

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

Scalable Genetic System Design Using Synthetic RNA Regulators

### Permalink

<https://escholarship.org/uc/item/1ts9005z>

### Author

Qi, Lei

### Publication Date

2012

Peer reviewed|Thesis/dissertation

# **Scalable Genetic System Design Using Synthetic RNA Regulators**

**By**

**Lei Qi**

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Adam P. Arkin, Chair

Professor Jennifer A. Doudna

Professor Xiaoliang Zhang

Professor John E. Dueber

Spring 2012

© 2012 Lei Qi  
All rights reserved

## Abstract

Scalable Genetic System Design Using Synthetic RNA Regulators

By

Lei Qi

Joint Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Adam Paul Arkin, Chair

Our ability to efficiently and predictably program cells is central to the fields of bioengineering and synthetic biology. Once thought to be a passive carrier of genetic information, RNA is now more appreciated as the main organizer of cellular networks. To harness the unique abilities of RNA molecules for programming cells, we show here how to rationally design novel synthetic RNA elements to recapitulate the functions of natural noncoding RNAs (ncRNAs), and how to assemble these synthetic elements into higher-order biological systems.

To create synthetic RNA elements, we start with two primary types of ncRNA-mediated natural systems. Both modulate RNA-level regulatory signals encoded in the 5' untranslated region, and are mediated by ncRNAs. In the first system the ncRNA represses transcription elongation, whereas in the second system the ncRNA inhibits translation initiation.

To create orthogonal RNA elements that work independently in the same cell, we systematically modify the RNA-RNA interaction in the natural systems. Our characterization results in families of orthogonally acting RNA elements for both transcription and translation controls. Furthermore, we develop mathematical thermodynamic models to predict new RNA elements *in silico* for translation controls. To engineer synthetic RNAs to sense and integrate cellular signals, we design allosteric RNA chimera molecules by fusing ncRNAs to RNA aptamers. We demonstrate the design principles for creating such chimeric RNA molecules that can sense proteins or small molecules and control transcription or translation. We show that the design strategy is modular, which allows us to reconfigure different ncRNA mutants and RNA aptamers to engineer orthogonal RNA chimeras that respond to different ligands and regulate different gene targets. We further show that multiple RNA chimeras allow logical integration of molecular signals in the same cell for cellular information processing.

We assemble multiple synthetic RNA elements to create basic regulatory network motifs. These include independent control, logic control, and cascading control. We characterize the performance and properties of these engineered RNA circuits such as their time response, signal sensitivity, and noises across cell populations. We further explore a strategy that can effectively convert orthogonal translational regulators into orthogonal transcriptional regulators, which can be used to perform multi-input logic computation. In an effort to engineer feedback circuits, we demonstrate the use of translational repressor and activator based on RNA-binding proteins. The designed positive or negative feedback circuits form a basis for programming complex functions.

To improve the predictability of engineered biological systems, we develop a synthetic RNA processing platform from the bacterial CRISPR genetic immune pathway. The synthetic RNA processing system can efficiently and specifically cleave desired precursor mRNAs at designed loci. Using this system, we show that transcript cleavage enables quantitative programming of gene expression by modular assembly of promoters, ribosome binding sites, *cis* regulatory elements, and riboregulators. These basic components can be grouped into multi-gene synthetic operons that behave predictably only after RNA processing. Physical separation of otherwise linked elements within biological assemblies allows design of sophisticated RNA-level regulatory systems that are not possible without it. Thus, our results exemplify a crucial design principle based on controllable RNA processing for improving the modularity and reliability of genetic systems.

To sum, our work established bacterial ncRNAs as an intriguing engineering substrate for scalable genetic circuit design and for programming cells. We provide a set of engineering principles for designing synthetic RNA elements as well as using them to sense signals and form genetic circuits. Our RNA-based engineering platform provides a versatile and powerful strategy for designing higher-order cellular information processing and computation systems, which can be readily applied to practical applications including chemical production, environment remediation, and therapeutics.

**This dissertation is dedicated to my family: my father, mother, wife, and son  
for their love, support, and patience throughout my life**

## Table of Contents

Preface .....	xii
Acknowledgements.....	xiii
Chapter 1 Introduction.....	1
1.1 Introduction to Synthetic Biology.....	1
1.2 Introduction to Prokaryotic Gene Expression.....	3
1.2.1 Transcriptional control.....	3
1.2.2 Translational control.....	4
1.2.3 mRNA degradation control.....	5
1.2.4 Relationship between promoter, RBS, and UTR.....	5
1.3 Introduction to Noncoding RNA Regulators.....	6
1.3.1 Why choose RNA? .....	6
1.3.2 RNA secondary structure prediction.....	6
1.4 Engineering Goals of Synthetic Genetic Elements.....	7
1.4.1 Independence.....	8
1.4.2 Reliability.....	8
1.4.3 Tunability.....	8
1.4.4 Orthogonality.....	9
1.4.5 Composability.....	9
1.4.6 Specialty parts and generic parts .....	9
Chapter 2 Methods, Materials, Techniques, and Data Analysis.....	11
2.1 Molecular Cloning Techniques.....	11
2.1.1 Cloning with restriction enzymes .....	11
2.1.2 BioBrick.....	11
2.1.3 PCR and inverse PCR (iPCR) .....	12
2.1.4 Circular polymerase extension cloning (CPEC) .....	12
2.1.5 Random mutagenesis.....	12
2.2 Cell Manipulations and Culturing.....	12
2.2.1 Plasmids, chemicals, and growth media. ....	13
2.2.2 Strains and transformation.....	13
2.2.3 Use of high-throughput 96-well plates.....	14
2.3 In Vivo Fluorescence Assay.....	14

2.3.1	Fluorescent proteins.....	14
2.3.2	Tecan measurement (end-point assay) .....	14
2.3.3	Time-course Tecan measurement.....	15
2.3.4	Flow cytometry.....	15
2.3.5	Microscope.....	16
2.3.6	Miller assay .....	16
2.4	RNA Extraction and Reverse Transcription.....	16
2.5	Quantitative Real-Time PCR (qRT-PCR) .....	17
2.6	Northern Blotting.....	17
2.7	In Vitro RNA Synthesis .....	18
2.8	SHAPE Experiment.....	18
2.9	Data Analysis.....	19
2.9.1	RNA secondary structure prediction.....	19
2.9.2	A linear model for protein expression.....	19
2.9.3	Calculation of protein production rates.....	20
2.9.4	Statistics.....	21
2.9.5	Flow cytometry data analysis.....	23
Chapter 3	RNA-Mediated Transcriptional Regulators.....	24
3.1	Introduction to The pT181 System.....	24
3.2	Optimization of The Natural pT181 System .....	25
3.2.1	Antisense RNA abundance is important .....	27
3.2.2	The intrinsic terminator is important for attenuation .....	27
3.2.3	The <i>repC</i> fragment is important for the regulatory function .....	28
3.2.4	Antisense RNA termination is not important.....	28
3.3	Characterization Of Nucleotide Importance In Repression.....	29
3.4	Engineering Orthogonal Transcriptional Attenuators.....	30
3.4.1	Engineering the specificity of RNA-RNA interactions.....	30
3.4.2	Similar functions of engineered orthogonal variants.....	31
3.4.3	Modularity between transcription with translation.....	32
3.4.4	kinetic properties of RNA-RNA interaction.....	33
3.4.5	Orthogonality is not preserved in translational fusions.....	34
3.5	Engineering Transcriptional Attenuator-Based Synthetic Circuits.....	35
3.5.1	Independent control motif.....	35



3.5.2 Logic control motif .....	36
3.5.3 Cascading control motif.....	37
3.5.4 Discussions of engineering the cascading control motif.....	38
3.5.5 Properties of RNA cascades.....	41
3.5.6 Tandem identical pT181 attenuators.....	42
3.5.7 Engineering NAND logics .....	45
3.5.8 Engineering decoy circuits.....	47
3.6 Reliability of The Circuits Containing Tandem Attenuators.....	48
3.7 Other Transcriptional Attenuation Systems.....	49
3.8 Strengths and Limitations of ncRNA Transcription Systems.....	50
3.8.1 Utilities of RNA-based transcription attenuators .....	50
3.8.2 Expanding families of orthogonal regulators. ....	51
3.8.3 Remaining challenges .....	52
Chapter 4 RNA-Mediated Translational Regulators.....	54
4.1 Introduction to The IS10 System.....	54
4.2 Characterization And Optimization of The Wildtype IS10 System.....	54
4.3 Engineering Orthogonality.....	57
4.3.1 Design of mutant libraries.....	57
4.3.2 Measurement and analysis of mutant libraries.....	59
4.3.3 Families of mutually orthogonal mutants.....	61
4.3.4 Develop mathematical models for RNA-RNA interactions.....	64
4.3.5 Using mathematical models to predict orthogonal pairs.....	65
4.4 Engineering IS10-Based Synthetic Circuits.....	67
4.5 Strengths and Limitations of The System.....	68
Chapter 5 Engineering Antisense RNAs to Sense Other Molecules.....	71
5.1 Introduction to RNA Aptamers.....	71
5.2 Strategy to Engineer Allostery.....	72
5.3 Engineering pT181 To Sense Theophylline.....	72
5.3.1 Pseudoknot design.....	73
5.3.2 Strand exchange design.....	75
5.3.3 Engineering mutant pT181 antisense RNA to sense.....	78
5.3.4 Engineering theophylline-sensing pT181 activators.....	79
5.4 Engineering pT181 to Sense MS2 Coat Protein.....	80

5.4.1 Engineering wildtype pT181 antisense RNA to sense.....	80
5.4.2 Engineering mutant pT181 antisense RNA to sense.....	82
5.5 Engineering pT181 to Sense Other Molecules.....	82
5.6 Engineering IS10 Antisense RNA to Sense Theophylline.....	84
5.6.1 Using the pseudoknot strategy .....	84
5.6.2 SHAPE reactions to verify allosteric switching .....	86
5.6.3 Engineering theophylline-sensing IS10 activators.....	88
5.7 Engineering Circuits Using Synthetic ncRNA Sensors.....	88
5.7.1 Sensory-level NOR gate.....	88
5.7.2 Engineering a type of signal compensator.....	89
5.8 Discussion on The Design Principles.....	90
5.9 Strengths and Limitations of The Fusion Design.....	93
Chapter 6 Other RNA-Related Genetic Regulatory Systems.....	96
6.1 Leveraging Transcription and Translation Controls.....	96
6.1.1 Leader peptide mediated system provides a conversion strategy.....	96
6.1.2 Engineering IS10-tna system to convert translation to transcription.....	98
6.1.3 Engineering mutually orthogonal transcriptional regulators .....	100
6.1.4 Engineering expandable multi-input logic gates.....	101
6.1.5 Discussion of the conversion system.....	102
6.2 Protein-RNA Translational Systems .....	102
6.2.1 Introduction to protein-RNA interaction systems.....	103
6.2.2 Orthogonality of RNA-binding proteins.....	104
6.2.3 Engineering translational feedback circuits.....	105
6.2.4 Strengths and limitations of the protein-RNA systems.....	107
Chapter 7 Engineering a Synthetic RNA Processing Platform Based on CRISPR.....	109
7.1 Predictable Engineering of Biological Systems.....	109
7.2 Standardization of Genetic Parts.....	109
7.3 Introduction to Bacterial CRISPR Systems .....	111
7.4 Engineering Predictable Translational Units.....	111
7.4.1 RNA processing reduces variability of translation.....	112
7.4.2 RNA processing reduces variability from genomic UTRs .....	114
7.4.3 Similar effects for different translational units.....	116
7.5 Engineering Predictable Promoters.....	117

7.6 Engineering Composite UTR Functions.....	120
7.6.1 Engineering tandem UTR systems.....	120
7.6.2 Characterizing noise properties of pT181 and IS10.....	121
7.7 Engineering Synthetic Operons.....	122
7.7.1 Introduction to operons.....	122
7.7.2 Rational design strategy using RNA processing.....	123
7.7.3 Using synthetic operons to measure transcription polarity.....	124
7.8 Engineering Independent Translational Controls Into Synthetic Operons.....	125
7.9 Comparing CRISPR to Other RNA Cleavage Elements.....	127
7.10 Growth Effects of Csy4 Expression.....	128
7.11 Discussion on Utilities of RNA processing.....	129
Chapter 8 Conclusions.....	130
8.1 Summary of Contributions .....	130
8.2 Future Directions.....	132
References.....	134
Appendix A Definitions of Concepts Used In The Study .....	144
Appendix B Plasmids Used In The Study .....	145
Appendix C Sequences For Important Genetic Elements .....	150

## List of Figures

Figure 1.1	The progress of DNA sequencing, DNA synthesis, and synthetic biology.....	3
Figure 1.2	Multi-scale regulation of natural gene expression.....	3
Figure 2.3.2	Comparison between different 96-well plate readers .....	15
Figure 2.9.2	A model for gene expression and calculation of protein production rates .....	20
Figure 3.1	Mechanism of the pT181 system .....	25
Figure 3.2	Optimization of the wildtype pT181 system .....	26
Figure 3.3	Characterization of antisense RNA .....	29
Figure 3.4.1	Loop and swap mutations act synergistically to reduce attenuation crosstalk.....	31
Figure 3.4.2	Induction curves using an inducible promoter for WT and MT attenuators.....	32
Figure 3.4.3	Modularity of transcriptional control with RBSs.....	33
Figure 3.4.4	Calculation of antisense/attenuator hybridization energy .....	34
Figure 3.4.5	Translational fusion show high level of crosstalk .....	35
Figure 3.5	Random transcriptional networks can be constructed from basic motifs.....	35
Figure 3.5.1	Independent regulation of two fluorescent genes in the same cell .....	36
Figure 3.5.2	Composed orthogonal attenuators exhibits NOR logical expression .....	37
Figure 3.5.3	Engineered three-level RNA-mediated transcriptional cascade .....	38
Figure 3.5.4	RNA cleavage is critical to the cascade function .....	40
Figure 3.5.5	Characterization of RNA-mediated cascade .....	42
Figure 3.5.6	Tandem identical attenuators .....	44
Figure 3.5.7	Engineering NAND-like logics using the pT181 system .....	46
Figure 3.5.8	Construction of decoy circuits .....	48
Figure 3.6	Reliability of tandem attenuators .....	49
Figure 3.7	Testing the pIP501 transcriptional attenuation system .....	50
Figure 3.8	Utilities of engineered orthogonal RNA transcriptional regulators .....	50
Figure 4.1	Wildtype IS10 translational control system .....	54
Figure 4.2	Characterization of the IS10 system .....	56
Figure 4.3.1	Calculation of free hybridization energy .....	59
Figure 4.3.2	Construction and characterization of 23 sense-antisense pairs .....	61
Figure 4.3.3	Test of five mutually orthogonal pairs .....	63
Figure 4.3.4	Results of the PLS regression model .....	65

Figure 4.3.5	The mathematical model can predict orthogonal pairs .....	67
Figure 4.4	Two-color experiments using orthogonal antisense-sense pairs .....	68
Figure 4.5	The quantitative framework of modeling .....	69
Figure 5.1	A list of molecules with known RNA aptamer sequences .....	71
Figure 5.2	The proposed riboswitch-like design for the aptamer-ncRNA fusion .....	72
Figure 5.3	Maps for the aptamer-ncRNA fusion and IS10/pT181 reporter plasmids .....	73
Figure 5.3.1	Pseudoknot design for fusing pT181 antisense with theophylline aptamer .....	75
Figure 5.3.2	Strand exchange design for fusing pT181 ncRNA with theophylline aptamer....	77
Figure 5.3.3	Orthogonal theophylline-sensing pT181 fusions using strand exchange. ....	79
Figure 5.3.4	pT181-based activator in response to theophylline .....	79
Figure 5.4.1	Designed MS2 coat protein-sensing pT181 ncRNA fusions . ....	81
Figure 5.4.2	Orthogonal MS2-sensing pT181 fusions using strand-exchange .....	82
Figure 5.5	Engineering pT181 antisense to sense other ligands .....	84
Figure 5.6.1	Designed theophylline aptamer-IS10 ncRNA fusions .....	85
Figure 5.6.2	SHAPE data .....	87
Figure 5.6.3	IS10-based activator design in response to theophylline .....	88
Figure 5.7.1	Engineering a sensory-level NOR logic circuit .....	89
Figure 5.7.2	Engineering a synthetic protein concentration compensator .....	90
Figure 5.8	Discussion on allosteric switching properties .....	92
Figure 5.9	Modularity of aptamer-ncRNAs at the molecular and network levels .....	93
Figure 6.1.1	Mechanism for the wild-type tna leader-peptide element .....	97
Figure 6.1.2	Characterization of IS10-tna fusions .....	99
Figure 6.1.3	Performance of orthogonal antisense-mediated regulators .....	100
Figure 6.1.4	Performance of NOR gates using converted orthogonal attenuators .....	102
Figure 6.2.1	RNA binding motifs for proteins.....	104
Figure 6.2.2	Orthogonality between RNA-binding proteins .....	105
Figure 6.2.3	Negative feedback and positive feedback based on RNA-binding proteins .....	107
Figure 7.3	Schematic of the CRISPR system in <i>P. aeruginosa</i> UCBPP-PA14 .....	111
Figure 7.4.1	The CRISPR RNA processing system improves the predictability of RBSs ....	113
Figure 7.4.2	RNA processing reduces variability of genomic UTRs on translation .....	115
Figure 7.4.3	RNA processing improves the predictability of translation.....	117
Figure 7.5	RNA processing improves the predictability of promoters.....	120
Figure 7.6.1	RNA processing allows design of complex UTR regulations .....	121

Figure 7.6.2	Noise properties of transcriptional control and translational control .....	122
Figure 7.7.2	RNA processing allows design of predictable synthetic operons .....	124
Figure 7.7.3	Measurement of transcriptional polarity effects in the synthetic operon .....	125
Figure 7.8	Construction of complex UTR controls in synthetic operons .....	126
Figure 7.9	Comparison of the efficacy of RNA cleavage elements .....	127
Figure 7.10	Visualization of cells (-) or (+) Csy4 expression under the microscope .....	128
Figure 8.1	Summary of genetic parts in this dissertation and their applications .....	132

## Table of Tables

Table 1.3.2	Summary of used RNA secondary structure prediction algorithms .....	7
Table 1.4	Summary of engineering goals of the synthetic genetic parts .....	7
Table 4.3.1	List of 23 pairs of mutant antisense and sense RNAs .....	59
Table 4.3.3	Mutually orthogonal families of antisense-sense pairs .....	62
Table 4.3.4	Compilation of features used in the PLSR model .....	64
Table 4.3.5	Predicted percentage repressions for 56 pairs .....	66

“If you haven’t found it yet, keep looking. Don’t settle. As with all matters of the heart, you’ll know when you find it. And, like any great relationship, it just gets better and better as the years roll on.”

- *Steve Paul Jobs*



## Preface

I came to Berkeley in 2006, majored in physics. I wasn't sure what I would like to do or what I could do, but I wanted to do something cool, something fewer people worked on, and something that would keep me dreaming. Berkeley was a great place, as there were so many great things going on. Few people would argue there how rebellion, non-canonical, ridiculous you or your work are. So it was pretty much like walking into a marvelous gigantic restaurant, figuring out what you were going to pick.

I first picked biophysics and more specifically, single molecule microcopy. I chose it because I knew I was good at physics, and I knew biology was cool. Indeed, biophysics was cool but soon I realized that it was not my type. My reason was: I wanted to try something more practical.

With this in mind, I entered an electrical engineering lab, working on fabricating nanowires. The goal was straightforward: how could you make nanoscale conductive wires that were longer and thinner than all your competitors? After three months, I decided to change again. My reason was: it was practical (relatively), but I wanted something more creative.

That's basically how I went to Adam. As that time, I had no idea about what was synthetic biology. But I had an easy calculation: engineering + physics = mechanics and electronics; engineering + chemistry = chemical engineering and synthetic polymers; engineering + biology = next big cool thing. So I asked Adam. Surprisingly, he took me on and assigned me a cool project. At the same time, Julius joined the lab and we both worked closely together. Two young men, who had never worked in the wet lab before, started figuring out every experiment details from running an electrophoresis gel to perform a PCR reaction. Of course, we were lucky that there were great experimentalists around to help us in the Arkin lab.

How Time flies. Now I am sitting here writing, summarizing what I have done in the past five years. I must admit that I am lucky to witness one of the most exciting fields, synthetic biology, to unfold and grow up. At the same time, I am glad that I could contribute to this process a little bit with my work and efforts. What I mostly enjoy about the field is the tremendous amount of creativity that is needed. Most of the time, we try to guess the design rules used by Nature in living organisms. We argue what might work and what might work better. We follow our intuitive and guess, and try to recapitulate what Nature does in lab. While I hope the work presented here will be useful to most people, I think people will soon find it as simple and basic as ABC, as the rapid progresses in synthetic biology will quickly build more sophisticated rules for more complicated biological systems based on the work, like this one, which provides a set of basic methods and toolboxes.

## Acknowledgements

I am obliged to my advisor, Adam Arkin, for the support and inspiration to finish the work. I will always admire Adam's vision, creativity, and cheerful minds towards research and science. From him I learned try to solve problems while enjoy thinking. I appreciate his open minds, especially that he was willing to take me aboard as a lab member at the first place at the time when I knew nothing about wet lab or synthetic biology.

I would like to thank my committee members: Jennifer Doudna, John Dueber, and Xiaoliang Zhang for all the great advice and help throughout the years. They offered me opportunities to tackle problems from different angles and perspectives. Working with Jennifer Doudna and Rachel Haurwitz has been one of my most enjoyable and productive collaborations, and I appreciate that they are encouraging all the time and their joyful ways of thinking. Thanks to Christopher Anderson for his enlightening and critical insights of the project. Thanks to Jan Liphardt and Patsy Babbitt who help define the project with so many excellent questions. Also, I would like to thank Steve Chu and Andrey Revyakin who led me to learn single molecule microscopy and start dreaming. I would like to thank Donald Glaser for the opportunity to stay with his family for years and for the numerous illuminating discussions on science and life.

I am grateful to Julius Lucks, who was my mentor and led me to enjoy research. Without the frequent interactions with Julius, much work couldn't have been done. I benefited a lot from working together with Vivek Mutalik, who is a very passionate and professional scientist. Many thanks to Chang Liu, whose creativity and imaginative thinking is always inspiring. I always enjoyed the interesting discussions with Xiling Shen, who is a magician knowing how to blend engineering with biology. Thanks to Denise Wang, who work with me and help with cloning and many experiments. Special thanks to Gwyneth Terry for providing tremendous support in the lab to make things possible. I would like to thank many other Arkin lab members: Weston Whitaker, Morgan Price, Adam Deutschbauer, Jeffrey Skerker, Michael Samoilov, Joao Guimaraes, Wenjun Shao, Stefano Cardinale, Guillaume Cambray, Gavin Price, David Chen, and Kelly Wetmore.

Special thanks to my parents, Guangyu Qi and Guilian Zhang for their love, support, and help ever since the day I was born. My son Calvin gives me joy everyday and makes me a proud dad. Most of all, I thank my wonderful wife Wenbo who has helped me throughout my whole PhD study at Berkeley, encouraged me, and motivated me to do better. I appreciate her efforts in raising the kid and providing me a warm and loving family so I could focus on the research many nights and weekends to finish this dissertation.

# Chapter 1 Introduction

## 1.1 Introduction to Synthetic Biology

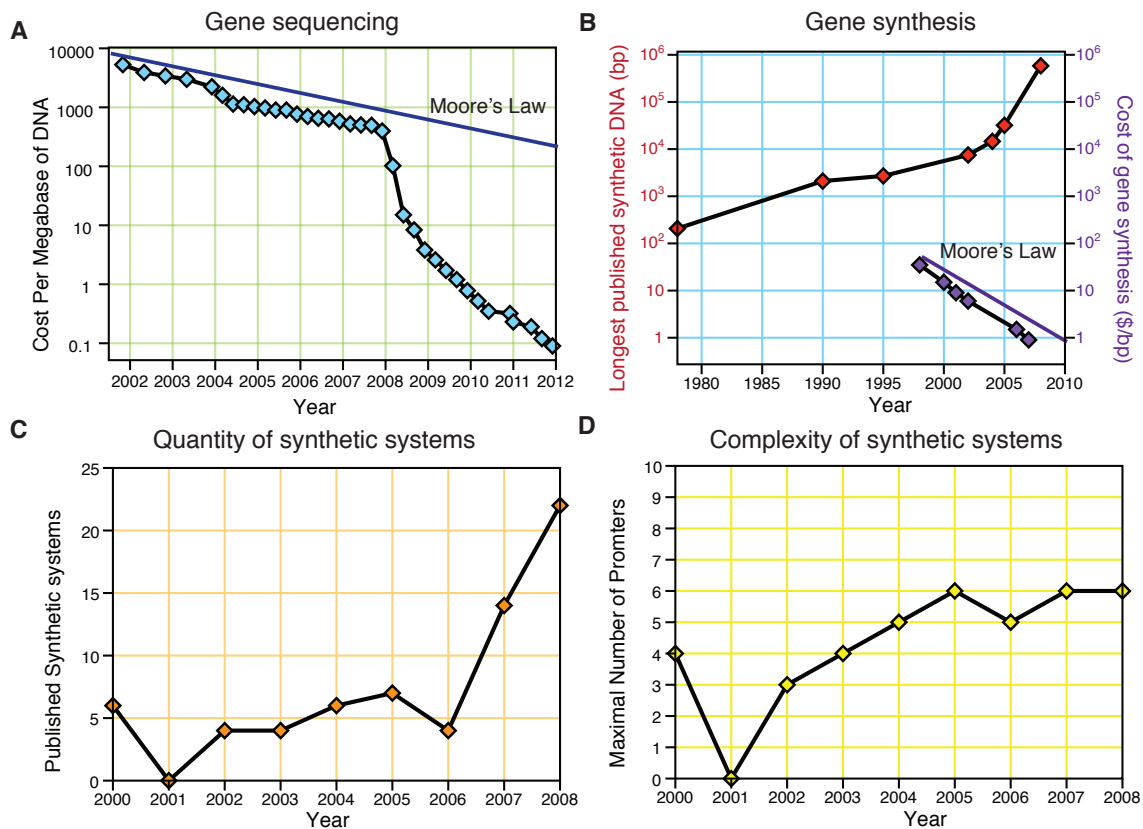
Life is arguably the most fascinating and delicate creation of Nature. While it is still challenging to define what is life in unequivocal terms that are sufficiently broad and general to encompass all life forms that we have encountered in the past and we might discover in the future, human never stop the endeavors of understanding the nature of life. One commonly held belief ever since the early nineteenth century is that biology could be explained in terms of physics and chemistry<sup>1</sup>. Since 1930s, fueled by passion to explore the microscopic world scientists made a number of critical discoveries that together set the basis of today's molecular biology. Among these, protein was found as a fundamental building block of life. Later, deoxyribonucleic acid, or DNA, was shown as the fundamental genetic material that was inherited during replication instead of protein. The most famous picture of this magic molecule came from James D. Watson and Francis Crick's 1953 paper that claimed DNA as a double helix. A few years later, another important molecule, RNA, at the time considered purely as an intermediate of the biological processes, was reported<sup>2</sup>. These discoveries together led Francis Crick in 1958 to propose the "central dogma of molecular biology", which stated that genetic information flow from DNA to RNA, and then from RNA to proteins<sup>3</sup>. A revolutionary perspective proposed by François Jacob and Jacques Monod was that life was controlled by a "genetic program" that was determined by interactions between molecules, including DNA, RNA, proteins, and numerous metabolites that constitute the cell<sup>4</sup>. Following this notion, to change the identity of a cell or its behavior or function, we need to change its genetic program. This particularly inspired the development of modern genetic engineering, which focused on the manipulation of genes, especially on the DNA level. With the advent of recombinant DNA techniques, including characterization of restriction enzymes, ligases, the invention of polymerase chain reaction, and utilization of mobile vectors such as plasmids, genetic engineering technologies and tools have revolutionized every field in the life science research, and vitalized the modern biotechnology industry.

Stepping into the twenty-first century, DNA sequencing and DNA synthesis techniques together brought another revolution to biomedical research and biotechnology. A rule of thumb in the history of computer hardware industry is Moore's Law, which states that the number of transistors (thus the computing power) that can be fabricated on an integrated circuit doubles every two years<sup>5</sup>. How about the field of DNA sequencing? Shown in **Figure 1.1A** is a plot of the cost for sequencing every megabase of DNA over time<sup>6</sup>. While before 2008, the industry more or less followed the Moore's Law, the field exhibited a super-exponential curve after that. This acceleration implies the strong supply and demand of increasing DNA sequencing power to meet the needs of applications in medicine, agriculture, and environment. Interestingly, similar trends were observed for the DNA synthesis<sup>7</sup>. Not only the cost of DNA synthesis drops exponentially, the longest length of synthetic DNA increases exponentially (**Figure 1.1B**). Taking together, our abilities to read and write genetic information into DNA and living organisms are evolving rapidly that we might be able to manipulate a living organism on the genome level technically in the near future<sup>8</sup>.

In contrast, our abilities to design meaningful biological systems with practically useful functions have been far falling behind. In one report, Weiss *et al.* showed that the number of synthetic systems designed and reported in publications merely increased moderately in a linear

manner over the past decade<sup>9</sup> (Figure 1.1C). The complexity of these designed systems, measured by the number of promoters used, plateaued since 2006 (Figure 1.1D). One reason to account for this discrepancy between synthesis and design is that we do not fully understand the design principles of complex biological systems<sup>10-12</sup>. Since the function of any designed biological system is determined by the intricate connections between components, when the number of components in the system gets large enough, their mutual interactions become overwhelming. The problem is further complicated by the cellular context dependence of the synthetic system in the cellular milieu, its stochastic behaviors, and evolution over time. One solution to the problem, proposed by synthetic biologists, is to characterize, modularize, and standardize basic genetic elements as building blocks<sup>11</sup>. These basic genetic elements, called genetic parts, should be completely predictable that can be described using mathematical models. Further, their performances should be independent of the cellular contexts or dependent but nevertheless fully understood. The long-term dream in synthetic biology is that once we have large numbers of well-defined genetic parts and design rules, we should be able to build useful cellular systems complex enough for many desired biological functions, and ultimately a living synthetic organisms. However, the available genetic parts that allow us to scale up synthetic biological systems are very limited in the present<sup>13,14</sup>. Hence, in this dissertation, I focus on the design of novel synthetic elements and the use of these elements to forward engineer higher-order systems.

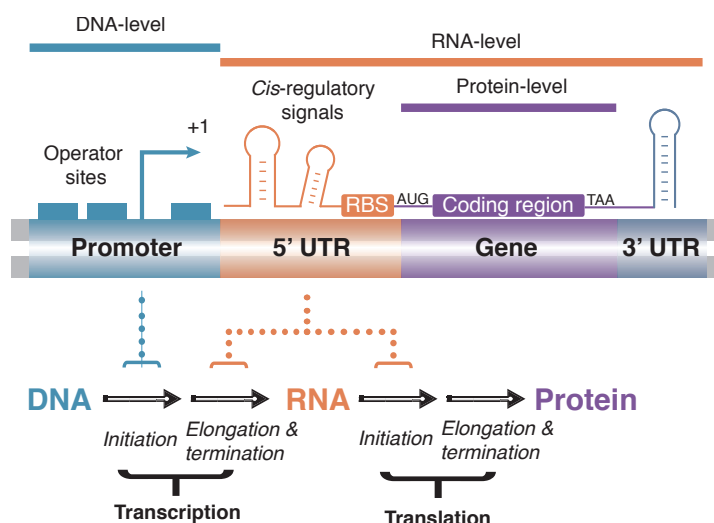
In our opinion, synthetic biology can be defined as “**an engineering science that aims to recapitulate life and higher-order biological functions from the bottom-up approach**”.



**Figure 1.1 The progress of DNA sequencing, DNA synthesis, and synthetic biology.** (A) The cost of DNA sequencing per megabases decreases faster than the Moore's Law<sup>7</sup>. (B) The cost of gene synthesis decreases following the Moore's Law, and the longest published DNA length increases exponentially<sup>7</sup>. (C) The number of reported synthetic systems increases linearly<sup>9</sup>. (D) The complexity of the synthetic system measured by the maximum number of promoters used plateaued in recent years<sup>9</sup>.

## 1.2 Introduction to Prokaryotic Gene Expression

Synthetic biology aims to create new life identities with novel functions for applications in therapeutics, cellular computation, and chemical production. Complex biological functions arise from interaction between genes in so-called genetic circuits or networks<sup>15</sup>. Thus, one major goal in synthetic biology is to rationally engineer large-scale genetic circuits that are modular, predictable, and programmable using basic genetic parts as building blocks. These genetic parts should be able to regulate different levels of gene expression, sense different types of signals, and work orthogonally without affecting each other's function in the same cell. To engineer synthetic systems, let's first look at how natural systems work (**Figure 1.2**). Since eukaryotic systems are significantly different from prokaryotic systems, here we focus on prokaryotes, especially *Escherichia coli* as our model organism. Natural bacterial gene regulation occurs on multiple levels. On the DNA level, regulators such as promoters control the transcriptional initiation. On the RNA level, regulators such as the ribosome binding sites (RBS), *cis*-regulatory elements of 5' and 3' untranslated regions (UTRs), RNA structures, internal RBS<sup>16</sup>, and codon usage<sup>17,18</sup> together control various events including transcriptional elongation, translation initiation and elongation, and mRNA degradation. On the protein level, regulators such as the kinase/phosphatase, chaperons, and proteases together control protein folding, transportation, and modification. Here we will do a brief introduction to different steps of transcription, translation, and mRNA degradation.



**Figure 1.2 Multi-scale regulation of natural gene expression.** Gene regulation occurs on levels of DNA, RNA, and proteins. Following the central dogma, promoters mainly regulate the transcription initiation, and the *cis*-regulatory signals encoded on mRNAs control transcription elongation, translation initiation, and mRNA degradation.

### 1.2.1 Transcriptional control

Promoters have been a hotspot for genetic engineering in prokaryotes. A promoter is a region of DNA that facilitates the transcription initiation of a particular gene. Two types of promoters are

generally used. Promoters can be constitutive (continuous activity) or inducible (activity is activated or repressed by other molecules). The general architecture of promoters consists of -10/-35 elements, operator sites for transcription factors binding, and a transcriptional start site (TSS). The upstream sequence (~hundreds base pairs) can also affect promoter activity<sup>19</sup>.

Promoters and their transcriptional activators or repressors have been widely used to construct synthetic circuits<sup>20-22</sup>. These promoters generally exhibit a high ON/OFF range, high cooperativity, and excellent modularity so they can be easily swapped to control different genes. Commonly used inducible promoters are engineered ones from natural promoters, including *tet* promoters<sup>23</sup> (repressed by TetR protein, and induced by anhydrotetracycline, or aTc), *lac* promoters<sup>23</sup> (repressed by LacI, and induced by Isopropyl  $\beta$ -D-1-thiogalactopyranoside, or IPTG), arabinose promoters<sup>24</sup> (activated by AraC, and induced by arabinose), and various promoters derived from quorum sensing systems<sup>25</sup> (such as the *lux* promoter, repressed by LuxR and induced by 3-oxohexanoyl-homoserine lactone (3OC6HSL); and the *las* promoter, repressed by lasR and induced by N-3-oxododecanoyl homoserine lactone (3OC12-HSL)). However, many of the promoters used in building synthetic circuits might contain poorly annotated TSS and obscure operator sites<sup>26</sup>, which append a sequence to the mRNA transcript. This might change the mRNA translation and stability unpredictably. We will discuss this in **Chapter 7**.

Transcription elongation control mainly occurs on the RNA level. Formation of RNA structure might affect the processivity of the RNA polymerase, causing it to pause or detach from the DNA template<sup>27</sup>. RNA polymerase has natural tendencies to fall off during elongation in a phenomenon called transcription polarity<sup>28</sup>. It was observed that distal genes are less transcribed compared to proximal genes to the promoter in an operon. As a consequence, bacteria tend to use smaller proteins instead of large ones.

Transcription termination in prokaryotes is usually controlled by two different mechanisms. In one mechanism, called *rho*-dependent termination, transcription is terminated by a Rho factor<sup>29</sup>. Rho factor is a protein that binds specifically to an exposed region on mRNA that is usually G/C-rich and unstructured. Bound Rho factor would catch up with RNA polymerase paused at terminator sites and unwind the DNA-RNA hybrid duplex structure. Besides the Rho factor, other protein factors such as Tau and NusA can also control termination in *E. coli*<sup>30</sup>. A second mechanism is called *rho*-independent termination, which occurs by intrinsic RNA structure formation<sup>31</sup>. In this mechanism, formation of a stable G/C-rich stem-loop structure could pause the RNA polymerase. This stem-loop structure is usually followed by a stretch of uracil (U) sequence (6~10 nt). The binding energy between adenine and uracil is weak, which destabilizes the RNA-DNA hybrid duplex and causes RNA polymerase to dissociate from the DNA template. It was reported that both the stem-loop structure and the length of uracil stretch were important for the termination efficiency<sup>32</sup>.

### ***1.2.2 Translational control***

Prokaryotic translation initiation happens on a purine-rich region on mRNA called Shine-Dalgarno (SD) sequence<sup>33</sup>. The SD sequence is important for ribosome 30S subunit binding, which is complementary to a pyrimidine rich region on the 16S rRNA component in the 30S subunit. Usually, the region flanking the SD sequence is called ribosome binding site (RBS). However, this concept is not well defined, as there is no clear definition which 5' and 3'

nucleotides should be included in the RBS. Several factors might affect the translational initiation rate. One is the RNA structure<sup>34</sup>. Bacteria utilize rich RNA structure control over RBS, which could either block or expose the RBS for different rate of translation under different conditions. The second is the distance between the SD and the start codon (AUG)<sup>35</sup>. The optimal distance is usually between 5~7nt, and too long or too short can all decrease the initiation rate.

Bacteria contain 5' and 3' untranslated regions (UTR), which are sequences flanking the protein coding region. Regulatory signals encoded in the 5' UTR usually controls transcription elongation by formation of intrinsic terminator<sup>36</sup>, translation initiation by controlling the SD sequence<sup>37</sup>, and mRNA degradation<sup>38</sup>. These 5' UTRs are hot regulatory spots for different protein factors, noncoding RNAs, and metabolites. For example, RNA-binding proteins can directly bind to 5' UTR and repress or activate translation<sup>39-41</sup>. Noncoding RNAs can interact with 5' UTR sequences usually via direct base pairing to affect a variety of transcript properties. In the past decade, riboswitch<sup>42</sup>, which is a type of *cis*-acting RNA element, can directly bind to small molecules and control gene expression. Furthermore, both natural and engineered ribozymes which are self-cleavage elements acting on mRNAs have been used to control gene expression<sup>43,44</sup>. To sum, it is convincing that besides the transcriptional regulatory network, prokaryotes also possess a sophisticated and delicate translational regulatory network that is interwoven by mRNAs, noncoding RNAs, RNA-binding proteins, and small molecules.

### ***1.2.3 mRNA degradation control***

Compared to transcription and translation, the process of mRNA decay is only poorly understood<sup>45</sup>. Two primary types of ribonucleases are responsible for mRNA degradation. One is called exoribonuclease, which can cleave mRNA from either 5' or 3' ends. The other is called endoribonuclease, which recognize internal mRNA sequence or structure and chew the mRNA from inside. Among all, RNase III and RNase P are probably the most studied endoribonucleases<sup>46,47</sup>. RNase III recognizes a specific structure with some nucleotide composition requirements<sup>48</sup>. RNase P is a ribonucleoprotein enzyme that is essential for the maturation of the 5' end of tRNAs<sup>49</sup>. However, much is still unknown about the detailed mechanisms about these ribonucleases. The mRNA degradation is also affected by RNA structures. In one study, Keasling *et al.* engineered an mRNA stabilization tag derived from the *ompA* 5' UTR. Attachment of the tag to transcripts significantly improved the transcript stability<sup>50,51</sup>. Compared to prokaryotes, transcript stability engineering is more popular in eukaryotes. Decay of eukaryotic mRNAs can be triggered by the removal of 3' poly-A sequence. As a result, ribozymes can flexibly control cleavage of 3' poly-A and gene expression<sup>52</sup>.

### ***1.2.4 Relationship between promoter, RBS, and UTR***

We should note here that regulations of transcription, translation, and mRNA degradation are not isolated<sup>53,54</sup>. One element might participate in multiple processes, and each process could be determined by interaction between different elements. For example, although promoters mainly determine transcription initiation, they might append extra sequences to the 5' end of transcripts, which could affect translation and mRNA decay. The case is further complicated by possible interactions between DNA, RNA, and proteins with the *trans*-acting host factors that are uncharacterized or poorly understood. As a consequence, many genetic systems require

extensive debugging and fine-tuning for optimal performances to account for these unpredictable coupling between genetic elements.

### **1.3 Introduction to Noncoding RNA Regulators**

In our study, we are particularly interested in using RNA regulators to program gene expression. RNA molecules are now recognized as the main controller and defender of genomes in all organisms. Several reasons promoted us to study the use of noncoding RNAs (ncRNA) as engineering substrates.

#### ***1.3.1 Why choose RNA?***

First, ncRNAs are ubiquitous regulators in all organisms ranging from bacteria to yeast to human<sup>37,55,56</sup>. In bacteria, there are numerous characterized ncRNAs that control a variety of important functions. Noncoding RNAs in bacteria account for about 3% of the genomic genes<sup>37</sup>. In eukaryotes, with the discovery of microRNAs, piRNAs, and long noncoding RNAs, it is appreciated that there are more noncoding genes than coding ones<sup>57,58</sup>. Comparing yeast to human, while the number of coding genes remained relative similarly (~20,000), the noncoding genes increased significantly, suggesting their roles in increasing the complexity of genetic networks and cellular functions. Second, many ncRNAs interact with mRNA targets simply by base pairing<sup>59</sup>. This simple rule makes these ncRNA highly amenable to rational design and engineering. Third, RNAs could potentially bind to a variety of different proteins and small molecules<sup>60</sup>. In this sense, they are versatile molecules, and can be potential modified for new functions. Fourth, RNA molecules possess unique and interesting physical and chemical properties. For example, their production, degradation, and folding are all faster than proteins. Thus, they can be very dynamic controllers for fast-responsive programs<sup>61</sup>. Finally, there are cutting-edge techniques to couple their sequence to structures and functions in the cellular milieu. For example, through Systematic Evolution of Ligands by Exponential Enrichment (SELEX) method<sup>62</sup>, we can relatively efficiently discover new RNA sequences that bind to desired molecules. To resolve RNA structures, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) can be used<sup>63</sup>. This technique has recently been developed to interface with the high-throughput deep sequencing technologies<sup>64</sup>, making it possible to resolve thousands of RNA structure in parallel in a short amount of time. To sum, we regard RNA molecules as useful engineering substrates, which can be more easily designed, evolved, and assembled as generic controllers to build gene connections in complex programs.

#### ***1.3.2 RNA secondary structure prediction***

One great advantage at the current stage for RNA design is the availability of numerous RNA prediction algorithms. A list of common used RNA prediction programs have been listed in **Table 1.3.2**. Many of the algorithms can decently predict RNA secondary structure at the equilibrium state<sup>65-69</sup>. Although different algorithms differ in the modeling details and sometimes just parameter values, most of them are based on minimal free energy constraint (MFE). At equilibrium states, thermodynamic folding of RNAs mainly determined by base-pairing rules is calculated of possible conformational space. The sub-optimal structures are adjusted by the presence of bulges, internal loops, non-canonical base pairing (such as G-U wobble), and loop length penalties. Alternative RNA secondary structure prediction programs are based on



comparative sequence analysis<sup>70</sup>. New RNA molecules are aligned to known RNA molecules whose structures have been resolved. With the advent of high-throughput RNA structure resolving techniques and availability of more RNA tertiary structures by crystallography and nuclear magnetic resonance microscopy, prediction of RNA structure will become easier, faster, and more accurate in the future.

Algorithm	Description	Kinetic?	Pseudoknot?	Notes
mFold	RNA structure prediction based on MFE	No	No	Mostly used in the current study
UNAFold	A software package for mFold, and could predict RNA-RNA hybridization	No	No	Can run on local machines based on Python scripts
RNAFold	Includes an implementation of the partition function for computing basepair probabilities and circular RNA folding.	No	No	Contain a downloadable package
NUPACK	Computes the non-pseudoknot partition function of interacting strands in diverse solution conditions.	No	No	Can run on local machines based on Python scripts
KineFold	Computes the kinetics of RNA sequences including pseudoknots using a partition function.	Yes	Yes	Calculates a series of folding during a kinetic process
RNAstructure	Predict structure based on constrained using experimental data, including SHAPE, enzymatic cleavage, and chemical modification accessibility.	No	No	Used in SHAPE data analysis and sequence alignment

**Table 1.3.2 Summary of used RNA secondary structure prediction algorithms.**

#### 1.4: Engineering Goals of Synthetic Genetic Elements

Lies at the heart of synthetic biology is the abilities to predictably design biological systems. This requires us to understand the mechanisms of genetic parts and their interactions precisely, foresee the cellular context influence on their performances, and achieve desired functions through effective wiring between genes. Here we define five basic properties of genetic parts that are needed to achieve this goal<sup>14</sup>: independence, reliability, tunability, orthogonality, and composability (**Table 1.4**). We expect parts with all these properties should lead to design of large-scale synthetic biological systems (scalability).

Part properties	Description	Examples?
Independence	A part does not interfere with the host genes and machineries and vice versa	tet promoters repressors in <i>E. coli</i>
Reliability	A part functions as intended despite variations in the contexts and environments	feedback circuits decreases intrinsic noise
Tunability	The performance of a part can be adjusted	RBS engineering
Orthogonality	Multiple parts derived from the parent part do not interfere with each other while exhibiting similar performances	Orthogonal Ribosome-RBS pairs
Composability	Multiple parts can be assemble as a single unit for a composite function	Chimeric proteins
Predictability	The characteristic performance of a part can be foreseen based on its mechanism or characterization	Characterized <i>lac</i> promoter transfer curve
Scalability	Multiple parts can be exploited to create distinct biological systems, whose scale and complexity can be extended	Multiple transcriptional repressor-promoter pairs

**Table 1.4 Summary of engineering goals of the synthetic genetic parts.**

### **1.4.1 Independence**

Independent parts do not interfere with their host circuitry and vice versa. In an early example involving complex function, the multi-gene nitrogen fixation system from *Klebsiella pneumoniae* was shown to function when transplanted into *E. coli*, although with diminished function<sup>71</sup>. That is, the transformed *E. coli* was able to fix nitrogen, implying that the system can operate with some degree of independence from the native context in which it was evolved. Independent parts do not interfere with each other, which are exemplified by transcriptional repressors that only affect cognate promoters and not other promoters<sup>72</sup>. However, part independence is far from guaranteed. For example, plasmids with different origins of replication might interfere with each other during segregation and partition, which often leads to unstable and unpredictable copy numbers in cells<sup>73</sup>.

### **1.4.2 Reliability**

A reliable biological part functions as intended, and exhibits robustness under noises in the cellular network and the cellular environment. Because of the fundamentally discrete and stochastic nature of chemistry, there is intrinsic noise in the dynamics of biochemical networks. As a possible route to engineering reliability, it has been shown that modifying molecular features or implementing feedback mechanisms into genetic circuits can alter the noise profiles of biological systems<sup>74,75</sup>. This sort of intrinsic noise, however, does not always lead to unreliable function but can actually be a source of reliability<sup>76</sup>. Intrinsic noise in physiological function can be leveraged to bank against uncertainty in the environment; for example, noisy promoters connected to an antibiotic resistance gene were shown to confer an advantage over more stable promoters for cells exposed to acute bursts of the antibiotic<sup>77</sup>. The function of a synthetic part can also be affected by cell-to-cell variation in key cellular resources required for transcription, translation, and replication that, in turn, can be affected by changes in the cellular environment. Addition of extra circuitry to a cell might place extra burdens on the cell<sup>78</sup>. Use of selective markers can maintain a burdensome part in the right environment. In the absence of such markers, mutants of the engineered cell that inactivate or remove the part might outgrow the original. For instance, You *et al.* discovered that cells started to escape their population-controlled cell death circuit three to six days after introducing the circuit into the cells<sup>22</sup>, and Canton *et al.* found that the function of a genetic controller decayed after ~50 generations of being present inside the cells<sup>79</sup>. While these examples may represent an actual resource load on the cell, other examples include exogenous metabolic pathways that produce intermediates toxic to the cell<sup>80</sup>.

### **1.4.3 Tunability**

Tunability means the ability to make controlled adjustment to a part's function. For simple parts such as RBSs, this might mean varying its sequence to change both the RNA structure around the translation initiation site and its interaction with the ribosome. Similar strategies have been widely used to tune gene expression over a wide range<sup>81</sup>. In addition to tuning translation, mRNA degradation can also be tuned by modifying the exoribonuclease access sites<sup>51</sup>. Tuning a part function often alters the entire function of the circuit. Gardner and Collins experimentally demonstrated that tuning RBSs in certain versions of a genetic toggle switch can affect whether the switch displays graded or bistable behavior<sup>20</sup>. Dueber *et al.* exploited the affinities of protein-

binding domains to different peptide targets to construct proteins that function conditionally in the presence or absence of multiple environmental inputs<sup>82</sup>. Voigt *et al.* describe a theoretical ‘evolvable’ circuit motif in which it is possible to tune the promoter to switch the behavior of the circuit from a graded switch, to a bistable switch, to an oscillator, to a pulse generator<sup>83</sup>.

#### ***1.4.4 Orthogonality***

Orthogonal parts families are derived from parent parts that can be tuned to the point that they do not crosstalk with each other, while maintaining the same basic conceptual function. Examples include in one report, Reina *et al.* tuned three different protein-binding domains to bind to new targets two orders of magnitude higher than to their cognate peptides<sup>84</sup>. Recently, there have been multiple examples illustrating the design of RNA molecules for creating orthogonal parts families. By mutating the 16S ribosomal RNA, Rackham and Chin created mutant ribosomes that each acted on a specific target RBS, independently of each other and the natural host ribosomes<sup>85</sup>. The RNA complementarity rules have been used to design orthogonal translational systems that can only be recognized and controlled by specific RNA molecules<sup>86,87</sup>. Rinaudo *et al.* exploited to design RNAi logics that relied on orthogonally acting siRNA molecules that specifically target cognate RNAi sites on mRNA transcripts for Boolean operations<sup>88</sup>.

#### ***1.4.5 Composability***

Composability is the property that allows multiple parts to be combined to perform a predictable function. Parts that are physically composable can be placed on the same molecule, DNA, mRNA, or proteins, and display a composite predictive behavior. Promoters, RBSs, and coding sequences of proteins are one of the most common examples of physically composable parts that can be placed in series on a DNA molecule to control expression of the gene. Physically composable parts can also be combined to form a composite part with chimeric function. Win and Smolke composed self-cleaving RNA ribozymes with small-molecule sensing RNA aptamers<sup>52</sup>. When the aptamer domain is placed in the middle of the ribozyme sequence, changes in conformation of the aptamer domain upon binding a specific small molecule, either allow, or prohibit ribozyme cleavage. When this composite part is composed downstream of a gene, cleavage by the ribozyme results in transcript degradation for gene silencing. When two parts are not on the same molecule they can still be functionally composed. For example, a transcriptional repressor encoded on one DNA molecule can control a promoter on another DNA molecule to repress its activity<sup>21</sup>. This requires matching part–part interactions, a currently heterogeneous process requiring a great deal of tuning.

#### ***1.4.6 Specialty parts and generic parts***

In most engineering disciplines, there are specialty parts and there are generic parts. The specialty parts carry out application-specific function and are interconnected by generic parts – key parts that are used in nearly every design. There are many specialty parts in biology, for example, enzymes and molecular machines such as photosynthesis, motility, protein secretion, and nitrogen fixation, evolved over billions of years. For now, we should essentially use these ‘as-is’, using our understanding of how to tune these parts to match particular biological circuit designs<sup>89</sup>. In contrast, for the generic parts, we need an engineering science that can provide predictable and scalable design. The confluence of parts that are all independent, reliable,

tunable, orthogonal, and composable, as outlined above, leads to scalable families of parts that can be readily combined together to form predictable and possibly complex new functions in cells. A central challenge of synthetic biology is how to smartly choose parts to build desired biological function. An initial effort ought to be the generation of scalable parts families that control each stage of gene regulation such as transcription, translation, and protein interaction. These processes are central to nearly every application and generally provide the logic by which the application's key activities are deployed. There is evidence that early success will come from nucleic-acid-based gene expression regulators where Watson-Crick base pairing rules are a good starting point for design<sup>81</sup>. With sufficiently deep parts families covering a broad, but carefully selected array of function, we should have the tools to finally enable a predictable biological circuit design cycle, thereby dramatically increasing the efficiency, safety and complexity of genetic engineering. Eventually, understanding these basic concepts will help us transit from the engineering of small biological circuits and pathways, to genome scale designs that operate beyond the bioreactor across the population, and ultimately ecological levels, all the while enabling a deeper identification and understanding of the design principles of biology.

## Chapter 2 Methods, Materials, Techniques, and Data Analysis

### 2.1 Molecular Cloning Techniques

#### 2.1.1 Cloning with restriction enzymes

Restriction enzymes are widely used for routine genetic cloning<sup>90</sup>. Two different restriction enzymes are usually used to digest both the insert and backbone DNA fragments, which are purified, ligated, and transformed into *E. coli* cells for selection of the correct clones. For this purpose, efficient restriction enzymes that generate sticky ends (as compared to blunt ends) are preferred, because their ligation efficiency and specificity are usually higher. Most commonly used restriction enzymes in our study include AatII, Acc65I, AlwNI, AvrII, BamHI, BglI, BglII, DpnI, EcoRI, EcoRV, HindIII, NdeI, NheI, NotI, PstI, SpeI, SphI, XbaI, and XhoI. DpnI only digests DNA sequences that are methylated. This is useful for specifically cleave organism-derived DNAs as compared to PCR-generated DNAs. An inherent requirement for restriction enzyme-based cloning is that the restriction sites should be unique. Otherwise, unwanted digestion events will occur. However, this is often not true for large DNA fragments. To eliminate duplicated digestion sites, site-directed mutagenesis PCR is first performed. The modified nucleotide should be a silent mutation that doesn't change the amino acid composition of the encoded protein. With the cheaper and faster DNA synthesis service, multiple duplicated restriction sites can be eliminated at once by *de novo* gene synthesis. To perform double digestion, a common digestion buffer is needed. A longer digestion time is needed for optimal results, as double digestion usually exhibits lower efficiencies compared to single-enzyme digestion. The information for double digestion can be found at <http://www.neb.com/nebecomm/DoubleDigestCalculator.asp#.TzhXP1F0sUw>. A standard protocol is in 20  $\mu$ L total reaction volume, add 10  $\mu$ L of DNA, 2  $\mu$ L of 10X digestion buffer, optional 2  $\mu$ L 10X BSA, and 0.5  $\mu$ L of each restriction enzyme, and H<sub>2</sub>O. Then incubate the reaction at 37 °C for 2~4 h.

#### 2.1.2 BioBrick

BioBrick is a common method used in synthetic biology<sup>91</sup>. It uses two or more compatible enzymes. There are two main sets of compatible enzymes. One set is BglII and BamHI (usually called Berkeley format), and the other is SpeI, NheI, and XbaI (usually called MIT format). Ligation between the DNA cleavage sites that are generated by these enzymes results in a “scar”, which can no longer be further digested. Thus, a scar is an end-point for the cloning. Using the BioBrick method, serial cloning steps are performed using only two or three enzymes again and again, eliminating the need to find unique digestion sites. In terms of labour and efforts, BioBrick is a slow cloning method. Furthermore, it forms an extra scar site between cloned genes. Nevertheless, BioBrick is very useful in cloning of tandem repeats. In our work, we frequently need to assemble two or more near-identical DNA fragments together. This is impossible for PCR based methods. If we use normal restriction digestions, we would need multiple unique restriction sites, which is often hard for large DNAs. The most effective way to clone tandem identical DNAs together is BioBrick. To do this, the insert is usually digested with X and BamHI (X is an arbitrary unique enzyme), and the backbone is digested with X and BglII. Ligation between these two fragments forms a new BamHI site contained in the insert, and a new BglII site contained in the backbone. This allows the same procedure to be repeated again.

### **2.1.3 PCR and inverse PCR (iPCR)**

Polymerase Chain Reaction (PCR) is undoubtedly the foundation of modern genetic cloning. With the advent of cheaper oligonucleotide synthesis, PCR is becoming one of the most important techniques for almost every bioengineering applications. In our work, PCR-directed cloning is widely used to introduce mutations, small insertion fragments, and deletions. When two primers are used in opposite directions along the DNA template, it is called inverse PCR (iPCR)<sup>92</sup>. One drawback is that iPCR might introduce point mutations, which are hard to detect without fully sequencing of the whole product. However, the engineered Phusion DNA polymerase possesses extremely high fidelity<sup>93</sup> (single base substitution error rate is 1 out of 450,000 nucleotide polymerized<sup>94</sup>), eliminating errors for normal sized products. To perform iPCR, the oligonucleotide primers first need to be phosphorylated. This is done by adding 0.5  $\mu\text{L}$  100mM primer into a 10  $\mu\text{L}$  reaction containing 1  $\mu\text{L}$  T4 PNK buffer, 0.5 T4 PNK and  $\text{H}_2\text{O}$ , and incubate at 37 °C for 1h. The phosphorylated primers are then directly used for subsequent PCR reaction following standard PCR protocol<sup>92</sup>. Detailed protocol can refer to <http://www.neb.com/nebecomm/products/productF-541.asp>.

