


Figure 3.2. Ensemble fluorescence emission scan upon excitation at 488nm comparing fluorescence of cells with induced activating RNA-OUT (A1) to those in which the activating RNA-OUT (A1) molecule has not been induced. Emission peaks at 510nm and 610nm are indicative of relative abundance of eGFP and mBeRFP respectively.

In order to examine the robustness of the RNA-OUT's activation of translation, and to see if subpopulations of cells have translation activated to a greater or lesser extent than others within the population, flow cytometry was used. Measurement by flow cytometry and analysis of single cells (Figure 3.4) provided insight into the activating nature of the RNA-IN/RNA-OUT interaction. Flow cytometry gates were set by analyzing the forward-scattered and side-scattered light and drawing a border around those cells which appeared to not be dividing, nor simply cell debris as above to isolate only singlet cells from the 100,000 cells analyzed. After gating for single cells, a comparison of the level of eGFP expression was performed. Induction of the activating RNA-OUT resulted in an increase in mean eGFP fluorescence of 7.7%.

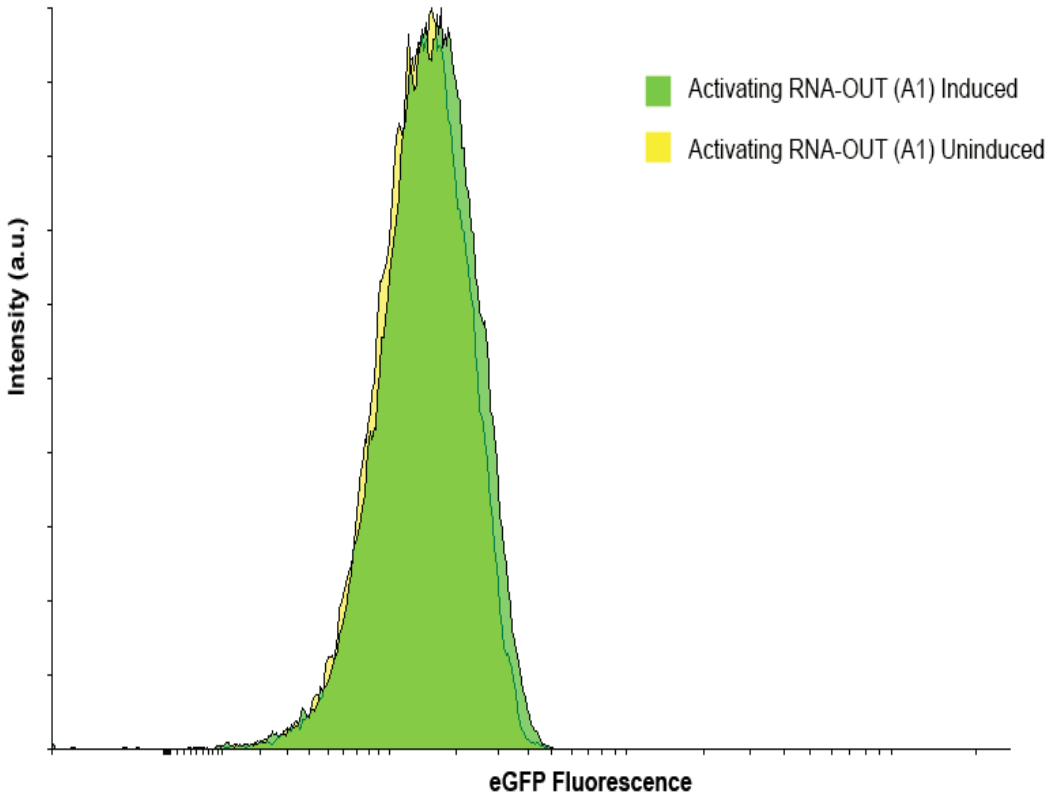


Figure 3.3. Flow cytometry comparison of eGFP fluorescence in BL21(DE3) *E. coli* harbouring both activating pRNA-IN and pRNA-OUT, both with and without induction of the trans-acting activating RNA-OUT (A1) molecule.

3.4 Discussion

In a bicistronic reporter design with two fluorescent proteins produced, a specifically designed *trans*-acting RNA molecule is able to increase translation of a specific fluorescent reporter relative to background fluorescence levels. While not a large increase, an increase in fluorescence is nonetheless observed. As can be seen in the ensemble fluorescence measurements, the designed *trans*-activating RNA is capable of up-regulating translation of a targeted gene expression. This demonstrates the potential to reverse the standard engineered down-regulatory *trans*-acting RNA interaction (19, 24). It allows bioengineers to utilize similar motifs in different contexts to achieve up- or down-regulation of translation using similar tools which are highly specific for either purpose

depending upon the desired output. The RNA-IN/OUT approach has proven to be a useful strategy for down-regulation in synthetic biology, however, a similar strategy using motifs conferring high specificity as found in the RNA-IN and RNA-OUT interaction for up-regulation has not been demonstrated previously.

The observed change in fluorescence is not a sizable one compared to other RNA-based translation activation systems. In comparison to prior reports of activation of translation in bacterial systems using toehold switches (9, 10), the level of activation is relatively small. It does appear in the designed system above, that the activation is sustainable and predictable across a population as seen in the ensemble fluorescence measurements. With a larger suite of RNA-IN/RNA-OUT activators, a larger dynamic range may be attained as more options with more possible combinations present new opportunities for activation. With more activation options at the disposal of a bioengineer, more specific tuning of translation may be possible. Additionally, the design of the reporters may be altered so that each mRNA is produced from a separate and identical promoter, with each still being regulated by a distinct *trans*-acting regulatory RNA. In this way, the interaction between the *trans*-acting RNA and messenger RNA may not be impeded by the considerable size of the long single transcript. As seen in prior works, the target site accessibility impacts the ability of a small RNA to exert control over translation of an mRNA (25–27). Perhaps measuring activation of only a single fluorescent protein, as done previously for translation repression is an alternative method that could be adopted (19). As seen previously, eGFP is stable *in vivo* and activating its translation should be apparent within the timeframe tested. While the increase in fluorescence is not indicative of a high level of activation as observed with sRNA activation, it is nonetheless a step

forward, as eGFP fluorescence was increased in both fluorescence spectroscopy and flow cytometry measurements. However, the low increase in reporter fluorescence indicates that the choice of inducible promoter contributed significantly to the amount of observed activation between induced and uninduced *trans*-acting RNA conditions. Induction with IPTG of the LacO-1 promoter has been documented to be leaky (28) under similar conditions, so future studies may benefit from selecting a tighter regulating promoter, such as pBAD (29). This will help in teasing apart the true dynamic range of translation activation driven by this particular system. In addition, the low activation observed in this case may be due to the initial structure of the mRNA disallowing activation. This could be improved by placing the interacting sequence in an unstructured region of the messenger RNA. These alternative RNA structures impact the free energy changes when compared to the traditional down-regulating design. In future, these free energy changes should be carefully considered in designing of new activating RNA-IN/OUT pairs. Finally, by making the *trans*-acting RNA unstructured relative to its mRNA counterpart, the stability and lifetime of the regulatory RNA may be altered, reducing their steady state levels available for activation. Therefore, further investigation into the stability of the newly engineered regulatory RNA should be considered as well.

It is worth noting that the design strategy of the RNA-IN/OUT pairs essentially inverted the structural motifs employed in the prior natural approach. It would be interesting for future work on this system to employ a diverse array of structures such as a pseudoknot with the RNA-IN/OUT motifs at the core to facilitate interaction in *trans*. This would allow for a more diverse library of RNA-IN/OUT pairs to be created, capable of down- or up-regulating translation to a predictable, dynamic extent.

References

1. Vigar, J. R. J., and Wieden, H. J. (2017) Engineering bacterial translation initiation — Do we have all the tools we need? *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 3060–3069

2. Topp, S., and Gallivan, J. P. (2008) Riboswitches in unexpected places--A synthetic riboswitch in a protein coding region. *RNA.* **14**, 2498–2503

3. Stoltenburg, R., Reinemann, C., and Strehlitz, B. (2007) SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **24**, 381–403

4. Famulok, M. (1994) Molecular Recognition of Amino Acids by RNA-Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder. *J. Am. Chem. Soc.* **116**, 1698–1706

5. Huang, Z., and Szostak, J. W. (2003) Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA.* **9**, 1456–1463

6. Dixon, N., Duncan, J. N., Geerlings, T., Dunstan, M. S., McCarthy, J. E. G., Leys, D., and Micklefield, J. (2010) Reengineering orthogonally selective riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2830–2835

7. Grundy, F. J., and Henkin, T. M. (2004) Regulation of gene expression by effectors that bind to RNA. *Curr. Opin. Microbiol.* **7**, 126–131

8. Espah Borujeni, A., Mishler, D. M., Wang, J., Huso, W., and Salis, H. M. (2016) Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Res.* **44**, 1–13

9. Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: De-novo-designed regulators of gene expression. *Cell*. **159**, 925–939

10. Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., Daringer, N. M., Bosch, I., Dudley, D. M., O'Connor, D. H., Gehrke, L., and Collins, J. J. (2016) Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*. **165**, 1255–1266

11. Callura, J. M., Cantor, C. R., and Collins, J. J. (2012) Genetic switchboard for synthetic biology applications. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5850–5855

12. Takahashi, M. K., and Lucks, J. B. (2013) A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucleic Acids Res.* **41**, 7577–7588

13. Ross, J. A., Ellis, M. J., Hossain, S., and Haniford, D. B. (2013) Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system.

RNA. **19**, 670–684

14. Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004) The bacterial Sm-like protein Hfq: A key player in RNA transactions. *Mol. Microbiol.* **51**, 1525–1533

15. Barik, A., and Das, S. (2018) A comparative study of sequence- and structure-based features of small RNAs and other RNAs of bacteria. *RNA Biol.* **15**, 95–103

16. Chae, T. U., Kim, W. J., Choi, S., Park, S. J., and Lee, S. Y. (2015) Metabolic engineering of Escherichia coli for the production of 1,3-diaminopropane, a three carbon diamine. *Sci. Rep.* **5**, 1–13

17. Gottesman, S., and Storz, G. (2011) Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harb. Perspect. Biol.* **3,** a003798

18. Liu, C. C., Qi, L., Lucks, J. B., Segall-Shapiro, T. H., Wang, D., Mutalik, V. K., and Arkin, A. P. (2012) An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat. Methods*. **9**, 1088–1094

19. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454

20. Ross, J. A., Wardle, S. J., and Haniford, D. B. (2010) Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Mol. Microbiol.* **78**, 607–621

21. Kleckner, N. (1990) Regulating tn10 and is10 transposition. *Genetics*. **124**, 449–454

22. Yang, J., Wang, L., Yang, F., Luo, H., Xu, L., Lu, J., Zeng, S., and Zhang, Z. (2013) mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *PLoS One*. **8**, 6–11

23. Iizuka, R., Yamagishi-Shirasaki, M., and Funatsu, T. (2011) Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem.* **414**, 173–178

24. Qi, L. S., and Arkin, A. P. (2014) A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nat. Rev. Microbiol.* **12**, 341–354

25. Tafer, H., Ameres, S. L., Obernosterer, G., Gebeshuber, C. A., Schroeder, R., Martinez, J., and Hofacker, I. L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.* **26**, 578–583

26. Urban, J. H., and Vogel, J. (2007) Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res.* **35**, 1018–1037

27. Hackermüller, J., Meisner, N. C., Auer, M., Jaritz, M., and Stadler, P. F. (2005) The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: A quantitative model. *Gene*. **345**, 3–12

28. Marbach, A., and Bettenbrock, K. (2012) Lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J. Biotechnol.* **157**, 82–88

29. Guzman, L.-M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter. *J. Bacteriol.* **177**, 4121–4130

4 2A Peptide Library Engineering

4.1 Introduction

4.1.1 Naturally-Occurring 2A Peptides

Among viruses, notably within viral family *Picornaviridae* (1), several mammalian pathogens have developed the ability to translate multiple proteins from their RNA genome without translation re-initiation events occurring (2). The strategy employed by *Picornaviridae* uses a short peptide linker between the 2A and 2B protein coding regions of the genome to facilitate ribosome skipping (Figure 4.1) (3, 4). While other such 2A peptides in separate viral families are implicated in viral structure, proteolysis and apoptosis (5), the most commonly studied 2A peptides come from the *Picornaviridae* family.

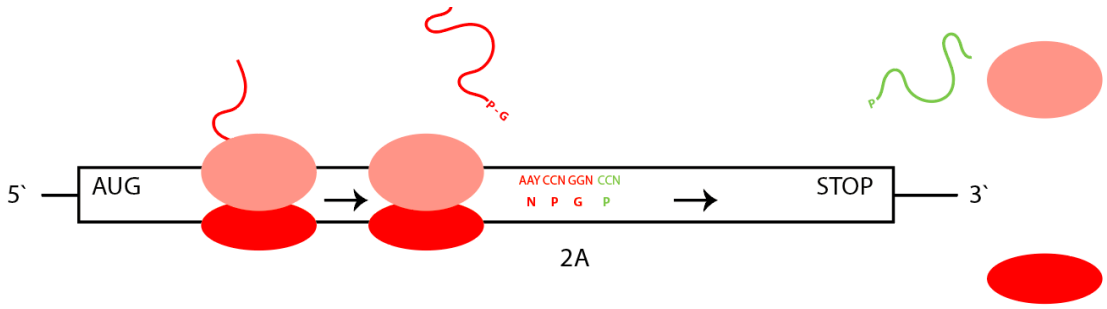


Figure 4.1. General schematic of 2A peptide operation. At the junction between 2A and 2B peptide regions, a ribosomal skipping event takes place and the two cistrons are separated (red, green). Highly conserved, and important residues are noted in the 2A region. Ribosomal small subunit (SSU) and large subunit (LSU) are indicated in red and pink, respectively.

The motifs across many different viral 2A peptides all contain a canonical C-terminal region consisting of the amino acids NPG|P, where the second prolyl residue is the first amino acid of the second peptide (Figure 4.1, Table 4.1, 4.2). Across a range of mammalian and insect viruses, this 2A-ribosomal skipping strategy is employed, and the evolution of such an approach appears to have occurred several times (3). It is not surprising that this strategy is found within rapidly-evolving viruses, with the goal of maximizing the economy of viral resources (3). The ribosomal skipping mechanism is proposed to be due to an interaction between the 2A fragment and the ribosome exit tunnel to alter the conformation of the peptide, disfavouring addition of the next proline in the A site (6, 7). While the exact mechanism of 2A peptide ribosome skipping is not well understood structurally, the hypothesized mechanism has only been demonstrated to operate in eukaryotic systems (7, 8). Due to the eukaryotic viral origin of these 2A peptide sequences, this is expected (3) . However, the similarity in ribosome structure and function across biological kingdoms (9) suggests that a similar sequence motif may be discovered within other kingdoms, including bacteria.

4.1.2 2A Peptides For Bioengineering

The applications of 2A peptides in eukaryotes have been numerous, with studies on their function in insects (10), yeast (8), zebrafish, mice, and a variety of mammalian and human cell lines (5, 11, 12). Broadly speaking, the use of 2A peptides has been focused on expressing multiple proteins in distinct and reliable ratios (13). However, this is not the only application, as they have also been utilized in the production of bicistronic mRNAs capable of producing 2 proteins in the context of gene therapy for cancer cells (11). Importantly, 2A peptide sequences offer solutions for bioengineers wishing to

express proteins in specific stoichiometries for metabolic engineering, biofuel production, or pharmaceutical applications. Not all 2A peptide sequences have been studied extensively, with members such as the foot-and-mouth disease virus 2A (F2A), porcine teschovirus-1 2A (P2A) and equine rhinitis A virus 2A (E2A) being most studied. No single study has investigated a large number of combinations of 2A peptide sequences , and no library of 2A peptides for expression of multiple genes has been developed or characterized to date. As bioengineering moves forward, standard (for ease of use between facilities and instrumentation), well described modular (for ease of use in exchanging segments when making genetic constructs) parts are essential to the performance and production of biological circuitry. Therefore, to allow for extension of the utility of this approach in bioengineering, it is imperative that a strategy be developed for engineering of novel, modular, standard 2A peptide sequences that exist outside of the natural setting. Furthermore, while 2A peptide activity has not been described in prokaryotes, the construction of a suitably complex library may indeed allow for the identification of sequences that operate in the same manner as a canonical 2A peptide sequence in eukaryotes. This in turn will also address the question if the mechanism applied by the 2A peptide is conserved among different kingdoms of life.

Having a well-described library of 2A peptides with a range of ribosome-skipping efficiency, some of which operate within prokaryotes would be a dramatic step forward for genetic engineering.

4.1.3 Design of 2A Peptide Library

The plasmid used to screen 2A peptide sequences for skipping activity in bacteria was designed to allow for potential functional 2A peptide identification and isolation based on

fluorescence output. In order to accomplish this, a dual fluorescent reporter system was constructed as an adaptation of plasmids discussed above (Figure 4.1). To design the 2A peptide library a multiple sequence alignment of known 2A peptides was constructed based on the sequences of 37 well-studied 2A peptides, all containing the canonical NPG|P motif (14, 15) (Figure 4.3). To develop the respective library, a sequence alignment of all 37 representatives, and a subset of 6 of the most commonly studied (16) was generated using Geneious 10.2.3. Two separate libraries of 2A peptides were constructed (Figure 4.3b, 4.3c). Each library was generated using overlapping primers and extension PCR, with two different size products generated. A 90nt library with high variability (1.8×10^9 possible sequences) resulting from the alignment of the 37 viral 2A peptides (Figure 4.3b, Table 4.1) and a 72nt library with lower variability (165,888 possible sequences) resulting from the alignment of 6 of the most commonly studied 2A peptide sequences (Figure 4.3) were constructed. The different library sizes were chosen due to the limitation of transformation efficiency (producing approximately 10^8 clones) and the desire to effectively sample the entire library. Sequence logos were generated from the multiple sequence alignment performed using the Berkeley WebLogo (https://weblogo.berkeley.edu/logo.cgi) design tool with default nucleotide image parameters (17).

Table 4.1 Sequences of 37 viral 2A peptides from *Picornaviridae* members and their consensus sequence for construction of 90nt library (Y=C or T; R=A or G; N=A, T, G, or C; W=A or T; S=C or G; M=A or C; H=A, C or T).

Virus	**2A Peptide Nucleotide Sequence**
Consensus	ATYRGYGTGTTYAAYAARAARC TAGAYGWYGRYAAGGCGRAGT AYGAYAWBCTHATGCTATGYG

	GYGAYGTHGARACGAAYCCNG GNCCN
Theiler's Murine Encephalomyelitis Virus (TMEV)	TTYMGNGARTTYTTYAARGCNG TNMGNGGNTAYCAYGCNGAYT AYTAYAARCARMGNYTNATHC AYGAYGTNGARATGAAYCCNG GNCCN
Theiler-Like Virus of rats (T-LV)	TTYWSNGAYTTYTTYAARCAYG TNMGNGARTAYCAYGCNGCNT AYTAYAARCARMGNYTNATGC AYGAYGTNGARACNAAYCCNG GNCCN
Thosea asigna Virus (TaV)	MGNGGNCCNMGNCCNCARAAY YTNGGNGTNMGNGCNGARGGN MGNGGNWSNYTNYTNACNTGY GGNGAYGTNGARGARAAYCCN GGNCCN
Seneca Valley Virus (SVV)	MGNGCNTGGTGYCCNWSNATG YTNCCNTTYMGNWSNTAYAAR CARAARATGYTNATGCARWSNG GNGAYATHGARACNAAYCCNG GNCCN
Saffold Virus (SAF-V)	TTYACNGAYTTYTTYAARGCNG TNMGNGAYTAYCAYGCNWSNT AYTAYAARCARMGNYTNCARC AYGAYGTNGARACNAAYCCNG GNCCN
Porcine Teschovirus (PTV)	GCNATGACNGTNATGGCNTTYC ARGGNCCNGGNGCNACNAAYT TYWSNYTNYTNAARCARGCNG GNGAYGTNGARGARAAYCCNG GNCCN
Providence Virus (PrV) III	ACNYTNATGGGNAAYATHATG ACNYTNGCNGGNWSNGGNGGN MGNGGNWSNYTNYTNACNGCN GGNGAYGTNGARAARAAYCCN GGNCCN
Providence Virus (PrV) II	AAYWSNGAYGAYGARGARCCN GARTAYCCNMGNGGNGAYCCN ATHGARGAYYTNACNGAYGAY GGNGAYATHGARAARAAYCCN GGNCCN
Providence Virus (PrV) I	YTNGARATGAARGARWSNAAY WSNGGNTAYGTNGTNGGNGGN MGNGGNWSNYTNYTNACNTGY