### **2.1.4 Circular polymerase extension cloning (CPEC)**

CPEC, or circular polymerase extension cloning<sup>95</sup>, is analogous to other DNA assembling techniques such as SLIC<sup>96</sup> and Gibson assembly method<sup>97</sup>. Like SLIC and Gibson assembly, CPEC is standardized, scar-less, and largely sequence-independent. Since there is no exonuclease chew-back in CPEC, small DNA fragments can be assembled directly. Furthermore, there is no dNTP addition step (unlike SLIC). Only a single enzyme (i.e. DNA polymerase) is required (unlike Gibson). Since the CPEC assembly reaction occurs at higher temperatures than the SLIC and Gibson methods, stable secondary structures at the ends of assembly pieces are relatively less of a concern. The main disadvantage is that it is more likely to result in polymerase-derived mutations than SLIC or Gibson, and mis-priming events are now possible anywhere along the sequences of the fragments to be assembled. CPEC generally consists of the following steps<sup>95</sup>. (1) PCR amplification of multiple inserts and the backbone vector using designed primers. (2) Add 1  $\mu\text{L}$  DpnI and 5  $\mu\text{L}$  Buffer 4 (NEB) directly into the PCR product without purification, and incubate at 37 °C for 2 h. (3) PCR purification of all inserts and the backbone digestion solutions. (4) Add 100 ng of the backbone and equimolar amounts of the other assembly pieces to a 25  $\mu\text{L}$  total volume assembly reaction mixture with 5  $\mu\text{L}$  5X HF Phusion Reaction Buffer, 1  $\mu\text{L}$  dNTPs (10 mM), 0.75  $\mu\text{L}$  DMSO, 0.5  $\mu\text{L}$  2U/ $\mu\text{L}$  Phusion Polymerase, and  $\text{H}_2\text{O}$ . Perform the PCR reaction in a thermocycler as follows: 3 min, 98 °C; 15 cycles of 30 s, 98 °C, 30 s, 55 °C, X s, 72 °C, X = length\* (kb) x 15; 10 min, 72 °C. (5) After the reaction is done, directly transform 5  $\mu\text{L}$  of the reaction into competent *E. coli*.

### **2.1.5 Random mutagenesis**

Random mutagenesis is useful in optimizing gene expression by randomly sampling the mutagenesis space. We can apply the method to promoters (-35 and/or -10 box) or RBS. Two phosphorylated oligonucleotide primers containing the desired mutations are used for iPCR on the template DNA. The product is self-ligated and screened using a 96-well plate reader.

## **2.2 Cell Manipulations and Culturing**

### 2.2.1 Plasmids, chemicals, and growth media.

Three plasmids are used for all cloning and transformation in our study. The first one is a high copy (50~100 copies per cell) plasmid (named ColE1 plasmid) with a ColE1 replication origin and an ampicillin resistance marker (*bla*)<sup>23</sup>. The second one is a medium copy (20~50 copies per cell) plasmid (named p15A plasmid) with a p15A replication origin and a chloramphenicol resistance marker (*cat*). The third one is a low copy (2~5 copies per cell) plasmid (name pSC101 plasmid) with a pSC101 replication origin and a chloramphenicol resistance marker (*cat*).

The antibiotics used are 100  $\mu\text{g ml}^{-1}$  carbenicillin (more reliable than ampicillin, Fisher, BP2648) and 34  $\mu\text{g ml}^{-1}$  chloramphenicol (Acros, AC22792). The inducers are 500  $\mu\text{M}$  Isopropyl-b-D-thiogalactopyranoside (IPTG, Fisher, BP1755) and 2  $\mu\text{M}$  anhydrotetracycline (aTc, Fluka, 37919) for maximal induction. Theophylline (T1633) and caffeine (C0750) are all purchase from Sigma-Aldrich.

LB Broth is prepared by mixing 10g Bacto tryptone (Difco), 5g yeast extract LD (BD), and 10g NaCl (Sigma) in 1000mL H<sub>2</sub>O. LB agar is prepared by adding 15g Bacto agar (BD) into the above LB Broth. M9 minimal media is prepared by mixing 11.28 g M9 minimal salts (Sigma Aldrich, M6030), 0.0001% thiamine (Supelco), 2 mM MgSO<sub>4</sub> (Sigma), 100  $\mu\text{M}$  CaCl<sub>2</sub> (Sigma), 20 mM glucose (Amresco, 0188) or 0.4 % glycerol (Fisher, BP2291). We also supplemented 0.02% casamino acids in the M9 minimal media, because Tg1 cells required proline to grow. EZ Rich Defined Media (Teknova, M2105) is prepared by mixing 100ml 10X MOPS (Teknova, M2101), 10ml 0.132M K<sub>2</sub>HPO<sub>4</sub> (Teknova, M2102), 100ml 10X ACGU (Teknova, M2103), 200ml 5X Supplement EZ (Teknova, M2104), 10ml 20% glucose (Teknova, G0520). All chemicals are sterilized either by autoclave or by using a filter system (Corning).

### 2.2.2 Strains and transformation

The *Escherichia coli* strain Tg1 (Zymo Research, T3017) with genotype F'[traD36 lacI<sup>q</sup>  $\Delta$ (lacZ) M15 proA<sup>+</sup>B<sup>+</sup>] supE  $\Delta$ (hsdM-mcrB)5 (rk<sup>-</sup> mk<sup>-</sup> McrB<sup>-</sup>) thi-1  $\Delta$ (lac-proAB) is used for most cloning tasks. This strain grows fast, and displays very high transformation efficiency. The *E. coli* strain Top10 (Invitrogen, C4040) with genotype F<sup>-</sup> mcrA  $\Delta$ (mrr-hsdRMS-mcrBC)  $\phi$ 80lacZ $\Delta$ M15  $\Delta$ lacX74 nupG recA1 araD139  $\Delta$ (ara-leu)7697 galE15 galK16 rpsL(Str<sup>R</sup>) endA1  $\lambda$  is used for most *in vivo* fluorescence assay tasks. For cloning of tandem repeats, Top10 cells are used instead of Tg1, because Top10 has a lower recombination rate.

For transformation of chemically competent *E. coli* cells, we followed the standard the protocol. In brief, 10  $\mu\text{l}$  cloning products or 1  $\mu\text{l}$  each plasmid are added to 10~50  $\mu\text{l}$  *E. coli* cells. The cells rest on ice for 5 min, followed by heat shock at 42 °C, 90s (30 s for Top10 cells). Then 30~150  $\mu\text{l}$  2YT or SOC media are added. The cells are recovered at 37 °C, 1000 r.p.m., for 1h, and plated on LB agar plates containing 100  $\mu\text{g ml}^{-1}$  carbenicillin and 34  $\mu\text{g ml}^{-1}$  chloramphenicol, and incubated overnight at 37 °C. Single colonies are either directly picked for DNA extraction, or for experiments. The experiment transformations are usually stored as glycerol stocks (50  $\mu\text{L}$  culture, 50  $\mu\text{l}$  50% glycerol) in 96-well plates for restreaking and experiment.

Electroporation usually exhibits higher transformation efficiency than chemical

transformations, but can be only performed at a lower-throughput manner. This is done by adding 1~2  $\mu\text{L}$  of cloning products or plasmids to 50  $\mu\text{L}$  electro-competent cells in chilled electroporation cuvettes, followed by using an electroporator (Harvard Apparatus, ECM 630) with parameters setup (1600~1800 V, 200 Ohm, 25  $\mu\text{F}$ ) optimized for *E. coli*. The SOC medium with volume of 250  $\mu\text{L}$  is added, and the cells are recovered at 37 °C, 1000 r.p.m. for 1h. Cells are then plated on LB agar plates with selectable antibiotics.

### ***2.2.3 Use of high-throughput 96-well plates***

We observed no difference of cell growths using 25 ml glass tubes (filled with 3~5 ml cell culture at 37 °C, 200 r.p.m.) or 2 ml 96-well plates (Costar 3960, filled with 200~350  $\mu\text{L}$  cell culture at 37 °C, 1000 r.p.m.). Thus, 96-well plates are useful to scale up experiments while not affecting cell growth.

## **2.3 In Vivo Fluorescence Assay**

### ***2.3.1 Fluorescent proteins***

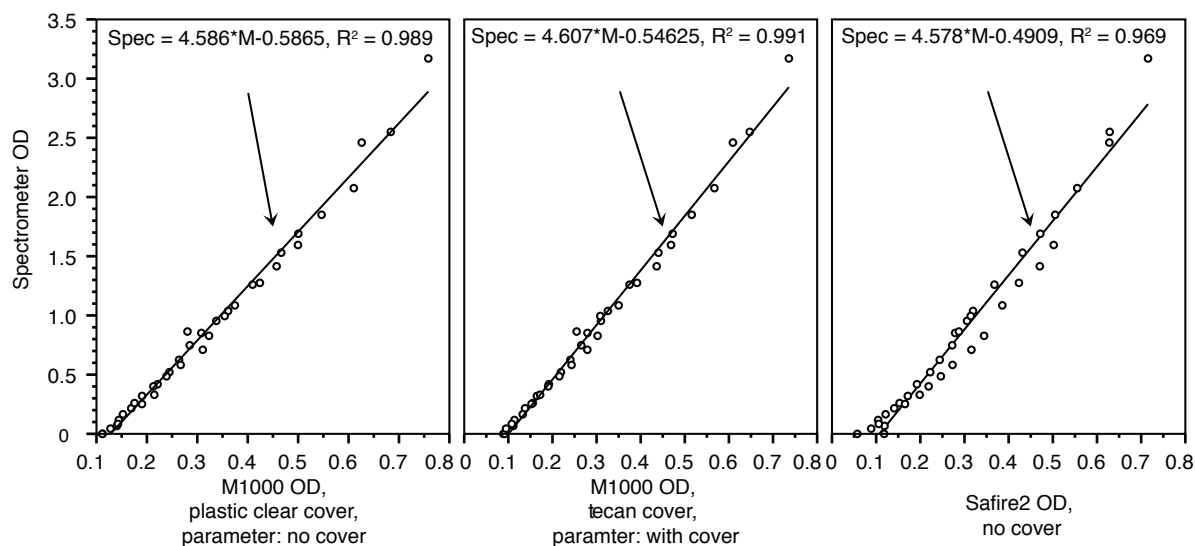
In our experiment, we used superfolder GFP (sfGFP)<sup>98</sup> and mRFP<sup>99</sup>. Both are monomers and fluoresce brightly. Both proteins are tolerant to fusions at their N-termini. The MS2 and PP7 coat proteins and Com phage protein have been fused to the N-terminus of either sfGFP and mRFP, and work well. Overexpression of mRFP might decrease the growth rate of *E. coli* cells.

### ***2.3.2 Tecan measurement (end-point assay)***

Colonies are picked into 300  $\mu\text{L}$  of LB containing 100  $\mu\text{g ml}^{-1}$  carbenicillin and 34  $\mu\text{g ml}^{-1}$  chloramphenicol in a 2 mL 96-well plate (Costar 3960). After growing 16 h at 37 °C and 1000 r.p.m. to saturation in a Labnet Vortemp 56 bench top shaker or Multitron II Incubator Shaker (ATR), 3  $\mu\text{L}$  culture is added to 297  $\mu\text{L}$  (1:100 dilution) of supplemented M9 Minimal Media or EZ Rich Define Media containing the same concentrations of antibiotics in a 2mL 96-well block. This is allowed to grow at 37 °C at 1000 r.p.m. until cells reach mid-log phase (usually between 3 h to 12 h depending on the metabolic burden and toxicity of the genes expressed). In the case of induction experiment, to achieve an even higher dilution rate, this plate is allowed to grow until saturation in about 16 h again. Another 1:100 dilution is performed, and cells grow to mid-log phase for fluorescence assay.

Then 50  $\mu\text{L}$  of cell culture is mixed with 100  $\mu\text{L}$  PBS (pH 7.4, Gibco, 10010-023) in 96-well Tecan assay plates (Costar 3603), and fluorescence (excitation at 485 nm, emission at 510 nm for sfGFP; excitation at 587 nm, emission at 610 nm for mRFP<sup>100</sup>) and ODs (optical densities, measured at 600 nm, were similar among experiments and fell in the linear range of our instruments (**Figure 2.3.2**) were then measured using a fluorescence plate reader (Tecan Safire2). The ratio of fluorescence to optical density (RFU/OD) was calculated and the background RFU/OD corresponding to the cells without fluorescent proteins was subtracted.





**Figure 2.3.2 Comparison between different 96-well plate readers.** Three machines are compared: spectrometer, Tecan (M1000), and Tecan (Safire2). The regression between OD<sub>600</sub> correlates with each other, confirming the linear relationship of the ODs for our experiments.

### 2.3.3 Time-course Tecan measurement

Overnight cell culture of 1  $\mu\text{L}$  is added to 149  $\mu\text{L}$  (1:150 dilution) fresh EZ Rich Defined Media supplemented with proper concentrations of antibiotics and inducers. The time course of fluorescence from the lag phase to the stationary phase is measured in a high-throughput fluorescence plate reader (Tecan M1000) for 24 h. The excitation and emission wavelengths used for sfGFP were 485 nm and 510 nm, and are 587 nm and 610 nm for mRFP. The OD is measured at 600 nm. The shaking period between consecutive measurements is 900 seconds with the shaking diameter of 2.5mm.

### 2.3.4 Flow cytometry

Cell cultures with volume of 5  $\mu\text{L}$  mixed with 250  $\mu\text{L}$  PBS containing 2 mg/ml kanamycin (Fisher, BP906) are transferred to 96-well plates, and assayed using the flow cytometer (Partec Cyflow Space) with a RobbyWell sample loader in the 5 parameters of time, forward scatter (FSC), side scatter (SSC), sfGFP fluorescence (excitation at 500 nm, emission at 527 nm), and mRFP fluorescence (excitation at 590 nm, emission at 610 nm). Data for at least 50,000 cellular counts (triggered by SSC) are collected for each sample in log<sub>4</sub> mode at 12-bit resolution. The mean autofluorescence distribution of the cells without fluorescent proteins is also measured. Since the RobbyWell uses a 5  $\mu\text{L}$  air bubble before and after each sample run for internal cleaning, we gate each sample in time to filter out signal due to these air bubbles. After applying the Time gate, an ellipse of cells is gated in a plot of FCS vs. SSC. The arithmetic mean of fluorescence distribution is calculated using FCS Express Version 3.0 (De Novo Software) with mean autofluorescence subtracted. In the case of 2-color flow cytometry, data is compensated during data acquisition (0% spectral overlap of GFP in RFP channel, 0.32% spectral overlap of RFP in GFP channel). These arithmetic mean values are exported into Microsoft Excel and analyzed. Sometime, these values are further converted into MEFL values using the calibration

procedure described below, using the calibration run for the day the experiment is run. It takes 3~4 h to perform one 96-well measurement using flow cytometry. To reduce the change in gene expression during this period, 10  $\mu\text{L}$  of this MM culture is added to 200  $\mu\text{L}$  of pH 7.4 PBS containing 2 mg/mL Kanamycin to stop growth and translation in a 200  $\mu\text{L}$  96-well flat bottom plate (Greiner, 655101), and stored at 4  $^{\circ}\text{C}$  before measurement. Flow Cytometry is calibrated using 8-Peak Rainbow Calibration Beads (Spherotech, 559123). A calibration curve is obtained by converting fluorescence intensity into Mean Equivalent of Fluorescein (MEFL) units<sup>74</sup>. The same gain settings as the experiments are used, and they are such that the top peak of the calibration beads was off-scale, and thus peaks 2~7 are used to create a calibration curve in some experiments.

### 2.3.5 Microscope

1 $\mu\text{L}$  of each sample prepared according to the cell culture methods was placed on a poly-L-lysine (Sigma-Aldrich) gel pad and observed under a Zeiss Axio Observer D1 microscope (excitation filter 470nm/40nm, emission filter 525nm/50nm). The brightness and contrast levels of all microscopic pictures were adjusted to the same level using ImageJ (U. S. National Institutes of Health, <http://imagej.nih.gov/ij/>). Fluorescence and ODs are estimated using the plate reader before the microscope measurement.

### 2.3.6 Miller assay ( $\beta$ -galactosidase assay)

Mid-log phase cell culture is used for the Miller Assay. The absorbance of cell culture at 600 nm ( $OD_{600}$ ) is measured. Then 20  $\mu\text{L}$  cell culture is mixed with 80  $\mu\text{L}$  Permeabilization Solution [100 mM  $\text{Na}_2\text{HPO}_4$ ; 20 mM KCl; 2 mM  $\text{MgSO}_4$ ; 0.8 mg  $\text{ml}^{-1}$  Hexadecyltrimethylammonium bromide (CTAB, Sigma H5882); 0.4 mg  $\text{mL}^{-1}$  Sodium Deoxycholate (Sigma D6752); 5.4  $\mu\text{L}$   $\text{mL}^{-1}$   $\beta$ -Mercaptoethanol (Sigma M7154)]. Pre-warmed the mixture and Substrate Solution [60 mM  $\text{Na}_2\text{HPO}_4$ ; 40 mM  $\text{NaH}_2\text{PO}_4$ ; 3 mg  $\text{ml}^{-1}$  2-Nitrophenyl  $\beta$ -D-Galactoside (ONPG, Fluka 73660); 2.7  $\mu\text{L}$   $\text{mL}^{-1}$   $\beta$ -Mercaptoethanol] at 30  $^{\circ}\text{C}$  for 30 min, then add 600  $\mu\text{L}$  Substrate Solution to mixture, and record the start time. After sufficient color developed, add 700  $\mu\text{L}$  Stop Solution [1M Sodium Carbonate ( $\text{Na}_2\text{CO}_3$ )], and record the stop time. Mix well and centrifuge, and measure the absorbance at 420 nm ( $OD_{420}$ ) of the liquid solution. Calculate the Miller Units according to

$$MillerUnit = \frac{OD_{420}}{OD_{600} \cdot V \cdot T}$$

Where V is the volume of cell culture, which equals 0.02 mL, and T is the reaction time, which is the difference between stop time and start time. For detailed protocol, refer to the online protocol at [http://openwetware.org/wiki/Beta-Galactosidase\\_Assay\\_\(A\\_better\\_Miller\)](http://openwetware.org/wiki/Beta-Galactosidase_Assay_(A_better_Miller)).

## 2.4 RNA Extraction and Reverse Transcription

Bacterial small noncoding RNAs usually have short half-lives (~1 min). We used TRIzol Max (Invitrogen, 16122-012) to extract small RNAs followed by purification with Chloroform-Isopropanol. For detailed procedure, refer to [http://tools.invitrogen.com/content/sfs/manuals/trizolmax\\_man.pdf](http://tools.invitrogen.com/content/sfs/manuals/trizolmax_man.pdf). The extracted RNA is

estimated by Nanodrop (Thermo Scientific) with purity tested using Agilent 2100 Bioanalyzer.

Reverse transcription is performed using High Capacity cDNA Reverse Transcription Kits (Applied Biosystems). In brief, 10  $\mu\text{L}$  of total RNA sample is mixed with 10  $\mu\text{L}$  of RT master mix prepared according standard commercial protocol. The reactions (in 96-well plates) are then placed in a thermocycler with conditions [25  $^{\circ}\text{C}$ , 10 min; annealing 37  $^{\circ}\text{C}$ , 120 min; 85  $^{\circ}\text{C}$ , 5 s]. For detailed protocol, refer to [http://tools.invitrogen.com/content/sfs/manuals/cms\\_042557.pdf](http://tools.invitrogen.com/content/sfs/manuals/cms_042557.pdf).

## 2.5 Quantitative Real-Time PCR (qRT-PCR)

Real-Time PCR is performed in 96-well plates using KAPA PROBE FAST qPCR kit (Kapabiosystems, KK4701). Probes and primers are designed using Primer Express® (Life Technologies), and ordered from Integrated DNA Technologies (IDT). The probe is on mini scale (0.5 nmole), with 5' mod 6-FAM and 3' mod ZEN-Iowa Black FQ. All primers and probes are diluted to 10  $\mu\text{M}$ . Standard curve was measured together with the samples in the same 96-well plate, with concentrations of template DNA ranging from  $10^3 \sim 10^{10}$  molecules  $\text{L}^{-1}$ . The template DNA is designed to be the same as the sample DNA. For the real-time PCR reaction, 2  $\mu\text{L}$  of sample DNA is mixed with 10  $\mu\text{L}$  2X Master Mix, 0.8  $\mu\text{L}$  of each primer, 1  $\mu\text{L}$  of probe, and 0.4  $\mu\text{L}$  of 50X Rox High in a total volume of 20  $\mu\text{L}$ . The reaction is performed in Applied Biosystems Step One Plus (Applied Biosystems), with conditions [95  $^{\circ}\text{C}$ , 20s; 95  $^{\circ}\text{C}$ , 1s, 60  $^{\circ}\text{C}$  30s, 40 cycles]. The results are then analyzed in Microsoft Excel. For detailed protocol, refer to <http://www.kapabiosystems.com/products/name/kapa-probe-fast-qpcr-kits>.

## 2.6 Northern Blotting

DNA probes and ssRNA ladder are biotinylated using BrightStar Psoralen-Biotin Nonisotopic Labeling Kit (Life Technologies, AM1480). The ssRNA ladder is purchased from New England Biolabs (N0362), and denatured at 100  $^{\circ}\text{C}$  for 20 mins followed by chilling in liquid nitrogen. RNase-free water is prepared by adding 200  $\mu\text{l}$  DEPC to 1 L  $\text{H}_2\text{O}$  to a final concentration of 0.02%. The mixture is incubated overnight at 37  $^{\circ}\text{C}$ , and autoclaved for 1h. RNase-free glassware is prepared by soaking in freshly prepared 0.1% DEPC- $\text{H}_2\text{O}$  for 1 h. Then DEPC- $\text{H}_2\text{O}$  is drained and the glassware is autoclaved for 15 min. Rinse glassware with RNaseZap (Life Technologies, AM9780) once and RNase-free water twice. The RNase-free plastic containers are prepared by spraying with RNaseZap, and incubating at room temperature overnight, and rinsing with 0.1% DEPC- $\text{H}_2\text{O}$ . The electrophoresis apparatus is cleaned by washing with detergent solution, followed by rinsing with  $\text{H}_2\text{O}$  and ethanol. The apparatus is dried and filled with 3%  $\text{H}_2\text{O}_2$  (Fisher, H325), incubated at room temperature for 10 min, and finally rinsed with 0.1% DEPC- $\text{H}_2\text{O}$ . DEPC is very toxic and could break down the plastic.

Formaldehyde denaturing gel is prepared according the NorthernMax Kit (Life Technologies, AM1940). One microliter of RNA samples and ssRNA ladders are denatured at 65  $^{\circ}\text{C}$ , 10 min, and loaded onto the formaldehyde gel. After electrophoresis, the gel is transferred to a BrightStar-Plus Positively Charged Nylon Membrane (Life Technologies, AM10102). The chromatography Blotting Paper (Whatman, 3MM Chr) is used to facilitate this blotting process for 3 h. Prolonger blotting process is avoided because small RNAs might be hydrolyzed. The blotted RNA is crosslinked by using a Stratalinker UV Crosslinker (Stratalinker, Model 1800) at Autocrosslink mode for 1 min. The membrane is then prehybridized in a hybridization oven at

42 °C, 12 r.p.m. for 1 h using ULTRAhyb Ultrasensitive Hybridization Buffer (Life Technologies, AM8670). Then biotinylated probes are added, and hybridization is performed at 42 °C, 10 r.p.m. overnight. After hybridization, unhybridized probes and ULTRAHyb hybridization buffer is washed with 20 ml NorthernMax Low Stringency Wash Solution (Ambion, AM8673) using a rotisserie at room temperature for 5 min, followed by washing with 20 ml High Stringency Wash Solution to conical tube (Ambion, AM8674) in the hybridization oven at 42 °C for 15 min. The washing step with High Stringency Wash Solution is repeated once. To detect RNA samples, the membrane is first washed with NorthernMax Wash Buffer, then NorthernMax Blocking Buffer. One microliter of Streptavidin Alkaline Phosphatase (Promega, V55910) is added, and incubated at room temperature for 30 min. Then membrane is then washed with NorthernMax Blocking Buffer and NorthernMax Assay Buffer, and incubated with CDP-Star (NEB, N7001), and incubated using a rotisserie at room temperature for 5 min. Finally, the membrane is visualized using X-ray films (Kodak), in a dark room for 30 min, which are then developed using Automatic X-ray film processor.

## 2.7 In Vitro RNA Synthesis

A DNA template for transcription of the desired RNA fragment is inserted in the context of the 5' and 3' flanking SHAPE structure cassette<sup>101</sup>. This is done by PCR [1 mL; containing 20 mM Tris (pH 8.4), 50 mM KCl, 2.5 mM MgCl<sub>2</sub>, 200 μM each dNTP, 500 nM each forward and reverse primer, 5 pM template, and 0.025 units/μL Taq polymerase; denaturation at 94 °C, 45 s; annealing 55 °C, 30 s; and elongation 72 °C, 1 min; 34 cycles]. The PCR product was recovered by ethanol precipitation and resuspended in 150 μL of TE [10 mM Tris (pH 8.0), 1 mM EDTA]. Transcription reactions (1.0 mL, 37 °C, 12-14 h) contained 40 mM Tris (pH 8.0), 20 mM MgCl<sub>2</sub>, 10 mM DTT, 2 mM spermidine, 0.01% (v/v) Triton X-100, 5 mM each NTP, 50 μL of PCR-generated template, 0.12 U/μL RNase Inhibitor (Promega) and 0.1 mg/mL of T7 RNA polymerase. The RNA products was purified by denaturing polyacrylamide gel electrophoresis (8% polyacrylamide, 7 M urea, 29:1 acrylamide:bisacrylamide, 32 W, 2 h), excised from the gel, and recovered by passive elution and ethanol precipitation. The purified RNA (~3 nmol) was resuspended in 200 μL TE.

## 2.8 SHAPE Experiment

Structure-selective RNA modification with and without theophylline is performed following the previous experimental design<sup>102</sup>. For the without theophylline reaction, 10 pmol of RNA was suspended in 12 uL of nuclease free water in a PCR tube. Using a thermocycler, the RNA was heated to 95 °C for 2 minutes, then placed on ice for 1 min. 6 uL of 3.3x folding buffer (333 mM HEPES (pH 8.0), 333 mM NaCl, 33 mM MgCl<sub>2</sub>) was added, followed by incubation at 37 °C in the thermocycler for 20 mins. 9 uL of this folded RNA solution was added to either 1 uL 10x 1M7 (65mM) (+ reaction), or 1uL neat DMSO (- reaction), and further incubated at 37 °C. Modified or control RNAs were then ethanol precipitated following<sup>103</sup>. For the with theophylline reaction, 10 pmol of RNA was suspended in 10 uL of water in a PCR tube. After identical heat denaturation steps, 2 uL of 9x theophylline (4.5 mM dissolved in water) was added, and 6 uL of 3.3x folding buffer was added followed by identical steps as above. The general procedure of primer extension and data analysis followed previous protocols<sup>103</sup>, using 5 pmol of RNA for dda

and ddT sequencing reactions. SHAPE intensities are converted into a pseudo-free energy change term in the RNAstructure program<sup>63</sup>:

$$\Delta G_{SHAPE} = m \cdot \ln[SHAPE_{reactivity} + 1.0] + b$$

The intercept,  $b$ , is the free-energy bonus for formation of a base pair with zero or low SHAPE reactivity, whereas  $m$ , the slope, drives an increasing penalty for base pairing as the SHAPE reactivity increases. These parameters dictate the strength of the experimental contribution to the energy function. The  $b$  and  $m$  parameters used were -0.5 and 3.4 kcal/mol, respectively.

## 2.9 Data Analysis

### 2.9.1 RNA secondary structure prediction

RNA secondary structure and folding thermodynamic energies are calculated using Mfold<sup>65</sup>.

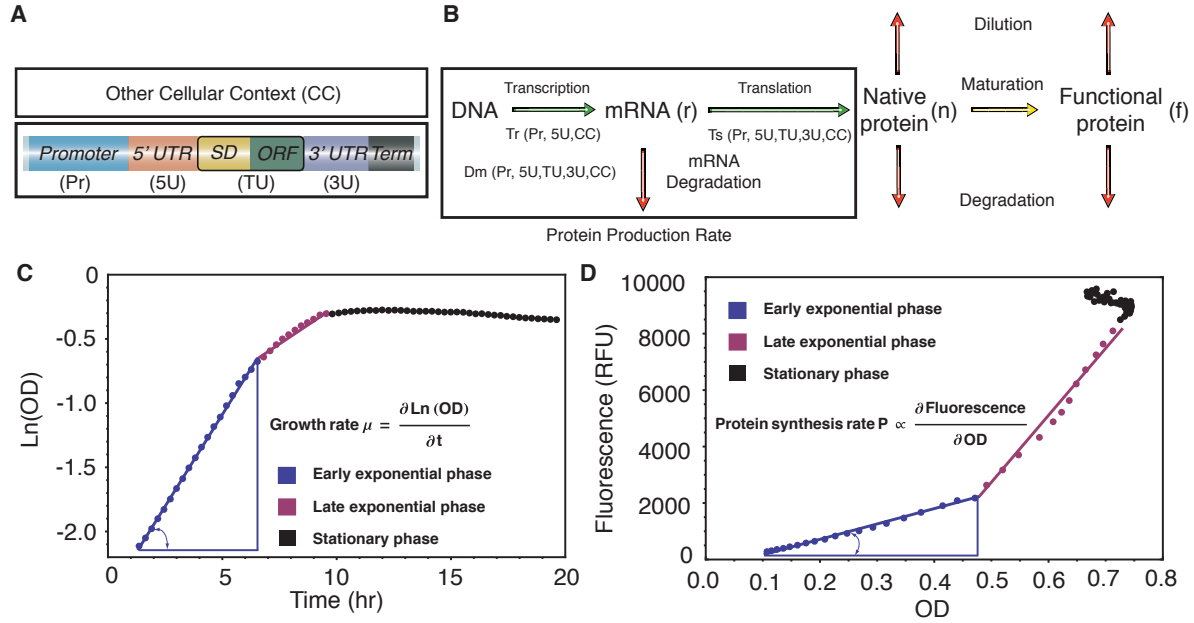
### 2.9.2 A linear model for protein expression

The mathematical model for protein expression followed the previous study<sup>104</sup>, by taking account of nonfluorescent protein production, protein maturation, degradation by proteases, and dilution by cell division.

As shown in **Figure 2.9.2A&B**, the amount of protein expression is determined jointly by multiple genetic components, including promoter (Pr), 5' UTR (5U), the translation unit consisting of the RBS and gene (TU), 3' UTR (3U), and cellular contexts (CC) such as abundance of RNA polymerase, ribosome and nutrients. Following the central dogma, each step of gene expression, such as transcription, translation, mRNA degradation, protein maturation and protein degradation/dilution, is a multivariate function of the sequences of the genetic elements. The protein production rate during the exponential phase can be expressed by:

$$\left. \frac{\partial P}{\partial t} \right|_{production,ss} = \frac{Tr \cdot Ts}{D_m}$$

Where  $Tr$ ,  $Ts$ ,  $D_m$  are all multivariate functions of different genetic elements. In the gene circuits, because changing the sequence of single components could affect all  $Tr$ ,  $Ts$ ,  $D_m$  in a perplexing way, and that's the main reason why gene circuits are not predictable.



**Figure 2.9.2 A model for gene expression and calculation of protein production rates.** (A) Multiple genetic components affect gene expression together. (B) The model for calculating protein production rate. (C) The growth rate is calculated from the plot of Log(OD) over time during the exponential phase (blue). The late exponential phase is shown in red and the stationary phase in black. (D) The quantity of is calculated from the plot of fluorescence (F) as a function of OD.

### 2.9.3 Calculation of protein production rates

The relationship between the amount of the nonfluorescent protein  $P$  and fluorescence  $F$  (assuming  $F$  is proportional to the amount of fluorescent protein which is true under our experimental conditions) can be described by<sup>104</sup>

$$\frac{\partial P}{\partial t} = \frac{\partial P}{\partial t} \Big|_{\text{production}} - v \cdot P - \mu \cdot P - D_P$$

$$\frac{\partial F}{\partial t} = v \cdot P - OD \cdot D_F$$

Where  $v$  is the maturation rate (per hour) of fluorescent protein, and  $\mu$  is the growth rate (per hour), and  $D_P$  and  $D_F$  are the degradation rate of nonfluorescent and fluorescent proteins, respectively. We calculate the growth rates ( $\mu$ ) during the exponential phase for each circuit by plotting Log(OD) over time (**Figure 2.9.2C**). In the term of  $OD \cdot D_F$ , it contains both effects of cell division and protein degradation. The cell division is described by

$$\frac{\partial OD}{\partial t} = \mu \cdot OD$$

So the quotient  $\partial F / \partial OD$  describes

$$\frac{\partial F}{\partial OD} = \frac{v \cdot P - D_F}{\mu}$$

At the steady state,  $\partial P / \partial t = 0$  and  $\partial F / \partial t = 0$ , and combining all equations we obtain

$$\left. \frac{\partial P}{\partial t} \right|_{production,ss} = \left. \frac{\partial F}{\partial OD} \right|_{ss} \cdot \mu \cdot \left(1 + \frac{\mu}{v}\right) + D_{P_{ss}} + D_{F_{ss}} \cdot \left(1 + \frac{\mu}{v}\right)$$

Assume  $D_{P_{ss}} \ll 1$  and  $D_{F_{ss}} \ll 1$ , which are true for both sfGFP<sup>98</sup> and mRFP<sup>99</sup>, we obtain

$$\left. \frac{\partial P}{\partial t} \right|_{production,ss} = \left. \frac{\partial F}{\partial OD} \right|_{ss} \cdot \mu \cdot \left(1 + \frac{\mu}{v}\right)$$

For each circuit, we measure the kinetic expression of both fluorescence and OD (**Figure 2.9.2D**). The plots of measured fluorescence (F) as a function of OD during the exponential phase show a linear curve for all circuits tested, which allows us to calculate the slopes of this linear region ( $\left. \frac{\partial F}{\partial OD} \right|_{ss}$ ). In all experiments, the maturation constants for sfGFP and mRFP are

$$m_{sfGFP} = 7h^{-1.98} \text{ and } m_{mRFP} = 3.5h^{-1.99}.$$

The data are analyzed using Mathematica 7 (Wolfram) by importing from Microsoft Excel exported files (.txt), and organized as three matrices [M×N]OD, [M×N]GFP, and [M×N]RFP. Where M is the number of cell cultures, and N is the number of time points. For each sample, we plot the Log(OD<sub>600</sub>) vs. t [time], and use native Mathematica 7 clustering function to cluster the whole curve into three segments:

$$\text{SegmentedCurve} = \text{FindClusters}[\text{CurveData}, 3]$$

We perform linear regression on each segment to calculate its slope (k) and linear regression fitness (R<sup>2</sup>). The three segments correspond to three growth phases, - lag phase, exponential phase, and stationary phase, respectively. The slope for exponential phase (k<sub>expn</sub>) is used to calculate of the growth rate during the exponential phase:

$$\begin{aligned} \text{SlopeData} &= \text{LinearModelFit}[\text{SegmentedCurve}, x, x] \\ \text{Rsquared} &= \text{Coefficient}[\text{Normal}[\text{SlopeData}, x], \text{SlopeData}["\text{RSquared}"]] \end{aligned}$$

We also perform linear regression on F as a function of OD<sub>600</sub> during the exponential phase and calculate the slope and regression fitness. The data are combined to compute the protein production rates as described.

#### 2.9.4 Statistics

Mean value and standard deviation are calculated as

$$Mean_F = \frac{1}{N} \sum_{i=1}^N Sample_{F_i}$$

$$Std_F = \sqrt{\frac{1}{N} \sum_{i=1}^N (Sample_{F_i} - Mean_F)^2}$$

Where N is the sample size,  $F_i$  are fluorescent values for each sample (i).

When the mean autofluorescence is subtracted, the compound standard deviation is calculated using standard propagation of errors as

$$Std_F^{corrected} = \sqrt{(Std_F)^2 + (Std_\phi)^2}$$

Repression percentage or transcription attenuation is calculated as

$$Repression\% = \left(1 - \frac{F^{(-)}}{F^{(+)}}\right) \cdot 100\%$$

Where  $F^{(-)}$  is the mean fluorescence without antisense RNA, and  $F^{(+)}$  is the mean fluorescence with WT antisense added. The standard deviation in this expression is

$$Std^{rep\%} = Repression\% \cdot \sqrt{\left(\frac{Std_{F^{(-)}}}{Mean_{F^{(-)}}}\right)^2 + \left(\frac{Std_{F^{(+)}}}{Mean_{F^{(+)}}}\right)^2}$$

Normalized fluorescence values are calculated in a similar fashion. Occasionally a colony is thrown out of analysis because of abnormal fluorescence results, for example, too large a spread in the FCS vs. SSC plot, or a low number of cellular counts observed causing a low signal to noise ratio in the fluorescence histogram.

The Relative Standard Deviation, which measures the variance (or noise) among the samples, is calculated as

$$RSD = \frac{Std}{Mean} \times 100\%$$

Linear regression, Student's *t*-test and *p*-values is carried out using Prism 5 (GraphPad Software). Spearman's rank correlation coefficients are calculated using Microsoft Excel.

For induction curve measurements, each sample with inducers is first normalized by the value without inducer present, then the normalized values for the same inducer concentration measurement is averaged and the error bars are calculated as the difference between multiple values. Mean fluorescence values with different inducer concentrations are used to determine the experimental steady-state induction curve and were fitted to the following Hill equation using Matlab<sup>74</sup>:



$$y = y_{\min} + \frac{y_{\max} - y_{\min}}{1 + \left(\frac{[\text{Inducer}]}{K}\right)^\eta}$$

Where  $y_{\min}$  presents leaky expression,  $y_{\max}$  presents maximum expression,  $[\text{Inducer}]$  represents the inducer concentration,  $\eta$  is the Hill coefficient, and  $K$  is the inducer concentration required for half repression of gene expression.

### 2.9.5 Flow cytometry data analysis

Flow cytometry data is acquired in log4 mode with 12-bits for each parameter (SSC, FSC, GFP, RFP, and Time). In the case of 12-bit data, each measured parameter falls into one of  $2^{12} = 4096$  possible channels, and each data point is recorded as a relative channel number,  $C$ , which falls in the range from 1 to 4096. Data is analyzed using FCS Express Version 3.0 (De Novo Software), where the relative channel number is converted to a relative intensity value,  $I$ , which falls in the range of 0~10,000 for log4 mode data. The relationship between  $C$  and  $I$  can be expressed as

$$C = \frac{r}{n} \cdot \log_{10}(I)$$

Where  $n = 4$  for log4 mode data, and the resolution  $r = 4096$  for 12-bit data.

A calibration curve represents the relationship between a known standard, and the measurement of that standard in the instrument. Let  $F$  be the known MEFL values of the bead peaks, and  $C$  be the measured relative channel number of each of the bead peaks. The calibration curve relationship is expressed as<sup>74</sup>

$$\log_{10}(F) = \alpha \cdot C + \beta$$

Where  $\alpha$  and  $\beta$  are coefficients to be fitted. Using the above relationship between  $C$  and  $I$ , this can be written as

$$\log_{10}(F) = \alpha \cdot \frac{r}{4} \cdot \log_{10}(I) + \beta$$

$\alpha$  and  $\beta$  are determined in Microsoft Excel using the LINEST least-squares fitting function, and had typical values of 0.001 and 1.26, respectively ( $R^2 = 1.00$ ). For cellular samples, measured fluorescence intensity values,  $I_c$ , are converted into MEFL units,  $F_c$ , using

$$F_c = 10^\beta \cdot I_c^{\alpha r/4}$$

## Chapter 3 RNA-Mediated Transcriptional Regulators

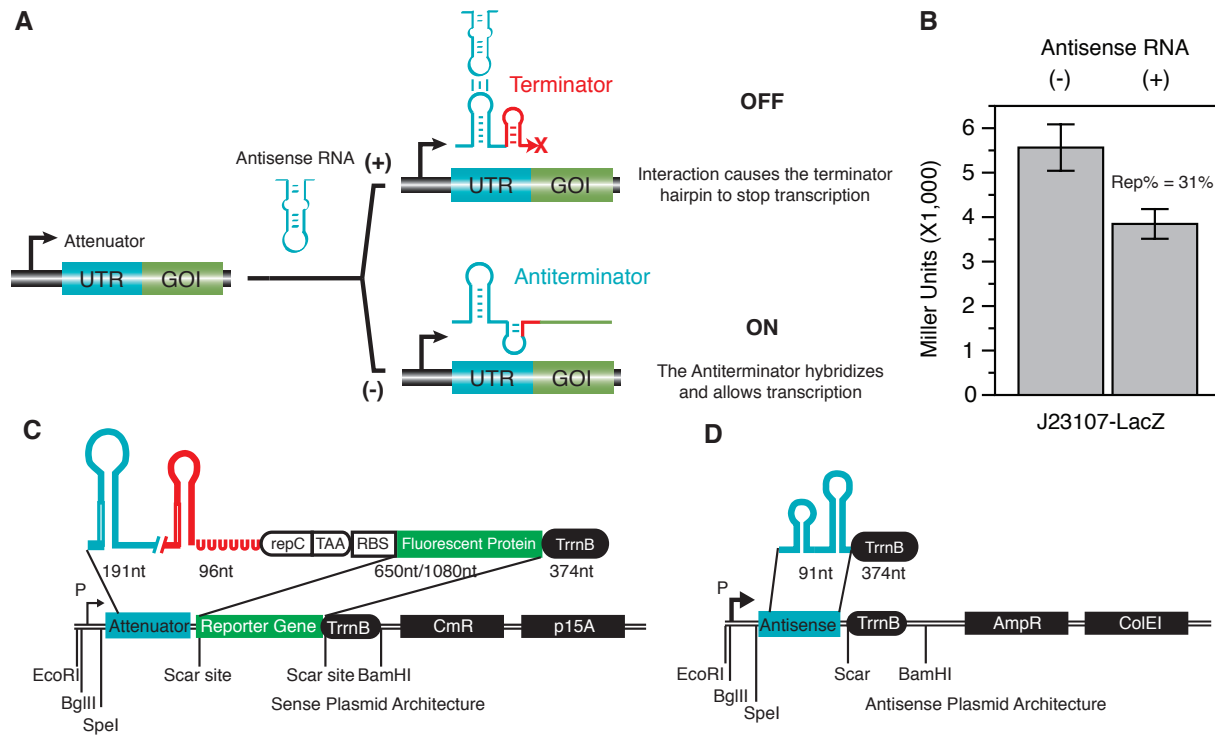
### 3.1 Introduction to The pT181 System

Antisense RNA-mediated gene regulation has been best characterized in accessory DNA elements including phages, plasmids and transposons. These mechanisms are important for the maintenance of copy number and transposability of these accessory elements<sup>105</sup>. Plasmid pT181 replicates via a rolling circle mechanism and is best characterized among *Staphylococcus aureus* pathogenic plasmids<sup>106</sup>. Previous studies have shown that the plasmid controls its copy number by an antisense RNA-mediated mechanism<sup>36</sup>. The synthesis of the replication rate-limiting RepC protein is controlled by the antisense RNA transcribed from the opposite strand. This antisense RNA interacts with the UTR region of the *repC* mRNA by base pairing (**Figure 3.1A**). When the antisense binds to the target region, it induces the formation of a transcription terminator upstream of the *repC* coding sequence (CDS). In this case, further transcription of downstream *repC* is inhibited, resulting in decreased expression of RepC protein and a lower plasmid replication rate. Otherwise, if the *repC* mRNA is not bound by the antisense RNA, full length mRNA is transcribed, which leads to increased RepC expression and a higher plasmid replication rate. Thus, the frequency of binding events between the antisense and mRNA target detects the existing plasmid copy number (the more plasmids the more antisense), allowing the plasmid to adjust its replication rate with a negative feedback control. The leaky and noisy binding between antisense-mRNA target is important to the wildtype system<sup>36</sup>, which allows better adaptation of the plasmid in diverse environments. However, for our engineering purpose, the noisy binding and leaky control of the downstream gene is undesired and should be reduced.

The pT181 system presents a well-studied RNA-mediated transcriptional elongation control system. This gives us a starting point to create more RNA-mediated transcriptional regulatory systems that can be used to build higher-order functions. We hypothesize that when the *repC* gene is replaced by a gene of interest (GOI), we can use the system to control transcription of an arbitrary gene. To verify this hypothesis, we followed previous studies<sup>107</sup>. First, we introduced a stop codon TAA to the 32<sup>th</sup> amino acid of the *repC* CDS (called repC fragment). Then we inserted a strong RBS (GGATCTAGGAGGAAGGATCT), a LacZ CDS, and a transcription terminator (TrnB, 368 bp) downstream of the repC fragment and placed the expression cassette on the p15A plasmid. A constitutive promoter J23107 ([http://partsregistry.org/Part:BBa\\_J23117](http://partsregistry.org/Part:BBa_J23117)) was used for the transcription of the sense cassette. Antisense RNA was placed on a high copy ColE1 plasmid under the control of a constitutive promoter J23100 ([http://partsregistry.org/Part:BBa\\_J23110](http://partsregistry.org/Part:BBa_J23110)) (**Figure 3.1C**). The  $\beta$ -galactosidase assay in *E. coli* Tg1 cells with and without the wildtype antisense RNA showed moderate repression, which is similar to previously described<sup>107</sup> (**Figure 3.1B**). This repression effect is weaker than that reported in its natural host *S. aureus*, which is more than 95%. The decrease in regulation probably come from several reasons such as insufficient expression of antisense RNA and inefficient intrinsic termination. Improvements of the repression will be discussed below.

We chose the pT181 system as the starting point to engineer novel RNA-mediated transcriptional systems for several reasons. First, the pT181 does not require other protein factors except for RNA polymerase, which avoids complicated protein-RNA interactions. This also likely makes it easier to transfer the system to other organisms. Second, the pT181 system has been well studied and characterized. The nucleotides that are important for the antisense-target

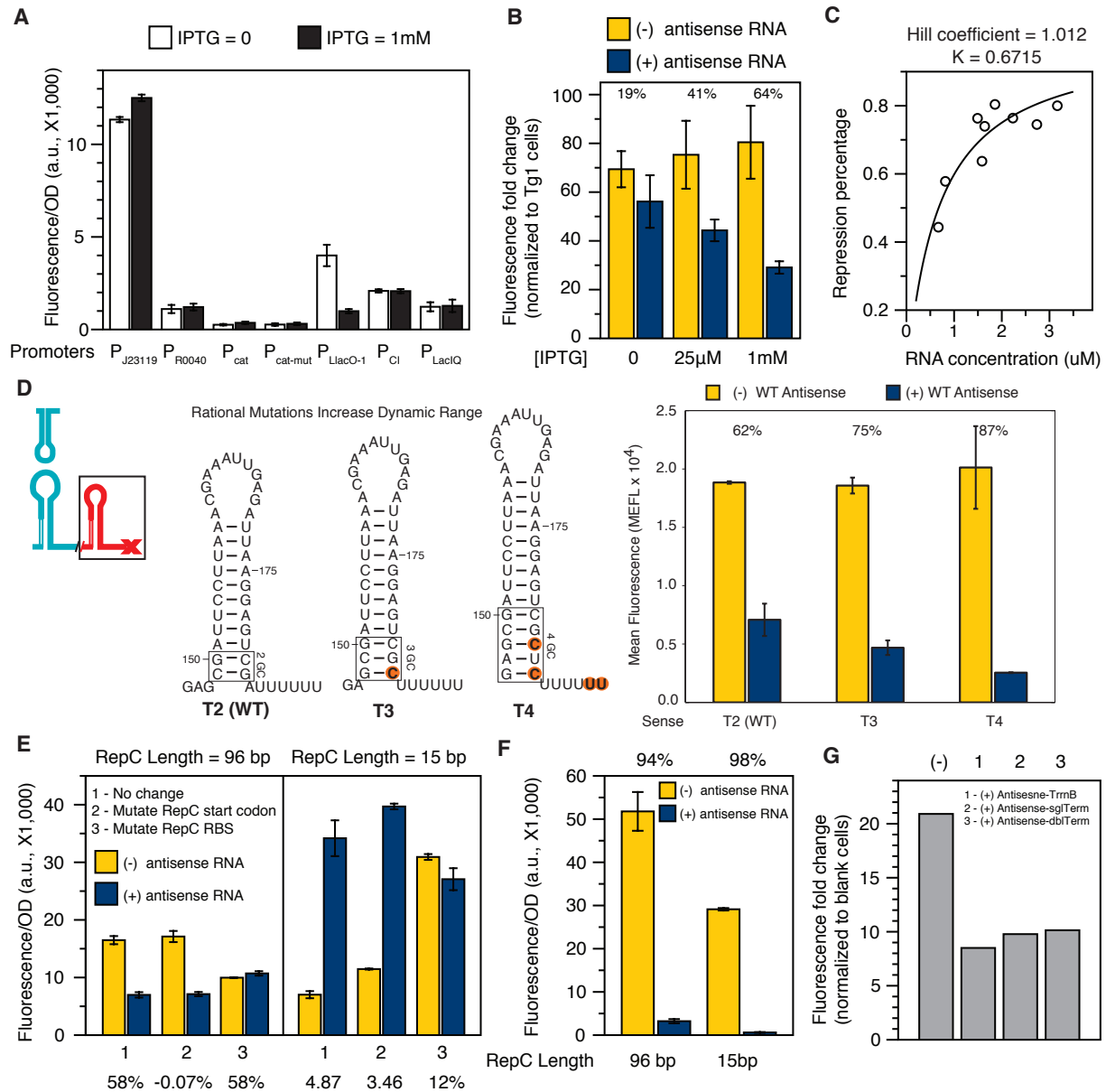
interaction have been reported<sup>106</sup>. Third, it has been shown that the system functions in both gram-negative and gram-positive model organisms, *E. coli* and *B. subtilis*. This sets a stage to further engineer the system with desired properties and novel functions.



**Figure 3.1 Mechanism of the pT181 system.** (A) Antisense RNA controls the pT181 UTR region, turning ON or OFF gene expression. (B) Initiation characterization of the wildtype pT181 system in *E. coli* using a LacZ gene as the reporter. (C) The reporter system containing pT181 sense UTR controlling a fluorescent protein on a plasmid with a p15A replication origin and a chloramphenicol resistance marker. (D) Antisense RNA expression system on a plasmid with ColE1 replication origin and an ampicillin resistance marker.

### 3.2 Optimization of The Natural pT181 System

The natural system only exhibited 31% of repression percentage in *E. coli* (Figure 3.1B), which is not useful for engineering other functions. It is important to optimize the ON/OFF dynamic range of the system as the first step. Previous studies showed that a high ratio between *in vivo* antisense and sense concentrations ( $[\text{antisense}]/[\text{sense}]$ ) is critical to efficient repression<sup>108</sup>. Furthermore, we examine if the intrinsic terminator can be improved, since this terminator might have been evolved for optimal function in *S. aureus* not in *E. coli*. Finally, we change the lengths of repC fragment, and study how this linker sequence affects regulation.



**Figure 3.2 Optimization of the wildtype pT181 system.** (A) Characterization of different constitutive and inducible promoters. (B) When we use the IPTG-inducible promoter P<sub>LlacO-1</sub> to induce antisense RNA, the reporter system shows increased repression with more IPTG supplemented. (C) We quantify *in vivo* antisense RNA expression using qRT-PCR. Fitness between repression percentages with antisense RNA concentration using a Hill equation shows Hill coefficient = 1, suggesting these is no cooperativity of the system. (D) Improvement of the intrinsic terminator strength. We modify the stem region G-C contents as well as the poly-U length, and obtain better ON/OFF range. (E) Different lengths of repC fragment affect functions of the system. (F) Translational fusion of the reporter gene to the repC fragment shows much bigger ON/OFF range, probably due to both transcription and translation controls. (G) There are no effects from different antisense RNA transcriptional terminators.

### 3.2.1 Antisense RNA abundance is important

Extensive previous work on the pT181 attenuator, performed both *in vitro*<sup>106</sup> and *in vivo*<sup>107</sup>, has elucidated the mechanism of the attenuator and its response to varying concentrations of antisense RNA. In particular, results from *in vitro* transcription attenuation assays demonstrate that the amount of observed attenuation depends on the ratio of [antisense]/[attenuator] and a high level of attenuation requires this ratio to be greater than one. In other antisense systems such as CopA-CopT, this ratio is reported to be in the range of 10~20 fold<sup>108</sup>. This is consistent with previous Northern blotting observations of attenuation *in vivo*<sup>107</sup>. In our original test platform, we used a high copy plasmid and a strong promoter to express antisense RNA, and a medium copy plasmid and a medium promoter for sense mRNA. However, the antisense promoter might not be strong enough to produce sufficient antisense RNA, especially given that the antisense generally possesses a very short half-life. Furthermore, in the wildtype (WT) system, antisense and sense are physically adjacent to each other, whereas in the synthetic system, they are separated on two plasmids. Thus, we first test that if the low repression is caused by insufficient antisense RNA expression.

To modify the promoters for both antisense and sense, we first characterized a set of promoters obtained from the MIT parts registry by cloning them to GFP. Among this group, J23119 ([http://partsregistry.org/Part:BBa\\_J23119](http://partsregistry.org/Part:BBa_J23119)) exhibits the strongest expression, which is 3-fold stronger than Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)-inducible P<sub>L</sub>lacO-1 (**Figure 3.2A**). We then chose J23119 to control the sense expression, and P<sub>L</sub>lacO-1 to control the antisense. Our results showed that the repression increased with IPTG concentrations (**Figure 3.2B**). The maximal repression percentage we obtained is  $64\% \pm 13\%$ . We also used J23119 to expression antisense, however, no difference in repression was observed using J23119 or P<sub>L</sub>lacO-1 for antisense (compare **Figure 3.2B** and **Figure 3.2D**). So, in later experiments, we used J23119 for both sense and antisense.

We measured the [antisense]/[sense] ratio using qRT-PCR for the WT antisense/WT-T4 attenuator system. Because of the perfect complementarity of the two RNA molecules, it is not possible to perform this measurement on strains harboring both sense and attenuator plasmids. Rather, separate measurements are performed on two separate strains. The resulting [antisense]/[sense] ratio observed when the WT antisense is transcribed from the J23119 promoter is found to lie in the range of 3.8~9.7, consistent with earlier experiments<sup>107</sup>. Since the same promoter is used to drive the transcription of the attenuator, this ratio most likely reflects the difference in plasmid copy numbers between the ColE1 and p15A origins and mRNA stability. Furthermore, we characterized the effects of different amounts of antisense RNA on attenuation by using the P<sub>L</sub>lacO-1 promoter to induce antisense and measuring antisense concentrations by qRT-PCR along with observed *in vivo* repression (**Figure 3.2C**). The trend of increasing antisense leading to increasing attenuation is clearly observable and consistent with the fluorescence assay data. Interestingly, when repression percentage is plotted versus [antisense], the effective Hill coefficient is near 1, implying there is no cooperativity in antisense-target interaction<sup>109</sup>.

### 3.2.2 The intrinsic terminator is important for attenuation

By increasing the concentration of antisense, we improved the percentage change to around 60%. To further improve attenuation, we tried to optimize the wild type attenuator. This involves optimization of both ON and OFF levels. Here we focus on tuning dynamic range by optimizing the OFF level, and optimizing the ON level is discussed below.

Watson-Crick base pairing of the pT181 attenuator with the complementary antisense RNA hairpin promotes the formation of an intrinsic terminator hairpin that causes polymerase to fall off and stop transcription. The termination efficiency of this type of rho-independent terminator is partially determined by the GC contents of the stem region and the length of the polyU. To improve attenuation, several mutations were rationally designed to increase the number of GC pairs at the base of the terminator stem, thus increasing thermodynamic stability in this region which is known to increase terminator efficiency<sup>32</sup>. Mutations that increase the number of pairs from two in the wild type (T2) to four (T4) increase attenuation to 84% (**Figure 3.2D**). Since each mutation preserves the absolute fluorescence (ON level) without antisense, these results clearly demonstrate the ability to tune attenuation in this system. Since wild type pT181 system in *S. aureus* exhibits 97% repression<sup>36</sup>, further optimization in the intrinsic terminator might yield an even better dynamic range. The effects of changing the length of polyU remain to be tested.

### ***3.2.3 The repC fragment is important for the regulatory function***

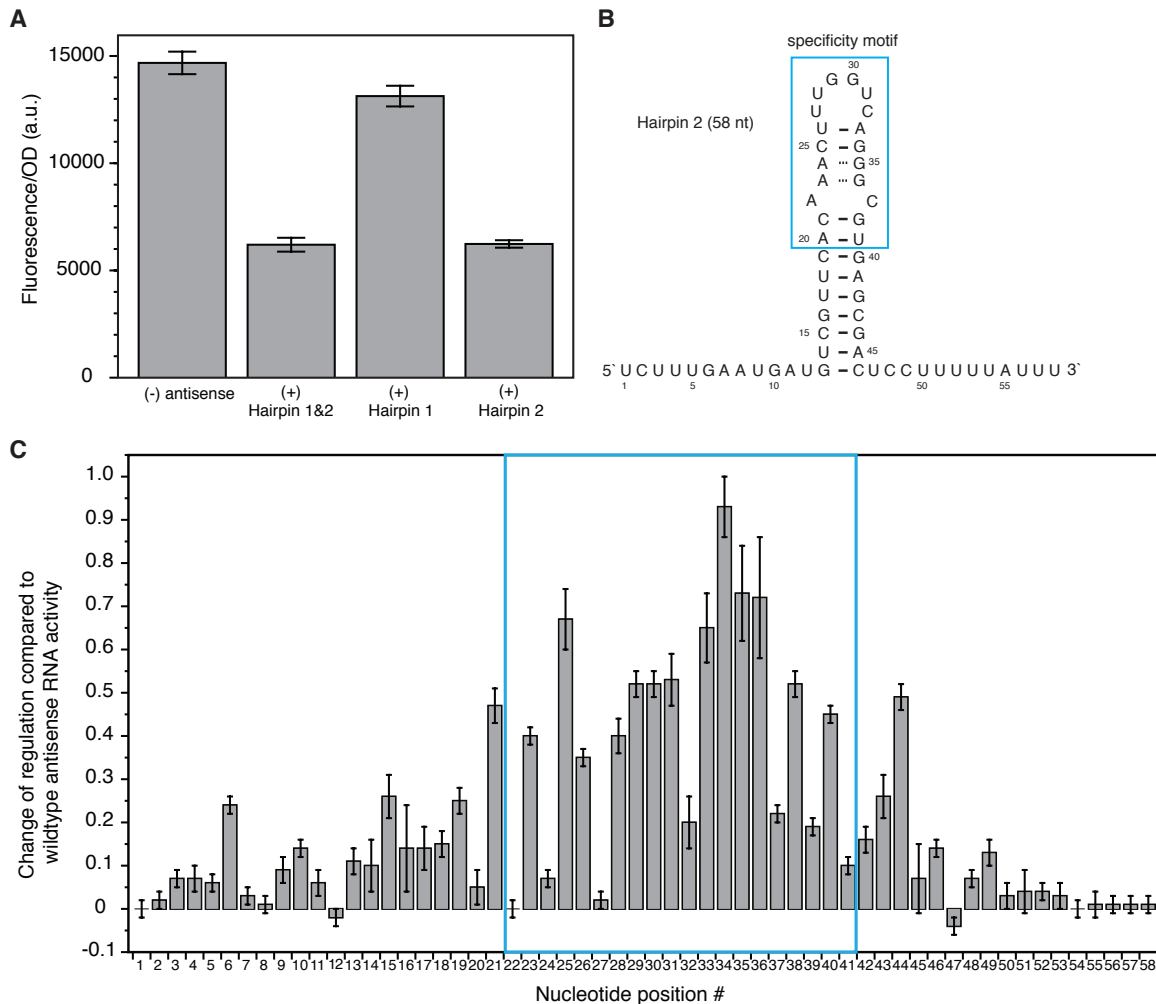
The repC fragment used is 96-nucleotide (nt) (32 amino acid) in length<sup>107</sup>. To investigate if the length of repC fragment is important for repression, we truncated the length to 15-nt (5 amino acids). Surprisingly, presence of antisense RNA activated the fluorescence expression by ~5-fold instead of repression (**Figure 3.2E**). Moreover, when we knocked out the proposed repC RBS that is embedded in the intrinsic terminator by mutating a duplet of GG to CC, we observed complete loss of regulation for both repC lengths. We also mutated the start codon of repC from ATG to TAA, but no changes were observed. A hypothesis is that the intrinsic terminator formation both terminates transcription and inhibits translation by sequestering the repC RBS, which regulates both transcription and translation. To test this hypothesis, we translationally fused sfGFP to the 32<sup>th</sup> or 5<sup>th</sup> amino acid of the repC fragment. In this case, the sfGFP utilizes the natural repC RBS for translational initiation. Consistent with our hypothesis, we observed a much stronger repression for both repC lengths, which was approximately 95% (**Figure 3.2F**) and consistent with the high repression percentage (97%) of the natural pT181 system<sup>36</sup>. Thus, we confirmed that pT181 is a hybrid system that exhibits both transcription and translation controls.

### ***3.2.4 Antisense RNA terminator is not important***

Antisense RNA is a 90-nt noncoding sequence. In our system, we used a *rho*-independent terminator (TrnB) to stop its transcription. Since these *rho*-independent terminators tend to attach an extra sequence to the 3' end of the transcribed antisense fragment, we tested if this resulted in lower repression. We changed the TrnB terminator to two different synthetic terminators, sglTerm and dblTerm, which have different sequences. No difference of repression was observed using these terminators, implying that the antisense RNA is robust to the 3' compositions (**Figure 3.2G**).

### 3.3 Characterization Of Nucleotide Importance In Repression

We thus obtained a fundamental pT181 system that allowed us to create new RNA-mediated transcriptional controls. The antisense RNA possesses two hairpins, and previous studies have shown that one hairpin is more important than the other one<sup>106</sup>. Here, we further truncated the whole antisense to two separate hairpins, and measured the repression from either hairpin. Our results showed that Hairpin 2 is enough for full repression, while Hairpin 1 almost had no effects (**Figure 3.3A**). To further understand the regulatory importance of each single nucleotide, we carried out single mutations to the single Hairpin 2 (**Figure 3.3B**). Each nucleotide was mutated to its complementary nucleotide (such as G to C, A to T, etc.), giving 58 mutant antisense RNAs each bearing a single mutation<sup>110</sup>. The regulatory activities of all 58 mutant antisense RNAs were then assayed with the pT181 sense reporter. By comparing with the WT antisense RNA, we found that nucleotides in the loop region is critical for repression while those in the lower stem can tolerate moderate modifications without affecting the regulatory function (**Figure 3.3C**).



**Figure 3.3 Characterization of antisense RNA.** (A) The activity of antisense RNAs consisting of hairpin 1, hairpin 2 or hairpin 1&2. Hairpin 2 displays similar repression as the full-length pT181 ncRNA (hairpin 1&2), confirming that this hairpin is critical for target recognition. (B) Secondary structure of the antisense RNA (Hairpin 2). (C) Point mutation study of the pT181

antisense Hairpin2 (58-nt). The activities of mutants that bear single mutations at each position are assayed and compared to that of the WT antisense RNA. The bar graph shows the change of activity relative to the WT antisense RNA, with the value of 0 meaning no change of activities (thus strong repression on the target) and the value of 1 meaning completely loss of repression. The target specificity region is shown in the cyan box.

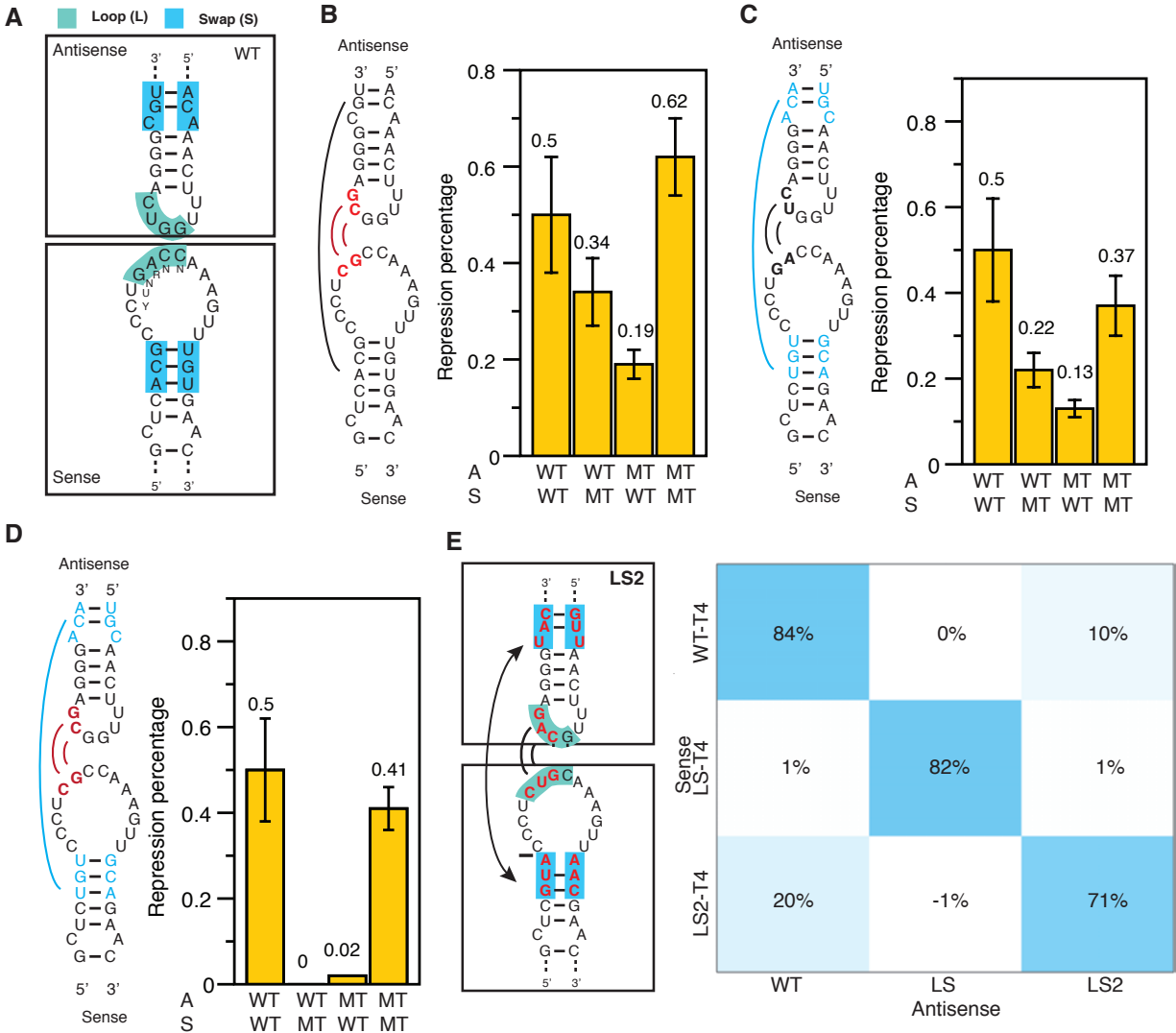
### 3.4 Engineering Orthogonal Transcriptional Attenuators

#### 3.4.1 Engineering the specificity of RNA-RNA interactions

Creating orthogonal antisense-attenuator variants requires changing the specificity of RNA-RNA interaction. An additional goal is to do this with as few mutations as possible so that orthogonal variants would have near-identical response to their cognate antisense. Our design strategy centered on making mutations in two specific regions of the pT181 RNA structures based on the mechanism in the related CopA/CopT RNA translation regulator<sup>111,112</sup>: the loop regions of the antisense and attenuator structures that are known to be important for initial RNA-RNA recognition, and the hairpin collars which are involved in stable antisense-attenuator complexes (**Figure 3.4.1A**).

The antisense and attenuator loops both have YUNR (Y - pyrimidine (C, U), R - purine (A, G), N - A, U, C, G) motifs, which are ubiquitous in recognition loops of natural antisense systems<sup>113</sup>. The YUNR motif has been shown as a useful design element in synthetic riboregulators<sup>86</sup>. We therefore search for mutations that preserve the motif, while otherwise disrupting interactions between non-cognate antisense-attenuator pairs. We found as few as a two-nucleotide change could decrease attenuation between non-cognate WT and MT antisense/attenuator pairs to between 19% (WT/MT) and 34% (MT/WT) (**Figure 3.4.1B**). In the CopA/CopT system, it was found that simply swapping base pairs of the hairpin stems disrupted non-cognate antisense-attenuator complexes, even if they had complementary hairpin loops. Swapping bases also has the advantage of preserving the RNA structures of the individual molecules, thus causing minimal disruption to the functioning of the mutant systems. We found that swapping three base pairs on the hairpin collar of either the pT181 antisense or attenuator substantially reduced crosstalk attenuation to 13% (WT/MT) and 22% (MT/WT) (**Figure 3.4.1C**). More significantly, when we combined the loop (L) and swap (S) mutations, they acted synergistically to completely remove crosstalk between non-cognate antisense-attenuator pairs (**Figure 3.4.1D**). To demonstrate the effectiveness of this design strategy, we expanded the searching region of YUNR motif. We included two immediate downstream nucleotides of the YUNR motif (YUNRNN), and constructed a larger mutant library. We only tested part the library, and found a third orthogonal pair LS2 (**Figure 3.4.1E**). This mutant shows almost identical ON and OFF levels to the WT and LS attenuators, and is completely orthogonal to the LS variant. We note that there is some crosstalk with the WT system.



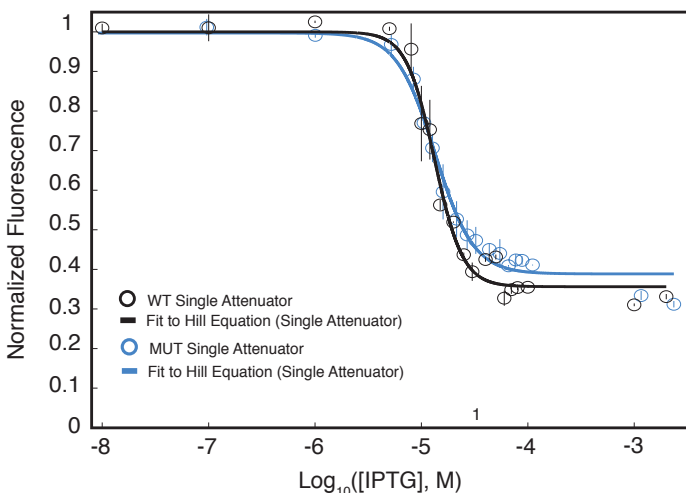


**Figure 3.4.1 Loop and swap mutations act synergistically to reduce attenuation crosstalk.** (A) Wildtype antisense RNA and sense RNA. Loop (L) and swap (S) mutations are highlighted in colors. (B) Loop mutations and orthogonality test. (C) Swap mutations and orthogonality test. (D) Combined loop-swap (LS) mutations and orthogonality test. Loop or swap mutations by themselves have a relatively high degree of crosstalk, and a noticeable asymmetry with WT-antisense/MT-sense showing less crosstalk than the converse. The combined loop-swap mutation brings crosstalk to near 0% with a much more symmetric response. This indicates that the combined mutations act synergistically to prevent attenuation when mixed antisense/sense pairs are used. (E) A three by three matrix of mutually orthogonal attenuators.

### 3.4.2 Similar functions of engineered orthogonal variants

A central aspect of our design strategy is to use a minimal number of mutations that change specificity without altering attenuation characteristics. To investigate the attenuation function of these engineered orthogonal variants, we measured the induction curves for both WT-T4 (wild type attenuator with T4 intrinsic terminator) and LS-T4 attenuators under various antisense

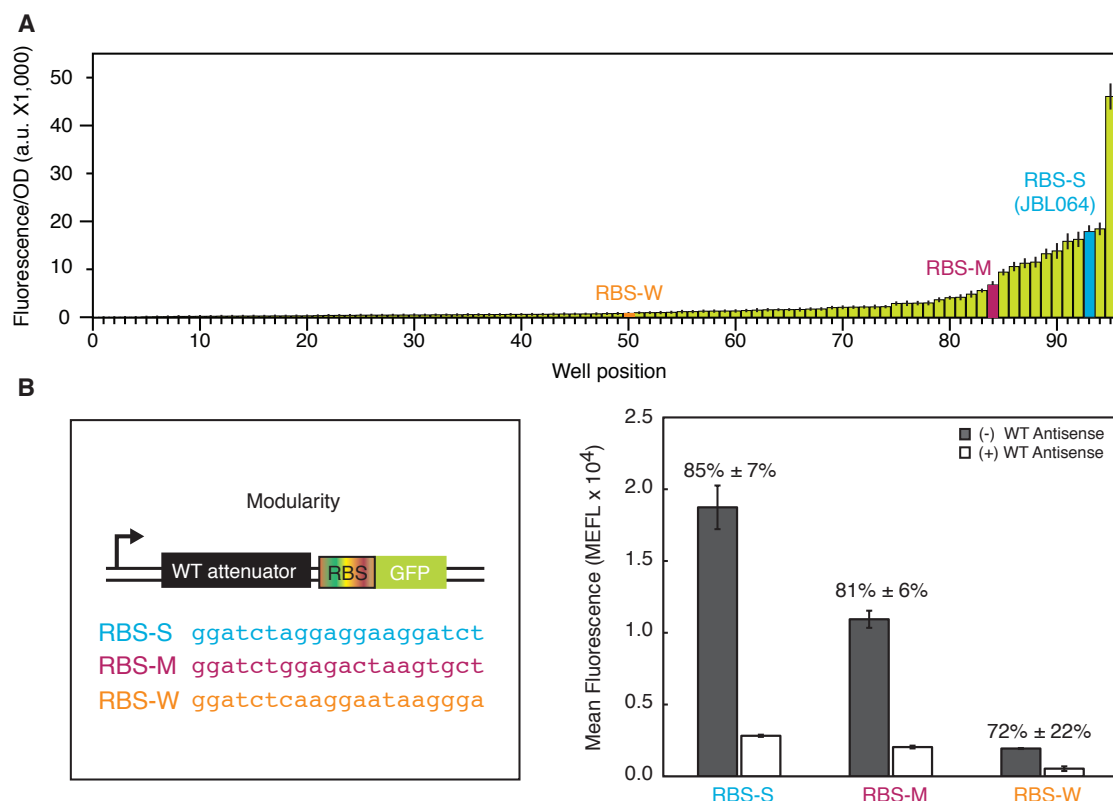
concentrations using the  $P_{\text{LacO-1}}$  promoter (**Figure 3.4.2**). We observed a high degree of similarity between each corresponding point of the WT-T4 and MT-T4 induction curves, implying the similarity in attenuation response to a wide range of antisense concentrations. These engineered orthogonal systems thus possess similar functions with different target specificity.



**Figure 3.4.2 Induction curves using an inducible promoter for WT (black) and MT-LS (blue) attenuators.** Measured fluorescence was normalized to that without IPTG for each attenuator (MEFL: single WT 12725, single MT 14561). The data are fitted to the Hill equation. The similar induction curve of both WT and MT attenuators demonstrates the homogeneity of their functions.

### 3.4.3 Modularity between transcription with translation

The attenuation system has been shown to act solely on the transcriptional level, which only changes the fraction of attenuated versus full-length transcripts and does not affect the translation of the latter<sup>106</sup>. Therefore, the RBS that regulates the translational initiation of the downstream CDS should be a modular component that can be tuned to change the absolute protein levels in the ON and OFF states, while preserving the repression percentage. To test this hypothesis, we randomly mutated a 14-nt region of the RBS, and assayed the fluorescence of 94 mutants without antisense. From the plotted expression rank order (**Figure 3.4.3A**), we chose three RBSs, a strong RBS (RBS-S) used in previous experiments and two successively weaker RBSs (RBS-M, RBS-W), and measured fluorescence with and without antisense (**Figure 3.4.3B**). All RBSs preserved the repression percentage within error bars, while providing almost an order of magnitude difference in ON levels. This demonstrates the decoupling of transcription and translation regulation in the engineered platform, as well as the ability to tune RBS to optimize the attenuation system for use in many applications<sup>114,115</sup>.

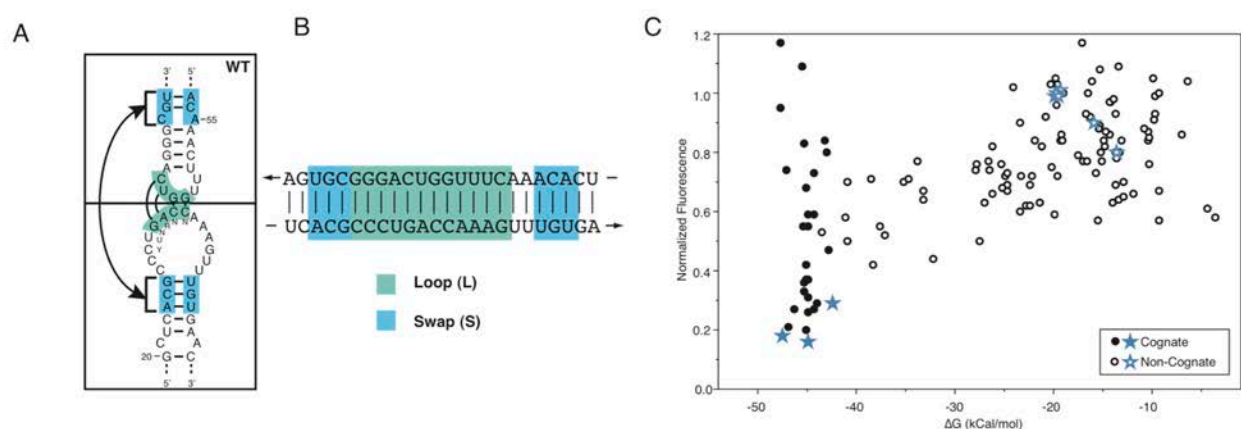


**Figure 3.4.3 Modularity of transcriptional control with RBSs.** (A) Screening of RBS strength. Fourteen nucleotides immediately downstream of the BioBrick scar site (GGATCT) are mutated, and their expression without antisense RNA is measured. Three RBSs, strong (RBS-S), medium (RBS-M), and weak (RBS-W), are chosen for measuring with antisense RNA. (B) RBS sequences for RBS-S, RBS-M, and RBS-W. (C) While absolute ON and OFF fluorescence levels decrease, transcription attenuation levels (shown as percentages above bars) stay in the same range and fall within error bars.

### 3.4.4 Kinetic properties of RNA-RNA interaction

It has been shown that the pT181 attenuation mechanism is based on an RNA-RNA kissing hairpin interaction<sup>106</sup>. It is also well-known that the regulatory decision for many antisense-mediated control systems occurs on timescales faster than the time needed for full antisense-attenuator duplex formation, and that in general the efficiency of these mechanisms are determined by the rate of antisense RNA binding. Nevertheless, it is reasonable to hypothesize that the thermodynamic binding free energy between a subset of the antisense-attenuator duplex may serve as a proxy for efficiency of attenuation, with lower free energy indicating stronger attenuation and thus lower observed fluorescence. To test this hypothesis, we first selected a region of the antisense-attenuator duplex known to form a hybrid 4-way junction in a related antisense-mediated system, CopA/CopT<sup>111</sup> (Figure 3.4.4A&B). Following our orthogonal mutant design principles, we created 29 mutant antisense/attenuator pairs by mutating bases in either the loop or swap regions previously identified. We calculated the free energy of each of these pairs, plus 110 non-cognate mixed pairs, along with the measured normalized fluorescence. Our results showed a weak trend between lower free energy and lower fluorescence as expected

(Figure 3.4.4C). This trend is even clearer for the WT, LS and LS2 attenuator/antisense pairs. However, the imperfect orthogonality between the WT and LS2 systems is not explained by this trend with the two highest free energies calculated for the 9x9 matrix (Sense WT-T4/Antisense LS2 = -15.9 kcal/mol, Sense LS2-T4/Antisense WT = -13.6 kcal/mol) corresponding precisely to the non-zero cross-talk values. This indicates that there are other, possibly structural, principles governing orthogonality. For cognate pairs, there is a very large spread in fluorescence observed over a relatively narrow range of free energy. This most likely reflects that certain mutations disrupt fundamental structural aspects of the pT181 system, which inhibits the mutants to attenuate even in the presence of their cognate antisense. We conclude that there is a loose association between thermodynamic binding free energies and observed attenuation. This is most likely reflective of the fact that this mechanism is driven primarily by the kinetics of antisense-attenuator interaction. Our conclusion is that while thermodynamics can guide the design, it does not completely predict the functions.



**Figure 3.4.4 Calculation of antisense/attenuator hybridization energy.** (A) Schematic of the kissing hairpin interaction between WT antisense and WT attenuator, highlighting the used in our mutational strategy. (B) The 24-nt hypothetical antisense (top)/attenuator (bottom) complex used in the thermodynamic binding energy calculations. (C) 29 cognate antisense/attenuator pairs are constructed by mutation and assayed for observed fluorescence. These values are plotted versus the predicted thermodynamic binding energy in the hypothetical antisense/sense complex as calculated by the DuplexFold program of the RNAstructure suite of programs<sup>66</sup>. Similar values are plotted for 110 non-cognate pairs. Stars represent values included for the WT, LS and LS2 pairs.

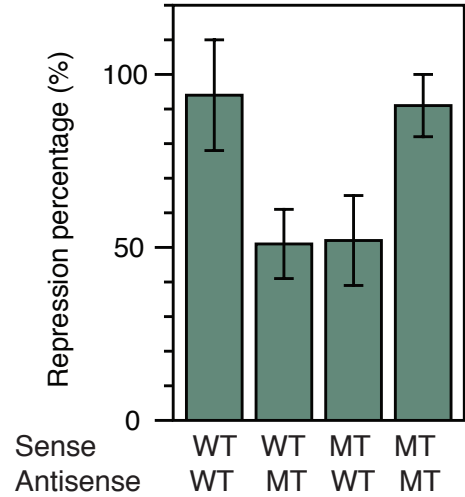
### 3.4.5 Orthogonality is not preserved in translational fusions

As described before, the WT pT181 system controls both transcription and translation. The question is when we translationally fuse the reporter gene to the engineered orthogonal variants, whether their orthogonality retains. To test this, we fused sfGFP to the 32<sup>th</sup> amino acid of the repC fragment to both WT-T4 and LS-T4 attenuators. Initiation of sfGFP translation starts at the repC RBS which is embedded in the intrinsic terminator. Measured fluorescence showed that, while both attenuators exhibit more than 95% repression, their crosstalk increases to almost 50% (Figure 3.4.5). This suggests that the translational control amplifies the transcriptional control

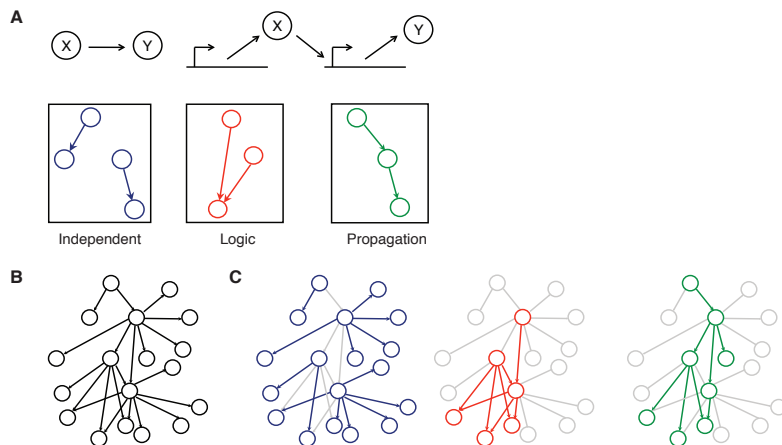
equally for cognate and non-cognate interactions, and affects the orthogonality. It remains a problem how orthogonality can be designed into the translational control.

### 3.5 Engineering Transcriptional Attenuator-Based Synthetic Circuits

The pT181 system controls the transcriptional elongation, and links cellular inputs to the RNA polymerase processivity during RNA synthesis. In contrast to the regulators of translational initiation, regulators of transcriptional elongation control both the production of coding and non-coding RNAs and can act on entire operons containing multiple genes. Furthermore, they are composable because when multiple transcriptional attenuators are linked in tandem in a 5'-UTR, the synthesis of the  $N^{\text{th}}$  attenuator is gated by the decision of the  $(N-1)^{\text{th}}$  attenuator; this predictably yields logic and higher-order functions. Thus, transcriptional controllers carry potential abilities to form fundamental regulatory motifs that lead to higher-order circuits. In this study, we investigate the use of orthogonal pT181 antisense-attenuator pairs to form three basic motifs. The first one is independent control, where two orthogonal antisense RNAs controls two genes independently in the same cell. Repression of one gene does not affect the other. The second motif is logic control, where two attenuators are assembled together in tandem to form a composite attenuator region. Since each attenuators exhibit a NOT function, this composite attenuator exhibit a NOR function using two antisense RNAs as inputs. The third motif is cascading control, where one attenuator, instead of controlling a downstream coding gene, controls the transcription of a second orthogonal antisense. This antisense can control its target, thus propagating the signal along the RNA cascade. As shown in **Figure 3.5**, we argue that these three motifs can theoretically be used to build an arbitrary gene network, which comprise a necessary and sufficient set of fundamental regulatory motifs.



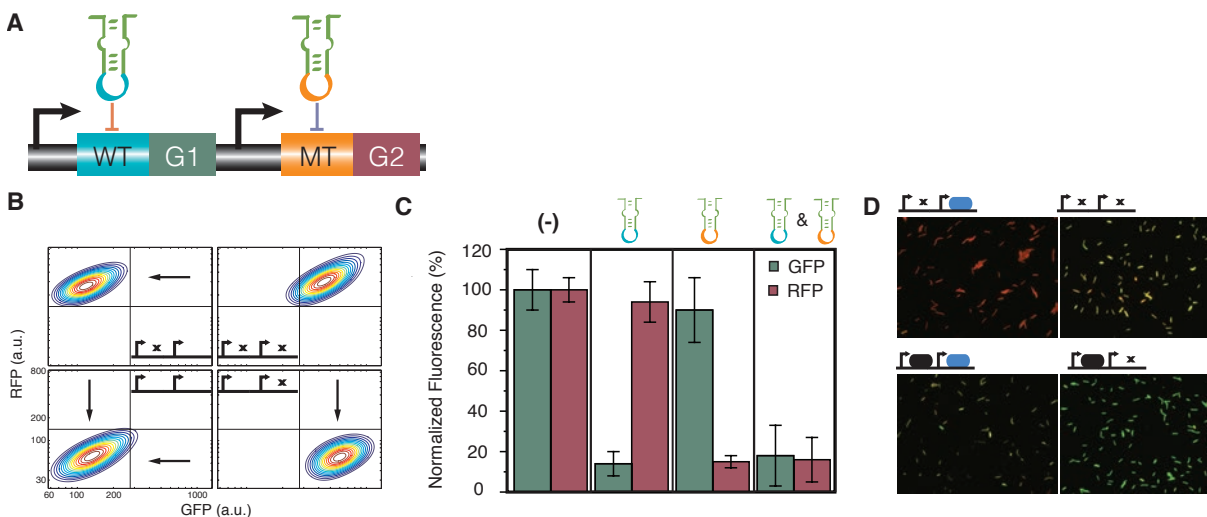
**Figure 3.4.5 Translational fusion show high level of crosstalk.**



**Figure 3.5 Random transcriptional networks can be constructed from basic motifs.** (A) Three fundamental network motifs. (B) A random transcriptional network. (C) Decomposition of the random network into three motifs.

#### 3.5.1 Independent control motif

Orthogonal antisense RNA-attenuator pairs, WT-T4 and LS-T4, are used to control two fluorescent proteins, mRFP and sfGFP in the same cell (**Figure 3.5.1A**). Two-color flow cytometry assay and microscopy (**Figure 3.5.1B-D**) both show that there is only minor crosstalk (6% - WT antisense, 12 % - LS antisense) when the attenuators are simultaneously used in the same cell. These crosstalk values are consistent with the previous orthogonality results. It is conceivable that with more orthogonal pairs, a large number of genes can be controlled independently and simultaneously.

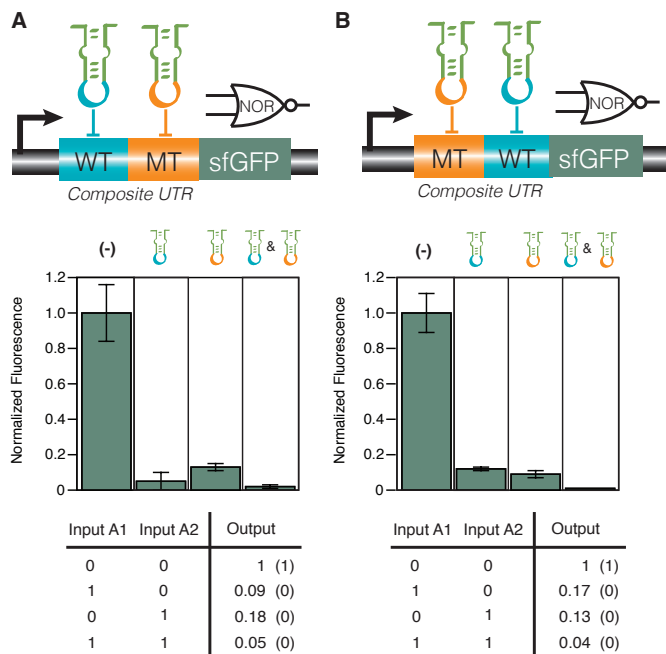


**Figure 3.5.1 Independent regulation of two fluorescent genes in the same cell.** (A) mRFP and sfGFP are controlled by the WT and MT attenuators in the same cell. (B) Representative 2-color flow cytometry percentile contours (darkest blue - 75% cells, red - 5% cells) for the four combinations of WT and MT antisense (inset - X represents no antisense). Arrows denote changes of the cellular density location in the RFP vs. GFP plane that indicate orthogonal changes in gene expression when different combinations of antisense are expressed. Fluorescence intensity is indicated as arbitrary units (a.u.) as determined by flow cytometry. (C) The bar plot shows the mean values of three biological replicates for each combination. (D) Representative microscopy images of cell cultures showing the expected changes in fluorescence intensity with different amounts of antisense expressed: no antisense, bright yellow (high GFP, high RFP); WT antisense, bright green (high GFP, low RFP); MT antisense, bright red (low GFP, high RFP); WT+MT antisense, dim yellow (low GFP, low RFP).

### 3.5.2 Logic control motif

Inspired by naturally occurring tandem riboswitches<sup>116</sup> and engineered tandem arrays of aptazymes<sup>44</sup> that integrate multiple molecular signals, we physically fused two orthogonal attenuators, WT-T4 and LS-T4, in tandem upstream of sfGFP, and measured attenuation from cognate antisense expression. We expect two orthogonal attenuators in tandem could integrate two antisense signals and only allow gene expression when neither antisense is present. This not-or-like (NOR) gene expression logic is indeed what we observed (**Figure 3.5.2A**). The similar NOR function was observed when we swap the positions, for example, LS-T4 was put upstream of WT-T4 (**Figure 3.5.2B**). Since NOR logic is functionally complete and can be layered

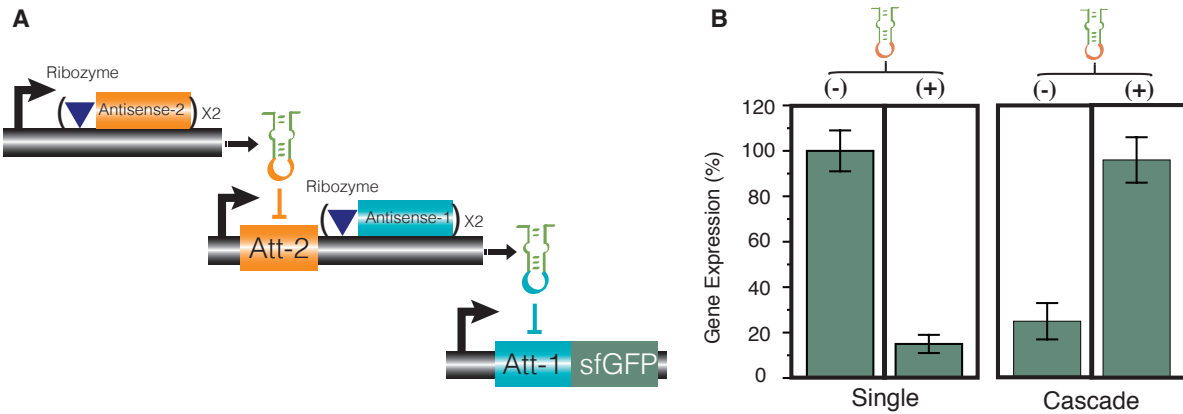
together to construct any other type of logic, our abilities to construct NOR gates has important implications for constructing higher-order synthetic devices<sup>117</sup>.



**Figure 3.5.2 Composed orthogonal attenuators exhibits NOR logical expression. (A) WT-MT composition. (B) MT-WT composition.** Data was normalized to 1 for the no antisense condition. The table shows the measured performance of the NOR logic gate, and the expected values for a perfect digital NOR gate.

### 3.5.3 Cascading control motif

The cascade is designed such that expression of sfGFP is controlled by Attenuator-1's (Att-1, WT-T4) interaction with Antisense-1 (Anti-1, WT), whose transcription is in turn controlled by Attenuator-2's (Att-2, LS-T4) interaction with Antisense-2 (LS) (**Figure 3.5.3A**). In this way, the antisense regulatory signal is propagated through a double inversion, which should produce a net activation of sfGFP expression. After trying several strategies as detailed below, we discovered that physical separation of the two regions on the same transcript using a self-cleaving ribozyme allows full function of the cascade. This hammerhead ribozyme is derived from small Tobacco Ring Spot Virus (sTRSV)<sup>118</sup>, and inserted into the region between Att-2 and Anti-1. The resulting three-level cascade could activate the sfGFP expression to 94% of its theoretical maximum (**Figure 3.5.3B**). While the functioning of the full three-level cascade is near optimal, the repression caused by the attenuator-ribozyme-antisense molecule is less than that caused by bare antisense. This could be due to the effects of attenuator auto-termination as discussed below. It should be noted that this three-level RNA regulatory cascade is the first of its kind reported, and is constructed simply by connecting our basic attenuators together.



**Figure 3.5.3 Engineered three-level RNA-mediated transcriptional cascade.** (A) The circuit design. The ribozymes is shown as the inverted triangle. The x2 symbol represents a tandem repeat of the module. The same promoter is used at each level. (B) The bar plot shows the expression of single attenuator (left) and cascaded attenuators (right) with and without Antisense-2. All fluorescence data was normalized to the single-level circuit value without antisense. The cascade shows 71% repression without antisense; with antisense, it is activated to 94% of the positive control level. Error bars represent the standard deviation from measurements of six biological replicates.

### 3.5.4 Discussions of engineering the cascading control motif

The construction of three-level transcriptional cascade has several important implications on RNA circuitry engineering. To create a functional cascade, it is necessary to compose Att-2 and Anti-1 on a single transcript in a way that Att-2 regulates the transcription of Anti-1. We tried two different approaches to address the problem. First, we inserted a linker sequence between Att-2 and Anti-1 to increase the distance between these two elements on the transcript. Second, we inserted a self-cleaving hammerhead ribozyme between the two elements, which is hypothesized to cleave and release Anti-1 to further propagate the regulatory signal. Each of these strategies is discussed below.

In the first strategy, we first tried to directly fuse Att-2 and Anti-1 on the same transcript. The 96-nt repC fragment was used the linker sequence between the two elements. We tested this fusion using our *in vivo* fluorescence assay on the Att-1-sfGFP reporter. The P<sub>L</sub>lacO-1 promoter was used to drive the expression of the Att-2-Anti-1 RNA fragment. Compared to expressing bare Anti-1, we noticed the fusion had a major decrease in repression on its target. We hypothesized that this decrease was caused by interference between Att-2 and Anti-1 because of their large degree of complementarity and their proximity on the same transcript. Using the secondary structure prediction algorithm Mfold<sup>65</sup>, we predicted that a direct fusion of Att-2 and Anti-1 would fold into a stable structure that likely diminished the activity of Anti-1 (Figure 3.5.4A). We therefore tried to increase the length of the linker sequence between Att-2 and Anti-1. Two longer lengths (417-nt, 876-nt) of repC CDS was used. As shown in Figure 3.5.4B, neither mitigated the intramolecular interactions between Att-2 and Anti-1. In fact, both linkers were found to cause nearly zero attenuation. However, this result does not rule out the possibility that a different choice of linker sequence could provide necessary insulation between Att-2 and Anti-1.



In an alternative strategy, a self-cleaving hammerhead ribozyme from sTRSV was inserted between Att-2 and Anti-1 (**Figure 3.5.4C**). The hammerhead ribozyme is assumed to help spatially insulate Att-2 and Anti-1 by directly cleaving the transcript between these two elements, releasing Anti-1 to act on the Att-1 target. Upon cleavage, the ribozyme leaves a 46-nt fragment<sup>119</sup> at the 5' end of the antisense RNA. It is possible that the extra fragment could interfere with the basic antisense-attenuator interaction to prevent attenuation and cause crosstalk between WT and LS antisense-attenuators pairs. To rule out this possibility, we repeated the single-color orthogonality experiment with the ribozyme sequence inserted to the 5' end of the antisense RNAs (**Figure 3.5.4D**). Our results show that the ribozyme preserves overall orthogonality, and even slightly improves attenuation effects between cognate antisense-attenuator pairs.

Next we tested the effectiveness of the ribozyme placed in between Att-2 and Anti-1 (**Figure 3.5.4E**). We observed that the ribozyme strategy yielded 54% repression percentage, which is a 23% improvement than without. However, this is still not as good as bare Anti-1, which shows more than 80% repression. There could be other factors causing the strategy to be sub-optimal: the ribozyme may not be efficiently cleaving the transcript; the ribozyme may be cleaving but is causing Anti-1 to have reduced activity due to misfolding or destabilization; or there is some inherent inefficiencies associated with placing Att-2 upstream of Anti-1 that reduces the transcription of Anti-1.

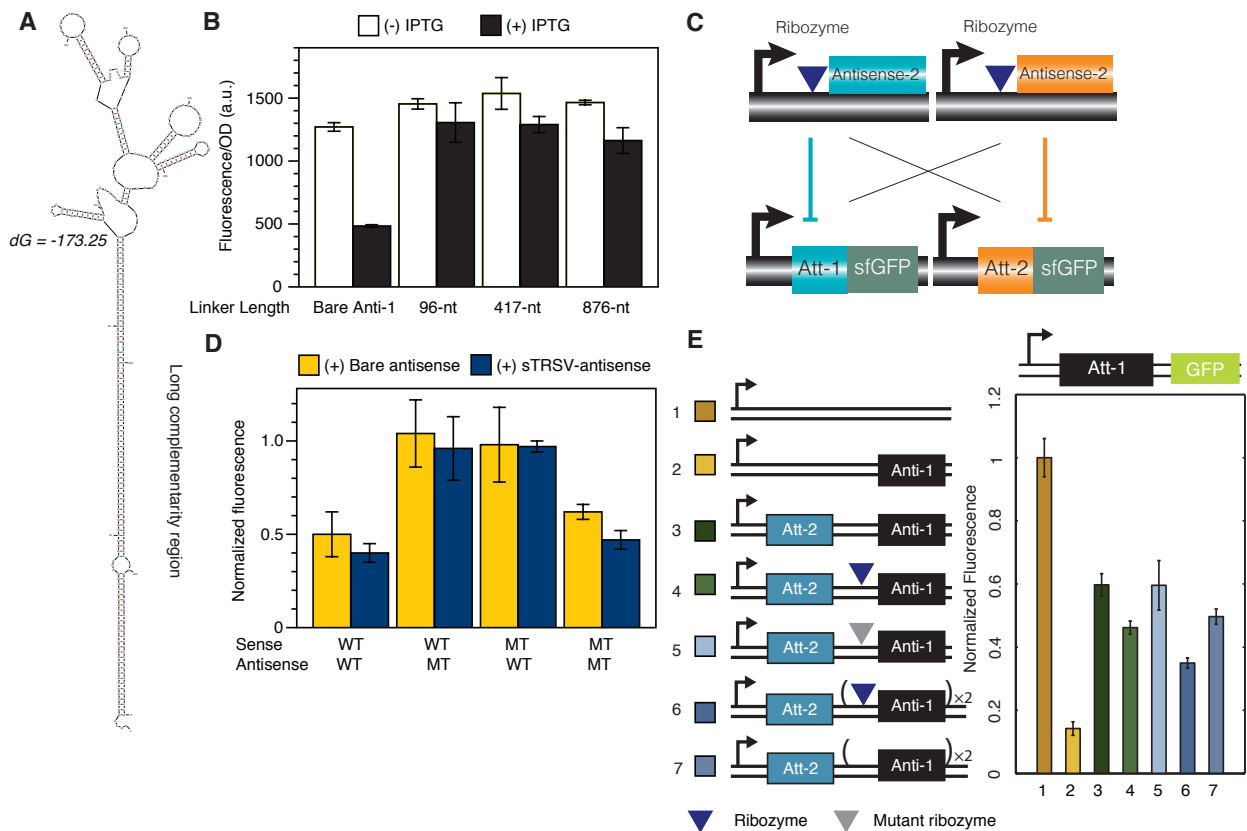
To test whether the inefficiency could be caused by the ribozyme not cleaving, we introduced a point mutation in the ribozyme that is known to silence its cleaving activity<sup>119</sup>. With this mutation, the attenuation level decreases to that without the ribozyme within error, indicating that the ribozyme is indeed active (**Figure 3.5.4E**). In addition, we created a construct with two copies of the same sTRSV-Anti-1 element in tandem, which showed a further increase in attenuation. In fact, this construct shows a 41% improvement over that without the ribozyme, nearly double the effect observed for a single sTRSV-Anti-1 element. Furthermore, removing the ribozymes from the tandem construct decreases attenuation. All of these observations indicate that the ribozyme is cleaving in this context.

Considering the second possibility, we examined the first pair of bars in **Figure 3.5.4D**. It is clear that the sTRSV-Anti-1 represses the target similarly as the bare Anti-1. First, this demonstrates that the ribozyme cleavage does not inactivate Anti-1. Second, since we observed that the amount of attenuation is a function of antisense concentration (**Figure 3.2C**), it follows that the concentration of antisense for these two constructs is approximately equal in these experiments. Under the simplest model of antisense production, the steady state antisense concentration should be equal to  $k_{\text{transcription}}/k_{\text{degradation}}$ , where  $k$  is a rate constant. Since the same promoters are used (making  $k_{\text{transcription}}$  identical), it then follows that the ribozyme cleavage does not substantially alter the antisense stability.

The remaining possibility then is that the inefficiency is associated with Att-2 being at the beginning of the transcript. As discussed below, we estimate that a single attenuator placed at the beginning of the transcript only allows 41% of the downstream transcript to be produced because of auto-termination. Thus the Att-2-sTRSV-Anti-1 construct is estimated to produce 41% of the antisense concentration that the bare Anti-1 construct produces. Since the bare Anti-1 construct provides 86% attenuation, an upper estimate of this effect predicts that the Att-2-sTRSV-Anti-1

construct should provide 35% attenuation, which is consistent with the 54% attenuation observed. While we cannot rule out other sources of inefficiency of the Att-2-Ribozyme-Anti-1 construct in attenuating an Att-1-sfGFP target, it appears that the dominant source is associated with the auto-termination effects of the Att-2 element limiting the amount of active Anti-1.

The attenuation of Anti-1 on its target Att-1 depends on their relative concentration ratio<sup>106</sup>. In a three-level cascade, random termination events, imperfect performance of the hammerhead ribozyme and misfolded antisense RNA could all lead to the loss of functional antisense RNA levels. To achieve a better performance of the cascade, an efficient signal amplification process is required to supplement the losses of propagated Anti-1 signal. However, unlike protein signaling, in which RBS and translational events naturally serve as amplification processes of the transcriptional events, no translational events are involved in the RNA-only transcriptional cascade. Here we demonstrate a tandem array of antisense RNA Anti-1 could serve as an amplifier as shown in **Figure 3.5.4E**. The tandem array is separated by the sTRSV hammerhead ribozyme that facilitates insulation between multiple antisense RNAs. This is confirmed by removal of the sTRSV that decreases attenuation of Anti-1. Thus, the strategy to tune and amplify the antisense signal by duplicating the sTRSV-Anti-1 module on the transcript in tandem increases attenuation.

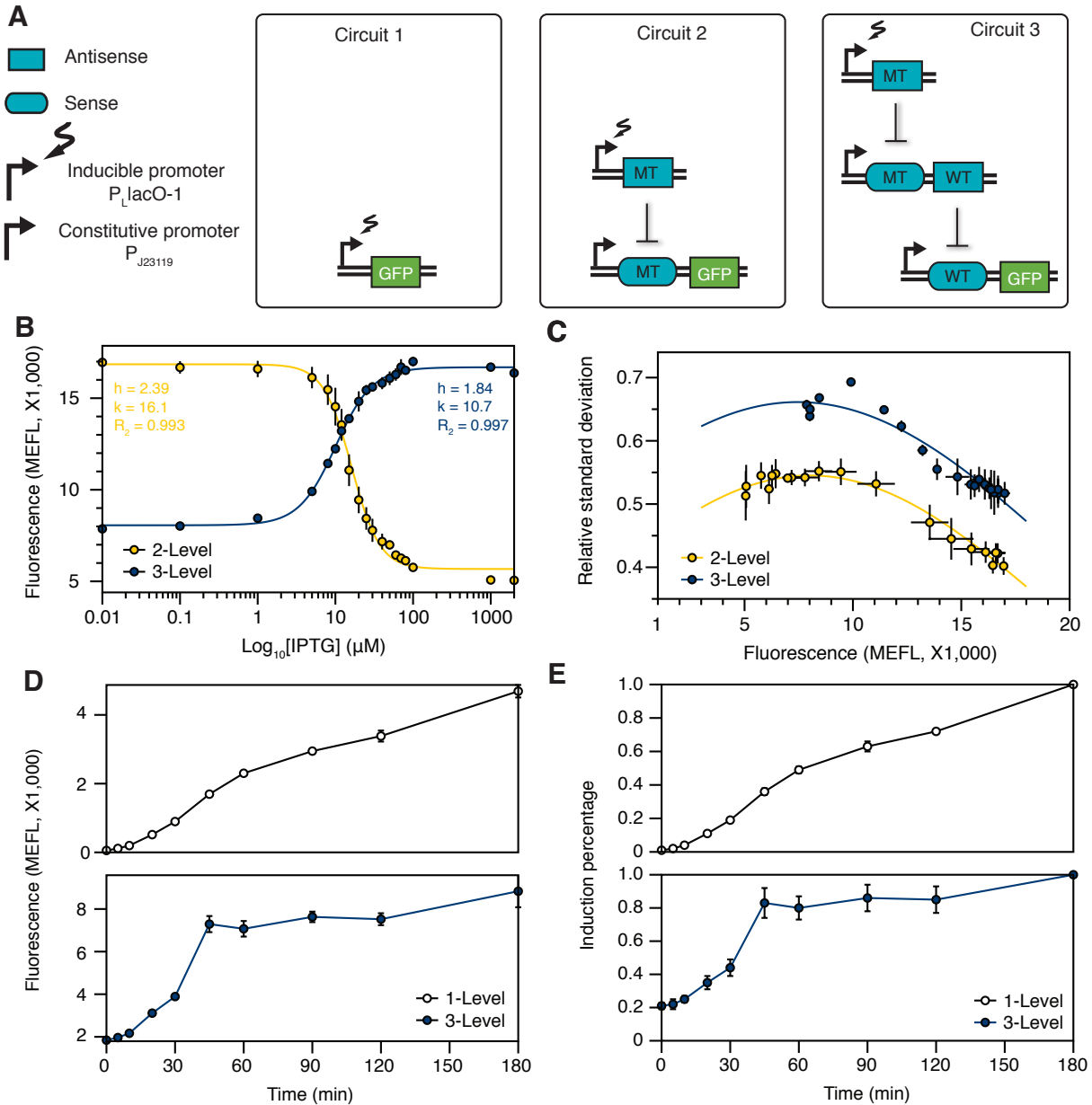


**Figure 3.5.4 RNA cleavage is critical to the cascade function.** (A) Lowest free energy structure of the 96-nt linker construct as predicted by Mfold. (B) Observed attenuation for the linker strategy. The  $P_{lacO-1}$  promoter is used to drive transcription of different antisense constructs, which attenuate a J23119-Att-1-sfGFP. Measured fluorescence divided by  $OD_{600}$  is

shown with and without 1 mM IPTG for different linker lengths. Bare Anti-1 is antisense transcribed without an upstream attenuator. Error bars represent standard deviation from triplicate measurements. The WT Att-1 sequence without terminator mutations is used in these experiments. (C) The single-color orthogonality experiment is repeated with the hammerhead ribozyme (inverted blue triangle) inserted before the antisense RNAs. (D) Average normalized fluorescence of the mixed (WT, MT) antisense/attenuator pairs shows a similar pattern of the orthogonality observed without the ribozyme. Furthermore, the ribozyme-antisense fusion has the same attenuation effects compared to the bare antisense within error. (E) The ribozyme insulation strategy. The J23119 constitutive promoter is used to drive transcription of different antisense constructs. Data is normalized to 100% for the positive control without antisense RNA. 1, No antisense; 2, bare Anti-1; 3, direct composition of Att-2 and Anti-1; 4, insertion of the sTRSV hammerhead ribozyme between Att-2 and Anti-1; 5, same as 4, but with a silencing point mutation<sup>119</sup> introduced into the wildtype hammerhead ribozyme (grey inverted triangle); 6, a tandem array of sTRSV-Anti-1 was used as an amplification strategy; 7, the tandem array of Anti-1 without sTRSV.

### 3.5.5 Properties of RNA cascades

Previous studies have shown that the protein-based transcriptional cascades had higher sensitivity, longer time delay and noisier expression compared to single-level circuits<sup>74</sup>. It is important to investigate if RNA cascades possess these properties. To study this, we constructed three circuits and compared their expression. Circuit 1 contained a sfGFP gene under the control of IPTG-inducible  $P_{LlacO-1}$ ; Circuit 2 contained a single antisense-attenuator pair with the antisense RNA controlled by  $P_{LlacO-1}$ ; and Circuit 3 is the three-level cascade circuit, with the top level antisense controlled by  $P_{LlacO-1}$  (**Figure 3.5.5A**). To estimate the sensitivity, we measured fluorescence of Circuit 2 and 3 in response to different concentrations of IPTG. Fitting of the data to the Hill Equation showed that the sensitivity of Circuit 3 decreased (Hill coefficient = 1.84 compared to 2.39 of Circuit 2) (**Figure 3.5.5B**). This is very different from the protein-based cascades<sup>74</sup>. Since there is no cooperativity in antisense-attenuator interaction, one explanation is that cascading multiple such interactions would not increase the Hill coefficient. However, it is not clear why the Hill coefficient observed decreased. Next, we measured the noise property by flow cytometry. The relative standard deviation (RSD) of the fluorescence distribution was calculated, and plotted against the mean value at each induction point for both Circuit 2 and Circuit 3. Consistent with protein cascades, the expression noisy of RNA cascades increases compared to a single attenuator system (**Figure 3.5.5C**). Finally, we measured the time response. It was shown theoretically that the time delay of a cascade is mainly determined by the degradation of propagating molecular signals<sup>120</sup>. Since RNAs possess much faster decay rate (1~2 min) than proteins (more than 30 min), we expect the RNA cascades show faster induction. Indeed, this rapid response was observed by assaying the temporal expression of Circuit 1 and Circuit 3 (**Figure 3.5.5D&E**). Our results showed that the time for half-activation for Circuit 3 is 30 min, but is more than 90 min for Circuit 1. To sum, we conclude that RNA cascades don't show ultrasensitivity as protein cascades do; they also amplify noisy; and they exhibit much shorter time delay in propagating signals. These biophysical properties might account for the rare cases of RNA cascades reported so far. However, since they have faster response, it explains why many RNA regulators are discovered in the SOS pathways such as heatshock, viral defense, and metal metabolism. Further understanding of the mechanisms that lead to these properties of RNA cascades is needed.