	GGNGAYGTNGARWSNAAYCCN GGNCCN
Porcine Rotavirus C (PoRV-C)	GGNAAYGGNAAYCCNYTNATH GTNGCNAAYGCNAARTTYCARA THGAYAARATHYTNATHWSNG GNGAYGTNGARYTNAAYCCNG GNCCN
Perina nuda Picorna-like Virus (PnPV) II	ACNMGNGGNGGNYTNMGNMG NCARAAYATHATHGGNGGNGG NCARAARGAYYTNACNCARGA YGGNGAYATHGARWSNAAYCC NGGNCCN
Perina nuda Picorna-like Virus (PnPV) I	GGNCARMGNACNACNGARCAR ATHGTNACNGCNCARGGNTGGG TNCCNGAYYTNACNGTNGAYG GNGAYGTNGARWSNAAYCCNG GNCCN
Operophtera brumata Cypovirus-18 (OpbuCPV-18)	ATHCAYGCNAAYGAYTAYCAR ATGGCNGTNTTYAARWSNAAYT AYGAYYTNYTNAARYTNTGYG GNGAYGTNGARWSNAAYCCNG GNCCN
Ljungan Virus (LV)	TAYTTYAAYATHATGCAYWSNG AYGARATGGAYTTYGCNGGNG GNAARTTYYTNAAYCARTGYGG NGAYGTNGARACNAAYCCNGG NCCN
Lymantria dispar Cypovirus 1 (LdCPV-1)	ATGACNGCNTTYGAYTTYCARC ARGCNGTNTTYMGNWSNAAYT AYGAYYTNYTNAARYTNTGYG GNGAYATHGARWSNAAYCCNG GNCCN
Kashmir Bee Virus (KBV)	ATHGGNTTYYTNAAYAARYTNT AYAARTGYGGNACNTGGGARW SNGTNYTNAAYYTNYTNGCNGG NGAYATHGARYTNAAYCCNGG NCCN
Infectious Myonecrosis Virus of penaeid shrimp (IMNV) II	MGNGAYGTNMGNTAYATHGAR AARCCNTTYGAYAARGARGARC AYACNGAYATHYTNYTNWSNG GNGAYGTNGARWSNAAYCCNG GNCCN
Infectious Myonecrosis Virus of penaeid shrimp (IMNV) I	TGGGAYCCNACNTAYATHGARA THWSNGAYTGYATGYTNCCNCC NCCNGAYYTNACNWSNTGYGG

	NGAYGTNGARWSNAAYCCNGG NCCN
Infectious Flacherie Virus (IFV)	CCNWSNATHGGNAAYGTNGCN MGNACNYTNACNMGNGCNGAR ATHGARGAYGARYTNATHMGN GCNGGNATHGARWSNAAYCCN GGNCCN
Israeli Acute Paralysis Virus (IAPV)	ATHGGNTTYYTNAAYAARYTNT AYMGNTGYGGNGAYTGGGAYW SNATHYTNYTNYTNYTNWSNGG NGAYATHGARGARAAYCCNGG NCCN
Human Rotavirus C (HuRV-C)	GGNGCNGGNTAYCCNYTNATH GTNGCNAAYWSNAARTTYCAR ATHGAYAARATHYTNATHWSN GGNGAYATHGARYTNAAYCCN GGNCCN
Foot-and-Mouth Disease Virus (FMDV)	CAYAARCARAARATHGTNGCNC CNGTNAARCARACNYTNAAYTT YGAYYTNYTNAARYTNGCNGG NGAYGTNGARWSNAAYCCNGG NCCN
Equine Rhinitis B Virus 1 (ERBV-1)	GARGCNACNYTNWSNACNATH YTNWSNGARTTYGCNACNAAYT TYWSNYTNYTNAARYTNGCNG GNGAYGTNGARYTNAAYCCNG GNCCN
Equine Rhinitis A Virus (ERAV)	MGNCAYAARTTYCCNACNAAY ATHAAYAARCARTGYACNAAYT AYWSNYTNYTNAARYTNGCNG GNGAYGTNGARWSNAAYCCNG GNCCN
Ectropis obliqua Picorna-like Virus (EoPV) II	ACNMGNGGNGGNYTNCARMGN CARAAYATHATHGGNGGNGGN CARMGNGAYYTNACNCARGAY GGNGAYATHGARWSNAAYCCN GGNCCN
Ectropis obliqua Picorna-like Virus (EoPV) I	GGNCARMGNACNACNGARCAR ATHGTNACNGCNCARGGNTGGG CNCCNGAYYTNACNCARGAYG GNGAYGTNGARWSNAAYCCNG GNCCN
Encephalomyocarditis Virus (ECMV)	GTNTTYGGNYTNTAYMGNATHT TYAAYGCNCAYTAYGCNGGNT AYTTYGCNGAYYTNYTNATHCA

	YGAYATHGARACNAAYCCNGG NCCN
Equine Encephalosis Virus (EEV)	MGNMGNYTNCCNGARWSNGCN CARYTNCCNCARGGNGCNGGN MGNGGNWSNYTNGTNACNTGY GGNGAYGTNGARGARAAYCCN GGNCCN
Dendrolimus punctatus Cypovirus-1 (DpCPV-1)	ATGACNGCNTTYGAYTTYCARC ARGCNGTNTTYMGNWSNAAYT AYGAYYTNYTNAARYTNTGYG GNGAYGTNGARWSNAAYCCNG GNCCN
Duck Hepatitis Virus-1 (DHV-1)	GCNTTYGARYTNAAYYTNGARA THGARWSNGAYCARATHMGNA AYAARAARGAYYTNACNACNG ARGGNGTNGARCCNAAYCCNG GNCCN
Drosophila C Virus (DCV)	CARGGNATHGGNAARAARAAY CCNAARCARGARGCNGCNMGN CARATGYTNYTNYTNYTNWSNG GNGAYGTNGARACNAAYCCNG GNCCN
Cricket Paralysis Virus (CrPV)	YTNGTNWSNWSNAAYGAYGAR TGYMGNGCNTTYYTNMGNAAR MGNACNCARYTNYTNATGWSN GGNGAYGTNGARWSNAAYCCN GGNCCN
Bovine Rhinovirus-2 (BRV-2)	YTNMGNYTNACNGGNGARATH GTNAARCARGGNGCNACNAAY TTYGARYTNYTNCARCARGCNG GNGAYGTNGARACNAAYCCNG GNCCN
Bovine Rotavirus C (BoRV-C)	GGNATHGGNAAYCCNYTNATH GTNGCNAAYWSNAARTTYCAR ATHGAYMGNATHYTNATHWSN GGNGAYGTNGARYTNAAYCCN GGNCCN
*Bombyx mori C*ypovirus 1 (BmCPV-1)	MGNACNGCNTTYGAYTTYCARC ARGAYGTNTTYMGNWSNAAYT AYGAYYTNYTNAARYTNTGYG GNGAYATHGARWSNAAYCCNG GNCCN
New Adult Dirrhea Virus (ADRV-N)	TTYTTYGAYWSNGTNTGGGTNT AYCAYYTNGCNAAYWSNWSNT GGGTNMGNGAYYTNACNMGNG

	ARTGYATHGARWSNAAYCCNG GNCCN
Acute Bee Paralysis Virus (ABPV)	ACNGGNTTYYTNAAYAARYTNT AYCAYTGYGGNWSNTGGACNG AYATHYTNYTNYTNYTNWSNG GNGAYGTNGARACNAAYCCNG GNCCN

Table 4.2 Sequences of 6 viral 2A peptides from *Picornaviridae* members and their consensus sequence for construction of 72nt library (Y=C or T; R=A or G; N=A, T, G, or C; W=A or T; S=C or G; M=A or C; H=A, C or T).

Virus	**2A Peptide Nucleotide Sequence**
Consensus	RMYGTGAHDCRGDCTAYRAATTATGAYCTRCTGAAGT TG--GCNGGAGACGTBGAGTCSAAYCCTCGGACCT
Thosea Asigna Virus	----------GAGGGCAGAGGAAG-TCTGCTAACAT-G---C- GGTGACGTCGAGGAGAATCCTCGGACCT
Porcine Teschovirus	----------GCTATAACTTCAG-CCTGCTGAAGCAG-- GCTGGAGACGTGGAGGAGAACCCTCGGACCT
Foot-and-Mouth Disease Virus	---GTGAAACAGACTTTGAATTTTGACCTTCTCAAGTTG- -GCGGGAGACGTGGAGTCCAACCCTCGGACCT
Equine Rhinitis A Virus	---------CAGTGTACTAATTATGCTCTCTTGAAATTG-- GCTGGAGATGTTGAGAGCAACCCTCGGACCT
Bombyx mori Flacherie Virus	ACTCTGACGAGGGCGAAGATTGAGGATGAATTGA-- TTCGTGCAGGA---ATTGAATCAAATCCTCGGACCT
Bombyx mori Cytoplasmic Polyhedrosis Virus	GACGTTTTTCGCTCT---AATTATGACCTACTAAAGTTG- TGC-GGTGATATCGAGTCTAATCCTCGGACCT

To observe the ribosome skipping event caused by a 2A peptide, a designed 72nt library of 2A peptides was inserted into a dual fluorescent reporter plasmid consisting of mBeRFP and eGFP coding sequences separated by a putative 2A peptide. Additionally, a RepA degradation tag (18) was placed at the N-terminus of the eGFP sequence. In this way, when a putative 2A peptide sequence does not result in ribosome skipping, an mBeRFP-eGFP fusion will be translated. However, when an active 2A peptide is present,

ribosome skipping will occur and a corresponding decrease in eGFP (due to the exposure of the N-terminal degradation tag) will take place. This can be measured using fluorescence spectroscopy (Figure 4.2).

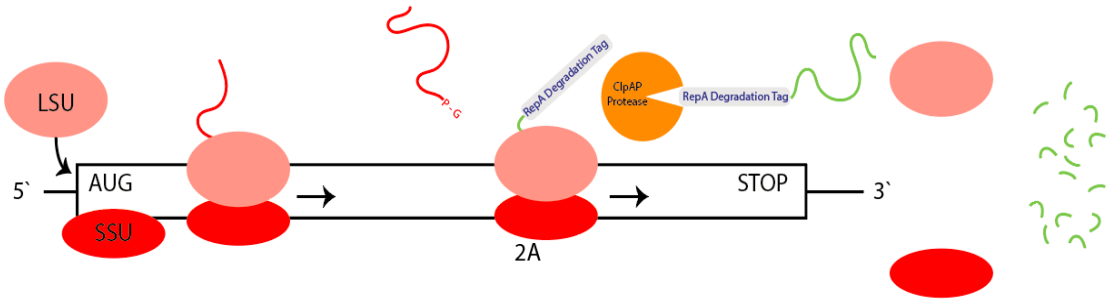


Figure 4.2. Design strategy for assessing the production of active 2A peptides in bacteria. Upon active 2A peptide translation, and skipping, an N-terminal degradation signal is exposed and degradation by ClpAP protease occurs. This results in a lowered fluorescent protein level which can be observed by fluorescence spectroscopy. Ribosomal small subunit (SSU) and large subunit (LSU) are indicated in red and pink, respectively.

4.2 Methods

4.2.1 Construction of Genetic Elements

The 2A library constructs were obtained as 4 separate primers (IDT). A 20µL overlap-PCR (using Phusion high-fidelity DNA polymerase HF mastermix, Thermo Fisher; 0.2µM primers; 20 extension cycles of 30 seconds) product containing the 2A peptide library and N-terminal degradation tag (Appendix D, G) was conducted in two separate parts. First, the 5` and 3` halves of the library were allowed to anneal by slow cooling (From 95°C to 20°C in 5°C decreasing steps each minute) in a Montreal Biotech Inc. Biometra T Gradient thermal cycler. Next, a PCR using the aforementioned protocol was performed using both the 3` and 5` halves to fill in the remaining nucleotides of the library. Subsequent clean up of the PCR was performed using BioBasic EZ-10 Spin Column DNA Miniprep Kit. A digestion of the PCR product with BamHI and NdeI was

performed. All digested products were gel-purified from a 1% agarose gel run at 10V/cm for 30 minutes. In designing the reporter plasmid, the previously constructed low-copy number pRNA-IN plasmid was modified. A restriction digest was performed on pRNA-IN with BamHI and NdeI (NEB) (1µg of plasmid DNA was digested in a 50µL reaction with 10 units of each restriction enzyme). The digestion products (both backbone plasmid and insert) were gel purified and a ligation was performed. *Escherichia coli* DH5α (100µL aliquots) were transformed by electroporation with the 1µL of the ligation product as described in previous chapters. Successful transformants (typically 25-50)underwent plasmid purification using the BioBasic EZ-10 Spin Column DNA Miniprep Kit.

4.2.2 Fluorescence Spectroscopy

Escherichia coli BL21-Gold(DE3) were transformed by electroporation with purified p2A-eGFP. Overnight cultures of 10 transformants were diluted to an OD_{600nm} 0.05 in 5mL fresh LB media with 50µg/mL kanamycin. Cultures were incubated at 37°C while shaking at 220rpm (Thermo Fisher MaxQ 4000) and grown until reaching OD_{600nm} 1.0. Cells were then harvested by centrifugation at 5000 x g for 2 min. The cell pellet was washed in 1 mL ice-cold sterile 1 X PBS before resuspension in 500 µL sterile 1 X PBS. The cell suspension was incubated for 90 min at 20°C to allow for the maturation and degradation of fluorescent proteins. Subsequently, the cell suspension was analyzed using a Cary Eclipse Fluorescence Spectrophotometer (Varian, Inc. Agilent Technologies) and its emission scan capabilities as well as the Varian Eclipse software suite. Upon blanking the spectrophotometer with PBS, relative fluorescence was calculated. Excitation of both fluorophores at 488nm and emission read from 500nm-700nm (1nm step size) allowed for

relative determination of fluorescent protein abundance in in transformants containing either the p2A-eGFP plasmid or the unmodified pRNA-IN from which it was derived. The peak fluorescence in both the p2A-eGFP containing cells and the pRNA-IN cells were used for comparison and assignment of an ensemble percent ribosome skipping value as calculated using the following equation (Equation 4.1).

$$Percent\ Ribosome\ Skipping = \left(\frac{Maximum\ Fluorescence_{\ pRNA-IN-containing\ cells} - Maximum\ Fluorescence_{\ p2A-containing\ cells}}{Maximum\ Fluorescence_{\ pRNA-IN-containing\ cells}} \times 100\%\right) \quad \text{(Equation 4.1)}$$

4.3 Results

Construction of 2A peptide libraries resulted in two separate complexities, one with 1.8 x 10^9 possible sequences and one with 165,888 possible sequences. The sequence logos for the peptide itself (Figure 4.3a) and the nucleotide sequences (Figure 4.3b-4.3c) used in construction are illustrated below. Colonies (approximately 10) were obtained for only the 72nt 2A peptide library.

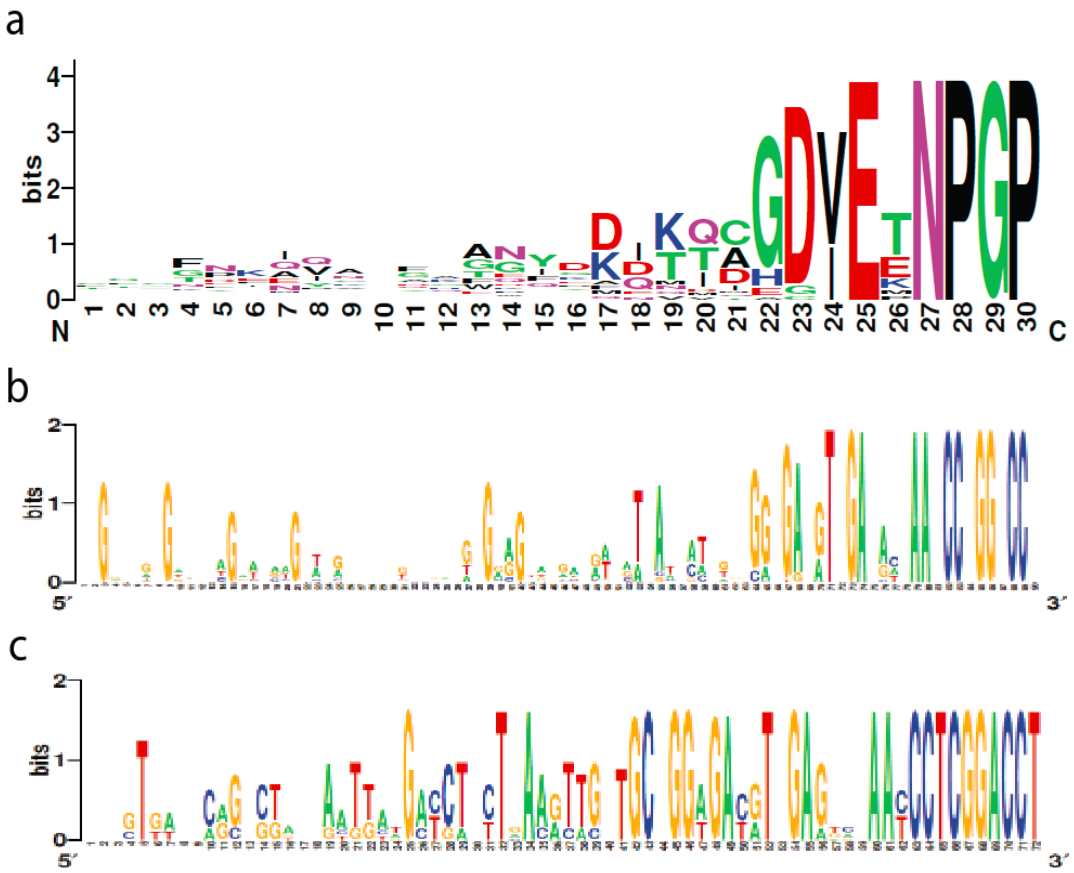


Figure 4.3. Sequence logos of 2A peptide libraries. Consensus peptide sequence from 37 well-studied 2A peptides (a). 90nt library constructed to generate maximal diversity in 2A peptide sequence (b). 72nt library constructed with more tightly prescribed sequence for initial characterization (c).

Fluorescence spectroscopic analysis of the 72nt 2A peptide library containing plasmid in *E.coli* was compared to the fluorescence profile of the plasmid on which it was based (pRNA-IN). The 2A peptide library containing plasmid also contained an N-terminal degradation tag fused to eGFP, so that any eGFP with exposed N-termini should be degraded within the 90 minute incubation step prior to the measurement, which is 60 minutes longer than Butz and colleagues (18), resulting in complete degradation. Transformants with the plasmid containing the smaller 2A library of 72nt displayed a 6.2% reduction in overall green fluorescence as compared to the reference pRNA-IN containing transformants (Figure 4.4). The decrease in mean fluorescence across triplicate trials is less pronounced at a decrease of 6.0% ± 2.5% (Appendix H)

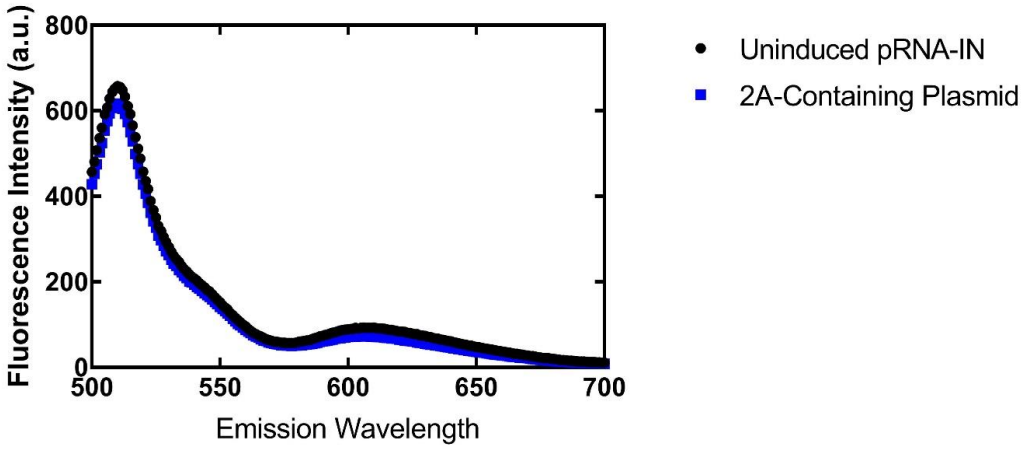


Figure 4.4. Ensemble fluorescence emission scan upon excitation at 488nm comparing fluorescence of cells with 2A library and N-terminally degradation tagged eGFP (blue) to those with pRNA-IN as a comparison (black). Emission peaks at 510nm and 610nm are indicative of relative abundance of eGFP and mBeRFP respectively.