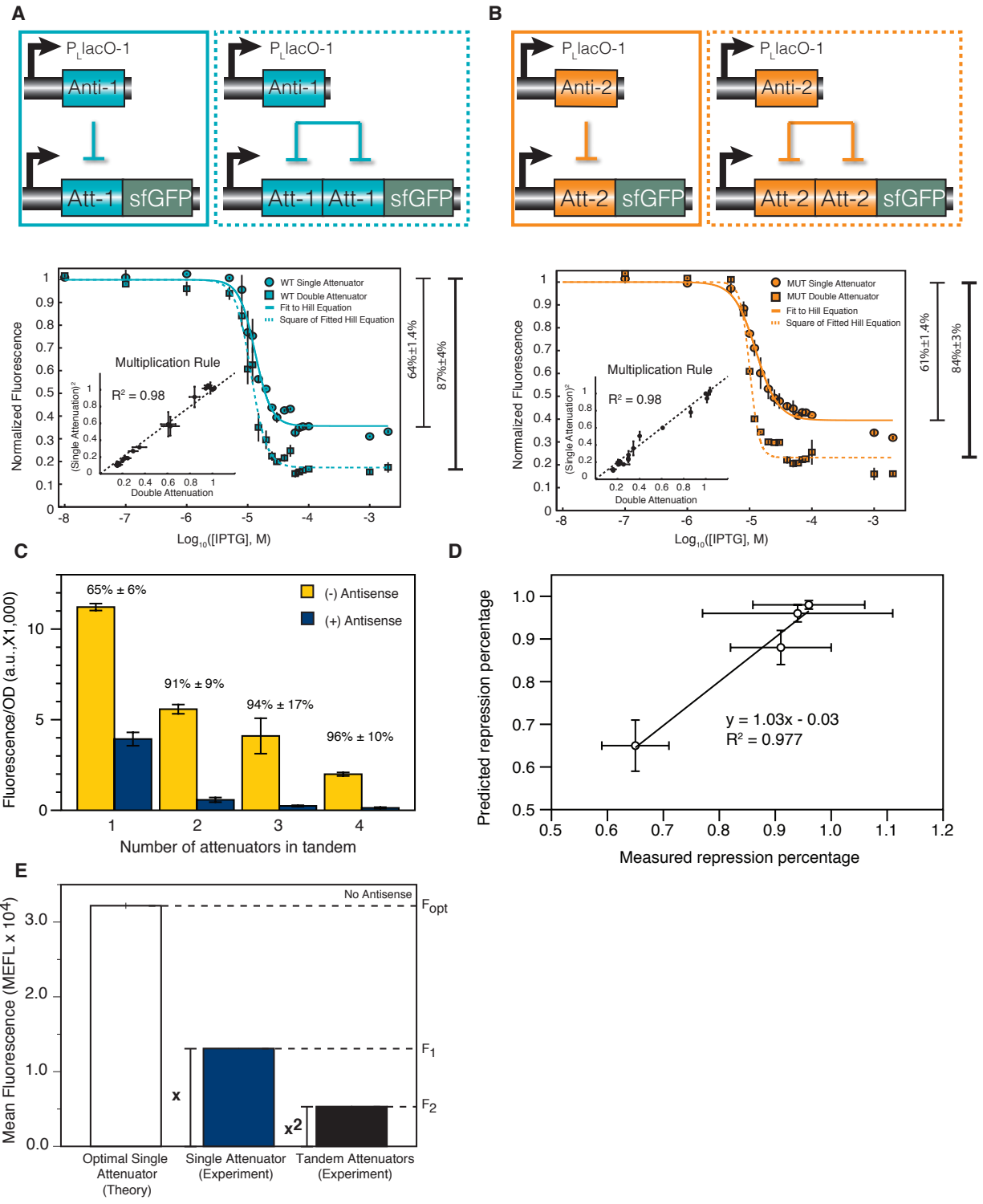
**Figure 3.5.5 Characterization of RNA-mediated cascade.** (A) Three circuits are measured and compared. (B) RNA cascade showed reversed function but with decreased cooperativity. The Hill coefficients measured are bigger than one because the  $P_{LacO-1}$  promoter is highly cooperative. (C) Characterization of noise. The cascade showed increased noise across populations. (D) The cascade circuit showed extremely fast response to external signals (supplemented with IPTG). (E) The data shows the induction percentages instead of absolute values of fluorescence. All experiments are done for biological triplicates.

### 3.5.6 Tandem identical *pT181* attenuators

We next compose two identical attenuators in series. We assembled two WT (WT-T4) (Figure 3.5.6A) or two MT (LS-T4) in tandem. The cognate antisense RNAs were induced from the

promoter P<sub>LlacO-1</sub>. Comparing to the single attenuators, the tandem attenuators show increased relative attenuation and steeper normalized induction curves (**Figure 3.5.6B**). To explain this effect, we hypothesize that attenuators in series function independently, much like the case for engineered tandem ribozyme devices. This implies that the overall attenuation of a tandem composite attenuator is a multiplicative function of the individual outputs. The insets plot this multiplication rule versus the observed double attenuation for each tested induction point. The fact that each value falls on a line of slope one shows that this simple multiplication rule is remarkably accurate for both WT and MT attenuators, reiterating their similarity of function. Then we compose three or four copies of the same WT attenuator in tandem, as expected, the trend continues, - the difference between ON/OFF increases with more attenuators composed together (**Figure 3.5.6C**); fitness between measured and predicted repression percentages shows almost perfect linear correlation (**Figure 3.5.6D**).

However, it is obvious that as more tandem attenuators are put together, the ON expression level drops. Due to the fact that the OFF level drops faster, the ON/OFF ratio increases as a consequence. This implies that even in the absence of antisense RNA, the attenuator exhibits auto-termination. Without antisense RNA, the attenuator must go through a dynamic structural transition to the terminator-sequestered state to allow further transcription<sup>106</sup>. Previous studies on the dynamic co-transcriptional folding of RNA ribozymes have indicated that these pathways are not deterministic and often become trapped in functionally misfolded states<sup>121</sup>. It is highly likely that the same is true of the pT181 attenuator, and that these misfolded states actually allow the terminator hairpin to form and thus cause auto-termination even in the absence of antisense RNA. **Figure 3.5.6E** shows the mean fluorescence of single and tandem attenuators in the absence of antisense RNA. If the attenuators were optimal and had no probability of auto-termination, we would expect these values to be nearly identical. However, in the presence of auto-termination, we expect the observed drop off due to the multiplicative auto-termination effect of tandem attenuators. This effect can be used to estimate the mean fluorescence of a single, optimal attenuator without antisense,  $F_{opt}$ . Let  $x = F_1/F_{opt}$  represent the fraction of RNA polymerase that passes through the sub-optimal attenuator. By the multiplication rule, the fraction of RNA polymerase that passes through the tandem attenuator,  $F_2/F_{opt} = x^2$ . Solving this equation, we find  $F_{opt} = (F_1)^2/F_2 = 32185$  MEFL, and  $x = 41\%$ . This estimation is likely an upper bound of the auto-termination effect since it does not take into account the limits of RNA polymerase processivity when transcribing the longer tandem attenuator sequence. Using the estimated RNA polymerase dropoff rate ( $= 10^{-3}$  per nucleotide) described in **Chapter 7**, we estimate  $x = (1-\lambda)^N \times F_1 \times F_2^{-1} = 0.55$ , and  $F_{opt} = 24007$  MEFL. This auto-termination effect does not affect the utility of our transcription attenuators in the basic circuit architectures since output levels can often be tuned with translational controls such as the ribosome binding site (**Figure 3.4.3**). Furthermore, it does not preclude more sophisticated circuit architectures as we have shown with the attenuator cascade. However, auto-termination is likely to reduce the concentration of Anti-1 in the cascade, and could be the main reason why the intermediate stage of the three-level cascade does not work as well as the construct that contained only one copy of sTRSV-Anti-1. Further optimization of the attenuator ON level then will involve studying the kinetic pathway by which the attenuator folds into its two alternative conformations to bias the fold to the read-through state in the absence of antisense RNA.



**Figure 3.5.6 Tandem identical attenuators.** (A-B) Induction curves of using the  $P_{LlacO-1}$  promoter to control single (circles, solid lines) and double (squares, dashed lines) attenuators in tandem. The wildtype (blue) and mutant (orange) tandem attenuators show similar attenuation over the full range of induction. Measured fluorescence was normalized to the case with [IPTG] = 0 for each attenuator (MEFL: single WT 12725, double WT 5744; single MT 14561, double

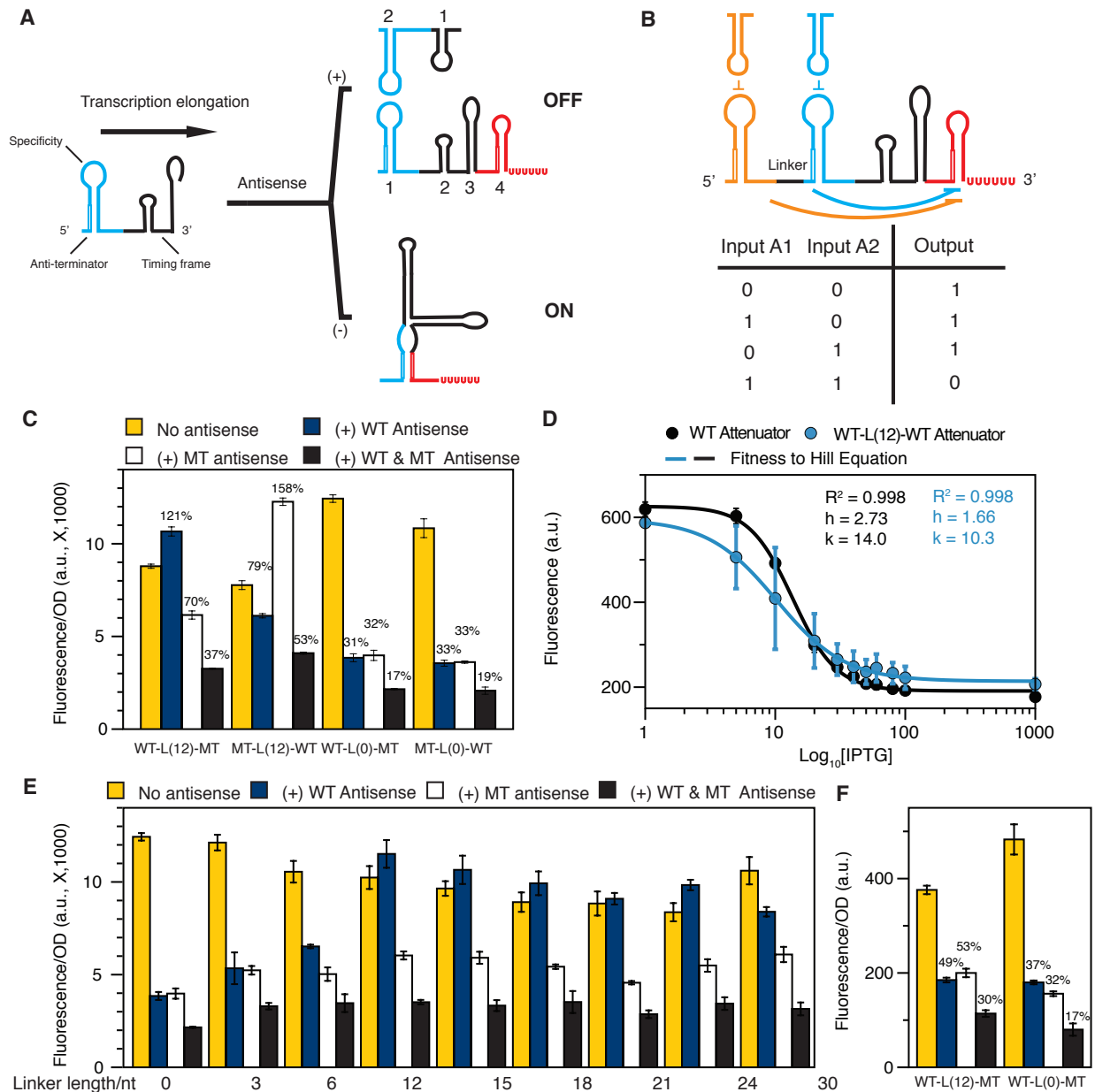
MT 5348). Hill equation fits to the single attenuator data are in solid curves. The dotted lines are calculated by squaring the fitted Hill functions. The inset plots show the comparison between the measured double attenuation and the square of measured single attenuation to verify the multiplication model. Error bars represent the difference between measurements on two biological replicates. (C) Assembly of 1~4 wild type attenuators in tandem. Data without antisense are in yellow, and those with are in blue. (D) Verification of multiplication rule for the tandem attenuators in (C). Fitness between measured and predicted repression percentages show a correlation of 0.977. (E) Estimation of auto-termination. Data show the experimental mean fluorescence for single (blue) and tandem (black) WT-T4 attenuators upstream of sfGFP without the antisense RNA. A theoretical estimation of fluorescence for an optimal attenuator without auto-termination is shown in white, and  $x$  represents the fraction of RNA polymerase that passes through the sub-optimal attenuator to produce a full transcript and thus sfGFP expression. By the multiplication rule discussed in the text, this fraction is  $x^2$  for the tandem attenuators.

### 3.5.7 Engineering NAND logics

The wild type pT181 attenuator consists of four hairpins (from 5' to 3', H1, H2, H3, H4) (**Figure 3.5.7A**). Previous studies have shown that, upon transcriptional elongation into H4, if there is no antisense RNA present, H1 refolds with part of the H4, which contains the intrinsic terminator, and promotes formation of a long stable RNA structure to precludes terminator formation<sup>106</sup>. Since H1 and H2 are complementary to the antisense RNA, and we have shown that the only hairpin 2 of antisense RNA is required for full attenuation, it is likely that only H1 of the attenuator (which binds to hairpin 2 of the antisense) is required for anti-termination activities. H2 and H3 of the attenuator are thus likely to provide a kinetic timing frame for the RNA-RNA interaction between the antisense and attenuator that allows transition between alternative RNA conformations. We further postulate that if we put two H1 with different binding specificities in tandem, we might be able to construct a NAND logic, which only repress the target when both antisense RNAs present (**Figure 3.5.7B**).

We started by fusing H1 of WT attenuator, WT(H1) to the 5' end of the MT attenuator, without linker (L(0), L for linker) or with a linker sequence of 12 nt (L(12)). Similarly, we fused H1 of MT to the 5' of the MT attenuator. We tested these fused attenuator constructs under different antisense RNA conditions. While the constructs with 12-nt linker region showed a NAND-like function, the constructs without the linker region showed a NOR-like function (**Figure 3.5.7C**). Moreover, WT antisense repressed the MT(H1)-L(12)-WT attenuator construct with higher efficiency than MT antisense. Similar results were obtained for the WT(H1)-L(12)-MT construct. We further fused WT(H1) to WT attenuator itself, and measured its induction curve with different IPTG concentrations to induce WT antisense expression. Compared to the wild type attenuator, the fusion construct showed reduced Hill coefficient (1.66 compared to 2.73), implying the two WT H1 fused together possess negative cooperativity. Note that the  $>1$  Hill coefficient reflects the cooperativity of the promoter used ( $P_{\text{LlacO-1}}$ )<sup>23</sup>. To study the effects on linker sequence on composite attenuator function, we systematically modified the linker length from 0 to 30-bp. Our results showed that with increasing linker length, the logic function gradually changed from NOR-like to NAND-like. More interestingly, the antisense RNA composition also determined the composite attenuator function. For experiments shown in **Figure 3.5.7C-E**, single hairpin 2 of the antisense was used. However, when full-length (hairpins 1&2) antisense was used, both WT(H1)-L(12)-MT and WT(H1)-L(0)-MT constructs showed

NOR-like functions (**Figure 3.5.7F**). This implies that the logic function not only depends on the interaction between H1's and H4, but also depends on the interaction between H1's and the antisense RNA. We note that all implemented NOR and NAND are not perfect, and need to be optimized for practical applications. Nonetheless, engineering the attenuator composition provide a useful method for programming of logic function, and is illuminating for further engineering with the pT181 system. We also point out that these complex logic functions are usually hard to obtain using screening or selection methods.



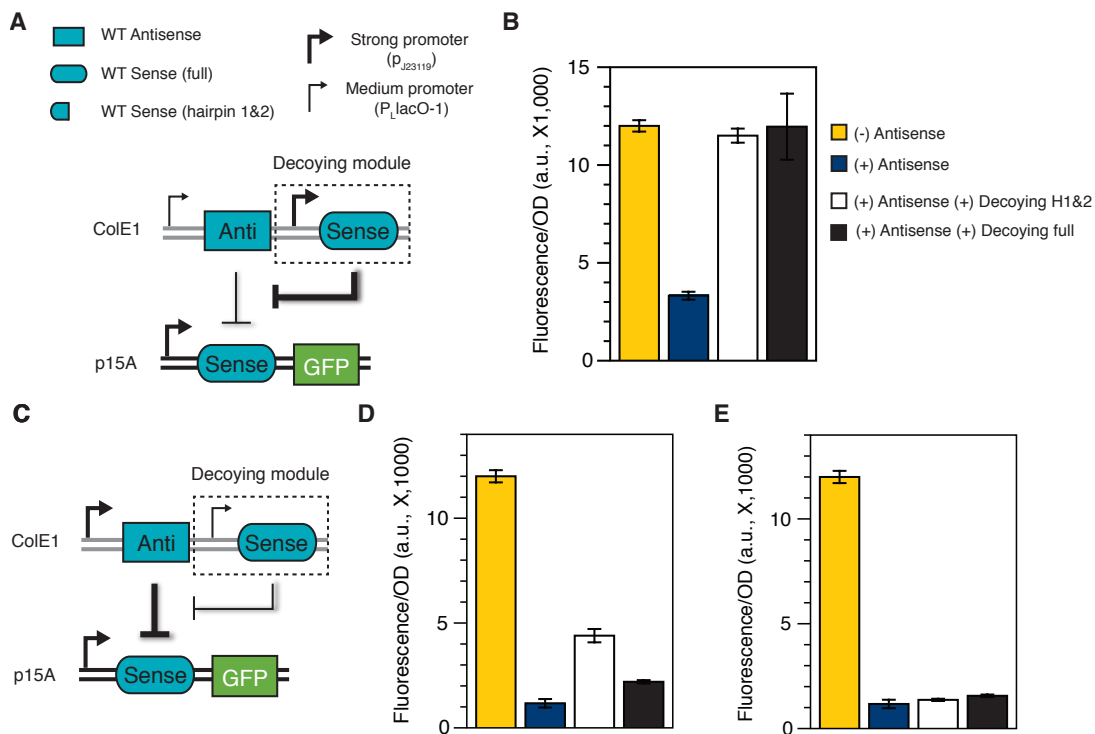
**Figure 3.5.7 Engineering NAND-like logics using the pT181 system.** (A) The mechanism of antisense-sense interaction. (B) Hypothesized NAND gate by engineering the sense UTR sequence. (C) Different lengths of the sequence between H1 and H2 lead to different functions. (D) The fusion construct shows reduced Hill coefficient (1.66 compared to 2.73), implying the



two WT H1 fused together inhibit each other. (E) Systematic modifications of the linker length show that the logic function gradually changes from NOR-like to NAND-like with increasing linker length. (F) When the full-length (hairpins 1&2) antisense RNA is used, both WT(H1)-L(12)-MT and WT(H1)-L(0)-MT constructs show NOR-like functions.

### 3.5.8 Engineering decoy circuits

Decoying a useful in sequestering excessive signaling molecules, and has been widely used in other engineering disciplines. In our system, antisense RNA interacts with the sense target *in trans*. We ask if we can design decoy functions into the system. To do this, we placed an extra copy of sense RNA under the control of the strong J23119 promoter on the ColE1 plasmid (Figure 3.5.8A). The antisense is controlled by the inducible P<sub>L</sub>lacO-1, which is 3-fold weaker than J23119. We designed two copies of sense RNA, one contained the first two hairpin of the attenuator (H1 and H2), and the other contained all four hairpins. In both cases, we observed that the attenuation was completely removed when the decoy circuit was present (Figure 3.5.8B). It is conceivable that for effective decoying, the decoy sense needs to be in excess relative to the antisense. To confirm this, we swapped the promoters for the antisense RNA and the decoy sense, such that J23119 controls antisense and P<sub>L</sub>lacO-1 controls decoy sense (Figure 3.5.8C). In this case, presence of the decoy sense only partially eliminated attenuation (Figure 3.5.8D). An interesting observation is that the H1&2 version of the decoy sense is working more effectively than the full length for reasons unknown. To verify the reduction of attenuation came from decoy sense, we removed IPTG induction on the decoy sense. In both cases, the attenuation was restored, implying the role of the decoy circuit (Figure 3.5.8E). These decoy circuits are useful in engineering the responses curves of the pT181 system, and might be useful for creation of complex functions such as bimodality.

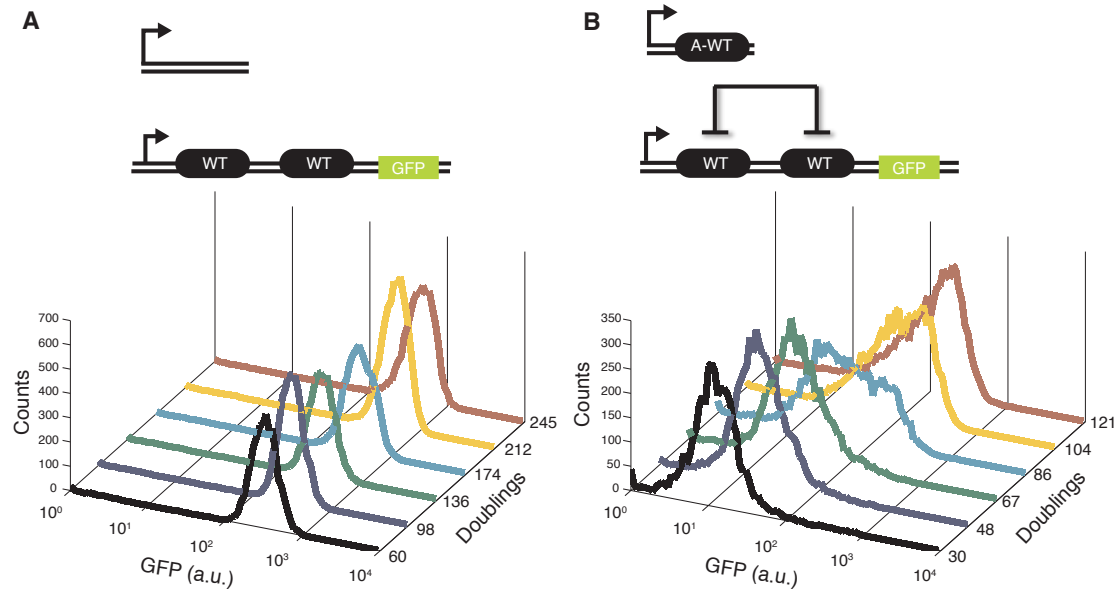


**Figure 3.5.8 Construction of decoy circuits.** (A) Decoy circuit design. (B) We observed that attenuation was removed when the decoying circuit was present. (C) We used an inducible promoter to control expression of the decoy sense. (D) Presence of the decoying sense partially removed the attenuation. (E) The attenuation was restored when we removed IPTG induction on the decoy sense.

### 3.6 Reliability of The Circuits Containing Tandem Attenuators

These tandem attenuators possess a large homologous region (~300 nt). Interestingly, they operate reliably over a number of generations, showing less susceptibility to genetic recombination than might be expected. Following the procedure of Canton *et al.*<sup>79</sup>, we measured the genetic and performance reliability of tandem WT attenuators without and with antisense RNA. **Figure 3.6** shows the progression of fluorescence distributions measured on successive days of continuous culture for these two combinations. There is no observable change in the GFP distribution with no antisense present for 245 culture doublings. Sequencing of the sense and antisense plasmids from the 245 doubling stock revealed only one out of the five colonies sequenced to have a 60bp deletion in between the two attenuators in the repC coding region, outside of the core attenuator sequences, and no changes in the antisense plasmids. Therefore both the performance reliability and the genetic reliability of the tandem attenuator in the absence of antisense RNA is more than 256. However, in the presence of antisense, there is a noticeable subpopulation shifted towards higher fluorescence starting at 86 culture doublings, which overtakes the culture by 121 culture doublings. Surprisingly, sequencing of both sense and antisense plasmids of the 121 doubling stocks indicated no genetic changes in either the tandem attenuator or the antisense RNA cassette. Thus the genetic reliability of the tandem attenuator in the presence of antisense RNA is more than 121 culture doublings, while the performance reliability is 86 culture doublings.

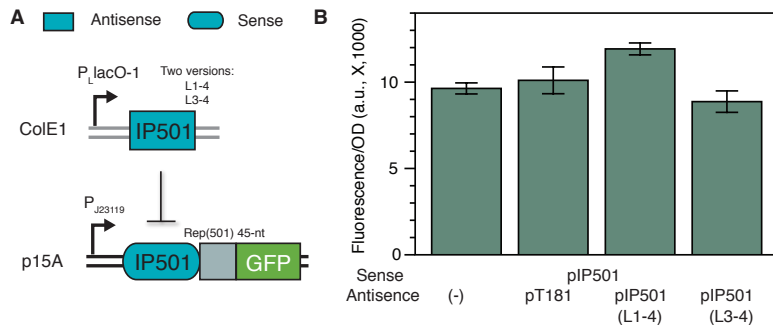
It is remarkable that despite the tandem arrangement of repeated 287-bp sequences in the attenuator, and the presence of 91-bp of complementary antisense RNA on a high copy plasmid, that there is perfect maintenance of the system on the DNA level for over 6 days of continuous culture growth. We noticed bimodality in some tandem attenuator constructs when they were grown in poorly aerated 96 well, 2 mL blocks. This bimodality could be caused by a number of factors including changes in average plasmid copy numbers, or general changes in cell physiology and the global gene regulation machinery. Since the genetics of the tandem attenuator with antisense RNA does not change over the timescale of this experiment, we hypothesize that the drift observed in the population fluorescence is due to a drift in the same physiological factors that caused our initial observation of bimodality. Since this drift occurs only in the presence of the antisense RNA plasmid, we further hypothesize that the effect is due to a drifting in the stable point for the copy number of the ColE1 plasmid used to express the antisense RNA. While imperfect performance reliability is not desirable, we note that the overall performance reliability of this sense and antisense combination is still better by 30 culture doublings (equivalent to 1.6 days of continuous culture) than that for a widely-used inducible promoter system characterized before<sup>79</sup>. Furthermore, unlike the previous work, there was no genetic recombination despite the tandem array of repeated sequence elements. Thus, the antisense RNA system represents an example of the uncoupling of performance and genetic reliability.



**Figure 3.6 Reliability of tandem attenuators.** *E. coli* cells containing the tandem attenuators upstream of sfGFP without or with antisense RNA were measured. Results show that Doubling times were calculated by successive OD measurements in the same conditions and were found to be 38 minutes (no antisense) and 77 minutes (with antisense). (A) Without antisense RNA; (B) with antisense RNA.

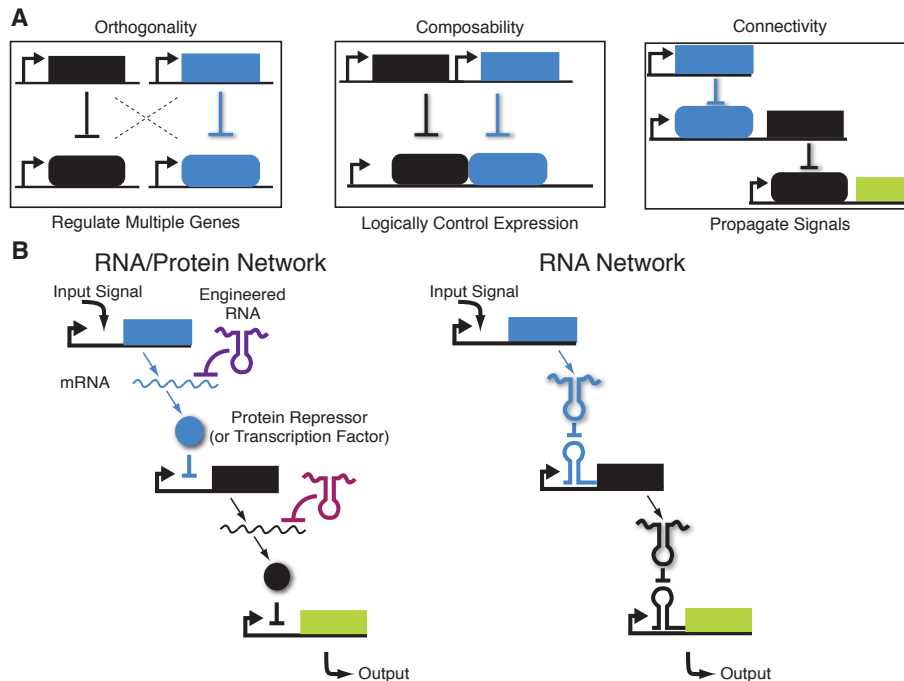
### 3.7 Other Transcriptional Attenuation Systems

Streptococcal plasmid pIP501 uses a similar antisense RNA-mediated mechanism to control its replication protein, repR<sup>107</sup>. However, the difference is that it also employs a transcriptional protein factor CopR, to repress the transcriptional initiation. The deletion of either control component causes a 10- to 20-fold increase in plasmid copy number<sup>105</sup>. Unlike pT181 antisense, the pIP501 antisense RNA has very long half-life, about ~30 min<sup>122</sup>. This exemplifies the important of the CopR protein repressor to balance the downward fluctuation of plasmid copy number. Further in vitro assays show that antisense-attenuator binding is sufficient for inhibition<sup>122</sup>. Only two hairpins of the antisense were needed to induce repression. Following these observations, we tested if the pIP501 system can be harness to control gene expression, and if this worked orthogonally to pT181. Similar to the pT181 system, the pIP501 antisense RNA is cloned under the IPTG-inducible P<sub>L</sub>lacO-1 promoter, and the attenuator together with a 45-nt repR fragment (15 amino acids) is cloned upstream of the fluorescent reporter gene (**Figure 3.7A**). A similar two-plasmid system, Cole1 and p15A, was used to express both. We tried two versions of the antisense RNA, - one contained the full-length antisense, and the other contained only the essential two hairpins. However, our system failed to exhibit repression. The pIP501 reporter system is always ON, even with the antisense RNA present (**Figure 3.7B**). It is not clear the problems with this synthetic platform, despite previous studies showed the system can be used to control gene expression *in vivo*<sup>107</sup>.



**Figure 3.7 Testing the pIP501 transcriptional attenuation system.** (A) The reporter system for testing pIP501. (B) The pIP501 reporter system is always ON, even with the antisense RNA present.

### 3.8 Strengths and Limitations of ncRNA Transcriptional Systems



**Figure 3.8 Utilities of the engineered orthogonal RNA transcriptional regulators.** (A) Schematic of three regulatory motifs that were implemented using orthogonal RNA regulators. (B) Cartoon of an example network (two-level transcriptional cascade) implemented as a hybrid RNA-protein network (left), or an RNA-only network (right). Both networks take a general input signal, propagate it through the cascade and ultimately create an output RNA signal that can be any non-coding, coding, or engineered RNA regulator. In RNA-protein hybrid networks, signal propagation requires the interconversion between mRNA and protein at each step of the network (arrows). In contrast, RNA networks that use regulators such as the attenuator in this work greatly simplify network designs by propagating all signals as RNA molecules, which feed directly into the next regulatory decision.

#### 3.8.1 Utilities of RNA-based transcription attenuators

We have demonstrated a design method to design orthogonally acting RNA-mediated transcriptional attenuators that can be configured to regulate multiple genes in the same cell,

logically control gene expression, and directly propagate RNA regulatory signals (**Figure 3.8A**). Each of these different functions can be achieved by a simple reuse and reconfiguration of the same set of orthogonal attenuators. There has been substantial work on other RNA-based regulatory mechanisms that can each perform some of these functions. However, the attenuators in this study provide the simplest route to achieving all of the functions within a single regulatory mechanism. The advantages are highlighted by considering two different implementations of transcriptional cascades as either hybrid protein-RNA networks, or purely RNA networks (**Figure 3.8B**). In the case of hybrid networks, any protein regulators must be translated from intermediate mRNAs, while in RNA networks this intermolecular conversion process is not required. This eliminates one molecular species with associated gene expression parameters such as half-life and maturation time, and one interconversion process for each network connection. Such benefits will be compounded as the number of network connections increase, and we believe the attenuators used in this study could become important components to simplify the design of large gene networks.

In addition to simplicity, the particular combination of functions displayed by the attenuators has recently been shown to be important for constructing complex genetic logics. As recently demonstrated using combinations of transcription factor logic gates and quorum sensing circuits to propagate the signals, NOR logic gates can be combined together in various ways to construct any type of logic function<sup>117</sup>. Since NOR-logic and signal propagation are two of the features of the attenuators, we anticipate our system to be useful when constructing computational circuits inside cells.

There may be other important applications of this system where the speed of signal propagation is a critical design requirement. The rate-limiting step in signal propagation through a cascade is the time required to degrade intermediate signaling molecules<sup>120</sup>. Since protein regulators often have half-lives greater than the doubling time, this is achieved by dilution through division, and protein cascades can only propagate as fast as one cell-cycle per step. Since the half-lives of the antisense RNAs that propagate the signals in this work are around 1~2 min, networks built out of the attenuators were shown to propagate signals faster. Faster transcriptional cascades may allow flexible programming of gene networks. These could be particularly useful if coupled to appropriate sensing mechanisms, such as two-component systems or riboswitches to control antisense production, to design fast responses to environmental signals. More work is needed to investigate the biophysical and physiological properties of RNA-based networks, such as the origin of propagated noise and the metabolic burden on hosts, and compares these properties with protein transcription factors and other RNA regulators.

### ***3.8.2 Expanding families of orthogonal regulators.***

Our fundamental approach to finding orthogonal regulators is to engineer them through rational mutagenesis of a carefully selected natural regulator. Previous studies that established the broad features of the pT181 attenuation mechanism placed great emphasis on the series of RNA structures responsible for antisense recognition and attenuation<sup>106</sup>. Therefore, as a design principle, we specifically focused on mutational strategies that would minimize disruption of the antisense and attenuator hairpin structures. In particular, the mutations of attenuator LS both conserve the YUNR motif in the loop of the hairpins, as well as the overall base-pairing pattern

in the stems of the hairpins. Our modeling efforts suggest that small perturbations in the sequence and structures of the antisense and attenuator RNAs can be tolerated, and that perhaps a larger sequence space could be sampled to find more orthogonal pairs. However, the fact that other RNA-mediated transcriptional attenuators have largely similar RNA structures suggests that there is a deeper structural principle to this type of gene regulation that is at the core of making quick decisions with RNA hairpin-hairpin interactions, and that completely arbitrary sequences would not yield functional attenuators<sup>122</sup>.

In this work we have begun to develop design rules for constructing more orthogonal attenuator/antisense pairs. While minimizing mutations is desired, we found that both loop and swap mutations are required for orthogonality. However, not every mutant that we created is orthogonal, or even functional, and there is still some degree of crosstalk between the LS2 and WT pairs. In previous work on translational regulation with engineered RNAs, it was found that lower thermodynamic RNA-RNA binding free energies were positively correlated with stronger repression<sup>86</sup>. However, in this work, we found only a loose correlation between lower calculated binding free energy and attenuation. Furthermore, there are many mutant cognate pairs that are predicted to have a low free energy of interaction but show very little attenuation. This suggests that these mutants may be misfolding, and that thermodynamic free energies can be used as a secondary design principle after the overall attenuator and antisense structures conform to the requirements of this system. More work is needed to uncover all the design principles behind orthogonality in this system.

### **3.8.3 Remaining challenges**

In addition to orthogonality, there are two specific remaining challenges associated with optimizing the attenuator ON and OFF levels, respectively. The attenuator ON level is determined by the strength of the promoter and the propensity for the attenuator to auto-terminate in the absence of antisense. Auto-termination manifests itself as the drop in fluorescence observed when two attenuators are placed in series, which can be used to estimate the amount of auto-termination due to a single attenuator to be 59%. When attenuators are used to control protein-coding regions that are later translated, these deficiencies can be compensated by tuning the strength of the RBS. However, in RNA-based circuits created by wiring together attenuators such as the cascade, auto-termination reduces the amount of antisense that can propagate the signal, and was found to be the likely cause for imperfect cascade performance. Auto-termination is likely a property of the dynamic refolding the attenuator undergoes as it is being transcribed, and decreasing it will probably require a deeper understanding of co-transcriptional RNA folding pathways.

Equally important is the OFF level attainable by the attenuator in the presence of antisense. The fact that we could improve this level by almost 100-fold with only 4 mutations suggests that the WT system is far from optimal in *E. coli* and more mutations along these lines could reduce the level further. Another limiting factor in the attenuator OFF level is due to the amount of antisense expressed. RNA-level measurements for this system show that the [antisense]/[attenuator] ratio is in the range of 3.8~9.7, confirming earlier work showing that antisense needs to be in abundance of sense for efficient attenuation. This has also been observed in previous work on engineering antisense-mediated transcription control, and may be a general feature of RNA-based regulation. Titration experiments also show that increasing the ratio

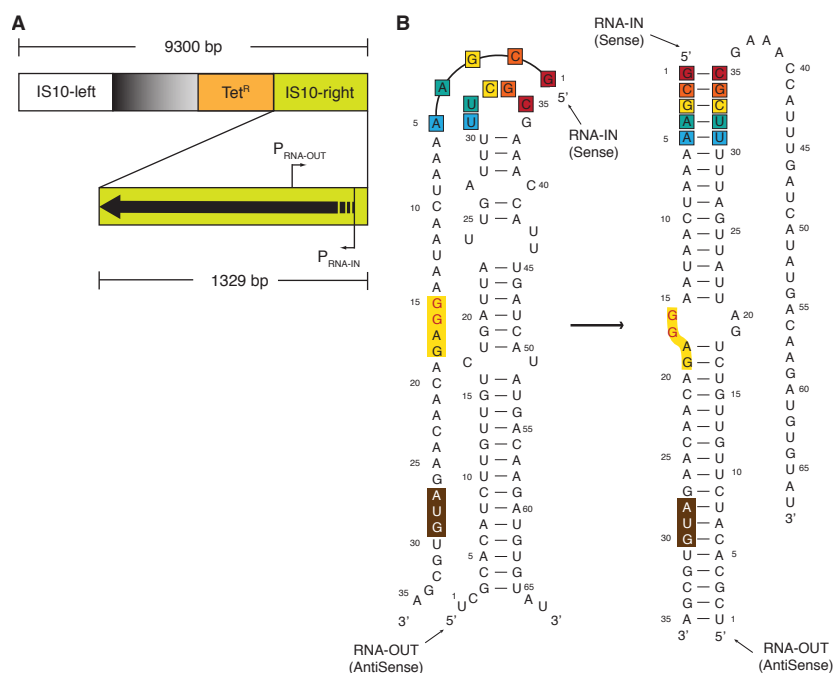
increases attenuation. Indeed, one way of improving the leakiness of attenuation, and thereby potentially some of the cell-to-cell variation that gives rise to the error bars in our experiment, would be to increase the strength of the promoter driving antisense transcription. We note that in our current configuration the [antisense]/[sense] ratio is achieved by the difference in plasmid copy number, and integrating our attenuation system into the chromosome or lower copy plasmids may require promoter tuning. Improving the OFF level in the presence of lower concentrations of antisense represents an important challenge in optimizing the system, and may be addressed by increasing the thermodynamic binding free energy between the attenuator and antisense.

Our engineering of the pT181 system to create orthogonally acting variants that can form diverse regulatory circuits adds to the growing repertoire of RNA synthetic biology. The work provides a versatile set of RNA-based transcriptional regulators that could change the way we think about designing and constructing gene networks. Our engineering strategy by constructing orthogonal variants of a natural RNA system with minimal changes so as to preserve overall function should be applicable to other gene regulatory mechanisms to further expand the diversity of genetic building blocks available. Furthermore, we anticipate that the ribozyme-mediated insulation strategy used in this work can be used as a general technique to compose diverse RNA regulatory elements on a single transcript, which could substantially increase the sophistication of RNA-based gene regulatory networks.

## Chapter 4 RNA-Mediated Translational Regulators

### 4.1 Introduction to The IS10 System

Insertion sequence IS10 is part of the *E. coli* composite transposon Tn10 (Figure 4.1A). The natural IS10 system regulates the copy number of Tn10<sup>123</sup>. As the Tn10 copy number increases, excessive antisense RNAs are produced, which repress the transposases gene and down-regulate the copy number. Similarly, up-regulation of the copy number occurs when the Tn10 copy number decreases. Thus, similar to the previous described pT181 system, IS10 antisense control exhibits similar physiological function. However, this function is implemented mostly on translation instead of transcription. IS10 encodes a single transposase gene whose expression is regulated by an IS10-encoded antisense RNA (RNA-OUT). RNA-OUT pairs with the 5' UTR sequence (RNA-IN) of the transposases mRNA over a 35-bp region of complementarity. The base pairing blocks RBS and the start codon, thereby preventing transposases gene translation<sup>123</sup>. RNA-OUT probably inhibits translation and triggers rapid degradation of the double stranded RNA by ribonucleases. Previous studies have shown that RNA-OUT forms a thermodynamically stable structure composed of a 21-bp stem region topped by a loop region<sup>124</sup> (Figure 4.1B). RNA-IN, on the contrary, remains largely unstructured. Further, it was shown that extra nucleotides appended to the 5' end of RNA-IN greatly reduce its regulation of the downstream gene<sup>125</sup>. More importantly, the interaction between RNA-OUT and RNA-IN is likely thermodynamic than kinetic. Thus, the interaction is probably more predictable based on the available thermodynamic calculation algorithms.



**Figure 4.1 Wildtype IS10 translational control system.** (A) Native insertion element IS10 controls translation of the transposase in transposon Tn10. (B) Schematic of sense RNA-IN and antisense RNA-OUT interaction. Colored boxes indicate the core recognition region between interacting species. The mutated SD region in RNA-IN mRNA is marked as yellow, and the AUG start codon is colored brown.

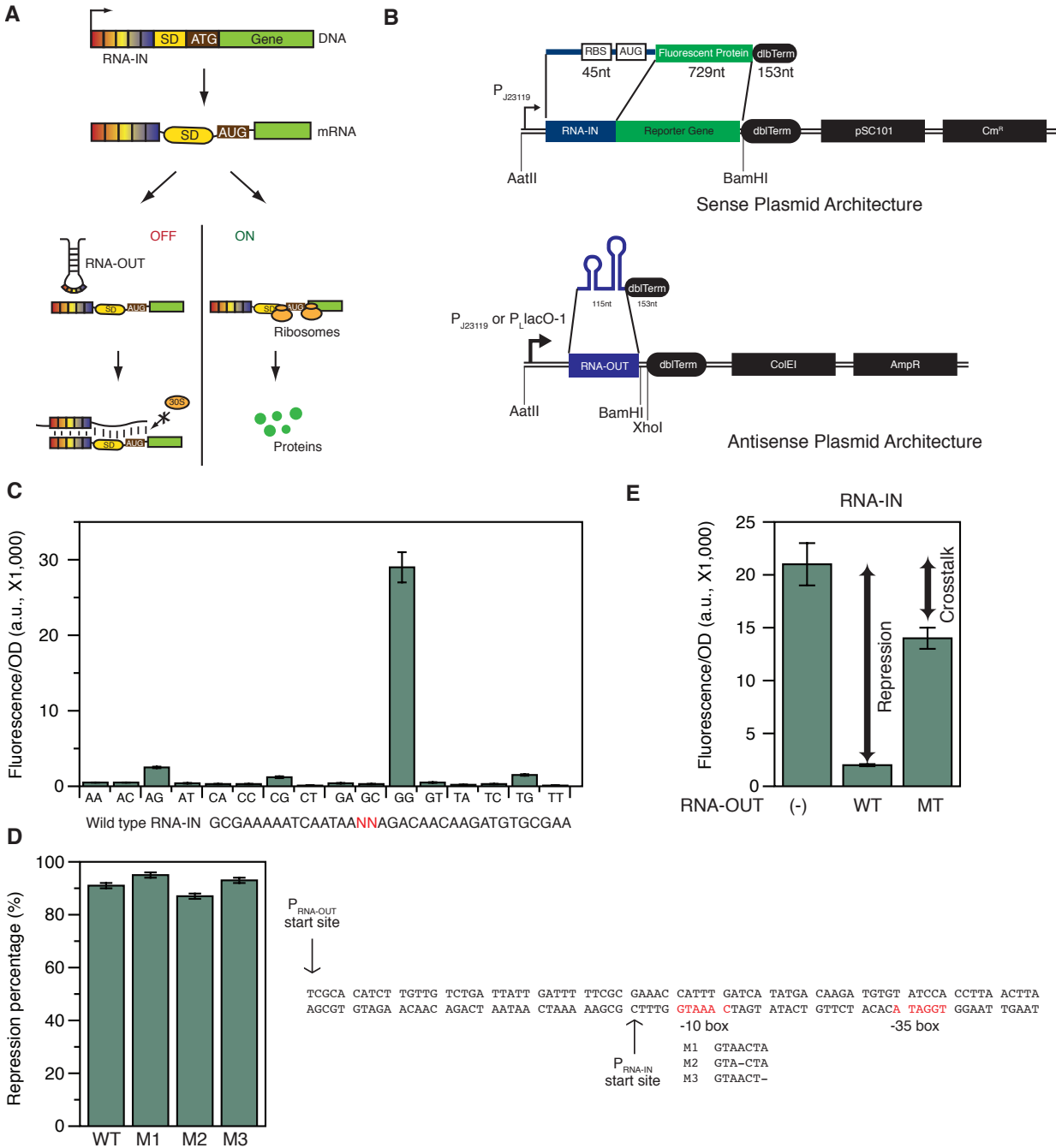
### 4.2 Characterization And Optimization of The Wildtype IS10 System

Our translational regulators are thus derived from the IS10 antisense control elements, wherein RNA-OUT base pairs to the translation initiation region of RNA-IN and inhibits transposase mRNA translation<sup>125</sup> (Figure 4.2A). The 5' end of the unstructured, unstable sense RNA-IN is



complementary to the top of the loop domain and one entire side of the stable RNA-OUT hairpin (**Figure 4.1B**). Earlier studies have suggested that the 5 base pairs in the 5' end of RNA-IN and the loop domain of RNA-OUT determine the initiation of the RNA duplex formation<sup>124</sup>. The loop domain of RNA-OUT contains a YUNR motif, and is predicted to promote specificity and rapid duplex formation with RNA-IN<sup>123,126</sup>. The first three base-pairs between RNA-IN and RNA-OUT are G-C pairs and the strength of hybridization free energy in this GC rich region seems to be critical for effective antisense interaction and molecular specificity<sup>127</sup>. We reason, therefore, that these specificity-determining interactions could be manipulated to create families of mutually orthogonal variants of the wild type system.

To simplify the engineering, wild type sense RNA-IN (+1 to +45) was translationally fused to sfGFP under the control of a strong constitutive promoter J23119 on a low copy pSC101 plasmid, with the start codon (AUG) embedded inside RNA-IN fused to the N-terminus of the 2<sup>nd</sup> amino acid of sfGFP. It is known that the expression of sense RNA-IN in the wild type system is very low due to a weak promoter and a non-consensus Shine-Dalgarno (SD) sequence. To improve the expression of RNA-IN in our synthetic platform, we generated a library of sense mutants by randomizing two base pairs in SD (+16, +17). We found that RNA-IN expression with GG mutant showed the strongest expression (due to the near consensus RBS sequence) compared to other mutants without RNA-OUT (**Figure 4.2C**). The wild type RNA-OUT (+1 to +115) was expressed from a high copy ColE1 vector under the control of an IPTG-inducible promoter P<sub>LlacO-1</sub>. A vector that lacked the RNA-OUT sequence was used as the control plasmid. The full length of RNA-OUT folds into two stable hairpins. The first hairpin is 65-nt, and the second one is 50-nt. To determine the effect of different lengths of RNA-OUT on translation repression, we measured repression of RNA-IN under either the full-length or the first hairpin (65-nt) of RNA-OUT. We observed that while the shorter version exhibited a strong repression, the full-length RNA-OUT performed even better. Thus, the full-length RNA-OUT has been used. The RNA-OUT construct contains an internal P<sub>RNA-IN</sub> promoter transcribing 36-nt sense RNA-IN from the opposite strand. Earlier work has shown that this promoter was very weak, whose transcriptional activity could be silenced by a point mutant in the -10 box of P<sub>RNA-IN</sub><sup>126</sup>. However, the high-copy plasmid for RNA-OUT in our assays might enhance the transcription from P<sub>RNA-IN</sub>, which may titrate away RNA-OUT from interacting with RNA-IN that was fused to sfGFP. To test this possibility, we inactivated the P<sub>RNA-IN</sub> promoter (**Figure 4.2D**). Our results showed that, inactivating the P<sub>RNA-IN</sub> promoter has a negligible effect on the overall percentage repression of RNA-IN-GFP by RNA-OUT as compared to that without inactivation of P<sub>RNA-IN</sub> promoter. Hence, we have retained wild type RNA-OUT with no modifications in the P<sub>RNA-IN</sub> promoter. However, we cannot rule out the possibility that some (or negligible amount) of the RNA-OUT might be titrated away by RNA-IN expressed from its opposite strand. To assess the performance of modified RNA-IN/OUT platform, we measured the repression percentage in *E. coli* strain TOP10. We observed more than 90% percentage of repression in the presence of RNA-OUT, confirming the effective inhibition of RNA-OUT on RNA-IN. Furthermore, a single nucleotide mutation in the loop region of RNA-OUT hairpin exhibited a much weaker repression, indicating the minimized platform is sensitive to changes in antisense/sense specificity<sup>128</sup> (**Figure 4.2E**).



**Figure 4.2 Characterization of the IS10 system.** (A) Schematic of the synthetic IS10 reporter system. Antisense RNA binds specifically to the target mRNA, which blocks the SD sequence and the start codon, inhibiting its translation. The binding sites are indicated by colored boxes. In the absence of antisense RNA, translation of mRNA produces GFP. (B) Two-plasmid system for engineering IS10 system. The architecture of each plasmid is shown. (C) Mutations of 2-nt in the SD sequence. The mutant with GG showed strongest expression, which was later used for all experiments. (D) Silencing of native antisense promoter from the (-) strand of DNA. The native promoter start sites for both sense and antisense RNAs are labeled with arrows. The -10 box and -35 box of the antisense RNA promoter are shown in red. Sequences of three -10 mutants are

shown. No difference was observed comparing the -10 element mutants to the wildtype. (E) Test orthogonality between a mutant antisense RNA and the wildtype mRNA.

### 4.3 Engineering orthogonal translational regulators

#### 4.3.1 Design of mutant libraries

To engineer mutually orthogonal sense-antisense pairs, we considered complementary mutations at the five nucleotides of all combinations in the 5' specificity region of RNA-IN and the corresponding nucleotides in the loop of RNA-OUT (**Figure 4.1B**). This leads to a set of 32 mutations in sense RNA-IN and antisense RNA-OUT (**Table 4.3.1A**). We reasoned that the possible number of orthogonal pairs could also be increased by inserting nucleotides within the recognition motif of this system thereby 'scaling-up' the RNA-RNA interaction region. We therefore considered insertion of 2 extra nucleotides AT, GC, TA and CG between positions +3 and +4 of RNA-IN (corresponding complementary nucleotides at positions +33 and +34 of RNA-OUT). We also hypothesized that compensatory mutations in the first 3 base pairs of the interaction region in these 'scaled-up' mutants would extend the number of orthogonal pairs and possibly improve regulatory efficiency. This resulted in 24 additional RNA-IN/OUT paired mutations for a total library size of 56 (**Table 4.3.1B**). This number may be further increased by considering all possible combination of (single, double, etc.) nucleotide insertions with all different first 3 bp combinations.

The RNA-IN/OUT interaction is largely governed by Watson-Crick base pairing and thermodynamically favored. Furthermore, our rationally constrained library of RNA-IN/RNA-OUT pairs is composed of mutants that have 5-bp variable sequence region surrounded by a common flanking sequence. We thus assumed that the specificity of interaction and stability of the duplex in our library of mutants could be explained, to a large extent, by differences in their hybridization free energies. The orthogonal pairs would be expected to have lower hybridization energy between the cognate pairs but high hybridization energy with non-cognate pairs. Thus, to predict which pairs in our virtual library would show highest specificity of interaction and lowest crosstalk with other members, we estimated the hybridization free energies using the Mfold software<sup>65</sup> for all 56 sense/antisense pairs in the library (total 3136 interactions) shown in **Figure 4.3.1**. To maximize the chance of mutual orthogonality, we selected 23 candidates from the total of 56 library members via a clustering procedure. Incidentally, only 5 RNA-OUT mutants out of the 23 conserve the YUNR motif, which also gave us a chance to test the importance of this motif in the functioning of the IN-OUT system.

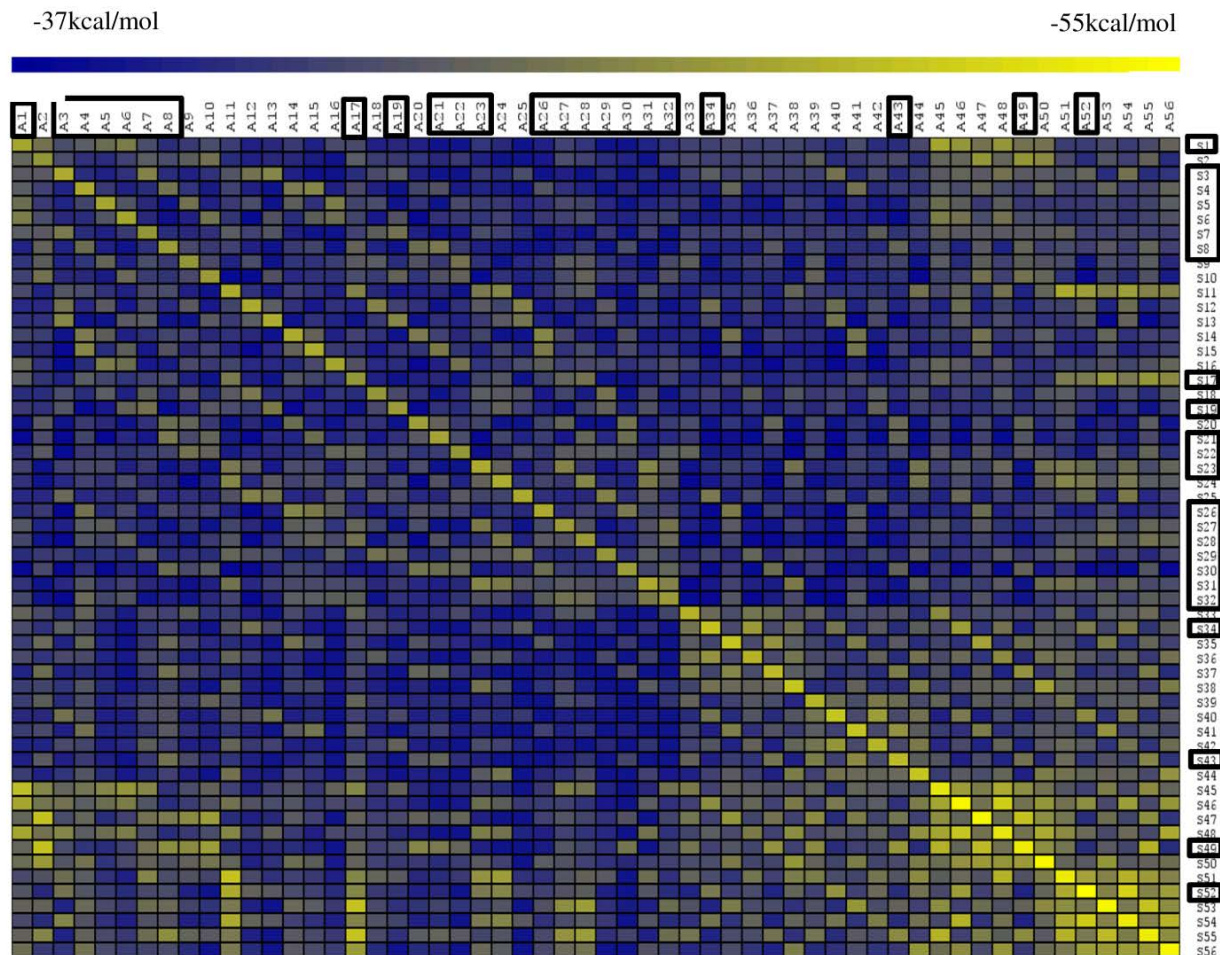
A

Sense	Sequence	Antisense	Sequence	Nucleotide swaps	Tested/possible
S1	GCGAA	A1	UUCGC	Wildtype	1/1
S3	GGGAA	A3	UUC <del>CC</del>	1	4/5
S4	G <del>CC</del> AA	A4	UUGGC		
S5	GCGUA	A5	UACGC		
S6	GCGAU	A6	AUCGC		
S7	C <del>GG</del> AA	A7	UUC <del>CG</del>	2	2/10
S8	C <del>CC</del> AA	A8	UUG <del>GG</del>		
S17	C <del>GC</del> AA	A17	UUG <del>CG</del>	3	6/10
S19	C <del>GG</del> AU	A19	AUC <del>CG</del>		
S21	C <del>CC</del> AU	A21	AUG <del>GG</del>		
S22	C <del>CG</del> UU	A22	AAC <del>GG</del>		
S23	G <del>GC</del> UA	A23	UAG <del>CC</del>		
S26	G <del>CC</del> UU	A26	AAG <del>CC</del>	4	5/5
S27	C <del>GC</del> UA	A27	UAG <del>CG</del>		
S28	C <del>GC</del> AU	A28	AUG <del>CG</del>		
S29	C <del>GG</del> UU	A29	AAC <del>CG</del>		
S30	C <del>CC</del> UU	A30	AAG <del>CC</del>	5	1/1
S31	G <del>GC</del> UU	A31	AAG <del>CC</del>		
S32	C <del>GC</del> UU	A32	AAG <del>CG</del>	Extra 2*	4/24
S34	G <del>GG</del> UAAA	A34	UUUAC <del>CC</del>		
S43	C <del>CC</del> AUAA	A43	UUUAG <del>GG</del>		
S49	C <del>CCCG</del> AA	A49	UUCG <del>GGG</del>		
S52	G <del>GGG</del> CAA	A52	UUC <del>CCCC</del>		
				Total	23/56

B

Mutant #	Sense Motif	Antisense Motif	# of Mutations
S1	GCGAA	TT <del>CGG</del>	0
S2	CCGAA	TT <del>CCC</del>	1
S3	GGGAA	TT <del>GGC</del>	1
S4	G <del>CC</del> AA	TACGC	1
S5	GCGTA	ATCGC	1
S6	GCGAT	TT <del>CCG</del>	1
S7	CGGAA	TT <del>GGG</del>	2
S8	C <del>CC</del> AA	TAC <del>GG</del>	2
S9	CCGTA	ATC <del>GG</del>	2
S10	CCGAT	TT <del>GCC</del>	2
S11	GGCAA	TAC <del>CC</del>	2
S12	GGGTA	AT <del>CC</del>	2
S13	GGGAT	TAG <del>GC</del>	2
S14	GCCTA	AT <del>GGC</del>	2
S15	GCCAT	AAC <del>GC</del>	2
S16	GCGTT	TT <del>GCG</del>	2
S17	CGCAA	TAC <del>CG</del>	3
S18	CGGTA	AT <del>CCG</del>	3
S19	CGGAT	TAG <del>GGG</del>	3
S20	CCCTA	AT <del>GGG</del>	3
S21	CCCAT	AAC <del>GG</del>	3
S22	CCGTT	TAG <del>CC</del>	3
S23	GGCTA	AT <del>GCC</del>	3
S24	GGCAT	AAC <del>CC</del>	3
S25	GGGTT	AAG <del>GC</del>	3
S26	GCCTT	TAG <del>CG</del>	3
S27	CGCTA	TAG <del>CG</del>	4
S28	CGCAT	AT <del>GCG</del>	4
S29	CGGTT	AAC <del>CG</del>	4
S30	CCCTT	AAG <del>GG</del>	4
S31	GGCTT	AAG <del>CC</del>	4
S32	CGCTT	AAG <del>CG</del>	5
S33	CCGTAAA	TTTAC <del>GG</del>	2+1
S34	GGGTAAA	TTTAC <del>CC</del>	2+1
S35	GCCTAAA	TTTAG <del>GC</del>	2+1
S36	CGGTAAA	TTTAC <del>CG</del>	2+2
S37	CCCTAAA	TTTAG <del>GGG</del>	2+2
S38	GGCTAAA	TTTAG <del>CC</del>	2+2
S39	CCGATAA	TTATC <del>GG</del>	2+1
S40	GGGATAA	TTAT <del>CC</del>	2+1
S41	GCCATAA	TTAT <del>GGC</del>	2+1
S42	CGGATAA	TTATC <del>CG</del>	2+2
S43	CCCATAA	TTAT <del>GGG</del>	2+2
S44	GGCATAA	TTAT <del>GCC</del>	2+2
S45	CCGCGAA	TTCG <del>CGG</del>	2+1
S46	GGGCGAA	TTCG <del>CCC</del>	2+1
S47	GCCCGAA	TTCG <del>GGC</del>	2+1
S48	CGGCGAA	TTCG <del>CCG</del>	2+2
S49	CCCCGAA	TTCG <del>GGG</del>	2+2
S50	GGCCGAA	TTCG <del>GCC</del>	2+2
S51	CCGGCAA	TTG <del>C</del> CGG	2+1
S52	GGGGCAA	TTG <del>CCCC</del>	2+1
S53	GCCGCAA	TTG <del>C</del> GGC	2+1
S54	CGGGCAA	TTG <del>CCCCG</del>	2+2
S55	CCCGCAA	TTG <del>C</del> GGG	2+2
S56	GGCGCAA	TTG <del>C</del> GCC	2+2

**Table 4.3.1 List of 23 pairs of mutant antisense and sense RNAs.** (A) The recognition motif of 23 pairs of sense and antisense RNAs are shown. Total number of possible mutants from a particular base swap and the number of chosen mutants for experimental characterization are also shown. (B) A list of all possible mutants including the experimentally tested 23 pairs.



**Figure 4.3.1 Calculation of free hybridization energy.** The entire interaction region (37 nt) of RNA-IN and RNA-OUT was used for the calculation. The calculated hybridization energy for sense and antisense RNA pairs varies between  $-37.8$  to  $-54.9$  kcal mol<sup>-1</sup>. As expected, we found that the cognate sense/antisense pairs along the diagonal have predicted free energies closer to  $-50$  kcal mol<sup>-1</sup> and, as a group they show far more stable hybrids than that of non-cognate pairs.

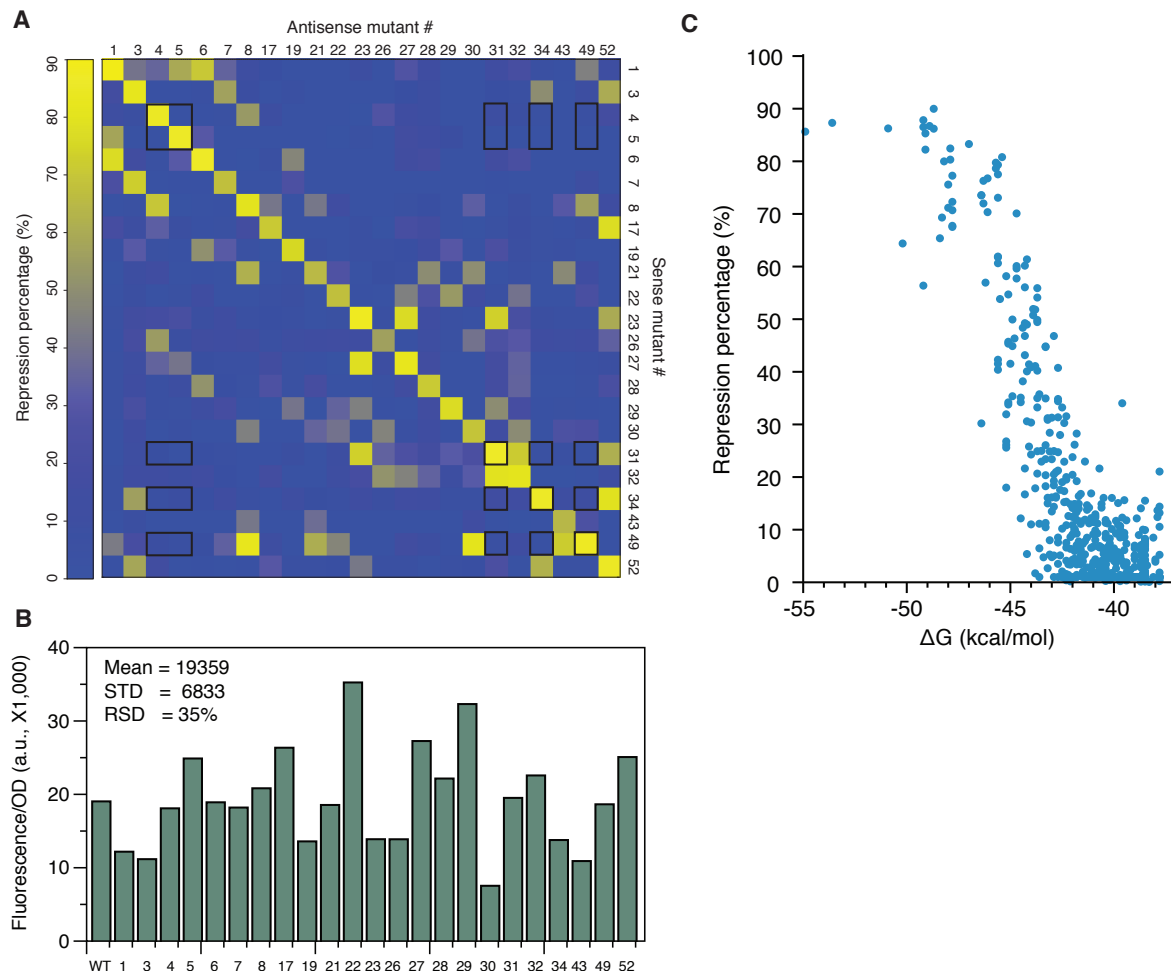
#### 4.3.2 Measurement and analysis of mutant libraries

We generated the 23 RNA-IN and RNA-OUT mutants by standard site directed mutagenesis. Cells bearing each of the 23 target RNA-IN plasmids were co-transformed with each of the plasmids expressing antisense RNA-OUT versions or the nonsense control. Expression of RNA-IN-GFP was quantified for each strain during the exponential phase and percentage repression was calculated. The matrix of percentage repression for the 529 combinations of sense and antisense mutants is shown as a heat map in **Figure 4.3.2A**. Most of cognate sense/antisense pairs show repression more than 80 %, while most non-cognate combinations show repression

smaller than 20 %. One intriguing result is that more than 70 % of cognate pairs show a repression level higher than 80 % and do not possess a YUNR motif in the antisense RNA species indicating that the YUNR motif is dispensable for the proper functioning of this system.

Earlier studies showed that single complementary mutations at the 3<sup>rd</sup> and 4<sup>th</sup> nucleotides at 5' end of RNA-IN and corresponding nucleotides in the loop region of RNA-OUT altered the sequence specificity of the antisense pairing reaction with their wild type counterparts<sup>125</sup>. Here, in addition to recapitulating this result, we observed that combinatorial complementary nucleotide swaps (1 to 5 nucleotides) clearly modified the interaction specificity with each other in addition to the wild type pair. The insertion of two extra nucleotides between the 3<sup>rd</sup> and 4<sup>th</sup> nucleotide within the recognition motif of RNA-IN and the corresponding positions of RNA-OUT retained the specificity of interaction between pairs, and displayed stronger interaction with few non-cognate partners. This indicates that the size of orthogonal mutant library could be greatly expanded by inserting new nucleotides into the recognition motif. This result is especially interesting since earlier work showed that replacing A-U bases at 4<sup>th</sup> and 5<sup>th</sup> position of RNA-IN and corresponding nucleotides of RNA-OUT with G-C bases showed decrease in the pairing interaction<sup>124</sup>. This indicates that the stretch of A-U's downstream of recognition motif may be essential for the rapid progression of the duplex. In summary, we observe that a wide range (negligible repression <10 % to 90 %) of target repression could be obtained by altering the complimentary nucleotides in the interaction region. We note here that the absolute levels for gene expression might vary across the mutants. As shown in **Figure 4.3.2B**, the fluorescence expression for the mutants without antisense RNA present exhibited a variance of 35 % relative standard deviation (RSD).

Though the hybridization energy has been used as a guiding feature for searching genome-wide noncoding RNA targets<sup>38</sup> and for engineering synthetic RNAs<sup>86</sup>, its role in imparting *in vivo* regulatory properties to RNAs has not been explored to a large extent. To determine how energetics of sense-antisense RNA interaction correlates with the experimental percentage repression, we plotted the calculated hybridization free energy for all 529 interactions against the experimental percentage repression (**Figure 4.3.2C**). Interestingly, we observed that interactions with a free energy more than -41 kcal mol<sup>-1</sup> are not active in repression, whereas most interactions with a free energy less than -46 kcal mol<sup>-1</sup> showed stronger repression (closer to 85%). These results indicate that there is a critical threshold free energy needed for the propagation of initial pairing interaction to a stable duplex formation and thereby causing efficient repression of target mRNA. Similar results have been reported for interaction of microRNAs with their targets in HeLa cell lines<sup>129</sup>.



**Figure 4.3.2 Construction and characterization of 23 sense-antisense pairs.** (A) Heat map of percentage repression profile of 23 RNA-IN mutants in presence of 23 antisense RNA-OUT mutants (total 529 data points). Cognate pairs are arranged diagonally and show maximum repression. Five mutually orthogonal pairs are shown as black boxes. (B) Basal level expression for 23 RNA-IN mutants without antisense RNA. (C) The scatter plot of calculated hybridization free energy ( $\Delta G$ , kcal mol<sup>-1</sup>) as a function of experimental percentage repression.

### 4.3.3 Mutually orthogonal mutants

Using the experimentally determined percentage repression data to quantify target and non-target specificity, we can identify families of RNA-IN/OUT variants that were expected to function orthogonally when placed in the same cell. Further, identifying non-cognate partners that show significant crosstalk aids in determining base-pairing features that impart the promiscuity. Thus, the definition of mutual orthogonality depends on thresholds of repression (%R) and cross-reactivity percentages (%C) for cognate and non-cognate pairs respectively that we deem acceptable for a specific application.

**A**

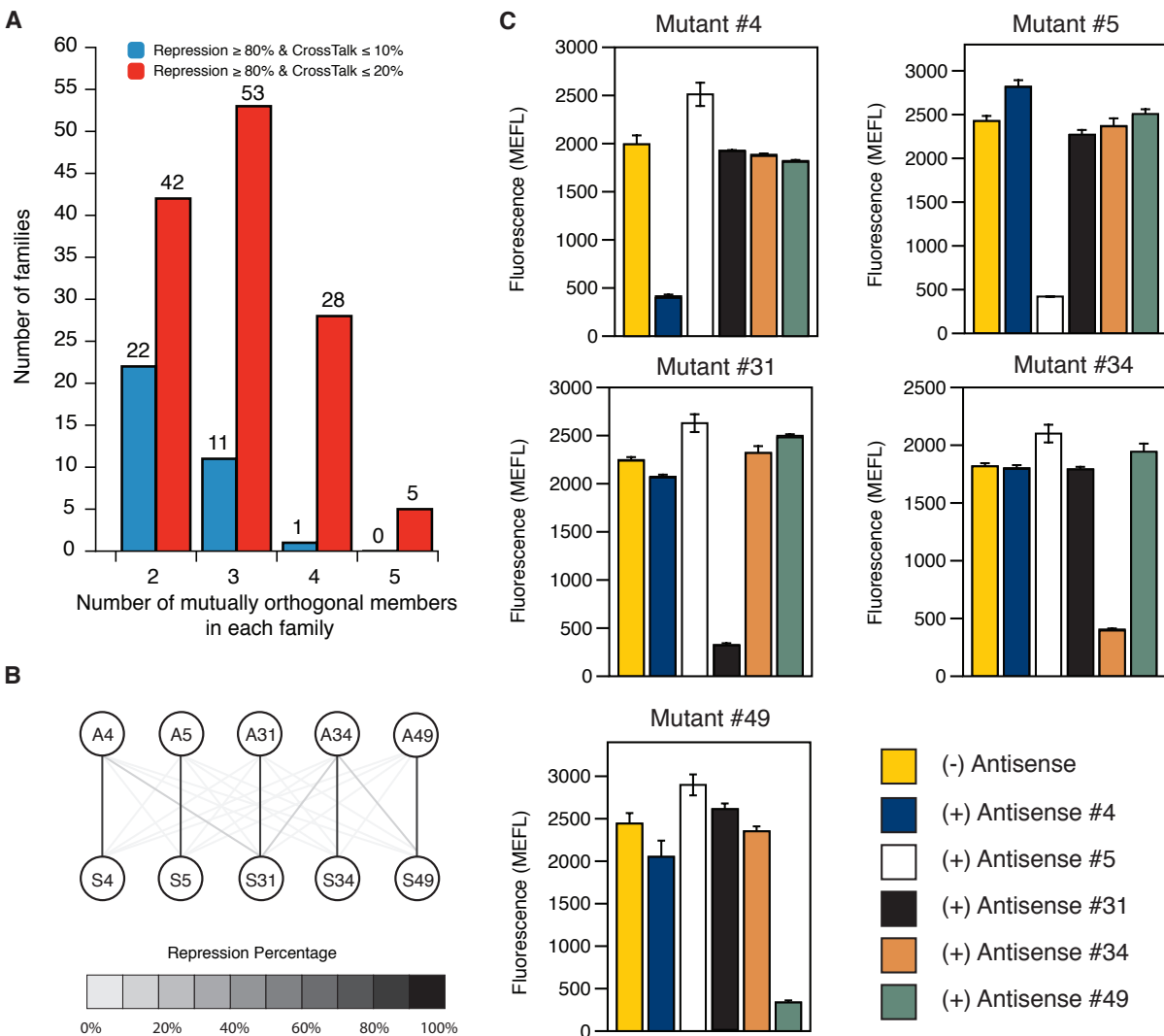
Mutually orthogonal Family Member size: 3					
1	S31	S34	S49		
2	S05	S49	S52		
3	S04	S06	S49		
4	S32	S49	S52		
5	S06	S34	S49		
6	S03	S04	S31		
7	S04	S06	S34		
8	S06	S49	S52		
9	S23	S34	S49		
10	S05	S32	S52		
11	S03	S04	S05		
12	S05	S31	S49		
13	S04	S31	S49		
14	S01	S23	S34		
15	S04	S32	S49		
16	S04	S31	S34		
17	S03	S04	S32		
18	S03	S05	S31		
19	S04	S05	S49		
20	S05	S31	S34		
21	S06	S23	S34		
22	S04	S05	S34		
23	S03	S05	S32		
24	S05	S32	S49		
25	S04	S05	S31		
26	S04	S34	S49		
27	S05	S34	S49		
28	S04	S05	S32		
29	S06	S23	S49		
Mutually orthogonal Family Member size: 4					
1	S03	S04	S05	S32	
2	S05	S32	S49	S52	
3	S05	S31	S34	S49	
4	S04	S06	S34	S49	
5	S04	S31	S34	S49	
6	S04	S05	S32	S49	
7	S04	S05	S31	S49	
8	S04	S05	S34	S49	
9	S03	S04	S05	S31	
10	S06	S23	S34	S49	
11	S04	S05	S31	S34	
Mutually orthogonal Family Member size: 5					
1	S04	S05	S31	S34	S49

**Table 4.3.3 Mutually orthogonal families of antisense-sense pairs.** Only orthogonal family with members more than two are shown. The criteria is >80% repression and <15% crosstalk.

The total number of observed mutants for different family sizes demonstrating 80% R and 10% or 20 %C with other members of the family (and orthogonal family) is shown in **Figure**



**4.3.3A.** At 80% R and 10% C we have more than 10 families of mutually orthogonal pairs and triplets, and one family of 4 orthogonal mutants, whereas at 20% C, we have more than 20 families made up of 2, 3 and 4 mutually orthogonal mutants and 5 families of 5 mutants. A list of mutually orthogonal mutants at thresholds of 80% R and 15% C are shown in **Table 4.3.3**. We picked up the group of 5 mutually orthogonal pairs (#4, #5, #31, #34, and #49), repeated the whole experiment by testing all combinations between antisense and sense using flow cytometry and obtained similar results (**Figure 4.3.3B**). To test if the orthogonality and repression profile of these mutants is retained with a sequence-divergent gene-of-interest, we fused five mutually orthogonal sense mutants to mRFP (52% sequence identity) and assayed them in the presence of corresponding antisense RNAs. The observed percentage repressions were quantitatively equivalent to that of sfGFP, demonstrating the modularity of the sense region and efficiency of antisense RNA (**Figure 4.3.3C**).



**Figure 4.3.3 Test of five mutually orthogonal pairs.** (A) Clustering of the 23X23 matrix resulted in groups of mutually orthogonal pairs. Two criteria are used for clustering: data for  $\geq 80\%$  repression and  $\leq 10\%$  crosstalk (blue bars) and data for  $>80\%$  repression and  $\leq 20\%$

crosstalk (red bars). **(B)** The observed network of interaction between mutually orthogonal cognate and noncognate sense and antisense RNAs (mutants 4, 5, 31, 34 and 49). The shading of links indicates the percentage repression (black = 100% repression; light gray = no repression). **(C)** Flow cytometry data for five mutually orthogonal pairs.

#### 4.3.4 Develop a mathematical model to explain observed orthogonality

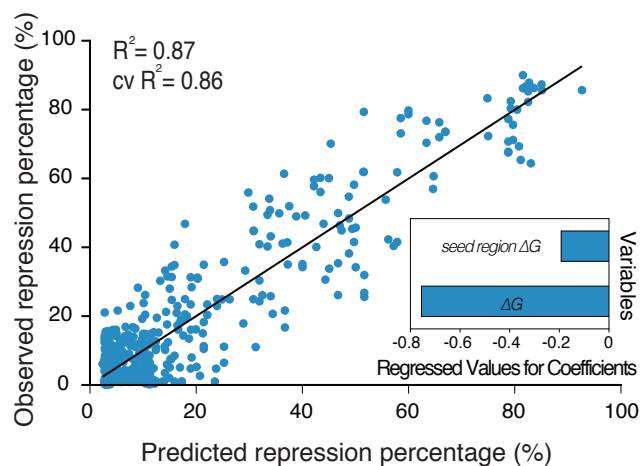
The correlation between the hybridization free energy and percent repression suggests that free energy is a good though not a perfect predictor of interaction specificity (**Figure 4.3.2C**). To find other features that determine the specificity of interaction between RNA-IN and RNA-OUT, we pursued modeling of sequence-function relationship for the *in vivo* experimental dataset.

# of Features	Feature	Description
1	$\Delta G$	Hybridization energy of sense-antisense duplex
2	Sense- $\Delta G$	Hybridization free energy of sense Minimum free energy (MFE) structure
3	Antisense- $\Delta G$	Hybridization free energy of antisense MFE structure
4	Seed_region- $\Delta g$	Hybridization energy of duplex in 5 nt seed region
5	Seed_unpaired	Number of unpaired bases in duplex of 5 nt seed region
6	motifAAI4	A4A5 (RNA-IN) – paired to (RNA-OUT) This feature is equal to 1 if the nucleotides AA (at position 4 and 5 in the RNA-IN molecule) are both paired in the sense-antisense duplex. Note that the pairing position on RNA-OUT is not fixed, that is A4A5 of RNA-IN can pair anywhere on RNA-OUT not necessarily at the same position as the WT pairs (Fig 1b Main text).
7	motifAAI5	A5A6 (RNA-IN) – paired to (RNA-OUT)
8	motifUAI5	U5A6 (RNA-IN) – paired to (RNA-OUT)
9	motifUUI4	U4U5 (RNA-IN) – paired to (RNA-OUT)
10	motifGCI1	G1C2 (RNA-IN) – paired to (RNA-OUT)
11	motifGCI2	G2C3 (RNA-IN) – paired to (RNA-OUT)
12	motifCGI1	C1G2 (RNA-IN) – paired to (RNA-OUT)
13	motifCGI2	C2G3 (RNA-IN) – paired to (RNA-OUT)
14	motifGGI1	G1G2 (RNA-IN) – paired to (RNA-OUT)
15	motifGGI2	G2G3 (RNA-IN) – paired to (RNA-OUT)
16	motifCCI1	C1C2 (RNA-IN) – paired to (RNA-OUT)
17	motifCCI2	C2C3 (RNA-IN) – paired to (RNA-OUT)
18	motifGAI3	G3A4 (RNA-IN) – paired to (RNA-OUT)
19	motifGUI3	G3U4 (RNA-IN) – paired to (RNA-OUT)
20	motifCAI3	C3A4 (RNA-IN) – paired to (RNA-OUT)
21	motifCUI3	C3U4 (RNA-IN) – paired to (RNA-OUT)
22	motifs- $\Delta g$	Hybridization energy of dinucleotide motifs (from #6 to # 21 above), considers just the free energy from bound nucleotides. This feature doesn't consider free energy from bulges.
23	Interior_loops_number	Number of bulges in duplex
24	Exterior_loop- $\Delta g$	Hybridization free energy for the exterior loop of sense-antisense duplex
25	s_paired_6	Paired status of nucleotide position 6 of sense MFE structure (change in structure when nucleotide 1-5 are mutated)
26	s_exterior_loop- $\Delta g$	Hybridization free energy for the exterior loop of sense MFE structure
27	s_exterior_loop_ss	Number of bases single stranded from sense MFE structure
28	s_seed_region- $\Delta g$	Hybridization free energy in 5 bp seed region of sense MFE structure
29	s_seed_unpaired	Number of bases unpaired in 5 bp seed region of sense MFE structure
30	as_seed_region- $\Delta g$	Hybridization free energy in 5bp seed region of antisense MFE structure
31	as_seed_unpaired	Number of bases unpaired in 5bp seed region of antisense MFE structure

**Table 4.3.4 Compilation of features used in the PLSR model.**

Based on prior work and inspection of the predicted RNA secondary structures and the form of the duplex, we selected a list of 31 possible features that might explain the observed patterns of repression<sup>124</sup> (**Table 4.3.3**). To formally select the most significant feature explaining

the repression data, we applied Partial Least Squares Regression with stepwise feature selection and outlier detection<sup>130</sup>. The analysis, after detecting 8 outliers out of 529 interactions, identified two main features that could explain 86% variation in the data after 10 fold cross validation: the hybridization energies of the entire 37-bp interaction region and a duplex seed region of 5 bps (**Figure 4.3.4**). The model suggests that the initial nucleation event at the GC rich 5-bp seed region and the subsequent helix progression is thermodynamically driven and determines the efficient repression of the target mRNA. These results recapitulate early studies that pointed out the importance of the 5 bp interaction region in determining the copy number control performance of RNA-IN/OUT system<sup>125</sup>. The unexplained 14% variance in the repression data may be due to other features/factors that are not included in this work, for example, the *in vivo* concentrations of interactant RNAs that influence the efficiency of antisense RNAs. We speculate that the peculiarity of the structure or the *in vivo* stability of duplex of these pairs may be the reason for their unpredictable performance. More detailed biochemical studies are needed to pursue these questions. However, the model, trained on the remaining 521 pairs, has sufficient explanatory power to support design of new pairs.

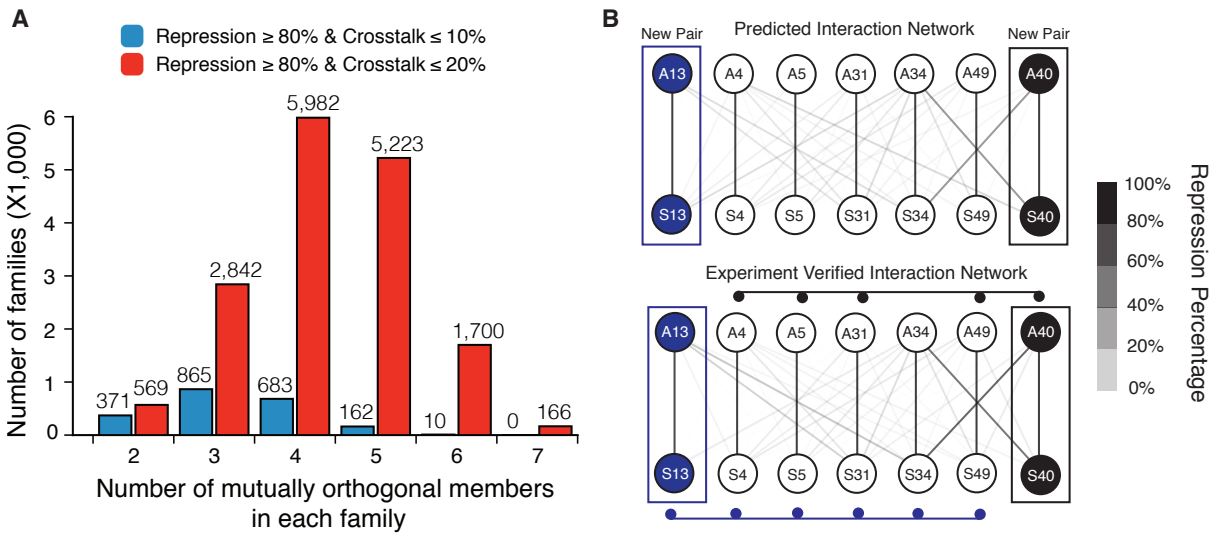


**Figure 4.3.4 Results of the PLS regression model.** A scatter plot of predicted versus observed percentage repression is shown. The coefficient of determination (a measure of the quality of the model,  $R^2 = 0.87$ ,  $P < 0.0001$ ) using a training set of 521 interactions is shown along with the ten-fold cross-validated model. Inset, weighted regression coefficients for the final two predictors.

#### 4.3.5 Using the mathematical model to predict orthogonal pairs

To validate the predictive capability of the model and forward engineer new orthogonal mutants we used the model trained on the 521 pair measures above to predict %R for all 56 mutant pairs we initially considered (**Figure 4.3.1**). This yielded a total of 3136 percentage repression predictions including the 529 experimentally tested pairs and is shown in **Table 4.3.5**. We estimated the total possible number of mutually orthogonal pairs in the 56 RNA-IN and OUT variants from these predictions and **Figure 4.3.5A** shows the total number of predicted orthogonal mutants for different family sizes at 80%R and different threshold of %C. At  $R\% = 80\%$  and  $C\% = 10\%$ , we have more than 300 families of mutually orthogonal pairs, triplets and quadruplets, more than 150 families of 5 mutants and 10 families of 6 orthogonal mutants. While at 20% C we have more than thousand families made up of 3, 4, 5 and 6 mutually orthogonal mutants and about 166 families of 7 mutants. To experimentally validate a subset of these predictions, we forward engineered two sense and antisense RNA pairs (mutant #13 and mutant #40) predicted to have desired strong %R and insignificant cross-talk with the experimentally discovered family of 5 orthogonal pairs, thus expanding the orthogonal family size. We characterized these four forward engineered mutants in the presence of their cognate and non-



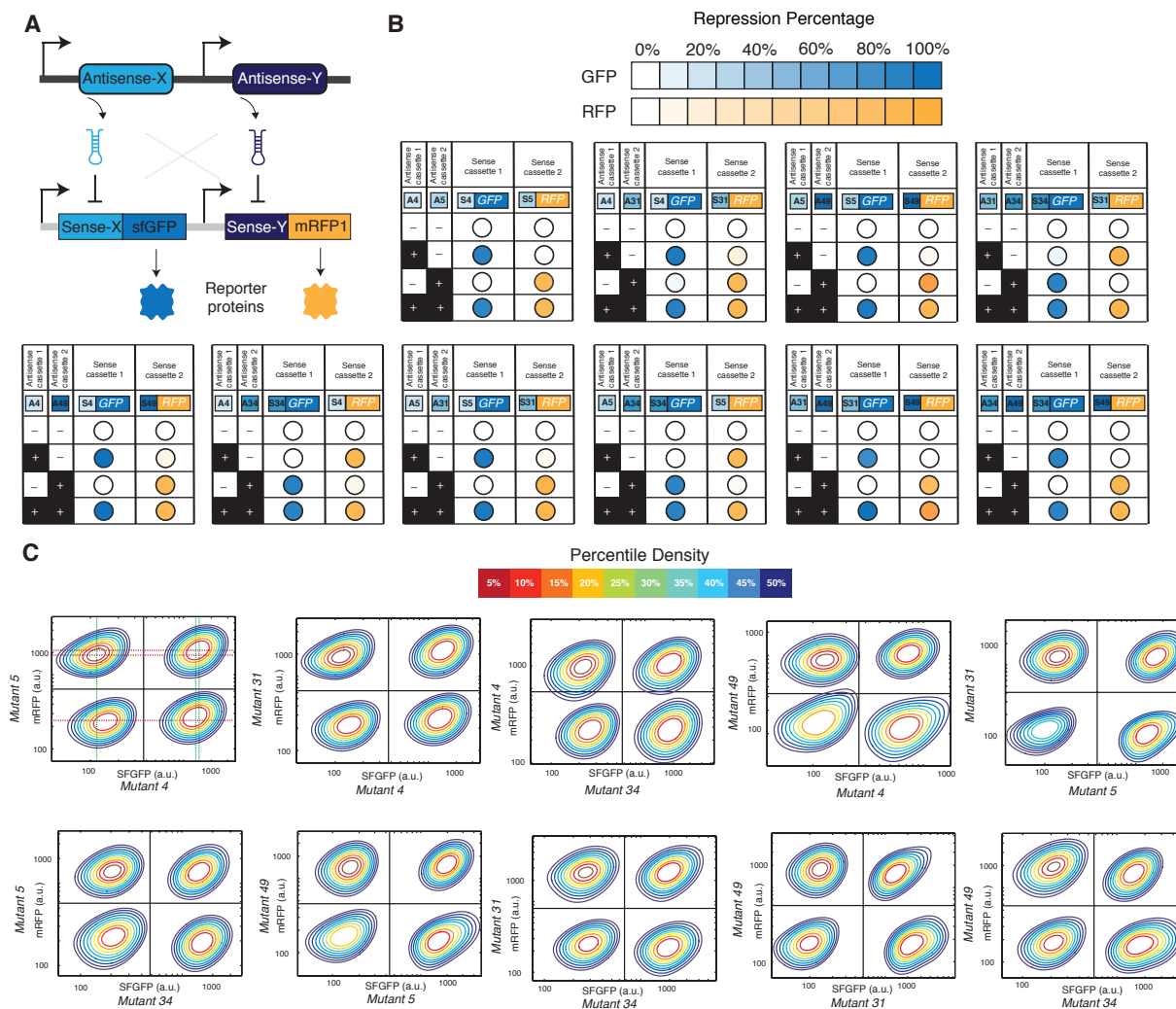


**Figure 4.3.5 The mathematical model can predict orthogonal pairs.** (A) Estimation of the number of families made up of two, three, four and five mutually orthogonal members at different thresholds of percentage repression and cross-talk. Two representative data sets are shown here: data for  $\geq 80\%$  repression and  $\leq 10\%$  crosstalk (blue bars) and data for  $>80\%$  repression and  $\leq 20\%$  crosstalk (red bars). (B) The predicted (top) and experimentally verified (bottom) network of interaction between mutually orthogonal cognate and noncognate sense and antisense RNAs (mutants 13 and 40 in addition to mutants 4, 5, 31, 34 and 49). The shading of links indicates the percentage repression (black = 100% repression; light gray = no repression). Model-predicted and experimentally verified enlarged orthogonal families are indicated with connections between orthogonal members, and new pairs are shown in blue and black.

#### 4.4 Engineering IS10-Based Synthetic Circuits

Ideally, a perfect test of mutual orthogonality of a family of mutants requires us to express all pairs in the same cell and demonstrate that induction of each antisense repressed only its cognate target but no other. Alternatively, we can expression every combination of two pairs in a single cell and show that induction of each antisense only affects its target whether or not the other antisense is expressed. This latter approach allowed us to construct multiple independent controls of a set of genes in the same cell. To demonstrate this, we picked five sense-antisense pairs shown in **Figure 4.3.3B**. The sense of each pair was translationally fused to either sfGFP or mRFP, and repression the presence of different combinations of antisense RNAs was assayed. Data from a representative two-color experiment using flow cytometry is shown in **Figure 4.4**. Our results of testing all combination of two out of five orthogonal pairs demonstrate that orthogonality is maintained when multiple translational regulators are used together in the same cell. In addition to these mutual orthogonal mutants, we find many examples of a single antisense RNA repressing multiple sense targets, and single sense targets recognized by multiple antisense RNAs. These promiscuous variants may be useful in understanding the specificity determinants in natural regulatory RNAs interacting and silencing multiple targets (prokaryotic small RNA global regulatory networks) and can also aid in designing synthetic circuits as signal propagation and/or signal integration modules<sup>88,131</sup>. However, that fact that translational

regulators can only control protein synthesis and not the production of RNAs and are constrained to act on single genes limits their application in constructing complex regulatory functions.

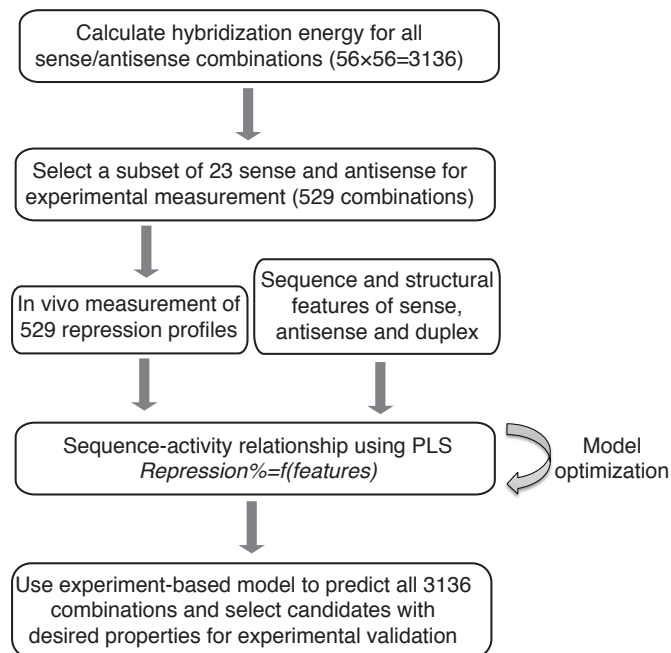


**Figure 4.4 Two-color experiments using orthogonal antisense-sense pairs.** (A) Schematic on the left shows the regulation of two target RNAs fused to sfGFP and mRFP1 by orthogonal antisense RNAs. (B) Ten panels on the right show the experimental percentage repression (shown as color intensity) for different combination of sense and antisense RNAs. The presence and absence of antisense RNA is indicated on the left and percentage repression is shown on the right for two target genes. (C) Representative two color percentile contour plots show unimodal expression (darkest blue - 75% cells, red - 5% cells) in all cases.

#### 4.5 Strengths and Limitations of The System

We have presented a quantitative framework that can describe RNA-RNA interactions by integrating *in vivo* reporter assay data with the sequence-activity modeling to generate a large panel of mutually orthogonal family of translation regulators<sup>132</sup>. We followed a design procedure as shown in **Figure 4.5**. We first calculated thermodynamic hybridization energy for 56 pairs of designed sense-antisense combinations (3136 interactions). Then we selected a subgroup of 23

pairs for experimental measurement (529 interactions). This dataset provided us to find 5 mutually orthogonal pairs with a given threshold (repression > 80%, crosstalk < 15%), and allowed us to develop and train a sequence-activity model based on PLS regression (521 interactions with 8 outliers excluded). Using the model, we identify duplex hybridization energy as the key determinant of their specificity, and search in the original 56 sense-antisense RNA pairs for novel orthogonal groups. This offers us a sixth pair to be added into the 5 mutually orthogonal group. The predictive model allowed us to quickly identify new orthogonal mutants that differ in specificity and crosstalk properties and that can either expand or create new variants of a regulatory family. Using this predictive and quantitative sequence-activity framework, we forward engineered two sense and antisense RNA mutants and experimentally validated more than 50 RNA-RNA interactions, which allowed us to discover and expand a family of 5 and 6 mutually orthogonal translational regulators that exhibit consistent and predictable performances. To our knowledge, this is the largest orthogonal family constructed from a single regulatory mechanism based on a predictive model (in contrast to commonly used screen-based approaches). In addition, we find that the potential mutually orthogonal families made up of 2, 3 and 4 regulatory members are in numbers of thousands simply by combining nucleotide swaps in the recognition region of sense and antisense RNAs. We also demonstrate that the specificity of interaction is preserved by inserting extra nucleotides or ‘scaling-up’ the core interaction region thereby allowing the enlarged sequence space for searching additional orthogonal regulators.



**Figure 4.5 The quantitative framework for modeling.**

The sequence-activity model provides some key insights into the operation of RNA-IN/OUT system and also suggests possible design principles in other antisense RNA regulated systems. We found that both hybridization energy of the entire duplex and that of the seed region (5-nt length) are the significant features explaining most of the *in vivo* reporter data and hence determine the sense-antisense RNA interaction specificity. We also observed that, the ubiquitously observed YUNR motif on variety of antisense RNAs<sup>113</sup> appears to be nonessential for retaining the specificity and efficiency of interaction in RNA-IN/OUT system. This is particularly interesting in few cases where antisense mutants with no YUNR motif exhibit the same %R as that of the WT. We speculate that having a stable stem structure with U-A rich region around the loop of antisense provides necessary flexibility to accommodate variations in the core recognition motif. This also allows bases following U residues to be directed outwards and hence available to interact with RNA-IN thereby retaining the efficiency and specificity of interaction in absence of YUNR motif. Though we have not studied the performance of loop region mutants in the context of alterations in UA-rich motif at the base of RNA-OUT loop

region, we speculate that they determine the rapid helix progression after the initial base pairing at GC rich motif. This agrees with our ‘scaled-up’ mutant data wherein insertion of two extra nucleotides between GC rich 3-nt core region and A-U rich downstream region is well tolerated.

This work reconfirms some of the earlier observations that the high specificity and efficiency shown by antisense RNA regulated systems cannot be generalized and explained by simple duplex hybridization free energy. The importance of sequence dependent hybridization energy in determining the specificity appears to be limited to systems in which an unstructured RNA end initiates binding within the loop of its counterpart and possibly forms duplex by single step strand exchange pathway. This does not appear to be the case with those systems in which sense and antisense RNAs initiate loop-loop kissing complex formation followed by the interaction at a distal site to overcome the topological limitation of helix progression (two-step pathway). In such cases the specificity appears to be the consequence of interactions between intricate three-dimensional structures of the interacting RNAs, such as CopT/CopA of plasmid R1, RNAI/RNAII of ColE1 plasmid, Inc/repZ of IncIalpha plasmid, RNAIII/rep of pIP501 plasmid, and pT181 system<sup>133</sup>. Though the path to inhibitory duplex formation is not well understood in trans-encoded RNAs, this result indicates additional structural features and accessibility of recognition domains determine the initial interaction and specificity.

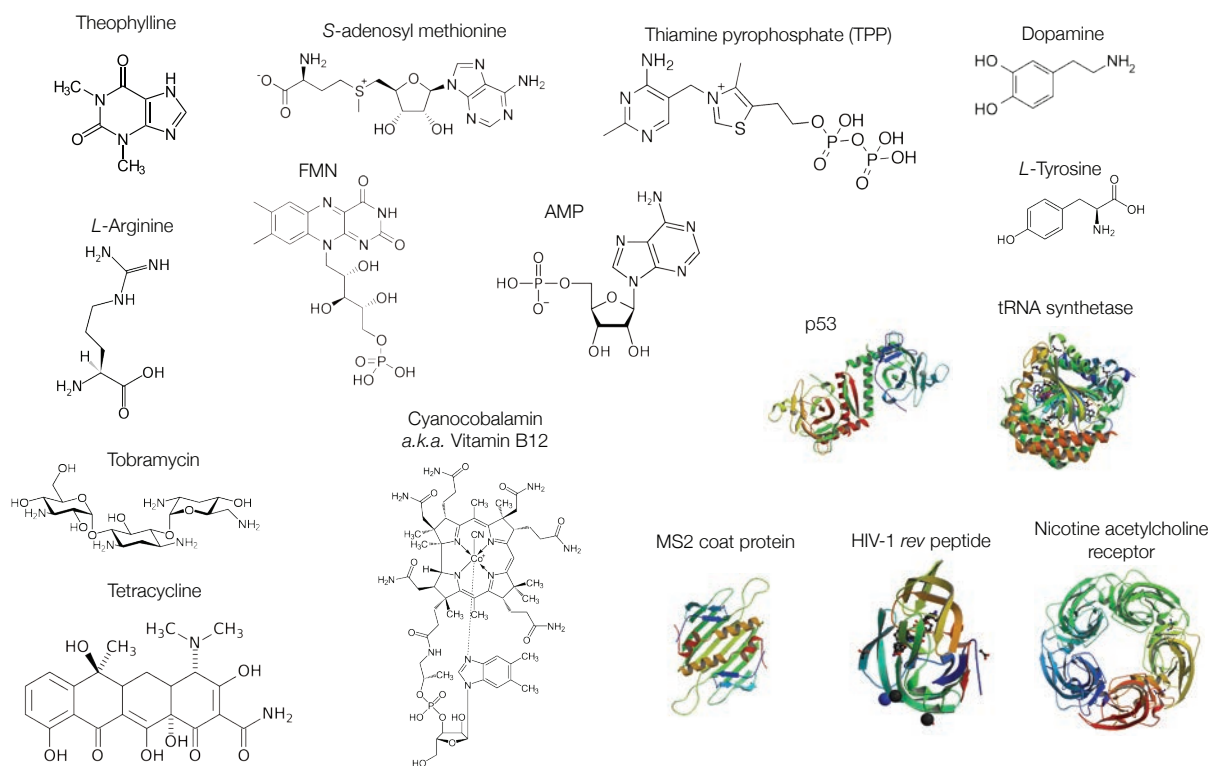
In the context of above complexities in uncovering key features determining specificity for mechanistic understanding and engineering of the system, our current work comes out as a major success. Having found hundreds of highly specific RNA regulators, we face one of the pertinent questions in this regard: do we need so many orthogonal regulators and for which applications? We envision that both the orthogonal regulators and the approach used to discover them have significance in variety of next generation synthetic biology applications. This includes, the use of orthogonal translation repressors for performing different modes of RNA computations in bacteria; for regulating expression of different genes in operons; to establish hierarchical order of regulation; to modulate stoichiometry of protein levels and imparting the pathway balance to explore the network architecture and understand the function of natural small-RNA regulatory networks by rewiring or building smaller synthetic circuits; to generate innovative functions; to understand the function of toxic genes; and to understand evolutionary design principles behind regulatory architectures. Even though the design specification for these various applications may not be very well defined/realized, having a compendium of well characterized parts to meet the diverse specification needs saves lot of resources from *ad hoc* approaches and provides a robust platform for further technological development (for example engineering the RNA components presented here to be sensitive to environmental signals<sup>44</sup>). This is particularly important since the real potential of using RNA-based components in genetic circuits is becoming realized because of their myriad of operational advantages compared to protein regulators. Therefore, we believe that the use of RNA components described in this work in combination with available transcriptional and post-transcriptional regulatory components and other recently developed orthogonal systems should provide an unsurpassed flexibility in designing and programming gene expression in a predictable manner.



## Chapter 5 Engineering Antisense RNAs to Sense Other Molecules

### 5.1 Introduction to RNA Aptamers

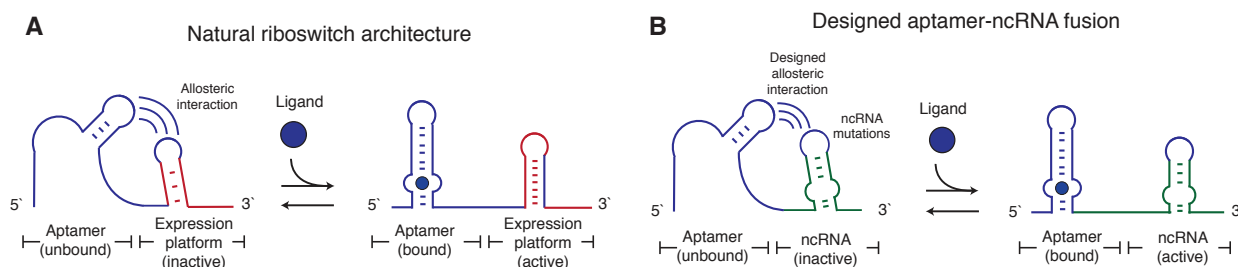
Nucleic acids called aptamers can bind to ligands such as small molecules, peptides, and proteins. These aptamers, either RNA or DNA, fold into well-defined 3D structures, which enable binding by complementary shape interactions with high affinity and specificity. They can be obtained by an *in vitro* selection method called SELEX (Systematic Evolution of Ligands by EXponential enrichment)<sup>62</sup>, or by discovery of natural riboswitches<sup>134</sup>. Since the binding events have a high inhibitory potential, are not toxic or immunogenic, and can be produced synthetically, aptamers are regarded as powerful genetic manipulation tools and therapeutic reagents<sup>60</sup>. So far, at least hundreds of molecules possess reported aptamer sequences, and a list of these molecules is shown in **Figure 5.1**. Theoretically, the SELEX method enables us to find aptamer sequences for an arbitrary molecular ligand. Combined with rapid progresses in RNA regulator engineering, this seems to provide an extremely powerful approach for regulation of gene expression in response to custom cellular inputs. However, most aptamers discovered so far mainly bind to nucleotide-like molecules, such as theophylline, AMP, FMN, etc. Aptamers that can bind to more complex compounds exist, including hoechst dyes and antibiotics, but finding their sequences is still challenging and remains a laborious process. On the contrary, discovering aptamers that bind to peptides or proteins might be easier. Thus, designing aptamer-based RNA devices to sense proteins might be a more feasible task in short-term applications.



**Figure 5.1** A list of molecules with known RNA aptamer sequences. The aptamers that bind to theophylline or MS2 coat protein are used for the study.

## 5.2 Strategy to Engineer Allostery

A large number of *trans*-acting antisense RNA molecules, including pT181 antisense and IS10 antisense, function solely as wires of gene circuits, linking the activity of one genetic element to another<sup>38</sup>. They therefore possess a well-defined and often highly tunable mechanism to regulate transcription or translation of target genes, but lack the ability to directly sense cellular signals. To expand the versatility of these non-coding RNAs (ncRNAs), we propose to engineer ligand-inducible switching capability into *trans*-acting ncRNA molecules by fusing them with molecule-sensing RNA aptamers. We expect the strategy of adding ligand sensing ability to ncRNA regulators will add an extra sensory layer to the increasingly sophisticated designs of circuits that utilize ncRNA regulations. Unlike small interfering RNAs (siRNAs) in eukaryotes, bacterial *trans*-acting ncRNAs usually possess highly structured conformations<sup>38</sup>. Examples of allosteric interactions between structured RNA elements and aptamers exist in natural *cis*-acting riboswitches<sup>134</sup>, which usually consist of two structural motifs fused together - an aptamer, and an expression platform that converts ligand binding at the aptamer into a structural rearrangement that can block or allow the formation of functional hairpins<sup>135</sup>. Our aptamer-ncRNA fusions are designed to recapitulate this ligand-dependent structural rearrangement, and we hypothesize that disruptions on ncRNA structure could be actuated through the designed fusion of RNA aptamers that would interfere with the ncRNA structure until bound to its cognate ligand<sup>110</sup> (**Figure 5.2**).

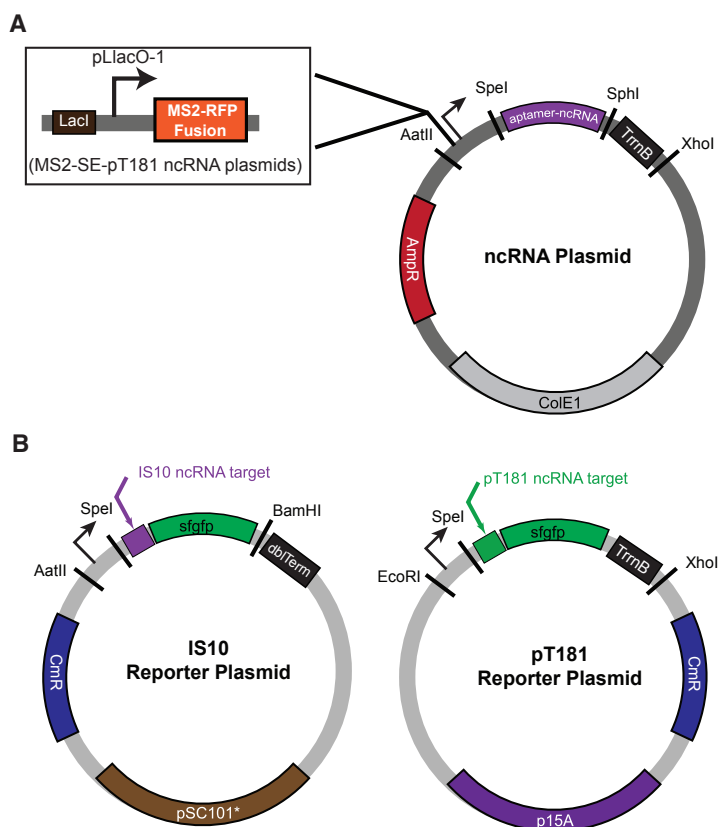


**Figure 5.2 The proposed riboswitch-like design for the aptamer-ncRNA fusion.** (A) A riboswitch usually consists of an aptamer (blue color) coupled to a structured expression platform (red color). The structural interaction between the aptamer and the expression platform could inactivate (left) or activate (right) the expression platform depending on the presence of ligand. (B) We propose to fuse the aptamer and the ncRNA (green color) in a similar architecture. The designed structural interaction between the aptamer and the ncRNA mutation regions (blue color) would inactivate ncRNA without the ligand (left). Ligand binding could eliminate such structural interactions and activate the ncRNA function (right).

## 5.3 Engineering pT181 To Sense Theophylline

To start, we utilized the well-known theophylline aptamer obtained by the *Systematic Evolution of Ligands by EXponential enrichment method* (SELEX)<sup>62</sup> as a demonstration for sensing small molecules. The theophylline-sensing aptamer is a single RNA hairpin that binds theophylline in an inner loop region with high affinity ( $K_d = \sim 300$  nM)<sup>136,137</sup>. The plasmids used for engineering aptamer-ncRNA fusions and for assaying fluorescence expression is shown in **Figure 5.3**. In all the designs, we introduced rational mutations into the desired region to create allosteric

interactions between ncRNAs and fused RNA aptamers. In the case of pseudoknot design, the mutations are designed in such a way that they can base pair with the loop region of the ncRNA to different levels of complementarity. For each mutant design, we used Mfold<sup>65</sup> to compute the substructures of the fusion molecule within 5% of lowest thermodynamic free energy. An arbitrary number of designs was attempted to cover a broad range of complementarity such that some designs showed uniformly inactivated substructures within 5% free energy difference (so the aptamer and ncRNA structures folded together), some showed uniformly activated substructures (so the aptamer and ncRNA structures folded separately), and some showed mixed substructures (candidates for switchable fusions). Then the computationally verified fusion molecules were subsequently experimentally tested for their conformational switching ability. In the design of the theophylline-sensing IS10 and theophylline-sensing pT181 fusions based on pseudoknot designs, the first attempted group of mutants all contained one functional design that was later assayed in detail. Similarly procedure was used in the case of strand exchange design. Most of the designed group contained functional fusions except the design of theophylline-sensing pT181 fusions based on strand exchange, in which the first attempted group consisting of seven mutants didn't show observable allosteric switching and a second group consisting of another 8 mutants was tested again until a functional mutant was found.



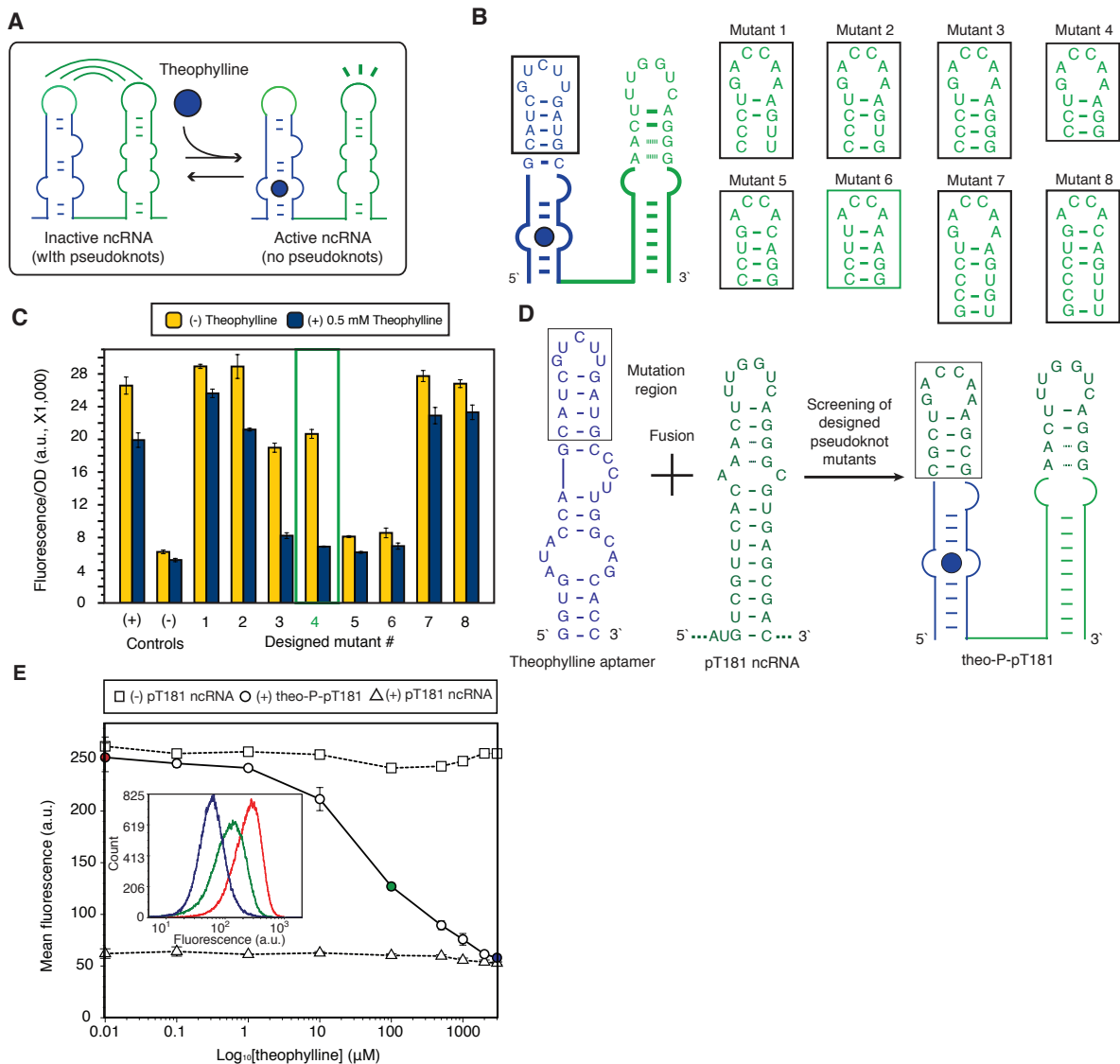
**Figure 5.3 Maps for the aptamer-ncRNA fusion and IS10/pT181 reporter plasmids.** The MS2-sensing pT181 ncRNA plasmids contain an MS2 coat protein or MS2-RFP protein fusion cassette under the control of an IPTG-inducible promoter,  $P_{LlacO-1}$ .