4.4 Discussion

2A peptides offer an intriguing method for production of multiple proteins in equimolar amounts at varying efficiencies (porcine teschovirus 2A has approximately 80% skipping

efficiency, while foot and mouth disease virus 2A has approximately 50% skipping efficiency) (12) within a cell. Indeed, 2A peptides have been widely studied within eukaryotic organisms to establish multi-gene expression systems (5, 12, 14, 16). However, 2A peptides have not been tested thoroughly in prokaryotic organisms (19), but when they have, they are shown to be ineffective in causing ribosomal skipping between two proteins (20, 21). This is perhaps due to the fact that there has not been systematic forward-engineering of *de novo* 2A peptide sequences performed to date. The approach outlined above generates a library of potential sequences based on what is known about 2A peptide sequence similarity, in order to increase the odds of successfully identifying sequence variants for bacteria such as *E. coli*. It has been hypothesized that the 2A peptide mechanism is due to the tight ribosome exit tunnel disallowing the addition of a proline and instead releasing the nascent peptide (7, 22). Given the conservation in ribosomal structure and function, and previous work on expression of eukaryotic proteins from bacterial ribosomes (23), it is possible and reasonable to expect that one 2A peptide sequence within a library of over 100,000 members may exhibit similar activity in both prokaryotes and eukaryotes. Indeed, preliminary evidence (Figure 4.4) illustrates that functional 2A peptides may be present and operate as desired in a bacterial chassis within the library. While the observed change in fluorescence is small, it is nonetheless informative as a proof of concept with respect to 2A peptides within a bacterial organism. As the measurement was an ensemble measurement, it cannot be said what proportion of the 72nt library results in functional 2A peptides, only that a slight reduction in eGFP corresponding to exposure of an N-terminal degradation tag has occurred. Alternatively, it may be the case that only a subset of the constructs are able to efficiently express the desired fusion protein, which bears further validation. Therefore, fluorescence activated

cell sorting should be employed in further studies to specifically isolate clones in which eGFP fluorescence is decreased relative to the control transformants. This will allow further validation of the active members of each library. Increasing the transformation efficiency of the libraries will be essential in capturing the entire sequence complexity present in the respective libraries.

Given that the majority of studies to this point have only interrogated known 2A peptide sequences from a small subset of *Picornaviridae* viruses in eukaryotic organisms, the 72nt library tested in this study represents a major step forward in the design of tools for multi-gene expression systems in eukaryotic organisms.. Additionally, as it is difficult to effectively sample every member of a library with such complexity, it would be useful to model the interaction between a 2A peptide sequence and the prokaryotic ribosome. Models for the ribosome-skipping mechanism in eukaryotes have been proposed (6, 7, 24), but given that 2A peptide activity has not been described or investigated thoroughly in prokaryotes, little explanation as to why 2A peptide separation may not operate in prokaryotes has been offered (7). However, through analysis of the precise mechanisms of 2A peptide activity, molecular dynamics simulations of this process may shed light on the process in both prokaryotic and eukaryotic systems of 2A peptide mediated multi-gene expression systems and facilitate improved rational design in the future.

References

1. Stanway, G. (1990) Structure , function and evolution of picornaviruses. *J. Gen. Virol.* **71**, 2483–2501

2. Forss, S., Strebel, K., Beck, E., and Schaller, H. (1984) Nucleotide sequence and genome organization of foot-and-mouth disease virus. *Nucleic Acids Res.* **12**, 6587–6601

3. Luke, G. A., de Felipe, P., Lukashev, A., Kallioinen, S. E., Bruno, E. A., and Ryan, M. D. (2008) Occurrence, function and evolutionary origins of "2A-like" sequences in virus genomes. *J. Gen. Virol.* **89**, 1036–1042

4. Gunišová, S., Hronová, V., Mohammad, M. P., Hinnebusch, A. G., and Valášek, L. S. (2017) Translation reinitiation in microbes and higher eukaryotes. *FEMS Microbiol. Rev.* 10.1093/femsre/fux059

5. Yang, X., Cheng, A., Wang, M., Jia, R., Sun, K., Pan, K., Yang, Q., Wu, Y., Zhu, D., Chen, S., Liu, M., Zhao, X. X., and Chen, X. (2017) Structures and corresponding functions of five types of picornaviral 2A proteins. *Front. Microbiol.* **8**, 1–14

6. Ryan, M. D., Donnelly, M., Lewis, A., Mehrotra, A. P., Wilkie, J., and Gani, D. (1999) A model for nonstoichiometric, cotranslational protein scission in eukaryotic ribosomes. *Bioorg. Chem.* **27**, 55–79

7. Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., and Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal "skip." *J. Gen. Virol.* **82**, 1013–1025

8. Sharma, P., Yan, F., Doronina, V. A., Escuin-Ordinas, H., Ryan, M. D., and Brown, J. D. (2012) 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res.* **40**, 3143–3151

9. Melnikov, S., Ben-Shem, A., Garreau De Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012) One core, two shells: Bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **19**, 560–567

10. Roosien, J., Belsham, G. J., Ryan, M. D., King, A. M., and Vlak, J. M. (1990) Synthesis of foot-and-mouth disease virus capsid proteins in insect cells using baculovirus expression vectors. *J. Gen. Virol.* **71**, 1703–11.

11. De Felipe, P., Martín, V., Cortés, M. L., Ryan, M., and Izquierdo, M. (1999) Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. *Gene Ther.* **6**, 198–208

12. Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K.,

Shin, B. A., and Choi, S. Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*. **6**, 1–8

13. Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., and Liu, J. (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 1–9

14. Szymczak, A. L., and Vignali, D. A. (2005) Development of 2A peptide-based strategies in the design of multicistronic vectors. *Expert Opin. Biol. Ther.* **5**, 627–638

15. Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M. D. (2001) The 'cleavage'activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like'sequences. *J. Gen. Virol.* **82**, 1027–1041

16. Wang, Y., Wang, F., Wang, R., Zhao, P., and Xia, Q. (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci. Rep.* **5**, 1–10

17. Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004) WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188–1190

18. Butz, M., Neuenschwander, M., Kast, P., and Hilvert, D. (2011) An N-terminal protein degradation tag enables robust selection of highly active enzymes. *Biochemistry*. **50**, 8594–8602

19. Dechamma, H. J., Kumar, C. A., Nagarajan, G., and Suryanarayana, V. V. S. (2008) Processing of multimer FMD virus VP1-2A protein expressed in E. coli into monomers. *Indian J. Exp. Biol.* **46**, 760–763

20. Ryan, M. D., and Drew, J. (1994) Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J.* **13**, 928–33

21. Donnelly, M. L. L., Gani, D., Flint, M., Monaghan, S., and Ryan, M. D. (1997) The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J. Gen. Virol.* **78**, 13–21

22. Lim, V. I., and Spirin, A. S. (1986) Stereochemical analysis of ribosomal transpeptidation conformation of nascent peptide. *J. Mol. Biol.* **188**, 565–574

23. Kolb, V. A., Makeyev, E. V., and Spirin, A. S. (2000) Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. *J. Biol. Chem.* **275**, 16597–16601

24. De Felipe, P., Hughes, L. E., Ryan, M. D., and Brown, J. D. (2003) Co-translational, intraribosomal cleavage of polypeptides by the foot-and-mouth

disease virus 2A peptide. *J. Biol. Chem.* **278**, 11441–11448

5 Synthesis and Conclusions

5.1 RNA-IN & RNA-OUT as a Flexible Translation Down-Regulatory Strategy

The orthogonal approach to translation regulation offered by the adaptation of the endogenous Tn*10* RNA-IN/OUT system affords specific control over translation (1). For bioengineering and synthetic biology purposes, it is imperative that well-characterized regulatory approaches be utilized to produce reliable quantities of downstream products or metabolites. In both the natural and engineered approach, the interaction between the RNA-OUT molecule and the mRNA of interest extends into the coding sequence. This is an appropriate approach for the natural system in which the mRNA being regulated is always the same. However, when one has the goal of regulating translation of several heterologous proteins within *E. coli*, an amendment must be made. As the interaction between RNA-IN and RNA-OUT molecules extends into the coding sequence, a reasonable alteration to the natural approach is to engineer the N-terminus of the coding sequence so that a single RNA-OUT can reliably regulate the translation of an mRNA with a matching RNA-IN motif and 5` regions. In this novel approach, an N-terminal hexahistidine or FLAG tag sequence was appended onto a fluorescent protein coding region and a complementary region was inserted into the 5` arm of the RNA-OUT hairpin. Testing this approach in a dual fluorescent reporter system expressed from a single cistron, it was noted in this work that the fluorescence of both fluorescent proteins decreased upon induction of the *trans*-acting regulatory RNA. This was not altogether

unexpected, as this double-stranded RNA molecule is a likely target for RNase III degradation *in vivo* (2–4). Furthermore, this was not investigated by previous studies (1), which raises the concern that the observed down-regulation is at least in part due to destabilization of the mRNA in addition to ribosome binding site occlusion. To further investigate this assertion, RT-qPCR was performed on each of the reporter genes and compared to an endogenous control. The relative transcript level of both reporters was decreased to a similar degree, indicating that in this case the RNA-OUT control mechanism is likely to be due to degradation of the RNA duplex and not simply occlusion of the ribosome binding site. To investigate if a specific *trans*-acting RNA molecule was still able to down-regulate translation of a single reporter, a similar fluorescence spectroscopy experiment was carried out with a single fluorescent reporter. In this case, as was previously reported (1), a high level of repression was observed. This higher level of repression compared to the dual fluorescent reporter system perhaps stems from an increased accessibility to the RNA-IN binding motif due to the transcript being shorter. The shorter transcript is less likely to be involved in long-distance base-pairing, especially with regions with similar nucleotide identity as in the two fluorescent protein coding sequences. Therefore, the *trans*-acting regulatory RNA is likely able to interact more readily with the shorter, less tightly structured mRNA.