### 5.3.1 Pseudoknot design

The pT181 antisense RNA is highly structured, which makes initial contact with the sense attenuator through a loop-loop kissing interaction<sup>107</sup>. Mutating the loop nucleotides can decrease its attenuation function<sup>106</sup>. Furthermore, these studies have shown that single nucleotide mutation

of the loop nucleotides resulted in almost complete loss of repression. On the other hand, previous studies have shown mutations in the loop region of the theophylline aptamer were tolerated as long as the loop structure was preserved<sup>137</sup>. This allowed us to mutate the loop of the theophylline aptamer to create a hypothesized pseudoknot interaction between the aptamer and the antisense RNA. We therefore sought to design a pseudoknot interaction between the loop regions of the aptamer and the ncRNA, such that the aptamer loop nucleotides could interact with the ncRNA loop nucleotides to disrupt its regulatory function. It was shown that a single hairpin of antisense RNA was sufficient to repress the target. To simplify our design, we used only this single hairpin to fuse to RNA aptamer. When fused to pT181 antisense RNA, we hypothesize that this pseudoknot would inhibit its regulatory activity in the absence of theophylline. In the presence of theophylline, we expect that the aptamer structure would be stabilized by the ligand, eliminating the pseudoknot interaction between the loops and restoring pT181 antisense RNA function (**Figure 5.3.1A**).

We rationally designed eight different aptamer-pT181 ncRNA fusion mutants bearing different aptamer loop mutations and tested them *in vivo* (**Figure 5.3.2B**). Each variant was transformed into *E. coli* Top10 cells with a reporter plasmid containing the pT181 attenuator target controlling the downstream fluorescent protein sfGFP. Colonies were picked and grown with and without theophylline, and fluorescence was measured using the plate reader with two controls: a positive control that lacked an aptamer-ncRNA sequence (high GFP), and a negative control that expressed the IS10 ncRNA not fused to any aptamer (low GFP) (**Figure 5.3.2C**). From this set, mutant #4 was distinguished by both obtaining nearly the same maximal GFP expression as the positive control and exhibiting the largest dynamic range of repression between the two theophylline conditions. To further characterize this mutant, called theo-P-pT181, we assayed GFP expression for varying amounts of theophylline using flow cytometry (**Figure 5.3.2D**). All fluorescence histograms showed a single peak. A large dynamic range was observed, which was close to the difference between positive and negative controls (**Figure 5.3.2E**). This data implied that, as designed, the fused aptamer interfered with the function of ncRNA only in the absence of theophylline, and that interaction of theophylline with the aptamer released the ncRNA hairpin to regulate its target.



**Figure 5.3.1 Pseudoknot design for fusing pT181 antisense with theophylline aptamer. (A)** Schematic of the design. **(B)** Rational mutations of the aptamer loop nucleotides (blue) are designed to base pair with the pT181 ncRNA loop nucleotides (green), yielding eight designs. **(C)** Fluorescence assay with the pT181 reporter plasmid using the plate reader as shown in the bar graph. Mutant #4 (in the green box), which is theo-P-pT181, was selected for further study. **(D)** Sequence of theo-P-pT181. **(E)** Flow cytometry results of theo-P-pT181 with varying concentrations of theophylline.

### 5.3.2 Strand exchange design

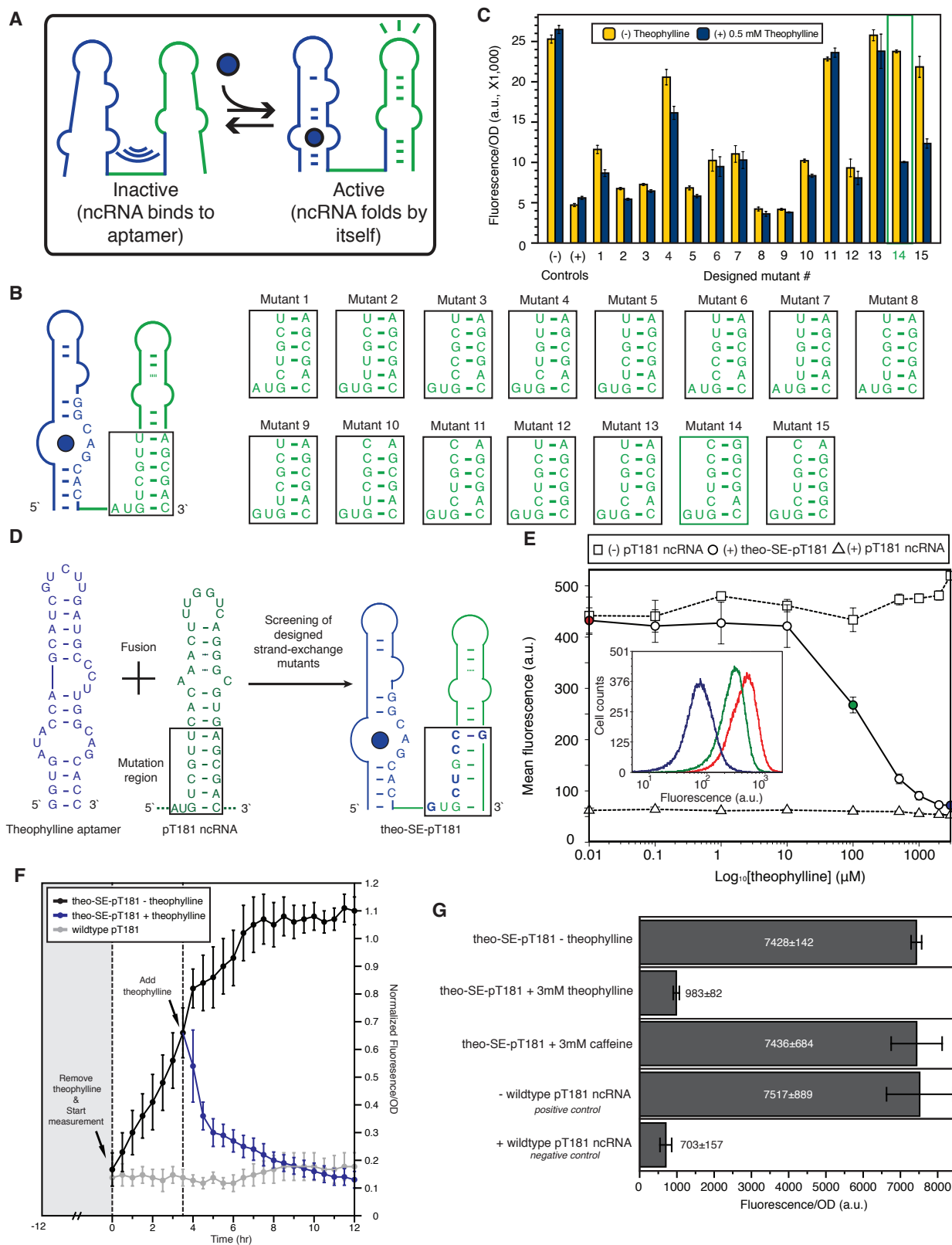
While the pseudoknot design strategy is effective, it is not inherently modular as the aptamer loop sequences must be adjusted to the ncRNA loop sequences. This is particularly relevant to the pT181 system, where the loop region of the pT181 antisense RNA has been exclusively mutated to create orthogonally acting ncRNA-attenuator systems. We hypothesize that we could engineer a more modular aptamer-ncRNA fusion by designing an interaction between the

aptamer and the ncRNA stem region, while leaving the ncRNA loop nucleotides for modifying target specificity (**Figure 5.3.2A**). In this design, named “strand exchange”, one strand on the ncRNA stem is mutated to exhibit exchange between two possible conformations of the aptamer-ncRNA fusion molecule: one where the strand is base paired to the aptamer causing the ncRNA to be non-functional (ncRNA inactivated), and one where the strand is base paired to the other strand of ncRNA to restore ncRNA function (ncRNA activated). Based on previous characterization of nucleotide important of the pT181 antisense, we decide to mutate the nucleotides in the lower stem that can tolerate moderate modifications without affecting antisense regulatory function. We hypothesize that theophylline binding would cause strand exchange between the conformations and bias the population of fusion molecules towards the activated conformation.

We therefore mutated this lower stem region of the pT181 antisense to base pair with the ligand-binding pocket of the theophylline aptamer. We designed fifteen aptamer-ncRNA fusion variants following the strand exchange strategy (**Figure 5.3.2B**), and measured their repression on the pT181 sense target with and without theophylline (**Figure 5.3.2C**). One mutant (#14) with the highest dynamic range, called theo-SE-pT181, was selected for assaying with varying amounts of theophylline using flow cytometry (**Figure 5.3.2D**). **Figure 5.3.2E** showed that repression by the theo-SE-pT181 fusion molecule (83.4%) almost covered the full dynamic range (86.6%) of the wildtype ncRNA regulator. The data implied that the strand exchange design was acting like an allosteric switch, with theophylline triggering a change in the conformations to allow the pT181 ncRNA to repress its target.

To confirm this switching behavior, we measured the temporal expression of the pT181 reporter gene under the control of theo-SE-pT181, which showed that its expression was gradually repressed to the level of the negative control around 3~4 h (**Figure 5.3.2F**). While this temporal measurement does not tell the switching speed of aptamer-ncRNA fusion, because gene expression rate is mainly limited by the slow protein degradation, it illustrates that the switching of designed allosteric device. The theophylline aptamer used in the study does not bind caffeine that differs from theophylline by a single methyl group<sup>136</sup>. We found that caffeine did not, as expected, repress gene expression in this system, implying that these aptamer-ncRNA fusions retained a high level of ligand specificity (**Figure 5.3.2G**).

To support our mechanistic hypothesis, we attempted to perform SHAPE on the theo-SE-pT181 fusion. However, the ncRNA hairpin prevented reverse transcriptase from transcribing the full-length molecule (**Figure 5.6.2C**), thus precluding direct measurement of the proposed structural transitions. As an alternative approach, we computed the thermodynamic free energies of substructures of four closely-related aptamer-pT181 fusion variants using the RNA secondary structure prediction algorithm Mfold<sup>65</sup> (**Figure 5.8**). The data suggested that the thermodynamic free energy of the two folds of the aptamer-pT181 fusion were balanced in such a way that allowed theophylline binding to bias the fold towards the active conformation. The computation results also suggested several principles in the design of switchable aptamer-ncRNA fusions as was detailed below.



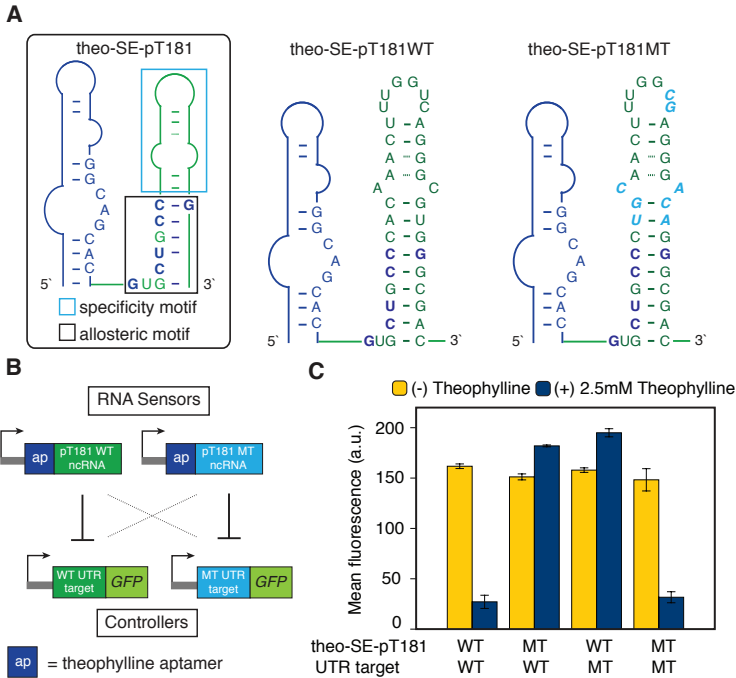
**Figure 5.3.2 Strand exchange design for fusing pT181 antisense with theophylline aptamer. (A) Schematic of the design. (B) Rational mutations of the aptamer loop nucleotides (blue) are**

designed to base pair with the pT181 ncRNA loop nucleotides (green), yielding 14 mutant designs. (C) Fluorescence assay with the pT181 reporter plasmid using the plate reader as shown in the bar graph. Mutant #14 (in the green box), which is theo-SE-pT181, was selected for further study. (D) Sequence of theo-SE-pT181. (E) Flow cytometry results of theo-SE-pT181 with varying concentrations of theophylline. (F) The temporal fluorescence expression of the pT181 reporter system under the control of theo-SE-pT181. Cells were incubated overnight with the presence of 3mM theophylline (shaded region). At time  $t = 0$ , cells were washed and resuspended in minimal media without theophylline, and temporal measurement of fluorescence was started (black). At  $t = 220$  mins, 3mM theophylline was supplemented (blue). The cells with constitutive wildtype pT181 ncRNA expression were shown as the negative control (grey). All data were normalized to the positive control without ncRNA expression. The error bars show the standard deviation of biological triplicates. The data suggest that the cells are fully repressed within 4 hours. (G) The theo-SE-pT181 distinguishes theophylline from caffeine. 3mM of theophylline and 3mM of caffeine were used. Average fluorescence/OD values of biological triplicates are shown. While theophylline addition significantly repressed the target, addition of the same concentration of caffeine had no effects.

### **5.3.3 Engineering mutant pT181 antisense RNA to sense**

We next study if the strand exchange strategy allows modularly fusing orthogonal pT181 antisense RNAs with RNA aptamers. Since the specificity mutations locate outside of the lower stem region of the ncRNA that base-paired with the aptamer, we used the same strand exchange design to produce fusions with the mutant pT181 antisense RNA, called theo-SE-pT181MT (**Figure 5.3.3A**). We tested the orthogonality of theo-SE-pT181WT and theo-SE-pT181MT against their cognate targets with and without theophylline (**Figure 5.3.3B**). Measured GFP expression for all four possible combinations of ncRNAs and their targets under different theophylline conditions showed that the theo-SE-pT181MT fusion responded to theophylline in the same way as the theo-SE-pT181WT fusion did (**Figure 5.3.3C**). Furthermore, both fusions were orthogonal relative to each other's target. These results demonstrated the modularity of the strand exchange design, and implied that these ligand-sensing fusion molecules could perform similarly as the pT181 antisense RNAs in sophisticated genetic networks, but with added functionality to fine-tune their activity with small molecules.



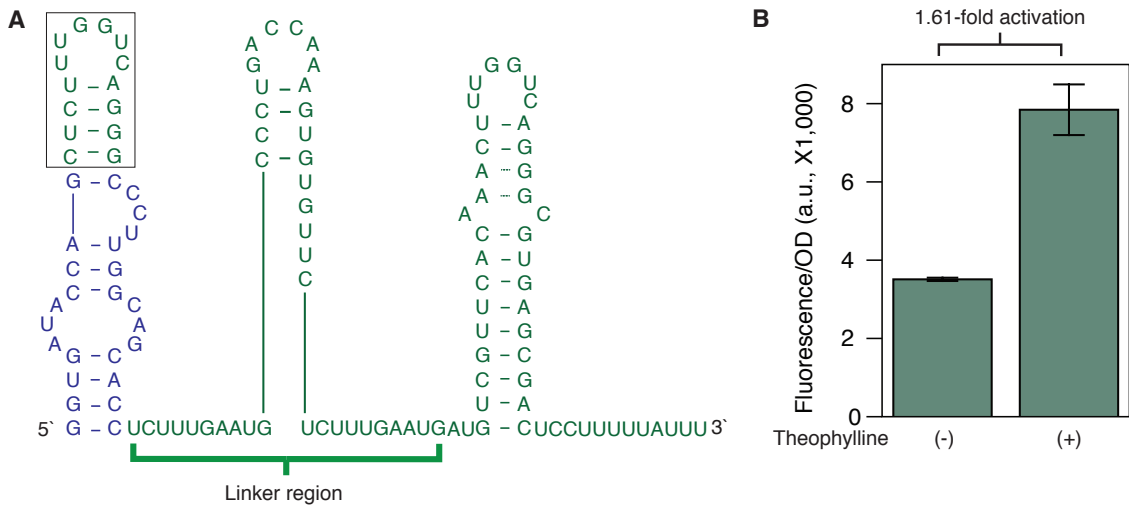


**Figure 5.3.3 Orthogonal theophylline-sensing pT181 fusions using strand exchange.**

(A) The region that determines the target-specificity of pT181 ncRNA is in the cyan box<sup>109</sup>, and allosteric region is in the black box. (B) Experimental setup to test theophylline-induced target specificity of theo-SE-pT181WT and theo-SE-pT181MT. (C) Fluorescence assay data with red bars showing (-) theophylline and blue bars (+) showing 2.5 mM theophylline.

### 5.3.4 Engineering theophylline-sensing pT181 activators

We tried to engineer fusion molecules that would activate gene expression in response to ligands. To do this, we inserted the +1 to +29 nucleotides of pT181 attenuator sequence into the theo-P-pT181 fusion as a linker region connecting the 5' aptamer and 3' antisense sequences (**Figure 5.3.4A**). However, the design didn't work well, and only exhibited moderate activation (1.61-fold) upon addition of theophylline (**Figure 5.3.4B**). Several ways might enable a better activator, including tuning the lengths of the linker region, screening for the linker compositions and selection. This is left for future study.



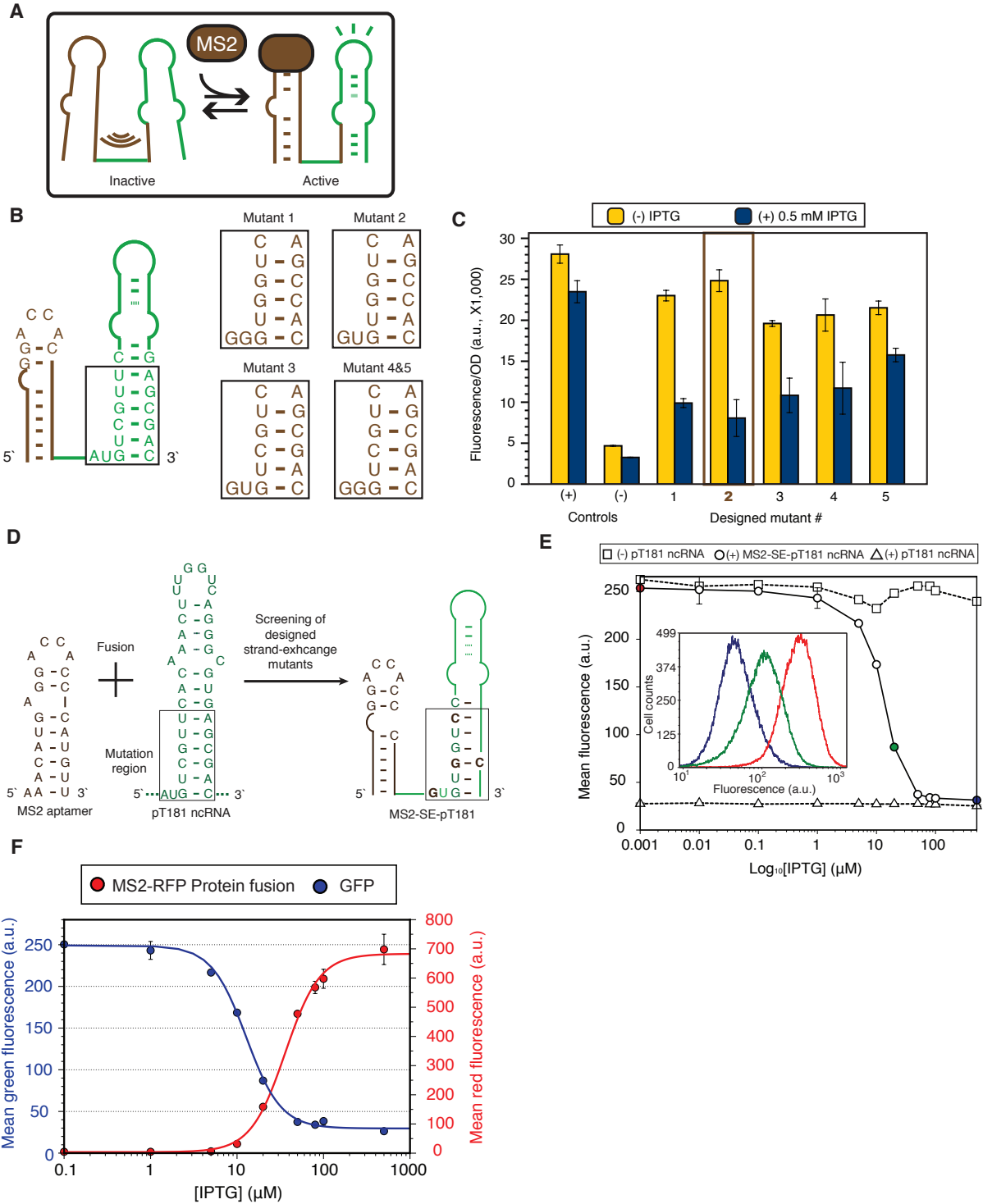
**Figure 5.3.4 pT181-based activator in response to theophylline.** (A) The best design. (B) Characterization of the performance.

## 5.4 Engineering pT181 to Sense MS2 Coat Protein

Equally important to sensing small molecules would be the ability of ncRNAs to sense protein concentrations and alter their regulatory functions accordingly. Here we used the MS2 coat protein sensing aptamer for several reasons. First, the natural MS2 coat protein aptamer is well-studied and binds to MS2 coat protein with high affinity ( $K_d = 20 \text{ pM}$ )<sup>138</sup>. Second, it has been shown that the MS2 coat protein could be fused to other proteins<sup>139</sup>, serving as an adapter to sense other intracellular proteins *in vivo*<sup>140</sup>. Third, it has recently been used to demonstrate a class of RNA control devices that can couple the abundance of desired proteins to targeted gene expression through alternative RNA splicing<sup>141</sup>.

### 5.4.1 Engineering wildtype pT181 antisense RNA to sense

Following the strand exchange design strategy, we mutated the lower stem of the pT181 ncRNA hairpin to base pair with the MS2 aptamer nucleotides that were critical for MS2 coat protein binding<sup>142</sup> (**Figure 5.4.1A**). We expected that without the MS2 coat protein, the aptamer and ncRNA would fold together to disrupt the antisense RNA structure and inhibit its function; when the MS2 coat protein was present and bound to the aptamer, the protein ligand would restore the pT181 antisense RNA structure and function. We designed five fusion mutants with different levels of base pairing between the ncRNA and the aptamer, and tested them in *E. coli* with the pT181 reporter plasmid (**Figure 5.4.1B**). We used the IPTG-inducible promoter  $P_{\text{LlacO-1}}$  to induce the expression of MS2 coat protein from the same ColE1 plasmids that harbored the fusion variants (**Figure 5.3A**). GFP expression was measured with and without IPTG for each variant. All designs were able to switch into an active state upon induction of the MS2 coat protein (**Figure 5.4.1C**), and one variant (#2) with the best performance, called MS2-SE-pT181 (**Figure 5.4.1D**), was selected for further study using flow cytometry (**Figure 5.4.1E**). The repression of MS2-SE-pT181 fusion molecule almost covered the full dynamic range of the wild type antisense RNA regulator. Once again, the data suggested that strand exchange design was acting like a switch, with the presence of MS2 coat protein triggering a change in the conformation of the fusion molecule to that allowed pT181 ncRNA to repress its target. Further, using the fusion protein of the MS2 coat protein and RFP, we observed a near-identical induction curve, indicating that MS2 protein could indeed serve as an adapter for sensing other intracellular proteins (**Figure 5.4.1F**).

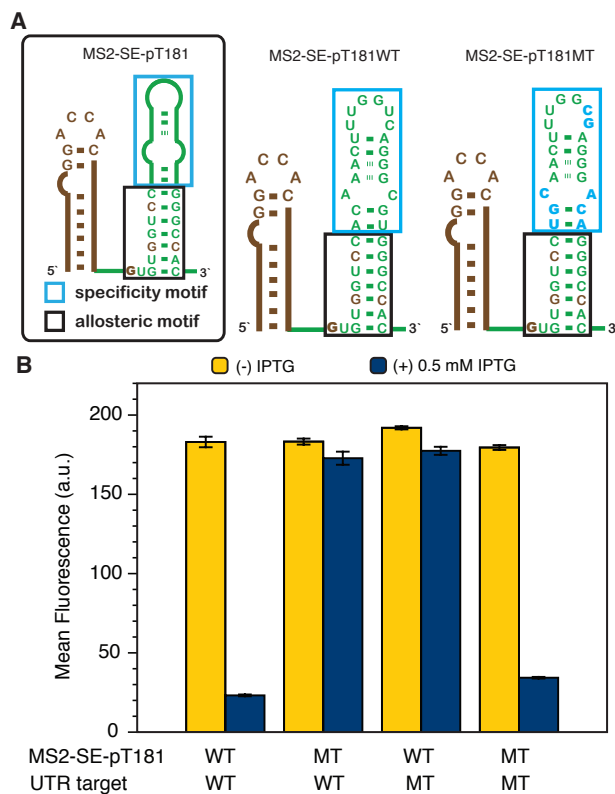


**Figure 5.4.1 Designed MS2 coat protein-sensing pT181 ncRNA fusions.** (A) Schematic of the design strategy. (B) Rational mutations of the ncRNA stem region (green) to base pair with the MS2 aptamer (brown) resulted in 5 mutant designs. Mutant #2, which is MS2-SE-pT181, in the brown box was selected for further study. Mutant #5 has the same ncRNA stem mutations as mutant #4 but with a different MS2 aptamer sequence. (C) Their regulatory activity was assayed

with the pT181 reporter plasmid using the plate reader as shown in the bar graph. (D) Sequences of the MS2-SE-pT181 aptamer-ncRNA fusion screened from five designed mutants. (E) Fluorescence assay of MS2-SE-pT181 ncRNA fusions with intracellular MS2 coat protein induced by IPTG. The repression percentage between 500  $\mu$ M and 1 nM IPTG is 87.5% compared to 89.1% between positive and negative controls. The inset shows the cytometry histogram of three IPTG concentrations. (F) The MS2-SE-pT181 ncRNA fusion can sense MS2-RFP protein fusions. The expression of MS2-RFP protein fusion is induced by IPTG and their expression is shown in red colors. Responsive to the MS2-RFP protein fusion expression, the expression of GFP controlled by pT181 ncRNA decreases as shown in blue colors. The correlation between the two curves demonstrated that the MS2-SE-pT181 fusion could serve as protein concentration sensor for arbitrary proteins by fusing them with the MS2 coat protein.

#### 5.4.2 Engineering mutant pT181 antisense RNA to sense

To confirm that the design of MS2-SE-pT181 was modular with respect to pT181 mutants with different target specificities, we constructed the MS2-SE-pT181MT variant by directly swapping the target specificity motif of the ncRNA, and tested the orthogonality with and without MS2 coat protein expression (Figure 5.4.2A). Similar to the theo-SE-pT181WT and theo-SE-pT181MT designs, the MS2-SE-pT181WT and MS2-SE-pT181MT fusions acted orthogonally and only showed repression on their cognate targets in the presence of MS2 coat proteins (Figure 5.4.2B).

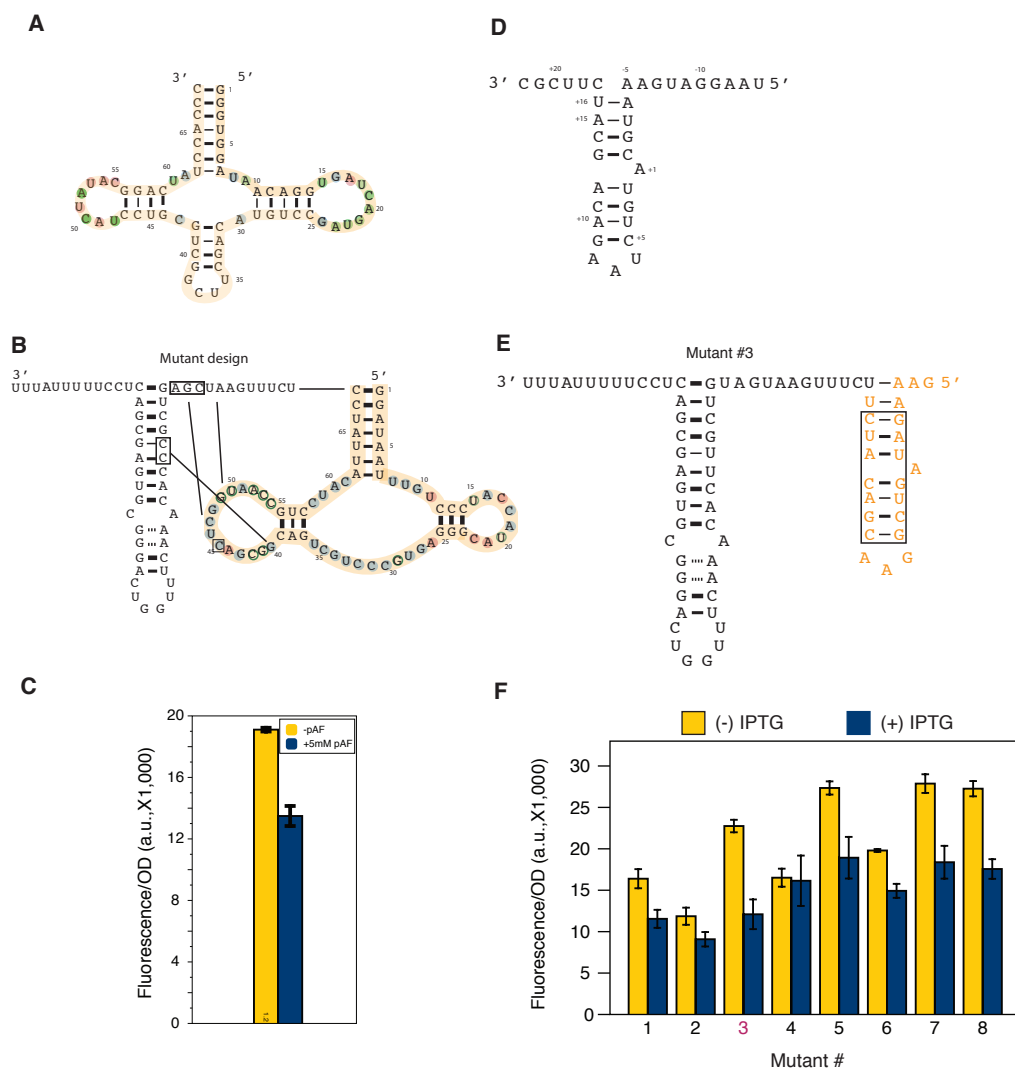


**Figure 5.4.2 Orthogonal MS2-sensing pT181 fusions using strand exchange.**

A different target recognition region of pT181MT ncRNA (in cyan box) was modularly fused to MS2-SE-pT181WT, resulting in MS2-SE-pT181MT. All four possible combinations between aptamer-ncRNA fusions and reporters were assayed with and without IPTG, and showed high specificity towards their own target.

### 5.5 Engineering pT181 to Sense Other Molecules

We tried to apply the strand exchange strategy to fuse pT181 antisense to an aptamer that could sense *p*-aminophenylaniline (pAF). The aptamer for pAF is relatively large and contain 3 stem-loops<sup>143</sup> (**Figure 5.5A**). We mutated the lower stem region of the pT181 antisense to make this region base paired to one stem-loop of the pAF aptamer. However, none of the designs exhibited satisfactory repression in response to 5 mM pAF. The fusion mutant with the best ON/OFF ratio is shown in **Figure 5.5B&C**. It is likely that the base paired stem-loop is not critical for pAF function. Another possibility is that binding between the ligand and the pAF aptamer reported is not strong enough to induce conformational change of the fusion molecule ( $K_d = 3600$  nM). Future studies are required to fuse this aptamer to the antisense RNA. We also tried to fuse the pT181 antisense to Q $\beta$  aptamer. The Q $\beta$  RNA aptamer is small (**Figure 5.5D**) and can bind to Q $\beta$  coat protein with high affinity ( $K_d = 4$  nM)<sup>144</sup>. Following very similar strand exchange strategy for fusing pT181 antisense with MS2 aptamer, we mutated the lower stem region of the antisense sequence. Out of eight fusion designs, one mutant (#3) showed highest level of allosteric switching (47% of repression percentage) (**Figure 5.5E&F**). In these experiments, similar to the case of MS2 coat protein, the Q $\beta$  protein is induced from a P<sub>L</sub>lacO-1 promoter with 0.5 mM IPTG. Further studies are required to optimize this fusion. We also tried to fuse pT181 antisense RNA to atrazine RNA aptamer<sup>145</sup>, but it didn't work well for all tested designs.



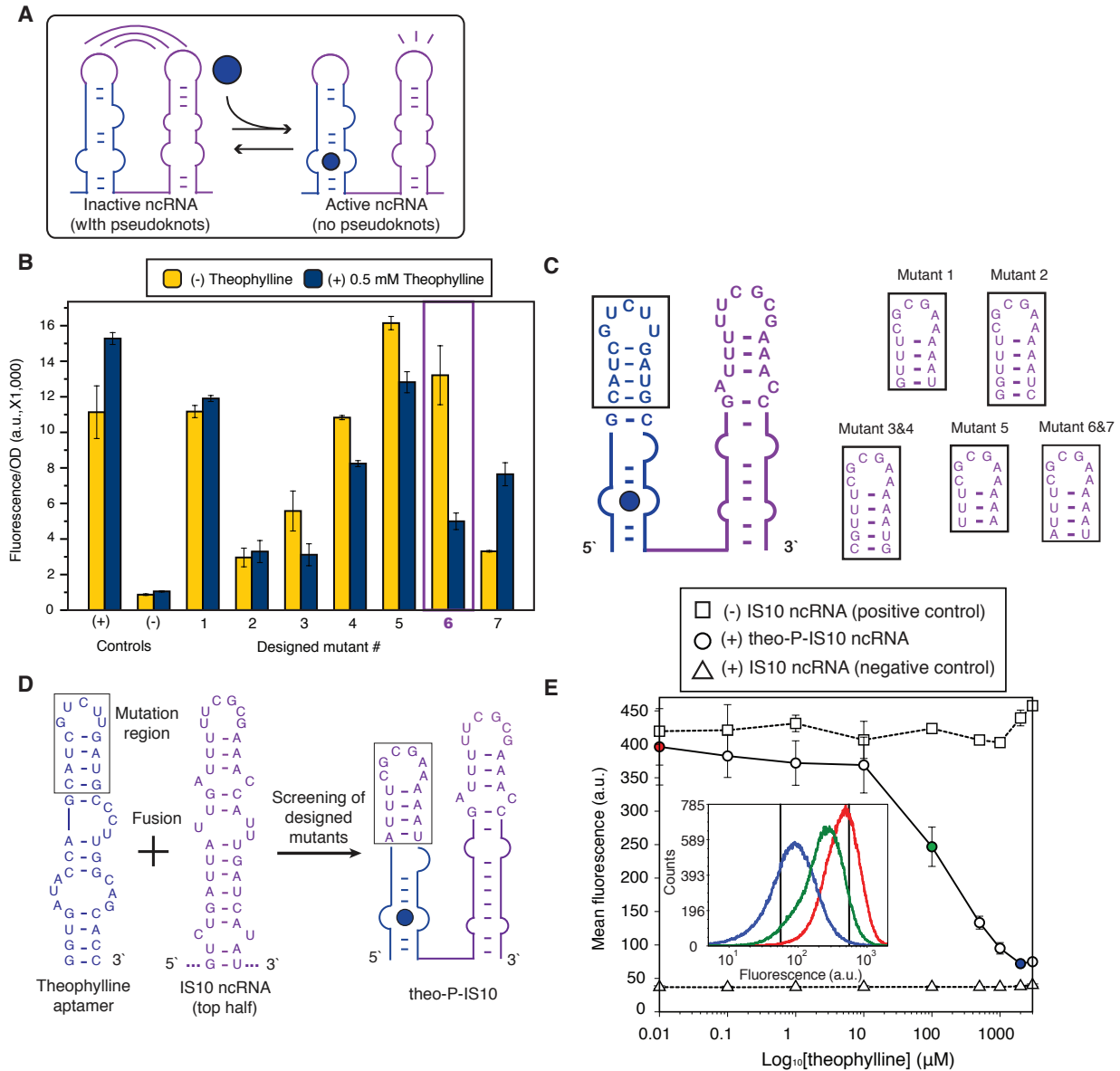
**Figure 5.5 Engineering pT181 antisense to sense other ligands.** (A) Selected RNA aptamer sequence that binds to pAF. (B) The “best” design of pAF-pT181. (C) Characterization of the “best” design of pAF-pT181. (D) RNA aptamer that binds to Q $\beta$  coat protein. (E) The “best” design of Q $\beta$ -pT181. (F) Characterization of the “best” design of Q $\beta$ -pT181.

## 5.6 Engineering IS10 Antisense RNA to Sense Theophylline

### 5.6.1 Use the pseudoknot strategy

To engineer IS10 antisense to sense theophylline, we used a single hairpin of the antisense RNA. In-depth biochemical studies have shown that the loop of IS10 ncRNA binds to the 5' most nucleotides of its target transcript, which nucleates further hybridization into the ribosome binding site of the target<sup>125</sup>. Furthermore, these studies have shown that single nucleotide mutation of the loop nucleotides resulted in almost complete loss of repression. On the other hand, IS10 antisense possesses a long stable stem, which makes the strand exchange harder. We therefore sought to design a pseudoknot interaction between the loop regions of the aptamer and the ncRNA, such that the aptamer loop nucleotides could interact with the ncRNA loop nucleotides to disrupt its regulatory function (**Figure 5.6.1A**).

We rationally designed seven different aptamer-IS10 ncRNA fusion mutants bearing different aptamer loop mutations and tested them *in vivo* (**Figure 5.6.1B**). Each variant was transformed into *E. coli* with a reporter plasmid containing the IS10 ncRNA UTR target controlling the downstream fluorescent protein sfGFP. Colonies were picked and grown with and without theophylline, and fluorescence was measured using the plate reader with two controls: a positive control that lacked an aptamer-ncRNA sequence (high GFP), and a negative control that expressed the IS10 ncRNA not fused to any aptamer (low GFP). Four variants (#3~#6) showed different levels of switching ability in the presence of theophylline (**Figure 5.6.1C**). From this set, mutant #6 was distinguished by both obtaining nearly the same maximal GFP expression as the positive control and exhibiting the largest dynamic range of repression between the two theophylline conditions (**Figure 5.6.1D**). To further characterize this mutant, called theo-P-IS10, we assayed GFP expression for varying amounts of theophylline using flow cytometry (**Figure 5.6.1E**). All fluorescence histograms showed a single peak. A large dynamic range (83.0%) was observed, which was close to the difference between positive and negative controls (91.0%). This data implied that, as designed, the fused aptamer interfered with the function of ncRNA only in the absence of theophylline, and that interaction of theophylline with the aptamer released the ncRNA hairpin to regulate its target.



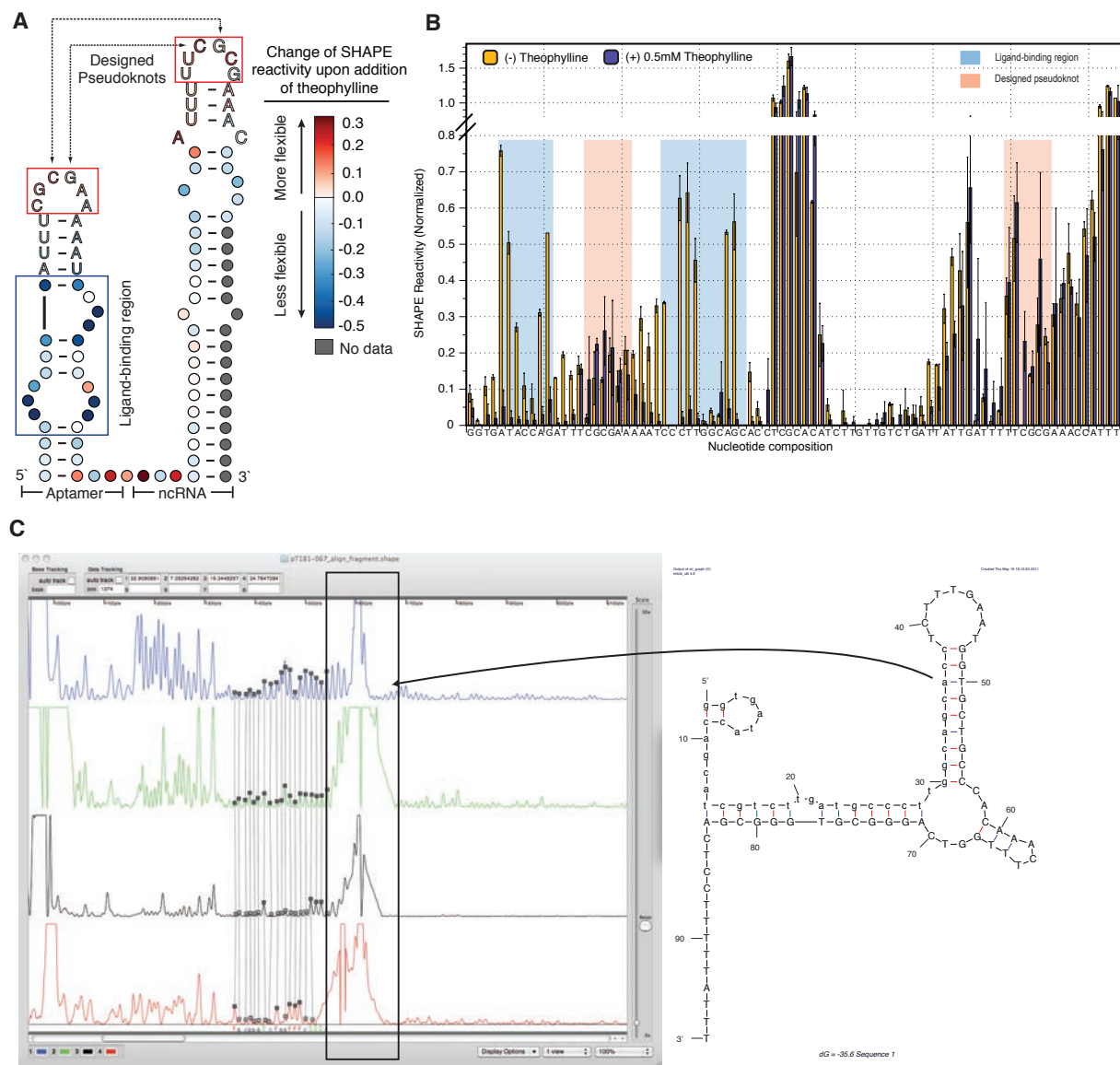
**Figure 5.6.1 Designed theophylline aptamer-IS10 ncRNA fusions.** (A) The proposed mechanism for allosteric switching of the fusion molecule. (B) Rational mutations of the aptamer loop nucleotides (blue) are designed to base pair with the IS10 ncRNA loop nucleotides (purple), yielding seven mutant designs. Mutant #6, which is theo-P-IS10 shown in the purple box, was selected for further study. Mutant #3 and #4 have the same loop sequences but with different ncRNA stem sequences. (C) The results of their regulatory activity with the IS10 reporter plasmid using the plate reader. Mutant #7 acts as an activator. (D) Sequence of theo-P-IS10 (P for pseudoknot). (E) Fluorescence assay of theo-P-IS10 using flow cytometry. The induction curves were plotted from the average values of three biological replicates at each theophylline concentration. The inset shows the cytometry histograms of three ligand concentrations (red – 0.01  $\mu\text{M}$ , green – 100  $\mu\text{M}$ , blue – 2 mM), with the two black vertical lines showing the mean values of the positive and negative controls. The repression percentage between 2 mM and 0.01  $\mu\text{M}$  theophylline is 83.0%, compared to 91.0% between the positive and negative controls.

### 5.6.2 SHAPE reactions to verify allosteric switching

To support our proposed allosteric switching mechanism, we carried out Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) experiments on the theo-P-IS10 fusion molecule to characterize its structure with and without theophylline. The SHAPE experiment uses a structure-dependent chemical probe, here 1-methyl-7-nitroisatoic anhydride (1M7), to modify RNA molecules preferentially at positions of high nucleotide flexibility<sup>146</sup>. After modification, RNAs are converted to cDNAs by a reverse transcriptase primer extension reaction, which is blocked by adducts. cDNA products are then analyzed by capillary electrophoresis to infer the location of adduct formation via cDNA length, and the intensity of adduct formation at each nucleotide via the amount of cDNA of a given length. These intensities are converted into reactivities, with higher reactivity interpreted as positions of lower or no RNA structure, and lower reactivity interpreted as positions of strong RNA structure or nucleotide constraint<sup>146</sup>.

We transcribed the theo-P-IS10 fusion molecule *in vitro*, which was then folded with and without theophylline, followed by SHAPE probing and analysis (**Figure 5.6.2B**). **Figure 5.6.2A** plots the difference in nucleotide reactivity between the two folding conditions overlaid on a secondary structure model of the theo-P-IS10 fusion. From this data, two reactivity changes can be discerned. First, as expected, the ligand-binding pocket of the aptamer shows a large drop in reactivity upon addition of theophylline, indicating that these nucleotides become constrained as they directly interact with theophylline. Secondly, both aptamer and IS10 loops show consistent but small increases in reactivity, suggesting that these nucleotides probably become unconstrained in the presence of theophylline. While the SHAPE analysis data is consistent with our hypothesis and *in vivo* fluorescence assay data that a designed pseudoknot interaction forms only in the absence of theophylline to prevent the ncRNA loop region from interacting with its target, we cannot rule out the possibility that multiple aptamer-ncRNA fusions dimerize with each other to affect their structures and functions in a synergetic way.



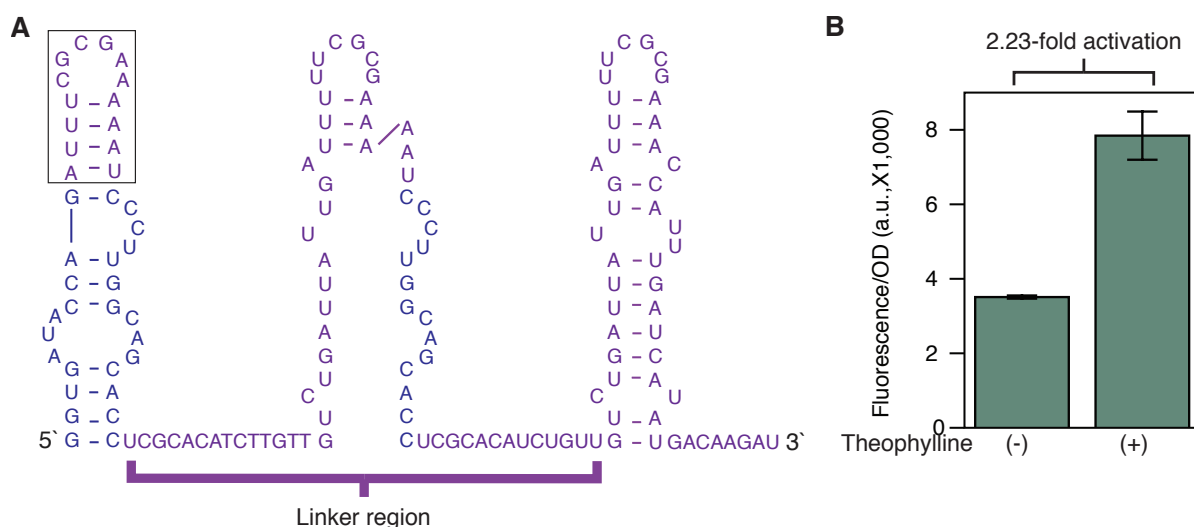


**Figure 5.6.2 SHAPE data.** (A) The difference in nucleotide reactivity with and without the ligand is overlaid on a hypothesized secondary structure model of theo-P-IS10<sup>124</sup>. Colors represent the changes of SHAPE reactivity upon addition of the ligand, with red colors showing positive changes (more flexible) and blue colors showing negative changes (more stable). The blue box is the known ligand-binding pocket. (B) SHAPE data for the theo-P-IS10 ncRNA fusion for (-) theophylline (orange) and (+) theophylline (blue). The nucleotide compositions are shown on the x-axis. The ligand-binding region<sup>137</sup> is colored in blue and the designed pseudoknot regions are colored in red. Higher values of SHAPE reactivity mean higher flexibility of the nucleotides, and lower values of SHAPE reactivity mean lower flexibility. (C) Raw SHAPE capillary electrophoresis traces data of theo-SE-pT181 (blue: (+) channel, green: (-) channel, black: ddT channel, red: ddA channel). The fusion molecule sequences proceed with 3' to 5' from left to right. The black box highlights the region that caused problematic reverse transcription reaction, and corresponded to the design strand-exchange region on the fusion molecule. The secondary structure of the fusion molecule was generated by Mfold<sup>65</sup>, and only

the substructure with lowest free energy is shown here. Capital letters on the secondary structure are ncRNA sequences, and small letters are aptamer sequences.

### 5.6.3 Engineering theophylline-sensing IS10 activators

We tried to design a type of activator in response to theophylline, which would sequester antisense RNA structure (thus activate gene expression) when theophylline binds to the aptamer. To do this, we inserted the +1 to +39 nucleotides of antisense RNA and the 3' end of the aptamer stem as a linker region into the 5' theo-P-IS10 fusion molecule (**Figure 5.6.3A**). Surprisingly, this designed exhibited more than 2-fold activation on gene expression (**Figure 5.6.3B**). While there is no structural data to explain the activation effects, we expect that the linker region might serve as a buffer sequence to couple RNA aptamer and antisense. In the absence of ligand, the linker binds to the RNA aptamer, thus the 3' antisense is activate to repress; when ligand is present, ligand stabilizes RNA aptamer and releases the linker region to disrupt the 3' antisense structure and function, turning on gene expression. We note that while this design is illuminating, it is not optimal, and further optimization is needed to create a completely activating switch (supposedly more than 10-fold activation).



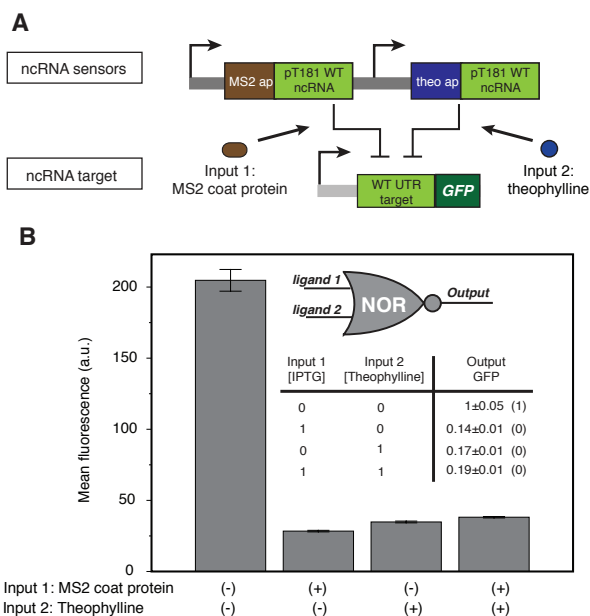
**Figure 5.6.3 IS10-based activator design in response to theophylline.** (A) The design. (B) characterization of its performance.

## 5.7 Engineering Circuits Using Synthetic ncRNA Sensors

### 5.7.1 Sensory-level NOR

To demonstrate the flexibility of using these engineered ncRNA sensors to regulate gene expression, we used the theo-SE-pT181WT and MS2-SE-pT181WT ncRNA fusion molecules to regulate the same gene via the wild type pT181 sense attenuator (**Figure 5.7.1A**). We expected that in the presence of either theophylline or MS2 coat protein, one of the aptamer-ncRNAs would be functional and repress the expression of the target gene. In this way, this system should be able to integrate two cellular signals and act like a NOR logic. Indeed, *In vivo* fluorescence data indicated that GFP expression was high only when there was no ligand present, and

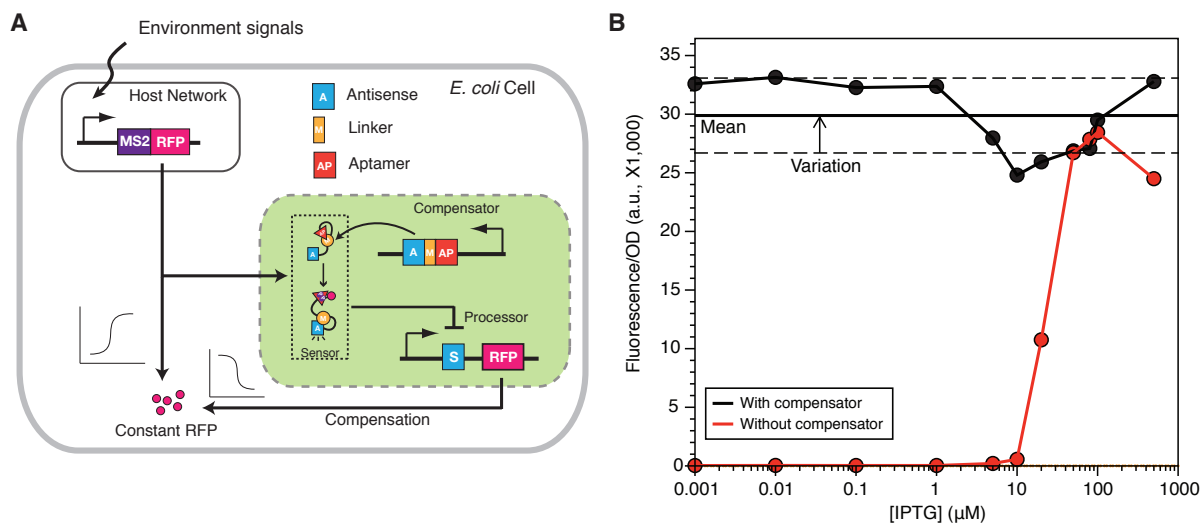
presence of either theophylline or MS2 coat protein induced by IPTG or both consistently repressed the reporter gene expression to a low expression (**Figure 5.7.1B**). This demonstrated the use of designed orthogonal aptamer-ncRNA fusions in logically integrating cellular signals on the sensory-level. Compared to the NOR logic using two orthogonal pT181 ncRNA-attenuator systems<sup>109</sup>, here we showed that only one pT181 ncRNA-attenuator pair was required for the same regulatory function, implying sensory-level engineering using orthogonal ncRNA sensors could further increase the scales and complexities of synthetic circuits.



**Figure 5.7.1 Engineering a sensory-level NOR logic circuit.** (A) The two aptamer-ncRNA fusions, theo-SE-pT181WT and MS2-SE-pT181WT, are used to control the same target and integrate ligand signals in a way that presence of any ligand represses the target gene expression. (B) Experimental results. A NOR truth table is shown in the plot.

### 5.7.2 Engineering a type of signal compensator

We engineered a type of endogenous protein concentration compensator based on the aptamer-ncRNA fusions. Our protein compensator buffers the variation of a target protein concentration due to environment changes, a property key to the reliable function of certain gene networks. In our model compensator, our target protein for concentration stabilization is mRFP. The compensator functions by fusing MS2 to the target protein and then compensating fluctuations in this protein by inverting its signal via an MS2-SE-pT181 module. Thus when the host-derived protein concentration increases, the expression of the “compensatory” protein is decreased and *vice versa*. In our demonstration, MS2-fused mRFP, the target protein, is expressed from the promoter  $P_{LlacO-1}$  in *E. coli*, and induction by IPTG simulates varying environmental signals whose fluctuation would cause unwanted variation in protein expression (**Figure 5.7.2A**). Our results show that, varying concentrations of IPTG from 1  $\mu$ M to 100  $\mu$ M caused more than 1,000-fold change in mRFP expression in cells without the compensator circuit. Expression of total mRFP remained relatively stable in cells with the compensator, with less than 10% variation measured by RSD values around the set mean value at steady-state (**Figure 5.7.2B**).