5.2 Inversion of the RNA-IN and RNA-OUT System Provides a Flexible Translational Activation Strategy

Given that RNA-IN/OUT has been demonstrated to be an effective and orthogonal strategy for translation down-regulation, it would be interesting to perform an inversion

of its typical function and turn this system into an activator of translation, adding another tool to the translation regulation tool-box. In doing this, an initially highly structured mRNA molecule had its structure disrupted and ribosome binding site revealed through an interaction upstream with an unstructured *trans*-acting RNA molecule. In performing this inversion, similar to toehold switch architecture (5) an initially highly structured mRNA molecule occludes the ribosome binding site, but upon binding of a cognate *trans*-acting RNA molecule, the structure is relieved and translation can proceed. This strategy was employed, using the RNA-IN and RNA-OUT recognition motifs to achieve an average of 4.1% (± 12.4%) increase in ensemble eGFP fluorescence upon induction of the *trans*-acting RNA species. This activation is notable; however, it could certainly be increased through optimization and testing of more members of the orthogonal library of RNA-IN/OUT motifs. The major portion that bears improvement is the initial inhibition of translation by the structured mRNA region prior to induction of an activating *trans*-acting RNA species. In this case, green and red fluorescence are seen even in the absence of the activating RNA, indicating that translation of the fluorescence reporters occurs in the repressed state. The activation would certainly be more pronounced if the initial background translation level was lower so as to facilitate a robust activation response. This initial low level may be difficult to achieve, however, as it may require a tightly structured mRNA which would be difficult for the *trans*-acting regulatory RNA to disrupt. Moreover, to mitigate the translation of reporters in the repressed state, it may be useful to place the induction of the *trans*-acting regulatory RNA under control of a more stringent promoter such as pBAD. It is certainly possible that the current system has leaky expression of the regulatory activator RNA and as such the baseline fluorescence level is skewed upwards. Additionally, only a subset of the possible RNA-IN/RNA-OUT pairs

were successfully cloned and analyzed. As such, potential for a larger dynamic range of translation activation using this approach may indeed exist, and should be the subject of future work. Inversion of the classical repression-only natural system represents a step forward in the use of this RNA-only translation-regulation system. Future investigation should be focused on expanding the library of working activating RNA-IN/OUT pairs of molecules as well as quantifying the response via RT-qPCR. This technique in tandem with the previous fluorescence characterization will allow for an accurate assessment of the transcript stability and comparison between protein production and transcript level.

5.3 Construction of a Synthetic 2A Peptide Library

2A peptides are naturally-occurring viral peptides which separate regions of the *Picornaviridae* polyprotein from one another (6). Via the short 2A peptide sequence, these viruses are able to translate multiple messages without separate re-initiation events. It is proposed that this is due to the 2A peptide sequence being in a conformation within the ribosomal exit tunnel such that the next amino acid (a proline) cannot be accommodated and added to the nascent polypeptide chain, and instead, the preceding polypeptide is released and translation continues with the following proline beginning a new polypeptide chain (7, 8). Importantly, these 2A peptides have thus far only been observed to be active in eukaryotic chassis organisms (9). This is not surprising, as the viruses from which 2A peptides arise infect eukaryotic, typically mammalian organisms. However, only a few commonly studied 2A peptides have been examined in bacterial systems with very inconclusive and non-replicated results showing cleavage of a protein into separate subunits (10). Given the unclear nature of the results obtained by prior groups, it makes sense to attempt to test many different 2A peptide sequences with

potential to elicit this ribosome skipping phenomenon within bacteria. In this thesis, two libraries of different lengths and complexities were created based upon multiple sequence alignments for 2A peptide sequences. These libraries were placed between two fluorescent protein coding sequences and the downstream fluorescent protein also had an N-terminal degradation tag appended on to facilitate measurement of ribosome skipping. In measuring the skipping activity of one such 72nt library, a 6.0% (± 2.5%) decrease in fluorescence (indicative of possible exposure of the N-terminal degradation tag) was observed. Because of the small number of colonies obtained in screening (approximately 10), this variability captures at most 10 out of a possible 165,888 variants. Therefore, a higher number of clones must be obtained to fully sample the entire 72nt library of 2A sequence variants. While the observed fluorescence change is not large, and not necessarily representative of the library as a whole, it does represent that a possible 2A peptide sequence that may be effective within a prokaryotic chassis. Given that there is no prior conclusive evidence of bacterial 2A peptide activity, the apparent activity shown here is unexpected. However, this measurement was performed on an entire population by fluorescence spectroscopy, which does not allow for the isolation of individual clones for sampling the entire diversity of the library. Therefore, individual clones and sequences should be obtained for further characterization of the 2A sequences developed here using the above library generating procedure. While this decrease in fluorescence is encouraging, it is possible that it is perhaps due to an error in experimental design. In a case in which a premature arrest of translation occurred in the region occupied by the 2A peptide library, the resulting red fluorescent protein signal would be indistinguishable from that resulting from an active 2A peptide sequence. Therefore, a control experiment may be designed in which a stop codon is placed between the two coding sequences to

serve as a baseline for the amount of red fluorescence, and readthrough green fluorescence one can expect in the absence of any active 2A peptide sequence. An alternative explanation for the fluorescence decrease is that this simply represents inter-clonal fluorescence variability between two *E. coli* samples, one containing the 2A library and one lacking the 2A library. With this initial information and data, a further study should be conducted to determine exactly which variants may contain 2A peptide activity, if any, by fluorescence activated cell sorting and sequencing of the identified variants.

5.4 RNA-Based Regulation of Translation Initiation

As has been illustrated throughout this work, use of RNA as a tool for modulation of translation (by RNA-IN/RNA-OUT interaction or possible 2A peptide activity) in bacterial systems is an exciting and worthwhile area of study. Using specific interaction motifs in the RNA-IN/RNA-OUT system, one can down-regulate translation irrespective of coding sequence. This facilitates the optimization of downstream genetic and metabolic engineering projects in which precise ratios of reactants must be produced. Interestingly, the down regulation can also be inverted while still maintaining the specificity of the RNA-IN/OUT interaction to act as an activator. Indeed, this RNA-based system now has been engineered for increased versatility and modularity for regulation of many different genes in either a positive or negative manner.

A strategy for developing new tools for genetic engineers in the form of large 2A peptide libraries has also been elucidated in this work. These 2A peptides allow for equimolar expression of separate proteins from a long multicistronic mRNA without the need for re-initiation to occur. Only ever documented in eukaryotes, this technology has been widely used (11–14). However, by constructing a sufficiently complex library, preliminary

evidence suggests that this 2A peptide “skipping” phenomenon may also occur within bacteria. The 2A peptide sequence can be said to be an RNA-based initiation regulator, as its sequence and structure specifically modulates the ribosome skipping.

Overall, this work has contributed an RNA-only approach to orthogonally down-regulate translation initiation via the RNA-IN/RNA-OUT system irrespective of coding sequence, extending the utility of this regulatory approach. Moreover, it has been demonstrated that this strategy can be used in inverse to activate translation, increasing the utility of the RNA-IN/RNA-OUT technology. Finally, this work has developed a framework for construction of an RNA-based strategy for production of multiple proteins from a single translation initiation event via a large 2A peptide library. Taken together, these novel applications and developments are exciting steps forward for synthetic biology and bioengineering. In particular, due to the specificity of these applications and interactions, the applications are broad. For instance, if a bioengineer wishes to express an entire suite of genes to achieve a specific reactant stoichiometry, a 2A peptide may be employed to yield a predictable amount of ribosome skipping and therefore desired ratio of protein produced. Similarly, an RNA-IN/RNA-OUT interaction may be used to specifically repress translation of specific mRNAs to yield a desired level of protein product. Overall, these RNA-based tools have useful applications which will certainly be expanded by further investigation.

References

1. Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454

2. Case, C. C., Roels, S. M., Jensen, P. D., Lee, J., Kleckner, N., and Simons, R. W. (1989) The unusual stability of the IS10 anti-sense RNA is critical for its function and is determined by the structure of its stem-domain. *EMBO J.* **8**, 4297–4305

3. Caron, M. P., Lafontaine, D. A., and Massé, E. (2010) Small RNA-mediated regulation at the level of transcript stability. *RNA Biol.* **7**, 140–144

4. Case, C. C., Simons, E. L., and Simons, R. W. (1990) The IS10 transposase mRNA is destabilized during antisense RNA control. *EMBO J.* **9**, 1259–1266

5. Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: De-novo-designed regulators of gene expression. *Cell*. **159**, 925–939

6. Forss, S., Strebel, K., Beck, E., and Schaller, H. (1984) Nucleotide sequence and genome organization of foot-and-mouth disease virus. *Nucleic Acids Res.* **12**, 6587–6601

7. Ryan, M. D., Donnelly, M., Lewis, A., Mehrotra, A. P., Wilkie, J., and Gani, D. (1999) A model for nonstoichiometric, cotranslational protein scission in eukaryotic ribosomes. *Bioorg. Chem.* **27**, 55–79

8. Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., and Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal "skip." *J. Gen. Virol.* **82**, 1013–1025

9. Sharma, P., Yan, F., Doronina, V. A., Escuin-Ordinas, H., Ryan, M. D., and Brown, J. D. (2012) 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res.* **40**, 3143–3151

10. Dechamma, H. J., Kumar, C. A., Nagarajan, G., and Suryanarayana, V. V. S. (2008) Processing of multimer FMD virus VP1-2A protein expressed in E. coli into monomers. *Indian J. Exp. Biol.* **46**, 760–763

11. Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A., and Choi, S. Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*. **6**, 1–8