**Figure 5.7.2 Engineering a synthetic protein concentration compensator.** (A) The protein concentration compensator contains a sensing unit based on MS2-SE-pT181 to drive compensatory expression of target protein copy in response to the variation in a target protein fusion concentration (MS2 coat protein fused to RFP). (B) Demonstration of the protein concentration compensator with IPTG as the simulated environmental signal. While *E. coli* cells without concentration compensator showed over 1,000-fold change in endogenous mRFP expression (red), *E. coli* cells with concentration compensator only showed less than 10% variation within the set mean value (black). The solid line is the mean value of mRFP and dotted lines shows mean value  $\pm$  standard deviation.

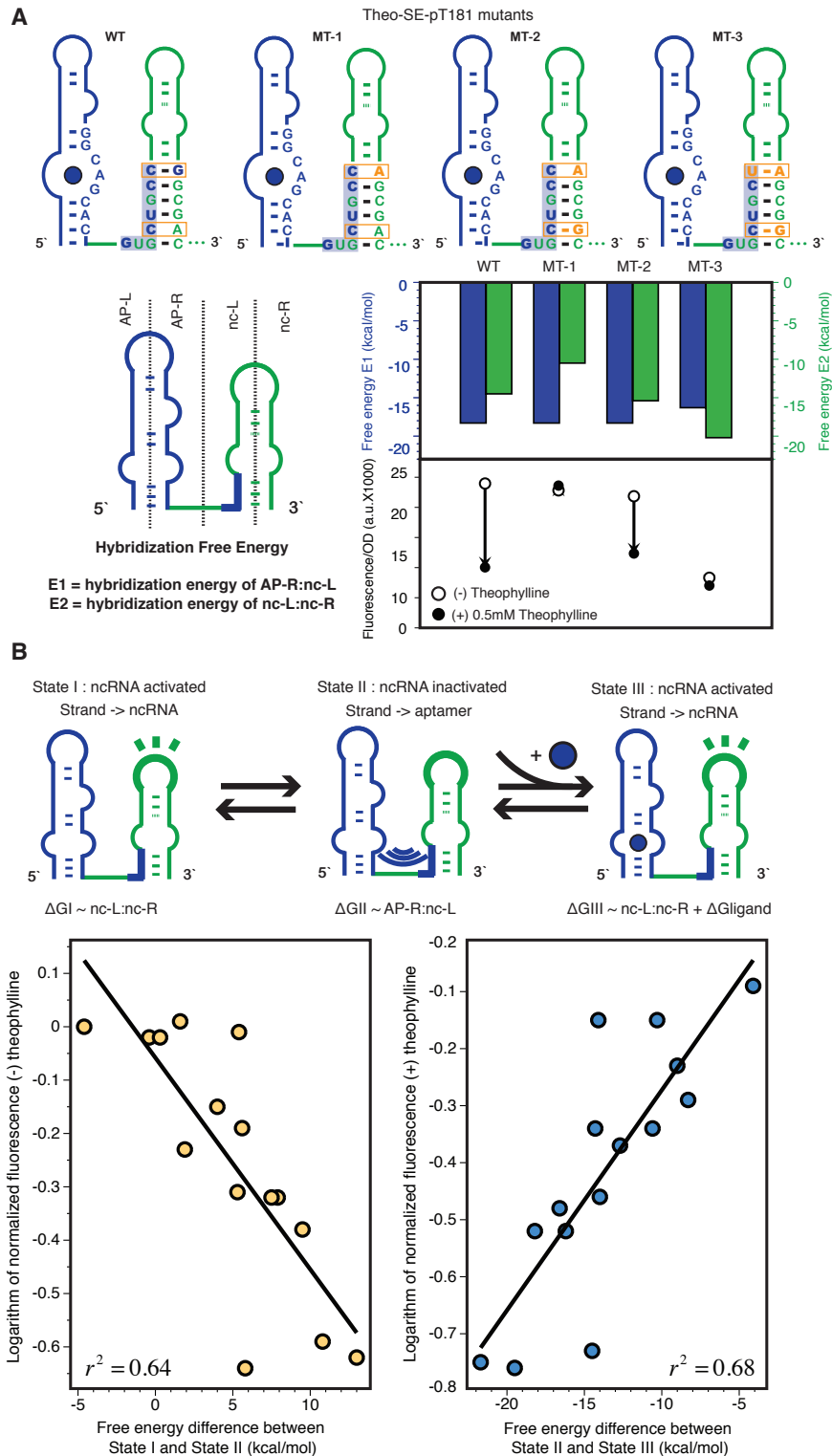
## 5.8 Discussion on The Design Principles

In this work, we have demonstrated two alternative designs to rationally engineer natural ncRNAs to sense ligands by fusing them with RNA aptamers. The IS10 and pT181 ncRNAs are both highly structured and likely to represent a class of ncRNA regulators whose function highly depends on the structure<sup>38</sup>. In our designs, we fused RNA aptamer sequences to the 5' end of the ncRNA molecules in the similar architecture as natural riboswitches. We introduced mutations into different locations in different designs to disrupt the ncRNA structure: the mutations were on the aptamer loop region to form pseudoknots with the ncRNA loop (pseudoknot design); or the mutations were on the ncRNA lower stem region to make this region exchangeable between alternative conformations (strand exchange design). In both designs, ligand binding eliminated disruptions on the ncRNA hairpin and activated its function.

In all designs, we have utilized the RNA structure prediction algorithm, Mfold<sup>65</sup>, to compute and verify that the disrupted conformation had the lowest thermodynamic free energy. Thus, without ligands, the structure of ncRNA was disrupted. Since secondary structure prediction algorithms cannot account for the free energies of ligand binding, we dissected the secondary structure of aptamer-ncRNA fusion to estimate its overall folding energy. We conceptually divided the aptamer-ncRNA fusions into 4 regions corresponding to the left (L) and right (R) halves of the aptamer and ncRNA hairpins (Figure 5.8A). For each variant, we calculated the free energy of the substructure that caused inactivation (R-aptamer/L-ncRNA) and compared it to the free energy of the intact ncRNA hairpin (L-ncRNA/R-ncRNA). Comparing

the computed folding energy of theo-SE-pT181 (WT) with those of three closely related variants (MT-1, MT-2, MT-3), we discovered that allosteric switching was not only determined by the interaction between the ncRNA and the aptamer, but also by the ncRNA folding itself. If there was over-disruption on ncRNA (MT-1), the fusion molecule was always inactivated; if the disruption was much weaker than the ncRNA folding energy (MT-3), the ncRNA was always activated. Only mutants with similar free energy between ncRNA disruption and formation showed allosteric switching (WT and MT-2). Surprisingly, one base-pair formation or elimination in the ncRNA stem could completely change the allosteric switching, indicating the allosteric properties of designed aptamer-ncRNA fusions were sensitive to nucleotide compositions.

Furthermore, we calculated free energy differences for all 15 theophylline-SE-pT181 fusion mutants between alternative conformations (**Figure 5.8B**). We fitted the measured fluorescence data to the calculated free energy differences, and observed almost linear correlation between the two quantities for both with and without the ligand. These experiments also suggest that the hybridization free energy between adjacent RNA strands can serve as a guide for designing allosteric properties of other aptamer-ncRNA fusions.

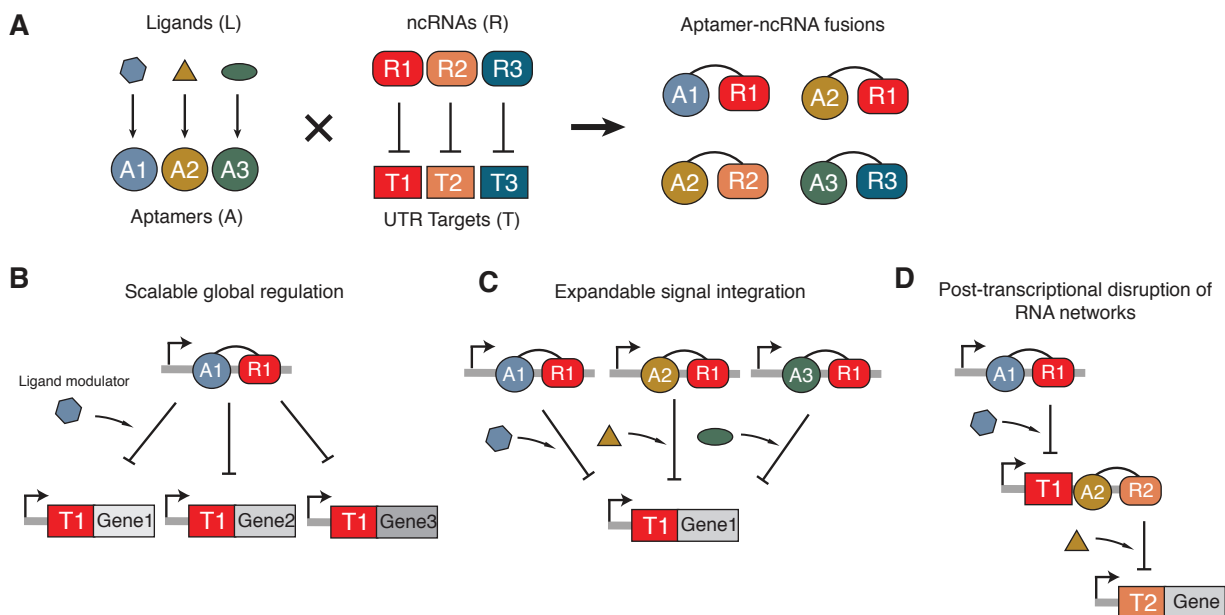


**Figure 5.8 Discussion on allosteric switching properties.** (A) The allosteric switching depends on the hybridization energy between adjacent RNA strands. Four closely related mutants of theophylline-sensing pT181 fusions based on the strand-exchange design. WT is the same as theo-SE-pT181, and the other three only differs one or two nucleotides from WT. The fusion

molecule is dissected into four adjacent strands, and the hybridization free energy between the adjacent strands is calculated using Mfold<sup>65</sup>. The calculated hybridization free energy terms for E1 (AP-R:nc-L) in blue colors and E2 (nc-L:nc-R) in green colors are shown on the top. Fluorescence assay of the regulatory activity on the pT181 reporter plasmid for four mutants are shown on the bottom. White circles: (-) theophylline; black circles: (+) 0.5mM theophylline. **(B)** Correlation between measured fluorescence data and calculated free energy differences for 15 theophylline aptamer-pT181 ncRNA fusions using the strand exchange design. A three-state model to account for the allosteric switching of the strand exchange design is shown on the top. Correlation data between the fluorescence data and the free energy difference is shown on the bottom, with the left plot showing the case for (-) theophylline, and the right plot showing the case for (+) 0.5mM theophylline. The  $R^2$  values for the plots are 0.64 and 0.68.

### 5.9 Strengths and Limitations of The Fusion Design

We have demonstrated a modular strategy for engineering ligand-responsive ncRNAs. This builds from a growing body of work on engineering gene circuits using RNA mechanisms that respond to ligands. Previous work has demonstrated that *cis*-acting RNA elements residing in the 5'<sup>42,145</sup> or 3'<sup>44,52</sup> UTR sequences can respond to small molecules, and there was pioneering work to engineer riboregulators in yeast to control translation in response to small molecules<sup>147</sup>. Our work here further expands the category of RNA molecules that can control gene expression in a ligand-inducible way to the naturally occurring *trans*-acting ncRNAs that modulate 5' UTR functions. This is particularly valuable for transcriptional ncRNA regulators, as transcriptional regulators have been shown as useful genetic parts that could be systematically tethered together to form logics and other higher-order regulatory systems.



**Figure 5.9 Modularity of aptamer-ncRNAs at the molecular and network levels.** **(A)** Combining orthogonal ncRNAs with different aptamers could provide a toolbox of orthogonally acting aptamer-ncRNA fusions that sense multiple ligands and control their cognate targets with high specificity. **(B)** The scalable global regulation allows regulation of multiple gene targets in

the same cell by responding to a global ligand modulator. (C) The expandable signal integration allows regulation of a single gene target by integrating multiple signal inputs in a logic way. (D) The posttranscriptional disruption of ncRNA networks allows us to debug and fine-tune the performance of individual regulators in complicated networks such as transcriptional cascades.

For the demonstration purpose, we fused theophylline- or MS2 coat protein-sensing aptamers to ncRNAs. So far, many RNA aptamers have been reported by *in vitro* selection<sup>148,149</sup> or through the discovery of natural riboswitches<sup>134</sup>. These aptamers can bind specifically to different ligands, including nucleotide-like molecules such as flavin mononucleotide (FMN)<sup>150</sup> and dopamine<sup>151</sup>, complex compounds such as tetracycline<sup>152</sup> and vitamin B12<sup>153</sup>, and proteins such as HIV-1 rev peptide<sup>154</sup> or nicotine acetylcholine receptor<sup>155</sup>. While our design strategy worked well for theophylline and MS2 coat protein aptamers that possess relatively simple structures, to design functional fusions between complex aptamers and ncRNAs, the detailed information about their structures and mechanisms is needed. We foresee the advent of high throughput RNA structure characterization methods such as SHAPE-Seq<sup>64</sup> will facilitate our design process. Such methods will identify the regions of aptamers that exhibit the appropriate flexibility for intramolecular interaction with the ncRNA following our design strategy. When connected to environmentally or medically relevant ligands, we expect these ligand-responsive RNA regulators offer novel capabilities to engineer cellular behaviors in response to different stimuli.

By focusing on switchable *trans*-acting aptamer-ncRNA fusions, we are extending this capability by introducing modularity on both molecular and network levels. First, at the molecular level, the modular design of aptamer-ncRNA fusions allows the ligand sensed to be switched with no further adjustment (**Figure 5.9A**). We have shown that transcriptional regulation by the same pT181 ncRNA molecule but modulated by the ligands of theophylline or MS2 coat protein. The only difference between the two experiments is the particular aptamer-ncRNA fusion molecule expressed, and the target of these aptamer-ncRNA fusions remains the same. To achieve this with *cis*-acting elements, the genetic context of the regulatory target would have to be changed by manipulating the 5' or 3' UTRs since *cis*-acting elements are inherently tightly coupled with their regulatory target. However, unlike *cis*-acting elements, the *in trans* nature of our regulatory elements allows the possibility of regulation of off-target genes. Nonetheless, many ncRNAs are highly specific or can be engineered to be so<sup>109,132</sup>. An assumption of our study is that the fusion of an aptamer to the ncRNA does not strongly affect the properties of components. This is supported in the present case by the maintenance of the quantitative behavior of the fused ncRNA compared to the wildtype ncRNA. However, it is possible that the fusion may interfere with the function and specificity of both components, - an eventuality that would need to be diagnosed by structural and functional studies. Although we only demonstrated the NOR logic by having two different aptamer-ncRNA fusions to control a single target, it is conceivable that other types of sensory-level logic such as NAND, AND, and OR can be theoretically implemented by placing the complex logic into the fusion molecules as suggested by other researcher<sup>44</sup>. For example, previous studies have demonstrated that AND and OR functions could be engineered by fusing two different aptamers to the same ribozyme<sup>44</sup>, and we expect a similar design could also work for the aptamer-ncRNA fusions. We plan to investigate this topic in our future study.



Second, at the gene network level, *trans*-acting aptamer-ncRNA fusions have the potential to globally change the regulation of many targets in response to customizable cellular signals (**Figure 5.9B**). Many natural ncRNAs have multiple targets in the cell<sup>37</sup>, and by engineering the capability to fine-tune the post-transcriptional function of these ncRNAs in response to ligands, this work creates the opportunity to globally switch on or off many targets in the cell simultaneously. In addition to adding flexibility in an engineering context, this could offer powerful capabilities to the scientific toolkit to study ncRNA biology through dynamic switching of global ncRNA regulators<sup>156</sup>. Since we have shown that our aptamer-pT181 designs are modular with respect to both aptamers and orthogonal pT181 variants, there are numerous places that we can integrate these fusion molecules into existing capabilities to create more sophisticated RNA/protein or RNA-only circuits with unique regulatory properties. Generally, transcriptional aptamer-ncRNAs could be used as alternatives to inducible promoters by having ncRNA responding directly to the ligands after being expressed from constitutive promoters; multiple signal integration at the sensor level can compute logics on external stimuli whose outputs could be fed into transcriptional logics for creation of advanced logic gates (**Figure 5.9C**); and aptamer-ncRNA fusion molecules themselves could serve as signal-transmitting molecules in RNA/protein or RNA-only networks to provide ligand-inducible control over network connection (**Figure 5.9D**). The ability to sense diverse signals (ncRNAs, proteins, and metabolites) and integrate them to execute gene regulatory programs is key to the detection or implementation of specific cellular responses in complex environments<sup>131</sup>. The modularity of aptamer-ncRNAs might offer powerful extensions to our capabilities for rationally engineering *trans*-acting RNA regulators-mediated genetic circuits, and the ubiquity of such mechanisms makes them potentially useful across a diverse range of organisms, from prokaryotes to humans<sup>147</sup>. Coupled with the rapid increase in our understanding of the regulatory roles of *trans*-acting ncRNAs, this work opens the door for engineering sensory-level gene regulations that are modular, flexible and versatile based on naturally occurring *trans*-acting ncRNAs.

## Chapter 6 Other RNA-Related Genetic Regulatory Systems

### 6.1 Leveraging Transcription and Translation Controls

In previous two chapters, we introduced two primary types of regulators found in 5'-UTRs that served as prototypes for designing new parts. One is the regulator of transcriptional elongation and the other is the regulator of translational initiation. To engineer these 5'-UTRs into useful regulatory parts that sense custom inputs and change the expression of desired genes, we generally need to satisfy two criteria:

*Criterion 1:* the regulatory parts must be easily engineered, in a way that yields large sets of orthogonal variants that respond to different custom inputs with similar functions;

*Criterion 2:* the parts must be composable such that they can be assembled predictably into useful higher-order functions.

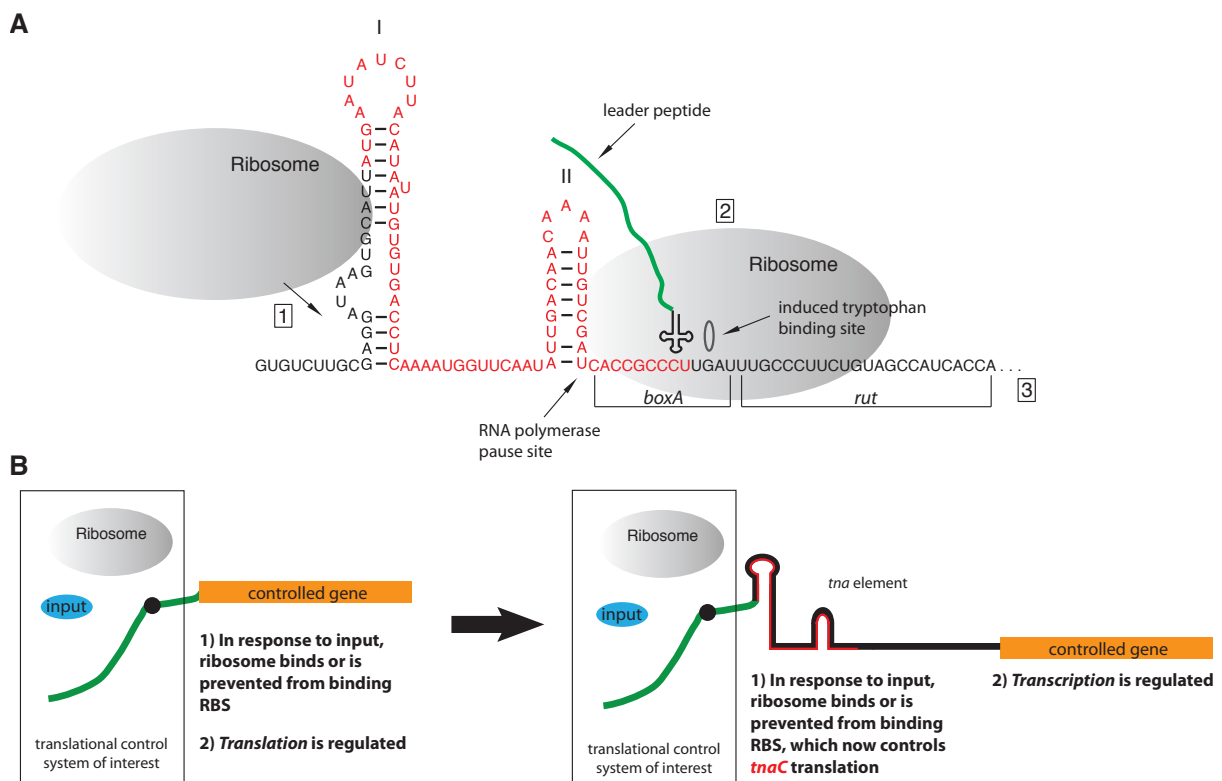
The regulators of transcriptional elongation such as pT181 system link cellular inputs to the processivity of RNA polymerase during RNA synthesis. One unique feature is that these regulators control both the production of coding and non-coding RNAs and can act on entire operons containing multiple genes. Furthermore, they are usually highly composable because when multiple cis-regulators of transcriptional elongation are linked in tandem in a 5'-UTR, the synthesis of the  $N^{\text{th}}$  regulator is gated by the decision of the  $(N-1)^{\text{th}}$  regulator. This yields predictable logic and other higher-order functions<sup>109</sup>. However, one limitation is that such regulators are difficult to engineer because their mechanisms involve action on a moving RNA polymerase, requiring the consideration of poorly defined kinetic and dynamic structural factors in their design. The existence of only a handful of synthetic regulators of transcriptional elongation testifies to this difficulty<sup>109,157</sup>. Therefore, regulators of transcriptional elongation satisfy *criterion 2* but not *criterion 1*.

The regulators of translational initiation such as IS10 system link cellular inputs to the accessibility of ribosome binding sites (RBSs) after RNA synthesis. From a parts perspective, these regulators are not only because they are well-represented in nature (*e.g.* riboswitches that sense small molecules<sup>42</sup>, antisense RNA repressors<sup>132</sup> and activators<sup>86</sup>, and RBSs responsive to proteins<sup>41</sup>, nutrients<sup>134</sup>, and temperature<sup>158</sup>), but also because RBS-based interactions can be tuned and even predicted *de novo* using thermodynamic models<sup>114,132</sup>. However, regulators of translational initiation cannot control the production of non-coding RNAs. They are constrained to act on single coding genes, and cannot be composed into complex regulatory functions as initiation at RBSs is a distributive process. Therefore, regulators of translational initiation satisfy *criterion 1* but not *criterion 2*.

To create genetic parts that satisfy both criteria, two categories of regulators should be combined in smart way. To do this, we leverage a natural leader peptide mediated system from the *tna* operon of *E. coli* as described below to convert translational regulatory events into transcriptional events<sup>159</sup>.

#### 6.1.1 Leader peptide mediated system provides a conversion strategy

Our conversion utilizes leader peptide mediated control. The system derived from *tna* operon contains, from 5' to 3', an RBS, the coding region for a short leader peptide (*tnaC*), a Rho factor-binding site, and a stretch of RNA required for Rho factor-mediated transcriptional termination. The controlled genes with their individual RBSs follow. In the *tna* element's mechanism, full translation of *tnaC* results in ribosomal stalling and blockage of a Rho factor-binding site adjacent to *tnaC*'s stop codon<sup>160,161</sup> (**Figure 6.1.1A**). Thus, unsuccessful *tnaC* translation results in Rho-mediated transcriptional termination, whereas successful translation of *tnaC* prevents Rho-mediated termination, allowing transcription of the controlled genes. In this way, the *tna* leader-peptide element couples translation of *tnaC* to transcriptional elongation of the downstream gene. Therefore, we hypothesized that replacing the native RBS and upstream context of *tnaC* with a desired cis-regulator of translational initiation would turn it into a cis-regulator of transcriptional elongation. This forms our general conversion strategy (**Figure 6.1.1B**).



**Figure 6.1.1 Mechanism for the wild-type *tna* leader-peptide element.** (A) The wild-type *tna* leader-peptide element's regulatory mechanism relies on the interaction between the last 12 residues of the leader peptide (TnaC) and the ribosome. This interaction causes a conformational change in the ribosome that creates a tryptophan-binding site. If this binding site is occupied by free tryptophan (in this work, free tryptophan is always present), ribosomal release is inhibited. Therefore, if translation of *tnaC* initiates (Step 1) and *tnaC* is fully translated, the ribosome stalls over the natural stop codon of *tnaC*, blocking the adjacent *rut* site (Step 2). This prevents Rho factor-mediated transcriptional termination, thus allowing the continuation of RNA polymerase into the controlled genes (Step 3). Nucleotides in red correspond to the *tnaC* open reading frame.

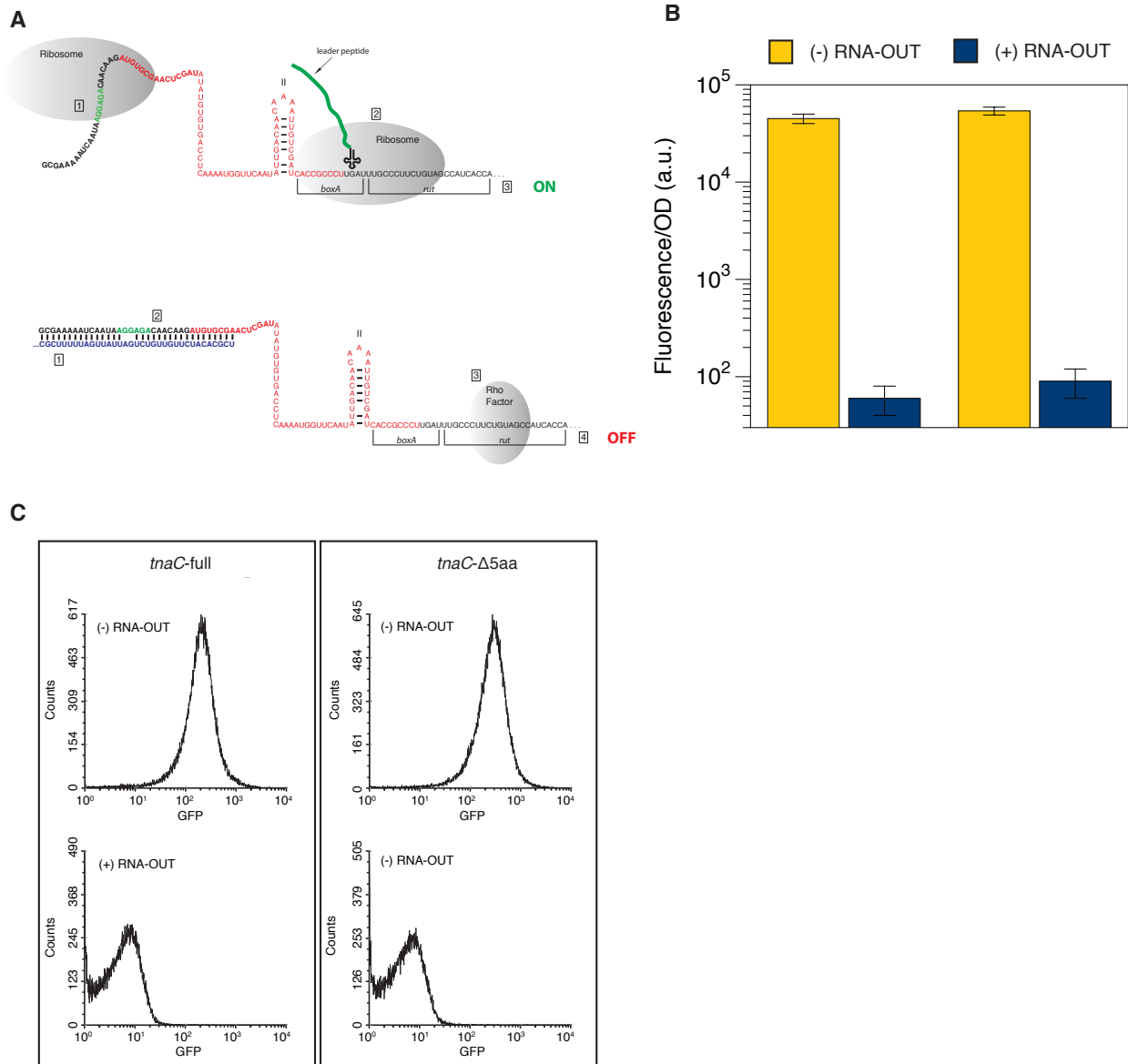
(B) Converting regulators of translational initiation into regulators of transcriptional elongation. The leader peptide's open reading frame (*tnaC*) is in red.

### 6.1.2 Engineering IS10-*tna* system to convert translation to transcription

To test our strategy, we combined the IS10 translational system with the *tna* leader peptide system. We fused RNA-IN to *tnaC* to effect control over its translational initiation by RNA-OUT. This was achieved by removing the first 24 or 39 nucleotides of the *tna* element and inserting the 42-nt RNA-IN unit into the same location. Two fusions resulted: in the first, the complete *tnaC* coding region was retained (*tnaC*-full); in the second, five N-terminal codons (this number corresponds to the five residues added by the coding fragment portion of the RNA-IN unit) were removed from the *tnaC* coding region such that the resulting fusion retained the length, though not identity, of wild type *tnaC* (*tnaC*- $\Delta$ 5aa). These two fusions were then cloned upstream of the sfGFP reporter gene containing its own RBS, which were placed under the control of a strong constitutive promoter on low-copy pSC101 plasmids.

To test whether our converted regulators properly control the transcription of GFP, *tnaC*-full and *tnaC*- $\Delta$ 5aa were tested in *E. coli* Top10 cells along with a second plasmid that expressed either RNA-OUT under the control of a strong constitutive promoter or no RNA-OUT. We expected that cells containing *tnaC*-full and *tnaC*- $\Delta$ 5aa but no RNA-OUT would show strong fluorescence because in the absence of RNA-OUT, translation of the *tnaC* variants should be uninhibited, resulting in transcriptional elongation into the GFP gene (**Figure 6.1.2A**). In contrast, we expected that cells expressing RNA-OUT would display low fluorescence because RNA-OUT inhibits translational initiation of the RNA-IN-*tnaC* fusions, resulting in transcriptional termination before RNA polymerase transcribes the GFP gene. These responses were indeed observed (**Fig. 6.1.2B**), unimodal (**Figure 6.1.2C**). Therefore, combination of IS10 system with *tna* leader peptide control enables successful conversion of translational initiation control into transcriptional elongation control.

Surprisingly, the converted regulators exhibited more than 500-fold attenuation in response to RNA-OUT, a change much higher than observed in other riboregulators. For instance, when RNA-IN directly controls translation of GFP, attenuation by RNA-OUT is about 20-fold<sup>132</sup>. This is of note because one common criticism of designing RNA-based regulators is their low dynamic range compared to regulators based on protein-promoter interactions. Yet in our case, it seems that conversion of the IS10 translational control system results also in an amplification of dynamic range, yielding a highly effective attenuator that rivals the efficiency of promoter-based transcriptional regulation while retaining the composability and scalability that make RNA parts desirable.

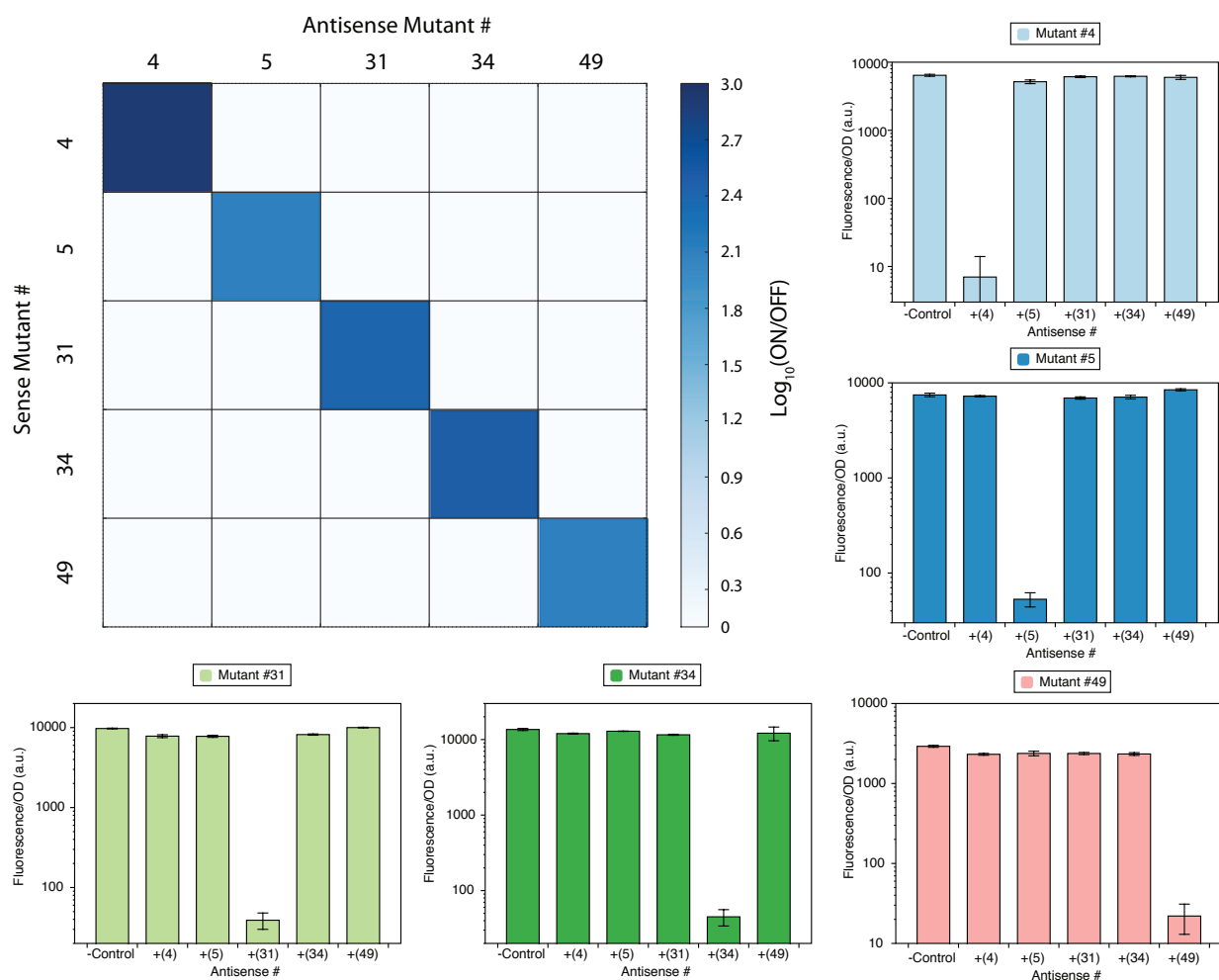


**Figure 6.1.2 Characterization of IS10-*tna* fusions.** (A) Specific conversion strategy for RNA-OUT repression. Nucleotides in green correspond to the RBS controlling translational initiation of *tnaC*. Nucleotides in red correspond to *tnaC*'s open reading frame. Nucleotides in blue correspond to RNA-OUT. When RNA-OUT is absent, the RBS contained in RNA-IN is free and the ribosome initiates translation (Step 1). The leader peptide is fully translated and the ribosome stalls over the natural stop codon of *tnaC*, blocking the adjacent *rut* site (Step 2). This prevents Rho factor-mediated transcriptional termination and RNA polymerase continues transcription into the controlled genes (Step 3). When RNA-OUT is present, the RBS contained in RNA-IN is sequestered by RNA-OUT (Step 1). The ribosome cannot initiate translation (Step 2), and the leader peptide is not synthesized. The *rut* site is therefore free, allowing Rho factor-mediated termination of the continuing RNA polymerase before it reaches controlled genes (Steps 3 and 4). (B) Performance of antisense-mediated regulators of transcriptional elongation achieved through conversion of the corresponding translational control systems. Fluorescence of cells containing design IS10-*tna* fusion is shown. Experiments were conducted in triplicate (error bars

are  $\pm$  standard deviation) on the same day. Data were collected using a fluorescence plate reader. (C) Representative cytometry histograms.

### 6.1.3 Engineering mutually orthogonal transcriptional regulators

We study to apply this conversion strategy to all five pairs and test if these pairs retain their orthogonality. To do this, these five mutually orthogonal variants of RNA-IN were fused to *tnaC* following the same construction as *tnaC*- $\Delta$ 5aa. The resulting plasmids were cotransformed into *E. coli* Top10 cells in conjunction with plasmids expressing RNA-OUT variants. For cells that contained cognate RNA-IN/RNA-OUT pairs, dramatic attenuation of GFP expression was observed in all five cases; for cells containing non-cognate RNA-IN/RNA-OUT pairs, little to no attenuation of GFP expression was observed for all 20 possibilities (**Figure 6.1.3**). Conversion of five mutually orthogonal attenuators of translational initiation into the corresponding regulators of transcriptional elongation was therefore successful, yielding the desired set of highly effective (dynamic ranges approaching 1000-fold) mutually orthogonal transcriptional attenuators.



**Figure 6.1.3 Performance of orthogonal antisense-mediated regulators.** RNA-IN-4, 5, 31, 34, and 49 are mutually orthogonal RNA-IN variants that respond to RNA-OUT-4, 5, 31, 34, and 49. Control refers to the presence of plasmid that expresses no antisense RNA. The heatmap plot

shows the ON/OFF dynamic range in logarithm. The bar plots shows GFP expression for each RNA-IN under different conditions of RNA-OUT. Experiments were conducted in triplicate (error bars are  $\pm$  standard deviation) on the same day. Data were collected using a fluorescence plate reader.

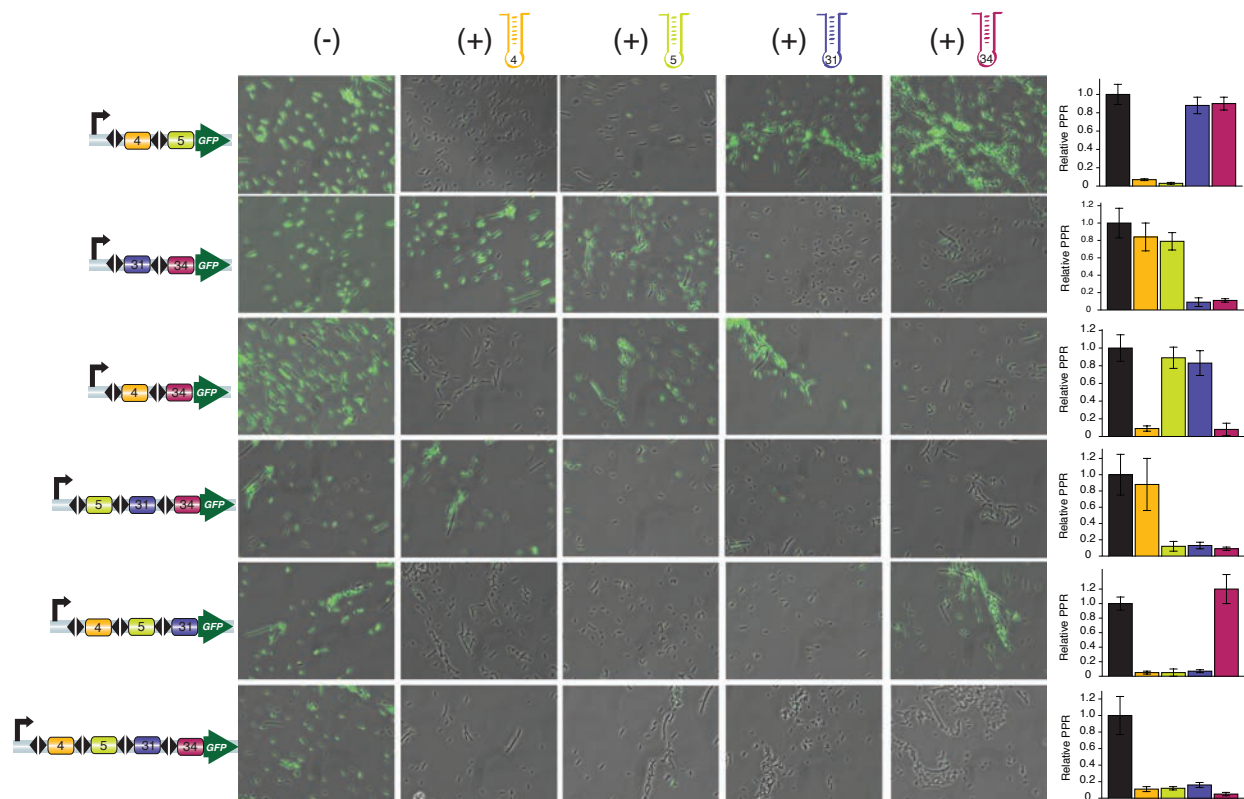
#### **6.1.4 Engineering expandable multi-input logic gates**

Although five mutually orthogonal pairs of have been engineered based on the IS10 system, they cannot be systematically composed into higher-order functions such as logic gates. An important property of regulators of transcriptional elongation is their composability into logics and higher-order functions. This means that our converted mutually orthogonal attenuators can be assembled into NOR gates of multiple inputs simply through tethering. To demonstrate this feature, we constructed a collection of NOR gates that integrate two, three, or four inputs by tethering together two, three, or four of our mutually orthogonal converted RNA-IN variants.

In constructing these NOR gates, we needed first to address the issue that RNA-IN variants require a flexible 5'-terminus for proper function. Therefore, we used a tethering strategy in which CRISPR repeats described in **Chapter 7** were inserted between linked RNA-IN variants such that once transcribed, a free 5'-terminus would be generated through cleavage by a coexpressed Csy4 protein, allowing the full regulatory effect of a given RNA-IN variant to be realized before the next regulatory unit may be transcribed. More generally, the insertion of CRISPR sequences between tandem regulatory units promotes independence of the joined units and therefore may be wise in all composition strategies regardless of whether a free 5'-end is required. This is discussed in more detail elsewhere. The resulting NOR gate compositions were cloned upstream of GFP, which acted as our reporter.

As shown in **Figure 6.1.4A**, all NOR gates behaved as desired. For example, the two-input NOR gates consisting of tandem RNA-IN variants were ON in the absence of the corresponding plasmid-encoded cognate RNA-OUTs, and OFF in the presence of either of two cognate RNA-OUTs, providing a set of universal NOR gates. Likewise, the three-input and four-input NOR gates responded only to cognate RNA-OUTs. In addition, these responses were unimodal, as evident from fluorescence microscopy and cytometry studies (**Figure 6.1.4B**). Therefore, our converted regulators are indeed predictably composable, yielding, in this case, a large number of synthetic NOR gates.

Two features of our composition strategy will guide future experiments. First, the systematic manner in which mutually orthogonal RNA-IN/OUT variants can be generated, converted into transcriptional regulators, and composed into universal NOR logics should facilitate the scalable NOR-gate-based assembly of multi-layered circuits, all within a single cell. Second, our tethered regulators effect the integration of sequential regulatory events into an overall decision and therefore may exhibit unique characteristics in terms of noise propagation, regulatory kinetics, and transfer function forms. We are currently exploring these areas with the aim of controlling each property through composition.



**Figure 6.1.4 Performance of NOR gates using converted orthogonal attenuators.** Fluorescence microscopy images of cells containing various plasmid pairs are shown. Bar graphs represent background-subtracted relative protein production rate (PPR) during log phase, as determined by tracking GFP fluorescence and optical density over time using a fluorescence plate reader. Experiments were conducted in triplicate on the same day.

### 6.1.5 Discussion of the conversion system

The unique versatility and composability of regulators of transcriptional elongation makes them ideal for building up custom regulatory functions for synthetic biology; but the less versatile regulators of translational initiation are easier to engineer and more common. We therefore developed a general strategy for turning regulators of translational initiation into regulators of transcriptional elongation. Using this strategy, we have constructed a series of highly effective RNA-based regulators of transcriptional elongation and have demonstrated the inherent composability of these regulators by assembling a collection of multi-input NOR gates. We believe that the continued application of this conversion strategy will lead to an explosion in the number of independent regulators of transcriptional elongation available to 5'-UTR engineering, ones that through simple composition can yield signal integration and higher-order functions for predictable biological design.

## 6.2 Protein-RNA Translational Systems

RNAs could interact with diverse protein factors in cells. Interaction between RNA and protein factors together determines the half-life, processing, localization, and translation of an RNA

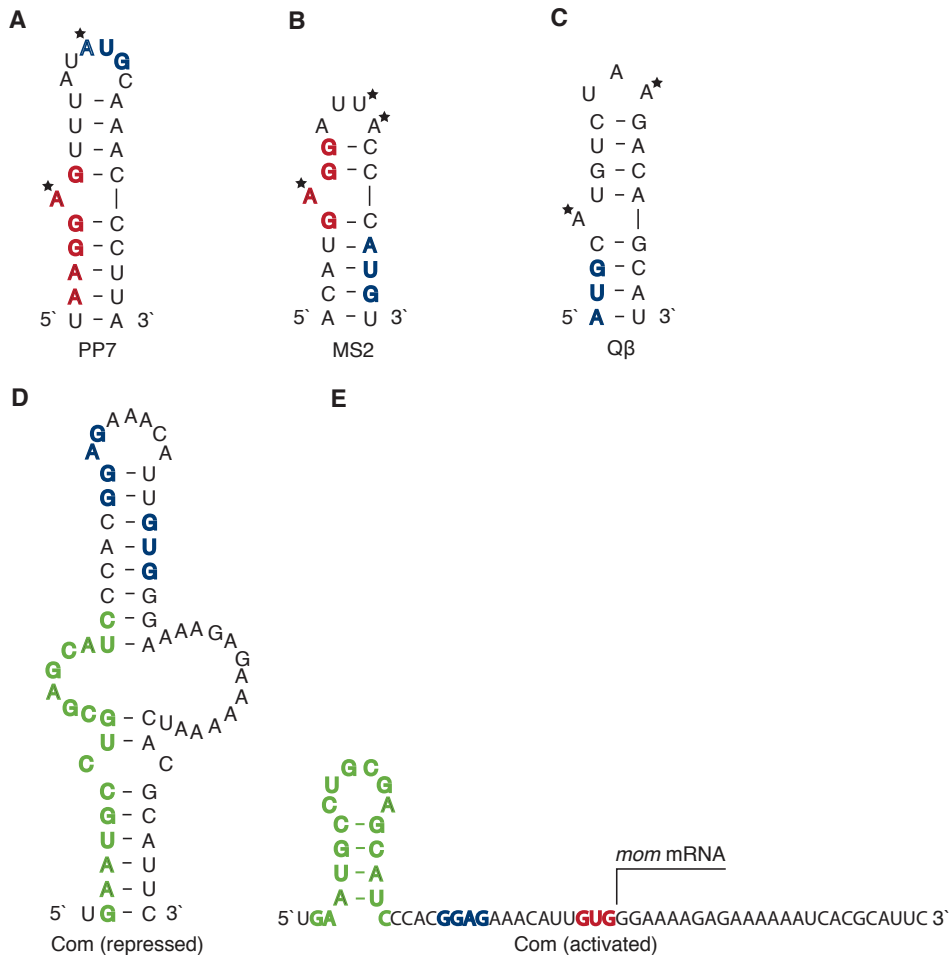


molecule<sup>162</sup>. Many proteins have been discovered from bacteria genomes and bacteriophages that bind to the 5' UTR region and compete with ribosome binding. They thus provide additional control over the target gene<sup>163,164</sup>. Similar to protein-binding operator sites on the DNA, the RNA sequence that can bind to certain proteins and modify the translation profile is often called translational operator sites. These operator sites usually contain sequence-specific and structure-specific motifs that binds to proteins with high specificity and affinity<sup>162</sup>.

### **6.2.1 Introduction to protein-RNA interaction systems**

Many RNA-binding proteins are discovered from phages. For example, MS2<sup>138</sup>, PP7<sup>165</sup>, Q $\beta$ <sup>166</sup> phages all contain coat proteins that serve both as structural proteins for capsid formation and as regulators to control viral replication and assembly. Interestingly, all of them recognize structured RNA hairpins and repress translation of downstream genes. Previous studies have shown that the length of stem of the hairpin, nucleotide compositions of the loop, and loops positioning on the stem together determine the binding affinity and specificity of an RNA hairpin to the coat protein. This is confirmed by creation of hairpin variants that exhibit a wide range of binding affinity<sup>138,167,167,168</sup>. Furthermore, in natural systems, many coat proteins bind to RNAs in dimers<sup>169</sup>.

Compared to translational repressors, only one translational activator, Com, has been reported from bacteriophage Mu<sup>170</sup>. The Mu phage contains a *com-mom* operon which is important for DNA modification called momification (*mom*-specific modification). Com protein activates Mom expression by binding to the intercistronic region between *com* and *mom*<sup>40</sup>. Com binding destabilizes the inhibitory structures around the Mom SD sequence and the start codon such that Mom translation can proceed. The natural RNA hairpins for coat proteins of MS2, PP7, and Q $\beta$ , and Com are shown in **Figure 6.2.1**. Since many of these proteins and their mechanisms have been well studied, here we further characterize these proteins and their targets and study their use as basic building blocks for constructing higher-order functions.

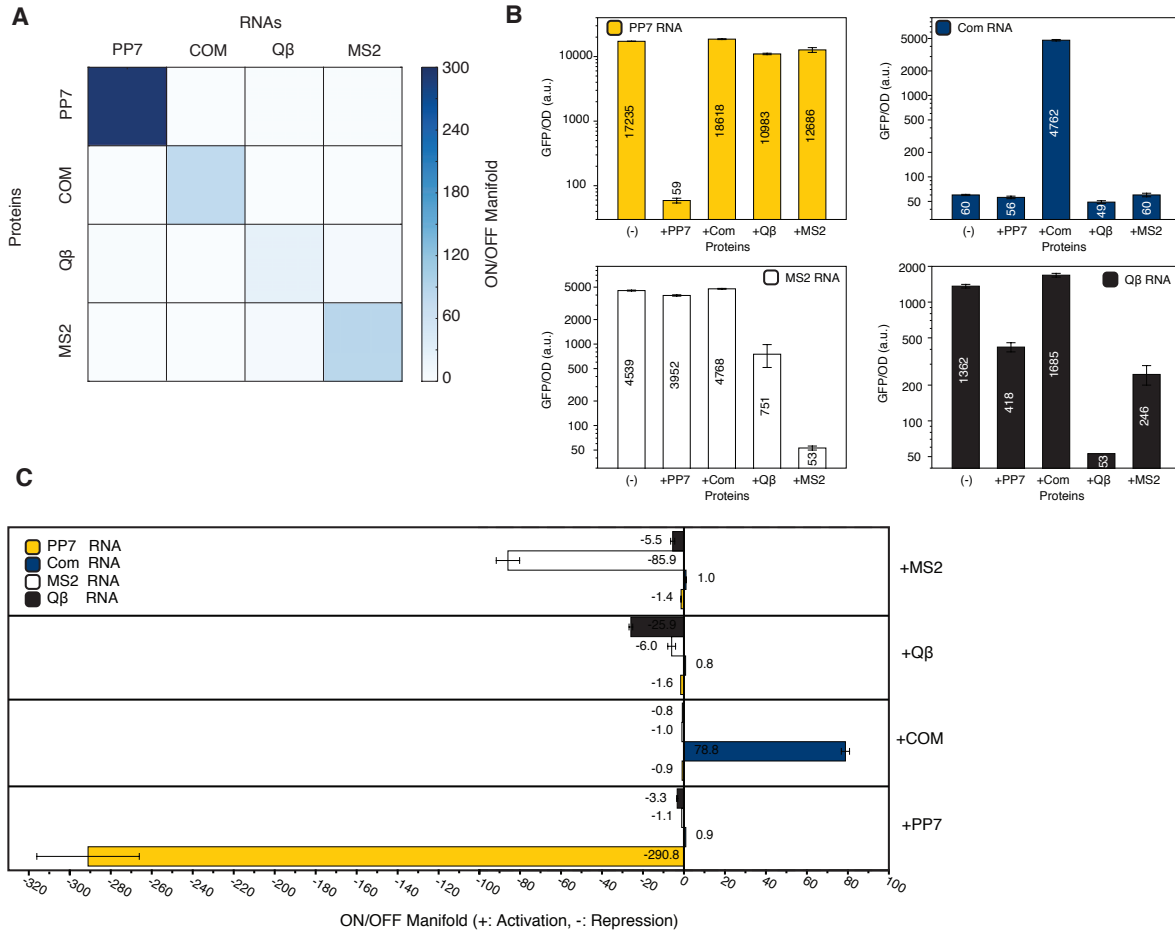


**Figure 6.2.1 RNA binding motifs for proteins.** RNA aptamers for (A) PP7 coat protein; (B) MS2 coat protein; (C) Q $\beta$  coat protein; (D-E) Com protein are shown. Shine-Dalgarno (SD) sequence is in blue, start codon is in red, and Com binding site is in green. Nucleotides important for PP7, MS2, Q $\beta$  recognition are labeled with a star. The repressed conformation of Com hairpin is shown in (D) and activated conformation is shown in (E).

### 6.2.2 Orthogonality of RNA-binding proteins

To test the orthogonality between RNA-binding proteins, we cloned MS2, PP7, Q $\beta$ , and Com proteins to a high copy ColE1 vector under the control of P<sub>LacO-1</sub>. Their cognate UTR sequences were translationally fused to a GFP report gene on a p15A plasmid. Combinations between proteins expression plasmids and reporter plasmids yields a 4x4 matrix with results shown in **Figure 6.2.2A**. Among these four pairs, the PP7 coat protein displayed the higher dynamic range (~300-fold), which is comparable to transcriptional protein repressors (**Figure 6.2.2B**). Com protein showed 100-fold activation on its target. MS2 and Q $\beta$  coat proteins exhibited weaker interaction, but still between 50~100 fold range. However, the Q $\beta$  coat protein is not orthogonal with PP7 and MS2 proteins, which is consistent that previous studies showing this coat protein recognizes a structural-specific instead of sequence-specific hairpin<sup>144</sup>. Thus, PP7, Com, and MS2 comprises a mutually orthogonal family that can be used together in the same cell. We note that in this case, two-plasmid system is not necessary. When we cloned both

protein expression cassette and reporter system onto the same p15A plasmid, we obtained almost identical regulatory function of these RNA-binding proteins.



**Figure 6.2.2 Orthogonality between RNA-binding proteins.** (A) Heatmap of mutual interactions between PP7, Com, Qβ, and MS2 proteins and their RNA targets translationally fused to a downstream GFP reporter gene. (B) The bar plots show the absolute values of fluorescence for the orthogonality experiment. (C) The plot shows the ON/OFF manifold change for the orthogonality experiment. Negative values represent repression while positive values show activation.

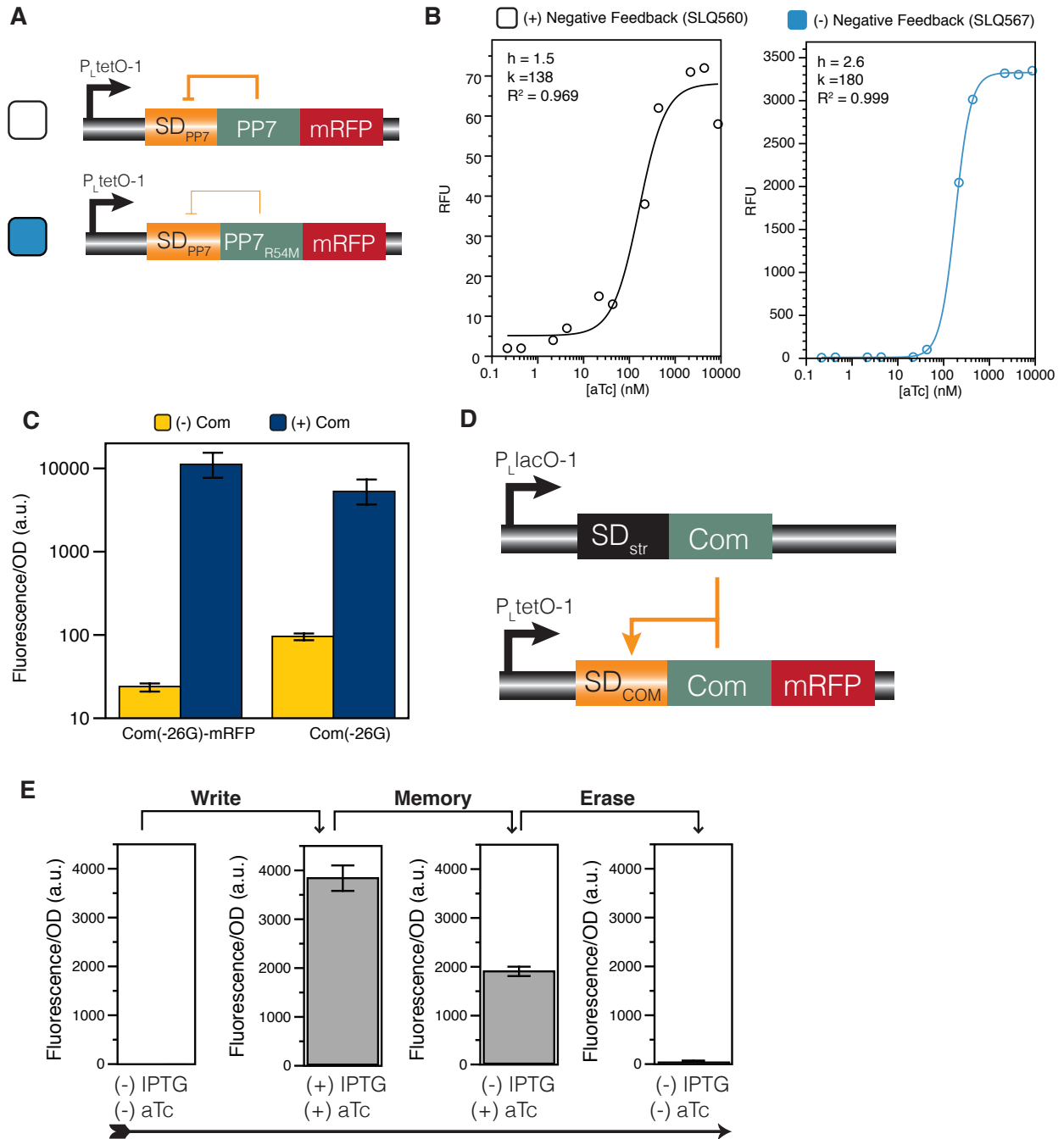
### 6.2.3 Engineering translational feedback circuits

We did preliminary studies on constructing circuits using PP7 and Com proteins and their cognate RNA targets. Based on their repression or activation activities, we tried to construct a negative feedback based on PP7 and a positive feedback based on Com.