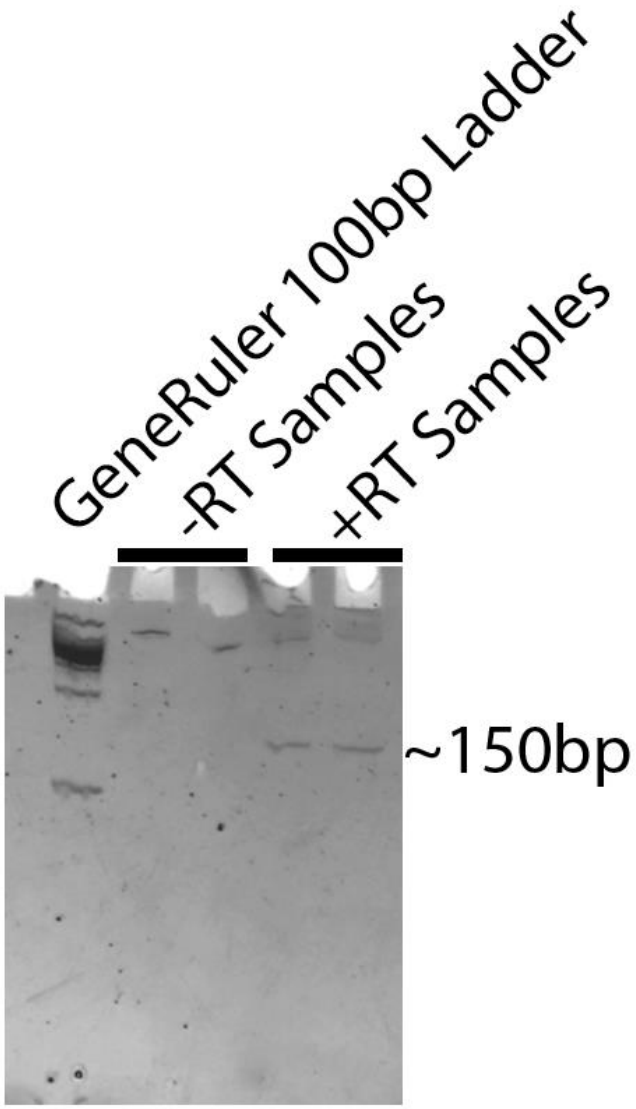
12. Wang, Y., Wang, F., Wang, R., Zhao, P., and Xia, Q. (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci.*

Rep. **5**, 1–10

13. Subramanian, V., Schuster, L. A., Moore, K. T., Taylor, L. E., Baker, J. O., Vander Wall, T. A., Linger, J. G., Himmel, M. E., and Decker, S. R. (2017) A versatile 2A peptide-based bicistronic protein expressing platform for the industrial cellulase producing fungus, Trichoderma reesei. *Biotechnol. Biofuels*. **10**, 1–15

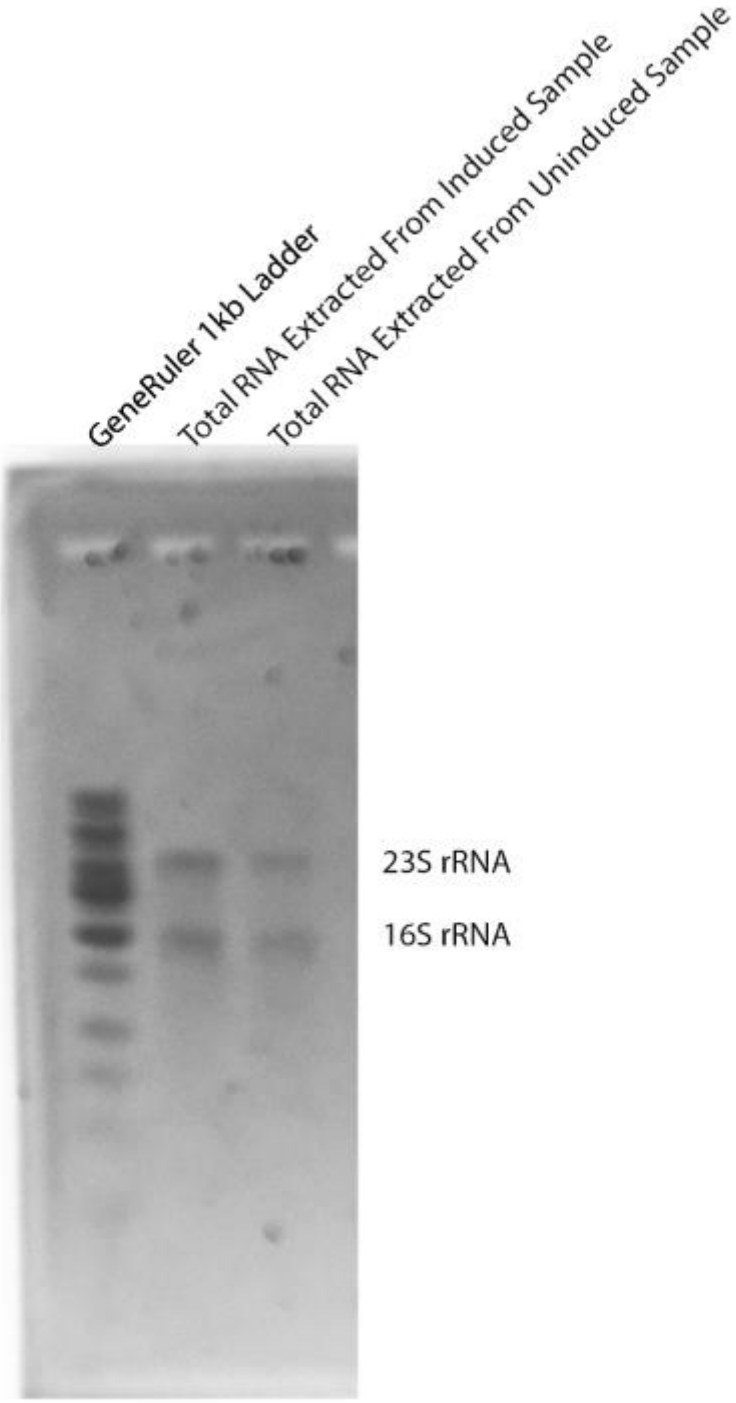
14. De Felipe, P., Martín, V., Cortés, M. L., Ryan, M., and Izquierdo, M. (1999) Use of the 2A sequence from foot-and-mouth disease virus in the generation of retroviral vectors for gene therapy. *Gene Ther.* **6**, 198–208

Appendices



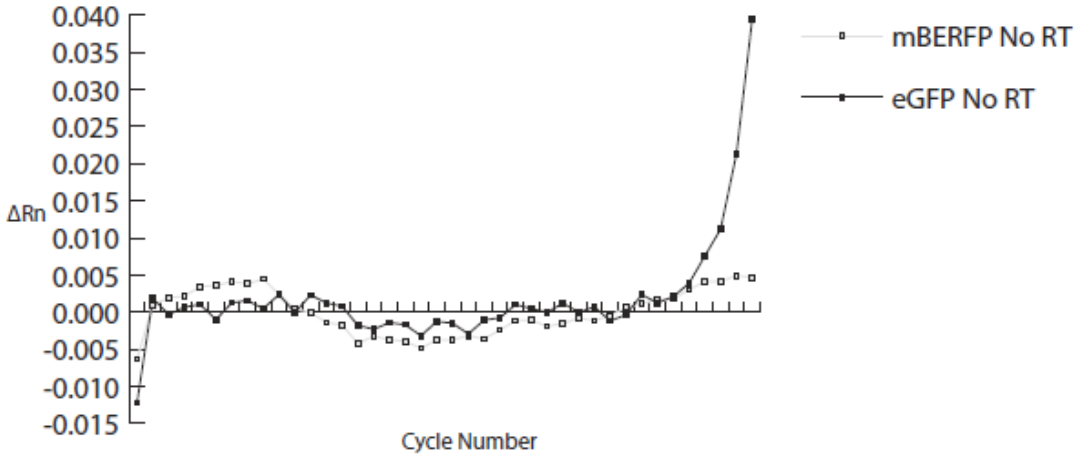
Appendix A. RT-PCR amplification and primer specificity testing.

Reverse transcription reactions using AMV Reverse Transcriptase (New England Biolabs) were performed on RNA extracted from an uninduced pRNA-IN/pRNA-OUT (A/S5) cotransformant BL21-Gold(DE3) cells. The primer pairs used amplified a region of the eGFP and mBeRFP coding sequences corresponding to a 150bp amplicon. As can be seen in the above 12% native polyacrylamide gel, the amplicons are present in the case of PCR after reverse transcription but absent in the non-reverse transcribed samples.



Appendix B. RNA quality assessment.

1% denaturing agarose gel of total RNA extracted from induced and uninduced BL21-Gold(DE3) cells cotransformed with pRNA-IN and pRNA-OUT. The presence of clear bands in the size range expected for 16S rRNA (~1.5kb) and 23S rRNA (~2.9kb) are indicative of a high-quality extraction.



Appendix C. RT-qPCR negative control experiments.

In the qPCR experiments, to prepare cDNA prior to PCR, a reverse transcriptase reaction was performed. To ensure that there was no contamination in the extracted total RNA, a No-RT control was performed in tandem. Shown above are NM values for transcripts for both fluorescent proteins used. Both are well below threshold for assignment of a Tc value.

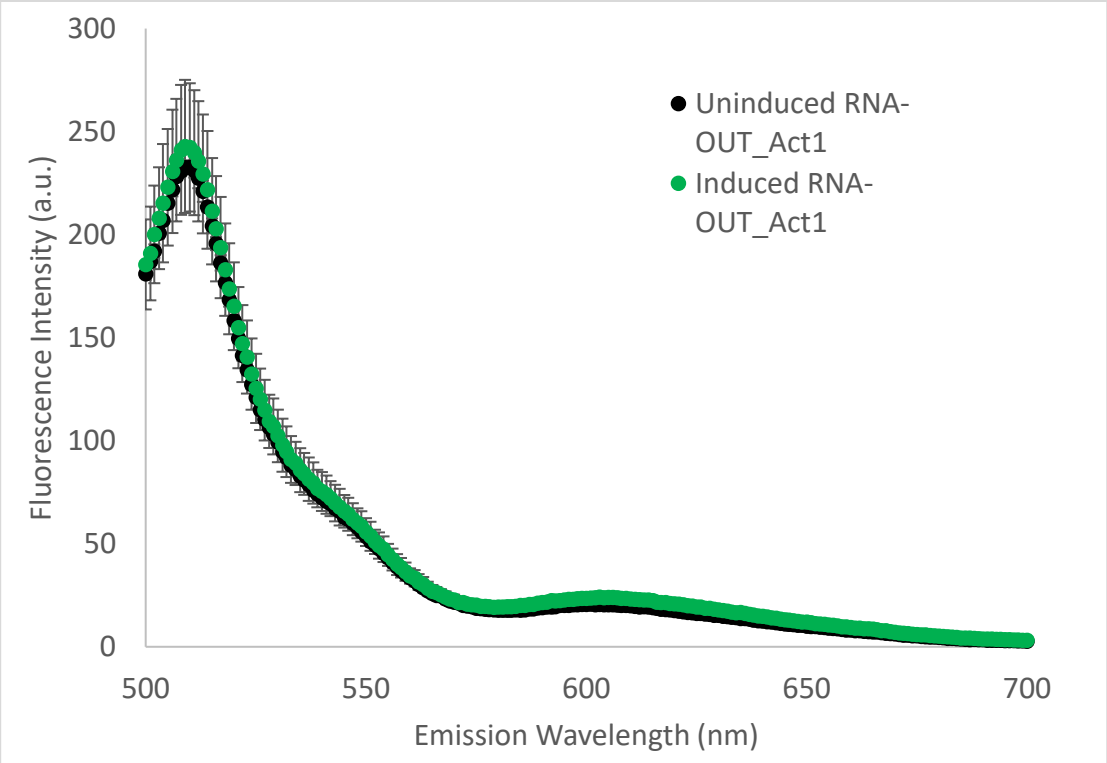
Appendix D. Preparation of 2A library DNA constructs.

Construction of 2A library DNA constructs by overlap-extension PCR of synthesized oligonucleotides. One library of 90bp size was constructed, while a second, smaller library of approximately 72bp was also constructed. Results visualized above on a 1% agarose gel.

Primer Name	Primer Sequence (5`- 3`)
pIN Fragment 1 FLAG-RNA-IN: S4	CGGTGAGCGAATTCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAA CAAAACACCTAATCTAGAGCCAAAAATCAATAAGGAGACAACAAGATGGA CTACAAAGACGATGACGACAAGAGCAGCGGCCTGGTGCCGCGCGGCAGCA TGGCTAGCATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTAC ATGGAGGGCACCGTGAACAACCACCACTTCAAGTGCACATCCGAGGGCGA AGGCAAGCCCTACGAGGGCACCCAGACCCAGAGAATCAAGGTGGTCGAGG GCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTACCAGCTTCATGTAC GGCAGTCACACCTTCATCAACCACACCCAGGGCATCCCCGACTTCTGGAA GCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACATACGAAG ACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGC CTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCCCATCCAACGGCCC TGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCCACACCGAGATGCTGT ACCCCGCTGACGGCGGCCTGGAAGGCAGAACCGCGCTGGCCCTGAAGCTC GTGGGCGGGGGCCACCTGATCTGCAACTTCAAGACCACATACAGATCCAA GAAACCCGCTAAGAACCTCAAGATGCCCGGCGTCTACTATGTGGACTACA GACTGGAAAGAATCAAGGAGGCCGACAAAGAGACCTACGTCGAGCAGCAC GAGGTGGCTGTGGCCAGATACTGCGACCTCCCTAGCAAACTGGGGCACAA ACTTAATTAAGGATCCTACAGCCGCG
pOUT-Fragment 2- RNA-OUT: A5	CGTACATCTAGCGGCCGCCGCAAAAAACCCCGCTTCGGCGGGGTTTTTTC GCATTAATAGATCTCGATCCCGCGAAAATAATACGACTCACTATAGGGGA ATTGTGAGCGGATAACAATTCCCCAAGCTTCCCGACATCTTGTTGTCTGA TTATTGATTTTACGCGAAACCATTTGATCACATGACAAGATGTCGATCTC GAGCGCAAAAAACCCCGCTTCGGCGGGGTTTTTTCGCACTAGTACTCACA CG
pIN Fragment 3 His-RNA-IN: S5	GGATCCGCGTAAAATCAATAAGGAGACAACAAGATGTCGGGCAGCCACCA TCATCATCACCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

Appendix E. Sequences used in construction of repressing RNA-IN/OUT plasmids.

Each of the above sequences was used in tandem with a backbone plasmid, either pBBRBB-eGFP or pBbE6a. The fragments above were used in order to construct full-length pRNA-IN or pRNA-OUT plasmids.



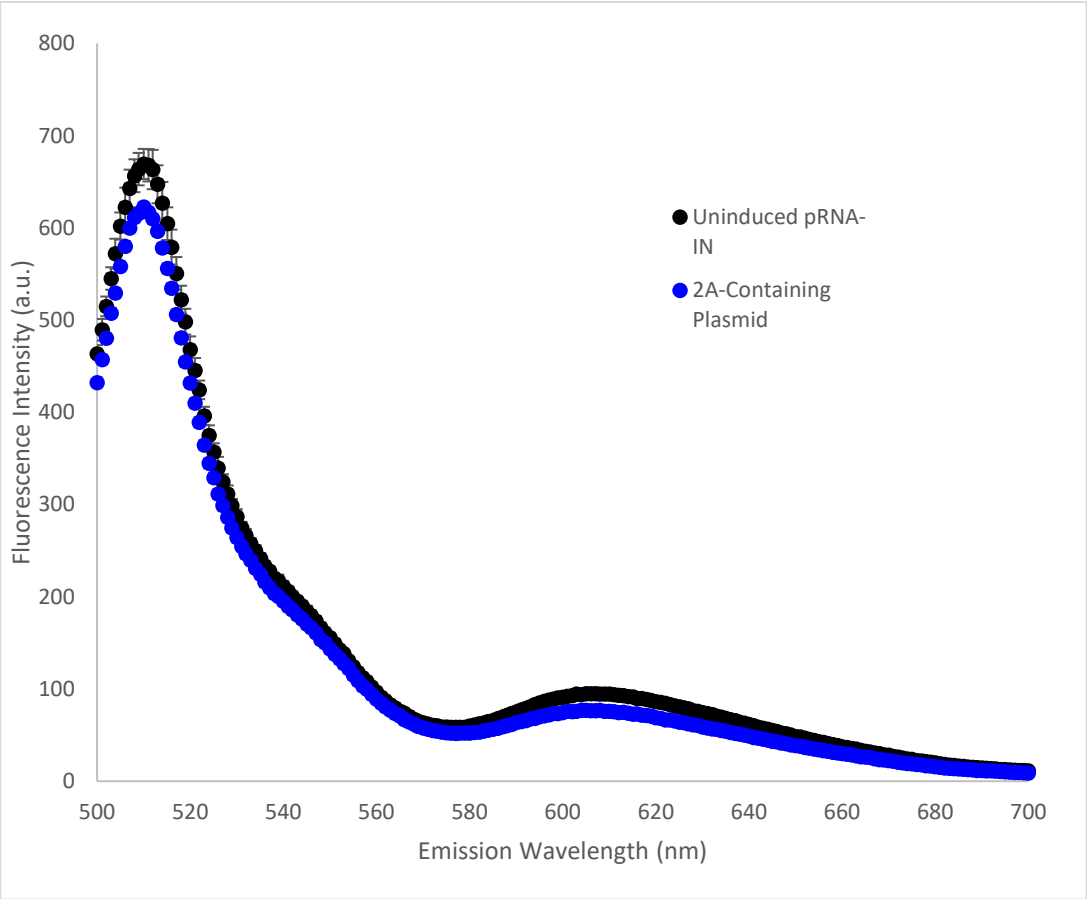
Appendix F. Average activation of translation by an activating RNA-OUT species.

Average of triplicate ensemble fluorescence emission scans upon excitation at 488nm comparing fluorescence of cells with induced activating RNA-OUT (A1) to those in which the activating RNA-OUT (A1) molecule has not been induced. Emission peaks at 510nm and 610nm are indicative of relative abundance of eGFP and mBeRFP respectively. Error bars indicate standard deviation.

Primer Name	Primer Sequence (5`- 3`)
oRH2A-seq1_1	GGATCAAAGCTTATYRGYGTGTTYAAYAARAARCTAGAYGWYGRYAAGGCGRAGTAYG
oRH2A-seq1_2	GRTTCGTYTCDACRTCRCCRCATAGCATDAGVWTRTCYTACTYCGCCTTRYCYWC
oRH2A-seq1_3	GYGGYGAYGTHGARACGAAYCCNGGNCCNATGAACCAGAGCTTTATTAGC
oRH2A-seq1_4	GACCAGCATATGTTCAATATCCGCATACAGAATATCGCTAATAAAGCTCTGGTTCATNGG
oRH2A-seq2_1	GGATCAAAGCTTRMYGTGAHDCRGDCTAYRAATTATGAYCTRCTGAAGTTGGTGCNGG
oRH2A-seq2_2	GGTCCGAGGRTTSGACTCVACGTCTCCNGCACCAACTTCAGYAG
oRH2A-seq2_3	CSAAYCCTCGGACCTATGAACCAGAGCTTTATTAGCGATATTC
oRH2A-seq2_4	GACCAGCATATGTTCAATATCCGCATACAGAATATCGCTAATAAAGCTCTGG

Appendix G. Sequences used in construction of 2A peptide library.

Each of the above sequences was used in an overlap PCR reaction to generate a full-length 2A peptide sequence and inserted into a dual fluorescent reporter plasmid. The fragments above were used to construct full-length p2A plasmids.



Appendix H. Average activity of a 2A peptide containing plasmid

Average of triplicate ensemble fluorescence emission scan upon excitation at 488nm comparing fluorescence of cells with 2A library and N-terminally degradation tagged eGFP (blue) to those with pRNA-IN as a comparison (black). Emission peaks at 510nm and 610nm are indicative of relative abundance of eGFP and mBeRFP respectively. Error bars indicate standard deviation.