To build a negative feedback, we first fused PP7 protein to the N-terminus of mRFP<sup>139</sup> (Figure 6.2.3A). The fusion protein was then translationally fused to the 3' end of the PP7 cognate hairpin which contained an intrinsic SD sequence and start codon. The whole cassette is controlled by P<sub>tetO-1</sub> promoter on a p15A plasmid. A PP7 mutant with R54M mutation was used as the negative control for the feedback, as this protein mutant can only bind to the RNA

target weakly<sup>165</sup>. We expected that with increasing concentrations of inducer aTc, the negative feedback circuit would show a much repressed induction curve, whereas the circuit without negative feedback would display a much higher dynamic range. This was indeed observed (**Figure 6.2.3B**). Furthermore, fitness of the experimental data to the Hill equation showed that the negative feedback also reduced the sensitivity of the response induction curve. Further mathematical modeling is needed to explain the observed phenomena.

We also constructed a positive feedback based on Com. The Com coding sequence contains an inherent Com binding site that might titrate the effective Com protein concentration. We first introduced a mutation into Com coding sequence (-26G) that was reported to prevent Com binding but did not change Com protein sequence<sup>40</sup>. Then this Com protein mutant was fused to the N-terminus of mRFP or sfGFP. We first verified that if the fusion proteins could still activate the target. Our results showed that Com-mRFP fusion could still strongly activate its target (**Figure 6.2.3C**) whereas Com-sfGFP lost its function (Data not shown). Interestingly, compared to the Com protein, Com-mRFP fusion displayed an even larger ON/OFF ratio for reasons unknown. Then fusion protein was translationally fused to Com hairpin such that the intrinsic SD controls the translation initiation of the fusion protein. The same P<sub>LtetO-1</sub> promoter was used to control its expression. We expected that with increasing concentrations of aTc, positive feedback would display a much fast induction and higher sensitivity. However, no activation was observed. This is probably because the Com hairpin is very tight and there is no leaky translation of Com-mRFP even with maximum transcription. Another possibility is that Com-mRFP is not very sensitive, and needs to accumulate to a relative high concentration to be functional. To solve this problem, we expressed an extra copy of Com from another plasmid under the control of P<sub>LlacO-1</sub> promoter (**Figure 6.2.3D**). We co-transformed both plasmids into *E. coli*, and tested the circuit under a series of IPTG and aTc conditions (**Figure 6.2.3E**). At first, no IPTG or aTc was present, and there was no mRFP expression. When we added both IPTG and aTc, after cells grew to exponential phase, we observed strong mRFP signals. We washed away IPTG but kept aTc, as expected, the mRFP signal retained. This signals was decreased by ~2 fold probably due to weaker activation effects from Com-mRFP fusion protein. Finally, we washed away aTc as well, and the mRFP signal disappeared after a few hours. Thus, this circuit is a memory device based on translational positive feedback: IPTG is a “write” signal, and aTc is an “erase” signal. To our knowledge, this is the first synthetic memory device based on translational feedback instead of transcriptional feedback<sup>171</sup>. We believe this device is modular as the P<sub>LlacO-1</sub> promoter can be replaced by a different promoter that can be activated by another protein. In this case, the circuit can detect and memorize the transient presence of the activator protein.



**Figure 6.2.3 Negative feedback and positive feedback based on RNA-binding proteins. (A)** A negative feedback circuit based on the PP7 protein and its RNA target. **(B)** Characterization of (+) negative feedback (left) and (-) negative feedback circuits (right). **(C)** Com-mRFP fusion protein can activate RNA target sequence. **(D)** Design of a positive feedback circuit based on the Com protein and its RNA target. **(E)** Measurement of the circuit with a series of IPTG and aTc inducers conditions in a sequential manner. This shows the circuit in (D) can behave as a memory device.

#### ***6.2.4 Strengths and limitations of the protein-RNA systems***

To sum, the RNA-binding proteins are promising building blocks for constructing complex circuits. One advantage of these regulators is that they display larger ON/OFF dynamic range compared to riboregulators. Also, these proteins are relatively modular and can be fused to other proteins. This allows detection of other protein concentrations and possibly facilitates rewiring of the original regulatory pathway. Existence of three mutually orthogonal protein-RNA pairs is a preliminary proof that these proteins interact with their target with high specificity. This mutually orthogonal list can be expanded given many RNA-binding proteins have been reported from different resources.

However, it is not straightforward to design protein-RNA interactions. Compared to Watson-Crick base pairing of RNA-RNA interactions, designing a protein-RNA interaction often requires knowledge of their tertiary structure and tedious mutagenesis study. Furthermore, compared to transcriptional repressors, translational repressors generally show lower sigmoidality, at least for the protein tested here. For example, Com protein binds to its target without forming multimers, implying the hill coefficient should be close to 1. A third disadvantage of these proteins is that they might introduce extra metabolic burden on the host cell and even toxicity. For example, high concentrations of MS2 coat protein will form giant protein clusters that are insoluble and harmful to cells.

# Chapter 7 Engineering A Synthetic RNA Processing Platform Based on CRISPR

## 7.1 Predictable Engineering of Biological Systems

Modular and predictable programming of gene expression is central to the engineering of new and useful biological systems for manufacturing, therapeutic and environmental applications<sup>11</sup>. To create genetic systems with desired qualitative and quantitative functions, two general approaches have been used. One approach is based on the screening or selection<sup>172</sup>, and the other is based on the rational design<sup>173,174</sup>. In the first approach, a large library containing random mutations of the genetic elements including promoters, 5' or 3' UTRs, RBSs, and ORFs is constructed, which is then screened or selected to find the mutants with desired phenotypes. However, the library size often increases rapidly with the number of mutagenized genetic elements, which makes the screening/selection process prohibitively inefficient. On the contrary, the second approach aims to design the performance characteristics of individual genetic elements, which could be predictably assembled into higher-order genetic systems. However, our ability to design genetic elements is largely based on our knowledge of their detailed functions and mechanisms, which is often missing. Furthermore, the designed elements that work well in one genetic context setting often fail in other setting, resulting in highly variable and uncertain performances of the same part<sup>175</sup>. To address this challenge, synthetic biologists have proposed to create standard genetic elements that work modularly and predictably in diverse circuits, pathways, and strains<sup>10</sup>.

However, it was extremely difficult to create such standard genetic elements. Previous studies have shown that gene expression is a sequential process composed of multiple events, including transcription, translation, mRNA decay, and protein degradation<sup>45,53,54,176</sup>. Biomolecules including DNA, RNA, and proteins as well as their structure and interactions with host factors and metabolites together determine when, where, the strength, and the kinetics of these gene expression events. This structural interaction and functional coupling could be one major cause of the uncertain performances. For example, the structural interaction between UTR and ORF could affect translation and mRNA degradation; on the other hand, promoters might contain operator sites embedded in the UTRs, which couples transcription, translation and degradation with each other.

We postulate that an efficient physical separation strategy within the biological assemblies such as DNA, RNA, and proteins would greatly enhance the modularity and predictability of components. Especially, given the central role of mRNAs in the whole process of genetic regulation, which involves in all events of transcription, translation, and degradation, insulation within the transcript could be a key solution to the rational design approach. In this study, we aim to create a novel RNA processing platform that will make the engineering of genetic systems more modular and predictable.

## 7.2 Standardization of Genetic Parts

Following definitions of standardization in other engineering disciplines, here we define a **standard** genetic element is an element whose function solely depends on its own sequence (thus structure) not on the sequences of other elements. In this sense, a standard genetic element

not only possesses defined 5' and 3' boundaries (since its sequences is 1-D) and adapts a standard physical format like BioBrick<sup>91</sup>, but also performs independently of their genetic contexts<sup>11</sup>.

A mathematical description of this definition is given below. The genotypic sequence of a **standard** part appears as a separable variable in the phenotypic function. A separable function, following quantum mechanics, is a function that can be separated into a combination of multiple sub-functions, each of which contains only one variable. Thus, if Y is a separable function of variables  $X_1, X_2, \dots, X_N$ , then Y could be expressed as

$$Y(X_1, X_2, \dots, X_N) = Y_1(X_1) \otimes Y_2(X_2) \otimes \dots \otimes Y_N(X_N)$$

Since the process of gene expression is composed of transcription, translation, mRNA degradation, protein degradation, and protein maturation, the native protein production rate at the steady state can be expressed as

$$\left. \frac{\partial P}{\partial t} \right|_{production,ss} = \frac{k_{Tr} \cdot k_{Ts}}{k_{Dm}}$$

Where  $k_{Tr}$  is the transcription success rate,  $k_{Ts}$  is the translation success rate, and  $k_{Dm}$  is the mRNA degradation rate. These four rates are functions of the sequences of genetic elements and cellular contexts such as temperature, pH, abundance of RNA polymerase, ribosome, RNase, protease, nutrients, and metabolites. Each of these rate terms can be expressed as

$$k_{Tr} = k_{Tr}(\text{Pr}_{DNA}, \text{UTR}_{DNA}, \text{TU}_{DNA}, Q_{CC})$$

$$k_{Ts} = k_{Ts}(\text{UTR}_{RNA}, \text{RBS} : \text{ORF}_{RNA}, Q_{CC})$$

$$k_{Dm} = k_{Dm}(\text{UTR}_{RNS}, \text{RBS} : \text{ORF}_{RNA}, Q_{CC})$$

Here, Pr is the promoter sequence, either in form of DNA or RNA; UTR is the UTR sequence, consisting of 5' UTR and 3' UTR; RBS:ORF is the sequence that is accessed by the ribosome; CC is the combined cellular context term; and Q means their quantity.

The term of  $k_{Tr}$  is usually not a separable function, because UTR could contain an operator site that affects promoter strength (mainly via transcriptional abortion). The  $k_{Ts}$  is not separable function because the UTR could contain extra sequences introduced by promoter, and the structural interaction between UTR and RBS:ORF could couple these two variables together. Similarly,  $k_{Dm}$  is not a separable function due to the coupling between UTR and RBS:ORF. Thus, natural genetic elements are not standard but **entangled**.

To make some of these variables separable, we need physical separation within the RNA molecules. For example, if we insert a cleavage element between the UTR and RBS:ORF, then the RBS:ORF sequence will be physically separated from the UTR sequence. In this case, the translation rate can be expressed as



$$k_{Ts} = k_{Ts}^{UTR} (UTR_{RNA}, Q_{CC}) \cdot k_{Ts}^{RBS:ORF} (RBS:ORF_{RNA}, Q_{CC})$$

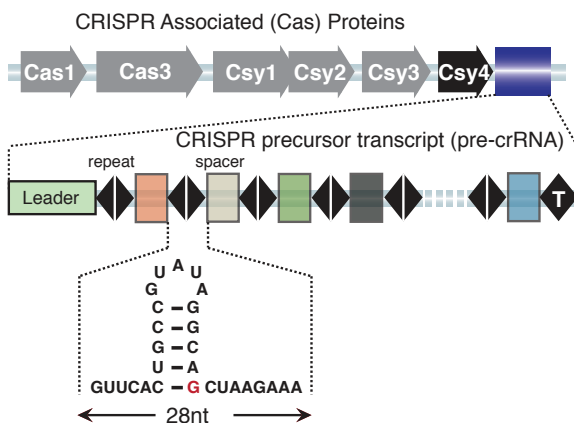
Similarly, if we cleave between the promoter and downstream sequences, the transcription rate can be expressed as

$$k_{Tr} = k_{Tr}^{Pr} (Pr_{DNA}, Q_{CC}) \cdot k_{Ts}^{Other} (UTR_{DNA}, RBS:ORF_{DNA}, Q_{CC})$$

Now the transcription and translation functions can be written as sub-functions that only involve a single variable. Following our definition, these genetic parts then become standardized. Of course, the term of cellular context still appears in every term, meaning that these functions are not separable regarding the cellular context.

### 7.3 Introduction to Bacterial CRISPR Systems

To engineer synthetic RNA processing platforms, we deployed a natural RNA cleavage mechanism from the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) system<sup>177</sup>. Many bacteria and archaea contain CRISPR/Cas systems that confer resistance to invasive genetic elements<sup>178,179</sup>, whose sequences are integrated into the CRISPR loci that are transcribed as long RNAs (pre-crRNAs) containing repetitive sequence elements<sup>180-183</sup> (**Figure 7.3**). A conserved property of evolutionarily diverse CRISPR systems is the ability of CRISPR-associated (Cas) proteins to efficiently process long pre-crRNAs into short crRNAs that serve as homing oligonucleotides to prevent the propagation of invading virus<sup>182-185</sup>. Recently, one such Cas-mediated transcript processing mechanism has been discovered in the pathogenic strain of *Pseudomonas aeruginosa* UCBPP-PA14<sup>186</sup>. In this system, the endoribonuclease Csy4 recognizes the 28-nucleotide repeat sequences within the pre-crRNA and cleaves immediately downstream of a 15-nt stem-loop in a sequence- and structure-specific manner. Since Csy4-based RNA cleavage is highly specific<sup>186</sup>, we believe it provides an elegant exogenous transcript processing system which is well-suited for controllable RNA processing in *E. coli* and likely a diverse of bacteria and archae hosts.



**Figure 7.3 Schematic of the CRISPR/Cas system in *Pseudomonas aeruginosa* strain UCBPP-PA14.** The CRISPR locus consists of a cluster of invader-derived spacer sequences (boxes) separated by 28-nucleotide repetitive elements (diamonds). After transcription, the Csy4 protein cleaves the folded repeat sequences after nucleotide G20 (red).

### 7.4 Engineering Predictable Translational Units

Untranslated regions (UTRs) of transcripts encode ubiquitous and often sophisticated regulatory elements that control the expression of physically adjacent genes both in natural and in synthetic

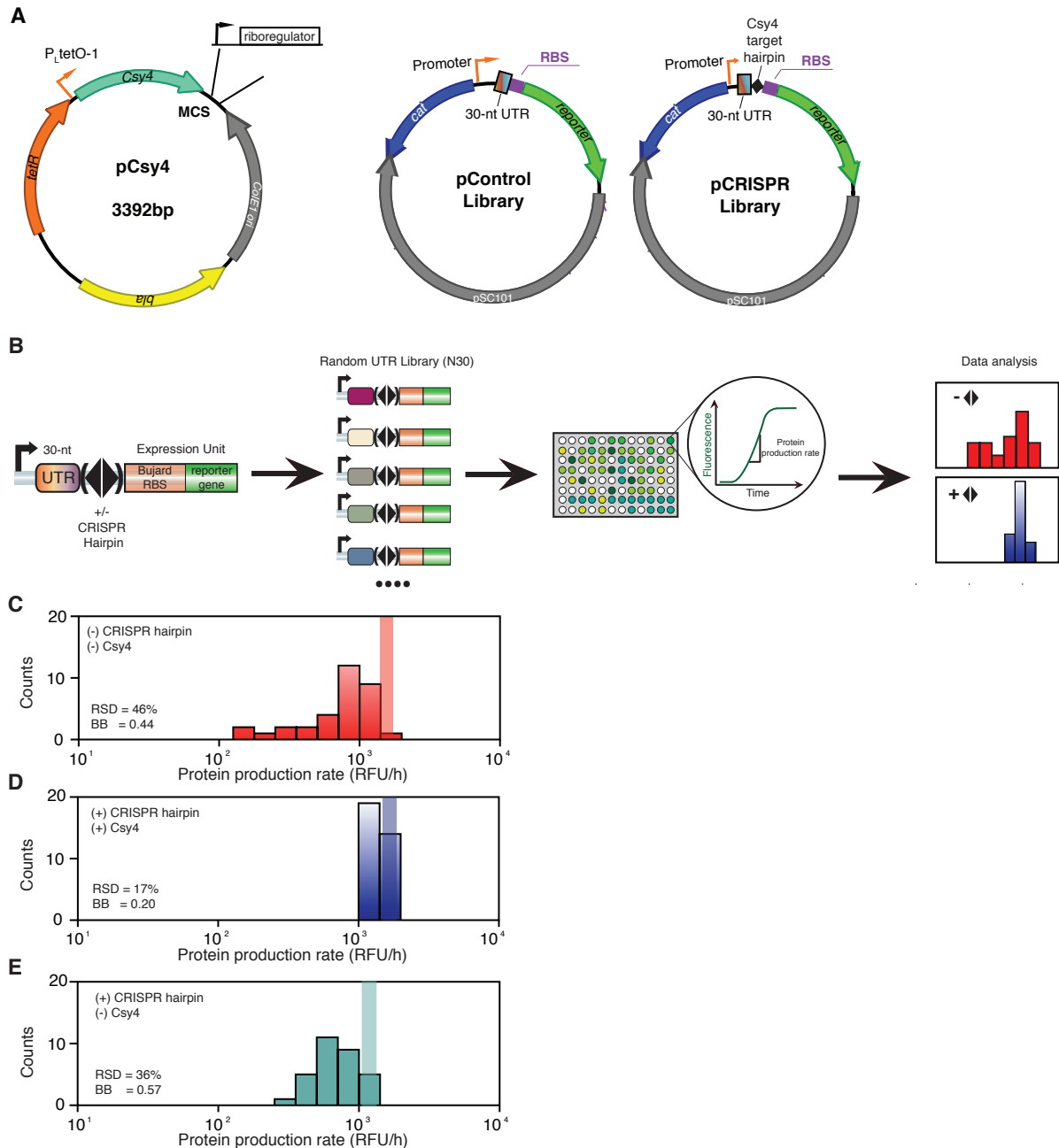
systems<sup>187</sup>. While UTRs exhibit versatile regulations, their function is often affected by structural interaction with both proximal and distal sequences on the same transcript. Such interactions could interfere with Shine-Dalgarno (SD) sequence accessibility, RNA polymerase and ribosomal processivity and transcript degradation in unpredictable ways<sup>188</sup>. Here we define a translational unit as the sequence that is processed by ribosome, thus it starts with RBS and ends with a stop codon.

#### **7.4.1 RNA processing reduces variability of translation**

We hypothesize that Cys4-mediated RNA cleavage would eliminate these context-dependent effects, making the translation of the downstream gene insensitive to the upstream UTR sequence. To test this hypothesis, we constructed two reporter gene libraries. The first, a control library, contained a random 30-nt 5' UTR region placed between the J23119 promoter with a known transcriptional start site and a strong Bujard RBS<sup>23</sup> fused to a green fluorescent protein, sfGFP<sup>98</sup>. The second library, the CRISPR library, contained the 28-nt CRISPR cleavage hairpin placed between the random UTR sequence and the Bujard RBS. Both libraries were expressed in *E. coli* Top10 cells from a plasmid conferring chloramphenicol resistance and containing the pSC101 replication origin (**Figure 7.4.1A**). The wild type Csy4 gene was cloned under a tetracycline-inducible promoter P<sub>LtetO-1</sub><sup>23</sup> on a plasmid encoding ampicillin resistance and ColE1 replication origin. If the CRISPR cleavage provides effective insulation, the variation in expression in the control library should be far larger than the CRISPR library when the Cys4 protein is expressed.

For both libraries, we measured the temporal fluorescence expression of a group of randomly picked constructs from the lag phase to the stationary phase to quantify the effects of random UTR on expression<sup>104</sup> (**Figure 7.4.1B**). The protein production rate (Relative Fluorescent Units per hour, RFU/hour) during the exponential phase was calculated using the temporal data for each clone. We estimated the variability of the measurement system as 10% by measuring the Relative Standard Deviation (RSD) of three biological replicates of the baseline circuits without the 30-nt UTR insertion (pControl( $\Delta$ 30UTR) and pCRISPR( $\Delta$ 30UTR)). This sets the lower limit for estimation of the RSD from the sampled libraries. Among the sampled constructs for the control library, the protein production rates were quite variable showing a 46% RSD (**Figure 7.4.1C**). The Ratio between the Mean Values (RMV) of the sampled circuits and the baseline circuit was only 0.56. Therefore, the random UTR had large effects on expression efficiency.

In contrast, the RSD for the production rates from the CRISPR library samples (when Cys4 was fully induced) was only 19%, a more than two-fold decrease relative to the control library (**Figure 7.4.1D**). In addition, the ratio between the mean values of the samples and the baseline circuit increased to 0.79. Presumably, transcript cleavage at the CRISPR hairpin by Csy4 protein separates the 30-nucleotide UTR sequences from the downstream sequence unit and reduces their interference with expression mechanisms such as translation initiation and RNA degradation. Without Csy4 expression, we observed an increased variance (36%) among the sampled circuits and a decreased relative mean value to the control (0.43), confirming the important role for transcript cleavage apart from any structural effect of the CRISPR hairpin (**Figure 7.4.1E**).

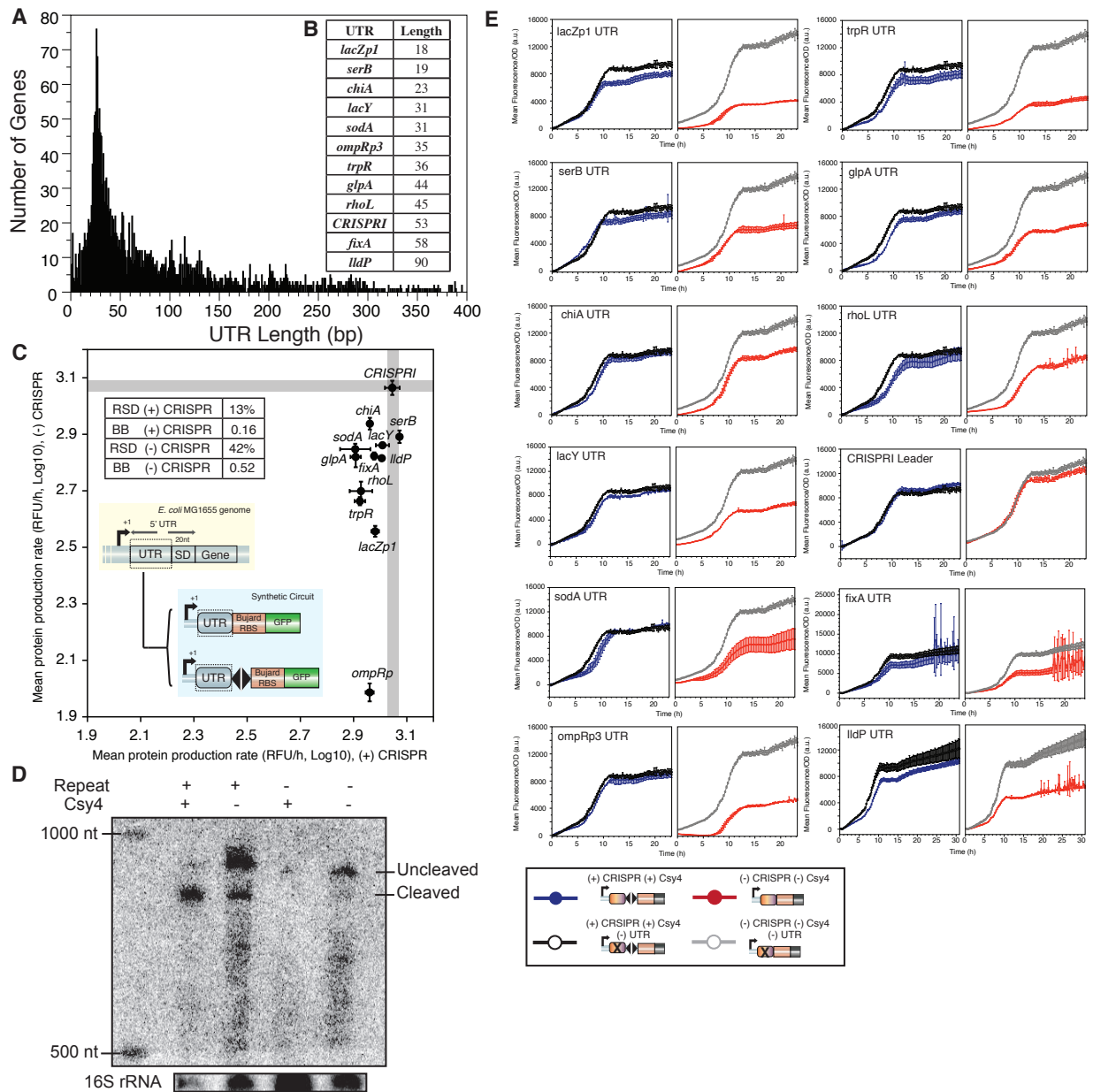


**Figure 7.4.1 The CRISPR RNA processing system improves the predictability of RBSs.** (A) The plasmid maps used in the study. The Csy4 protein was expressed from a ColE1-based plasmid under the control of a  $P_{tetO-1}$  inducible promoter. Combinations of antisense RNA expression cassettes were cloned onto the multiple cloning sites (MCS) downstream of the Csy4 protein. The vectors used for construction of control and CRISPR libraries for all experiments are shown on the right side. (B) Experimental procedure for measuring the effects of random UTRs on gene expression with and without the Csy4 cleavable element (diamond). (C-E) Statistical analysis of protein production rates. Red bars show the data without Csy4 cleavage elements (C), blue bars show data with Csy4 cleavage elements and Csy4 co-expression (D), and cyan bars show data with Csy4 cleavage elements but no Csy4 co-expression (E). The

expression of baseline constructs without 30-nt UTR insertions are plotted as shaded lines, with their widths representing the standard deviation of biological triplicates (measurement system variation). In all cases,  $P < 0.0001$  for the differences in RSD and BB using two-tailed Student's *t*-test.

#### **7.4.2 RNA processing reduces variability from genomic UTRs**

Genomic 5' UTR elements exist in varying lengths ranging from tens to hundreds of nucleotides<sup>189</sup> and encode a variety of structures (**Figure 7.4.2A**). To validate our strategy with genomic UTRs with different lengths, we identified UTR sequences from 12 genes with lengths ranging from 18 to 90 nucleotides from the *E. coli* MG1665 genome (**Figure 7.4.2B**), and inserted them into the circuits containing the Bujard RBS and sfGFP with or without the CRISPR hairpin to measure their effects on protein production rates. **Figure 7.4.2C** shows a 2D plot of mean protein production rates for each genomic UTR, with the x-axis showing the CRISPR hairpin present and cleaved and the y-axis without the CRISPR hairpin. The spread of points reflects the expression variability, with 13% along the x-axis and 42% along the y-axis. The mean production rate relative to the baseline construct without UTR insertion was 0.84 with cleavage and 0.48 without the cleavage. Thus, cleavage greatly reduces the effect of “natural” UTRs on downstream expression. We further measured the cleavage effects of Csy4 using Northern blotting on the construct containing the *lldP* UTR insertion. The Northern results were consistent with the Csy4 cleavage at the CRISPR cleavage hairpin, confirming that the reduction of variability came from the RNA cleavage (**Figure 7.4.2D**). A list of temporal expression data for the 12 genomic insertions is shown in **Figure 7.4.2E** for reference.

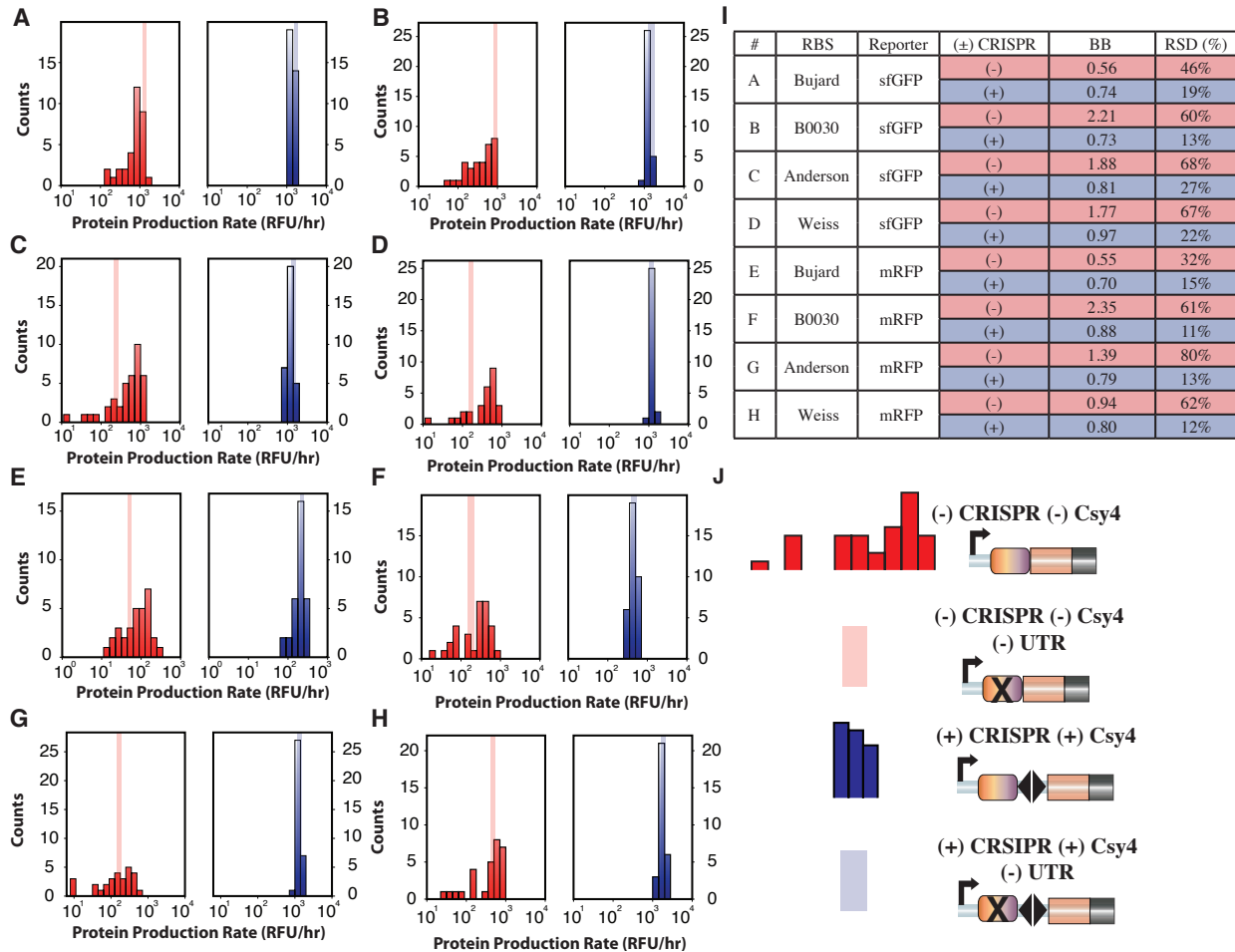


**Figure 7.4.2 RNA processing reduces variability of genomic UTRs on translation.** (A) The statistic histogram of the lengths of 5' UTR sequences in the *E. coli* MG1655 strain plotted using the data from RegulonDB<sup>189</sup>. (B) Temporal expression data for the 12 circuits with genomic UTR insertions. The blue lines show the data for the circuit with the CRISPR hairpin, and the red lines show those without the CRISPR hairpin. The control circuits without UTR sequence insertions are shown in black (with the CRISPR hairpin) and grey (without the CRISPR hairpin). All curves are averaged over three biological replicates and the error bars represent the standard deviation. (C) The 2-D plot of the mean protein production rates for 12 genomic UTR insertions, with the x-axis showing with RNA processing and the y-axis without. The grey lines are expression data of the baseline constructs, with the widths showing the standard deviation. The insets show the statistics and experimental procedure. The 20-nt fragment spanning the SD sequence was excluded. (D) Northern analysis of total RNA of *E. coli* cells to verify *in vivo* Csy4

cleavage. The cells in all columns contained an expression cassette that transcribed a segment of genomic *lldP* UTR (90 nt) and GFP coding sequence (744 nt). The first two columns further contained a 28-nt CRISPR cleavage hairpin inserted between the *lldP* UTR and GFP. When the Csy4 protein is co-expressed, cleavage should reduce the transcript size to 752 nt, which was observed in the first column (more than 70% cleaved). Whereas without Csy4, uncleaved transcript is 864 nt as shown in the second column (more than 65% remained uncleaved). In the absence of CRISPR repeat hairpin, no cleavage was observed as shown in columns 3 & 4. The leaky cleavage in the second column is probably due to an intrinsic transcriptional start site inside the *lldP* UTR sequence, which will be further investigated.

#### ***7.4.3 Similar effects for different translational units***

The particular expression unit in the library contained a Bujard RBS sequence and sfGFP. To test our strategy with other translational units, we performed the same experiments on eight different expression units by combining four RBS sequences, Bujard RBS, B0030 RBS, Anderson RBS, and Weiss RBS, with two fluorescent genes, sfGFP<sup>98</sup> and mRFP<sup>99</sup>. The CRISPR cleavage hairpin is inserted (pCRISPR libraries) or not (pControl libraries). We observed consistently lower variability of the clones with RNA processing than those without, on average a three-fold decrease of the RSD values (**Figure 7.4.3**). Furthermore, the RMV values between the mutants and the baseline controls without 30-nt UTR insertions were consistently higher and closer to 1.00. Thus, RNA processing reduces expression variability regardless of the particular translational units used.



**Figure 7.4.3 RNA processing improves the predictability of translation.** (A-H) Eight translation units composed of four RBS sequences and two reporter genes are tested with randomized 30-nucleotide UTR sequences with or without the Csy4 hairpin. The histograms in red show the constructs without the Csy4 hairpin, and the histograms in blue show those with. The mean values of the control construct without a 30-nucleotide UTR sequence are shown as lines and the width of the lines represents the standard deviation of three biological replicates. (I) The statistical summary of (A)-(H). (J) The legend.

## 7.5 Engineering Predictable Promoters

Promoters are essential gene regulatory elements, yet it remains a challenge to combine them with different RBS sequences in tandem to control the gene expression in a predictable manner. Most promoters used in applications have been derived from natural DNA sequences with poorly annotated transcriptional start sites and operator sites<sup>190,191</sup>. Assembling these elements into genetic circuits often introduces, among other things, extra UTR sequences that can affect all aspects of downstream gene expression including translational efficiency and transcript degradation<sup>192</sup>. As above, we proposed that insertion of a CRISPR hairpin between these *ad hoc* promoter elements and downstream cassette elements will insulate the two parts from one another. Transcript cleavage at the CRISPR hairpin should remove any UTR sequences derived from the promoters and prevent their structural interaction with downstream expression units.

Thus, the protein production rate for each construct should be directly proportional to the transcription rate. The relative strength of each promoter should remain constant relative to a standard across all compositions with downstream expression components.

To test this hypothesis, we once again created two libraries. Both derive from the combinatorial assembly of 7 popular promoters with 2 RBS sequences and 2 reporter genes to construct a small library of 28 constructs. The seven promoters were chosen in a way that reflects diversity, with some promoters only differ by the -35 or -10 box sequences, some contain extra UTR sequence after the transcriptional start site, and some contain operator sites. The CRISPR library contained an extra CRISPR cleavage hairpin inserted between the promoter and the RBS compared to the control library.

If RNA processing can efficiently separate downstream UTRs from upstream promoters, we expect that for the same UTR:RBS:ORF construct with different promoters, the protein production rate is proportional to the transcription rate. Because the compositions of mRNAs are the same after transcription, the translation and degradation rates are the same. When we take ratios between multiple promoters, we expect

$$\left. \frac{\partial P}{\partial t} \right|_1 : \left. \frac{\partial P}{\partial t} \right|_2 : \dots : \left. \frac{\partial P}{\partial t} \right|_N = k_{Tr1}(\text{Pr}_{DNA}) : k_{Tr2}(\text{Pr}_{DNA}) : \dots : k_{TrN}(\text{Pr}_{DNA})$$

To verify this relationship, we measured the mean protein production rate of biological triplicates for each construct in both libraries, and calculated the Relative Promoter Unit (RPU) for each promoter. The RPU is calculated by normalizing the production rate of each promoter to that of the promoter J23105<sup>193</sup>. Since

$$RPU = \left. \frac{\partial P}{\partial t} \right|_{promoter} / \left. \frac{\partial P}{\partial t} \right|_{reference}$$

The above relationship becomes

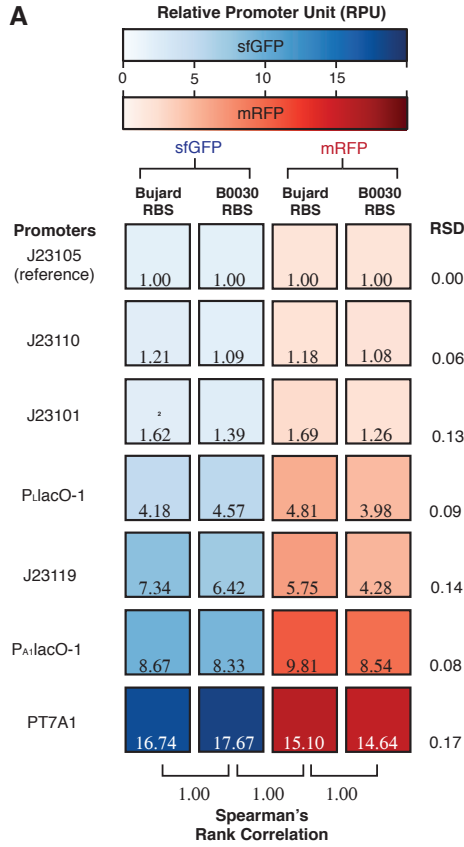
$$RPU_1 : RPU_2 : \dots : RPU_N = k_{Tr1}(\text{Pr}_{DNA}) : k_{Tr2}(\text{Pr}_{DNA}) : \dots : k_{TrN}(\text{Pr}_{DNA})$$

As expected, we observed highly variable RPUs of the promoters in the control library in different contexts (**Figure 7.5A&B**). In contrast, we observed consistently constant RPUs for each promoter in different contexts with RNA processing (**Figure 7.5C&D**). Furthermore, correlation between RPUs of the promoters with different expression units showed almost perfect linear correlation ( $R^2 = 0.97\sim 0.99$ ; the control library  $R^2 = 0.30\sim 0.84$ ), implying the promoter activity characterized with one expression unit could be used to predict its expression with other translation units, and allow us to characterize the “standard” RPU of promoters.

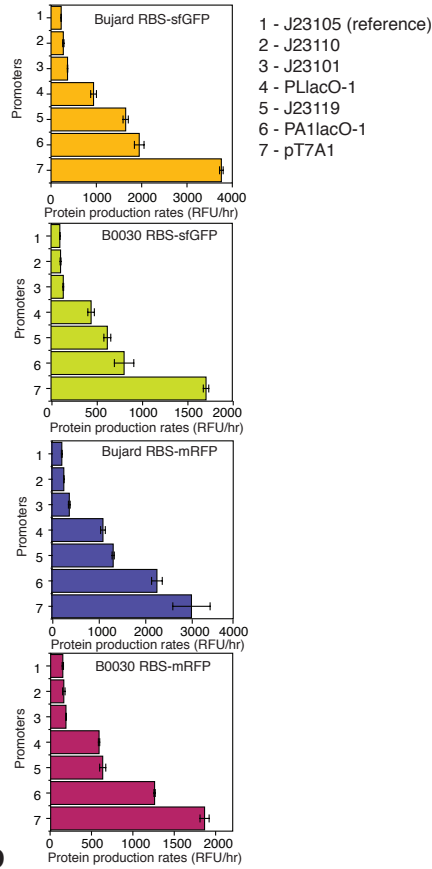
Thus, insertion of CRISPR cleavage hairpins between promoters and downstream sequences allows standard performance of promoters. Since promote engineering is central to genetic circuitry engineering and metabolic engineering, our ability to precisely and reliably design the activity of promoters will allow us to predict their behaviors in complex circuits.



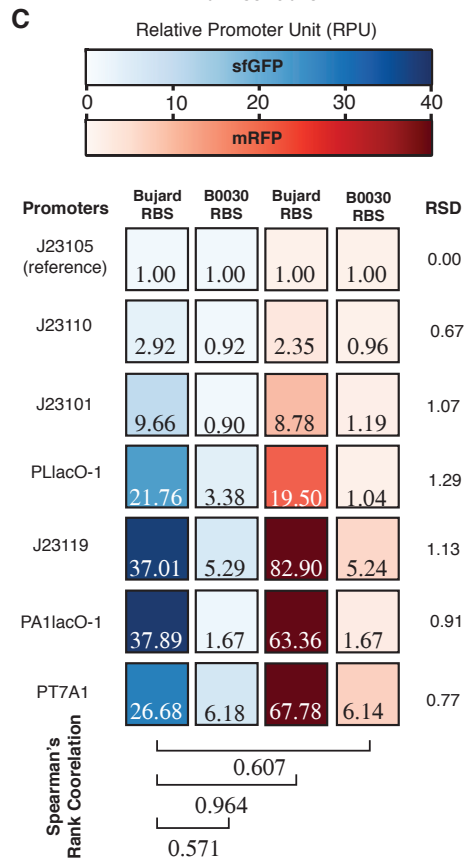
**A**



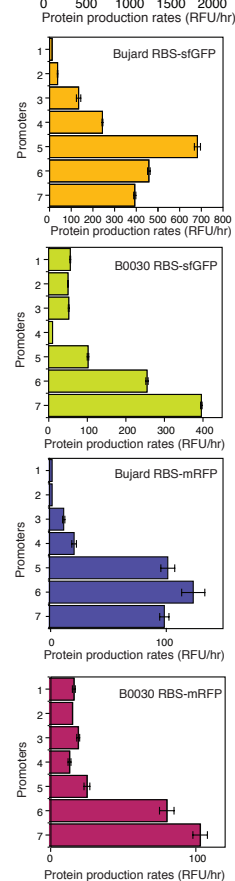
**B**



**C**



**D**

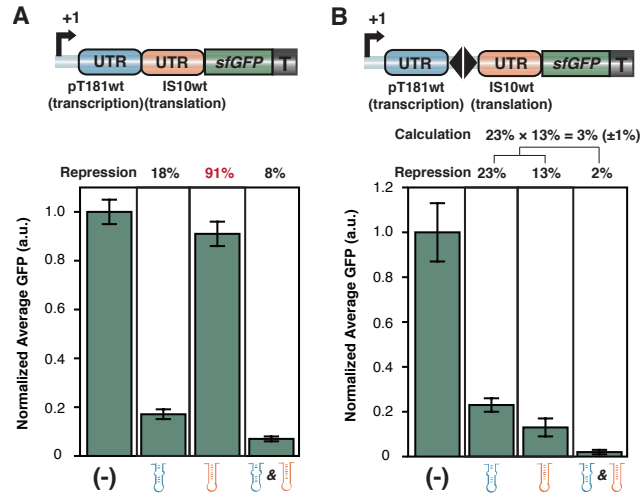


**Figure 7.5 RNA processing improves the predictability of promoters.** (A) Twenty-eight combinatory circuits composed of seven promoters, two RBSs and two reporter genes were constructed with the Csy4 cleavage hairpin inserted between the promoters and RBSs. The heatmaps were plotted from the RPU values of promoters, with each column normalized to the reference promoter J23105. The numbers after each row show the RSD values of RPUs (variance) across the contexts. Spearman's rank correlation coefficients between columns are labeled on the bottom. (B) Absolute protein production rates of gene circuits with RNA processing. From top to bottom, the contexts are Bujard RBS and sfGFP, B0030 RBS and sfGFP, Bujard RBS and mRFP, and B0030 RBS and mRFP. The numbers were used to calculate the RPUs of the promoters. (C) The heatmap shows the RPU of each promoter without RNA processing. (D) Absolute protein production rates without RNA processing.

## 7.6 Engineering Composite UTR Functions

### 7.6.1 Engineering tandem UTR systems

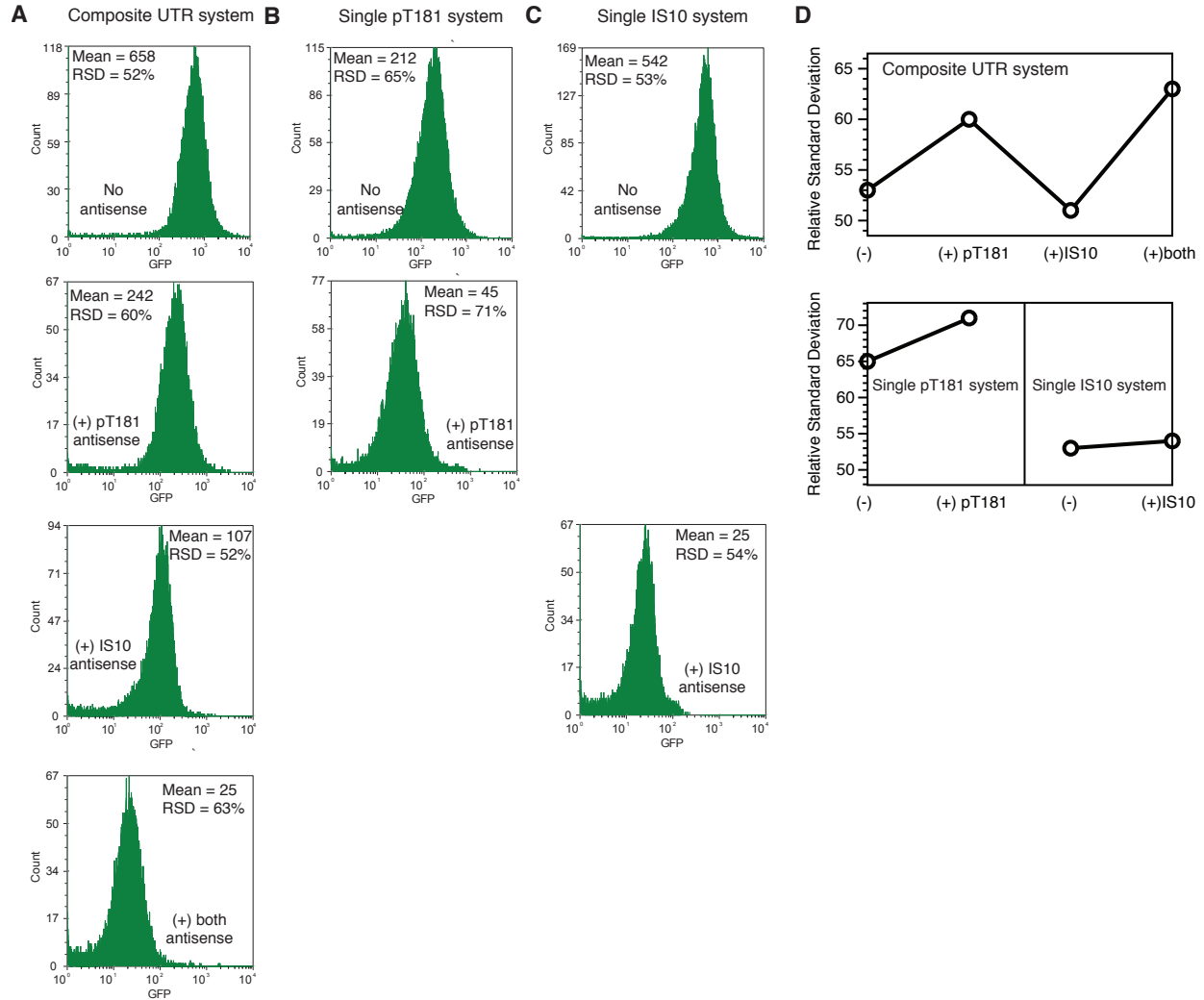
Post-transcriptional CRISPR hairpin cleavage facilitates inserting complex regulations into a single transcript that would otherwise be hampered by undesigned interactions. For example, RNA processing might allow engineering of complex UTR functions. Since *cis*-acting UTR elements can control transcription elongation (pT181wt<sup>109</sup>) or translation initiation (IS10wt<sup>132</sup>), we study if these two types of UTR regulations can be combined in tandem to exhibit multi-input regulation. To test this, we cloned the pT181wt UTR to the upstream of the IS10wt UTR, which together control the downstream sfGFP gene. We tested this construct under different conditions of antisense RNA: no antisense RNA, with either pT181wt or IS10wt antisense RNA, or both. While the pT181wt UTR could be repressed by its antisense, the IS10wt UTR failed to exhibit repression (**Figure 7.6.1A**), consistent with previous reports that nucleotides upstream of the IS10wt UTR could prevent antisense RNA from binding<sup>125</sup>. We then inserted the CRISPR cleavage hairpin between the two UTR elements, which should remove 5' contextual interference on the IS10wt UTR. Indeed, with RNA processing, each antisense RNA repressed the cognate UTR element effectively. Furthermore, the composite UTR functioned as a multiplication of the individual ones, implying that the two UTR elements acted independently due to cleavage (**Figure 7.6.1B**).



**Figure 7.6.1 RNA processing allows design of complex UTR regulations.** The transcriptional pT181wt and the translational IS10wt UTRs are used in tandem to perform a composite function, without (**A**) or with (**B**) RNA processing. All data are normalized to expression of the construct in the absence of all antisense RNAs. The numbers above the graphs show the repression percentages.

### 7.6.2 Characterizing noise properties of pT181 and IS10

We next used the composite UTR construct to evaluate the difference between the regulatory functions of pT181wt and IS10wt. We focused on the noise properties of these two types of regulation. Traditional methods using two different constructs to compare their regulatory difference is unlikely to provide useful insights into their noise properties, mainly because of the different reporter systems used. Our single composite UTR construct could remove the reporter layer difference in a way that any difference between the two regulatory functions based on this single construct could be attributed to the properties of antisense regulation. We measured the fluorescence expression under various antisense RNA conditions using the flow cytometer, and compared that to the single pT181wt and single IS10wt UTR constructs. Our results showed that the pT181wt system increased regulatory noise, while IS10wt system exhibited no change on the noise (**Figure 7.6.2A**). This trend can be confirmed using the composite UTR construct or the single UTR systems (**Figure 7.6.3B&C**). Furthermore, the pT181wt system caused a higher variability between cell populations than the IS10 system (**Figure 7.6.3D**). While we cannot distinguish where the noise properties come from, - the antisense RNA folding, RNA-RNA interaction, or the particular regulatory modes used (transcription or translation), - our results hinted on the deep quantitative properties of various *cis* regulatory systems. These properties should be considered when more quantitative properties are to be designed in addition to the fact that both systems caused around 10-fold repression. We hypothesize that further investigation on the different mutants of pT181 and IS10 systems should provide more insights into the problem.



**Figure 7.6.2 Noise properties of transcriptional control (pT181) and translational control (IS10).** (A) We assayed the fluorescence distribution across populations using the composite UTR system in different antisense RNA conditions. Addition of IS10 antisense RNA does not change expression variability but addition of pT181 antisense RNA increases variability. (B) Expression variability for pT181 system. (C) Expression variability for IS10 system. (D) Summary of RSD values for different reporter systems with different antisense conditions for (A-C).

## 7.7 Engineering Synthetic Operons

### 7.7.1 Introduction to operons

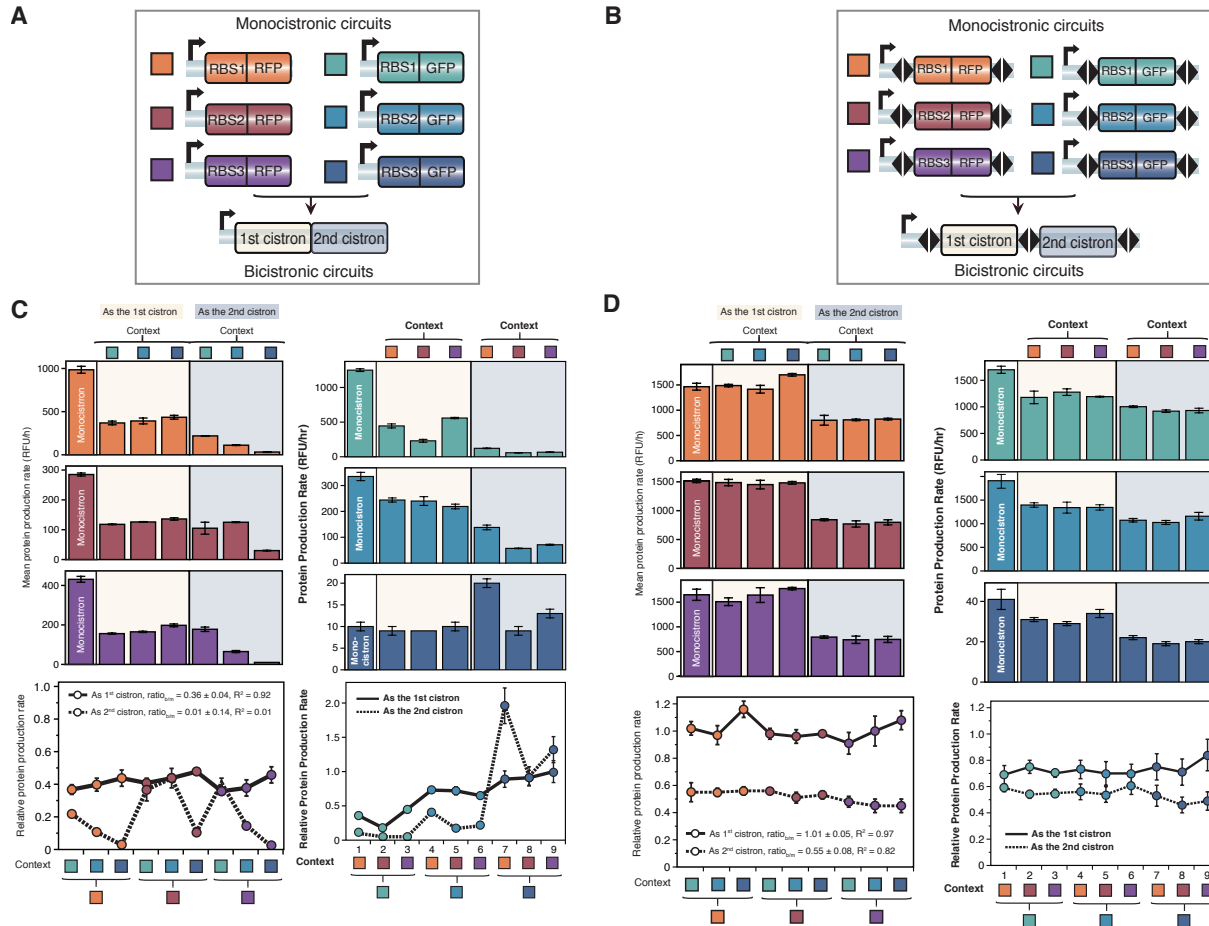
In a second application, we applied the transcript processing mechanism to engineer a reliably operating synthetic multigene operon system. Operons are advantageous for many applications due to their compactness and coordinated regulation<sup>194</sup>. However, rational engineering of synthetic operons remains a challenge due to the perplexing interactions between the cistrons<sup>195,196</sup>.

### 7.7.2 Rational design strategy using RNA processing

We hypothesized that introduction and cleavage of CRISPR hairpins between cistrons would alleviate context effects due to post-transcriptional interactions, although there would be no effect on transcriptional polarity. If the hypothesis is correct, then the correlation of the expression from a gene in a monocistronic operon with that in either the promoter proximal and distal positions in a bicistronic operon should be higher for a design in which cistrons are bounded by CRISPR hairpins.

We first characterized the protein production rates of six monocistronic circuits comprised of three different 5' SD/UTR sequences (“Bujard SD/UTR”, “B0030 SD/UTR” and “Anderson SD/UTR”) with two fluorescent proteins (sfGFP and mRFP). Then we combined these monocistrons into a library of bicistronic constructs such that each mRFP-based monocistron was composed with every sfGFP-based monocistron and vice versa, which produced a total of eighteen bicistronic circuits in permutation (**Figure 7.7.2A**). We found that if genes appeared in the first cistron, the expression was linearly correlated to that of the corresponding monocistronic circuit ( $\text{ratio}_{\text{RFP}} = 0.36 \pm 0.04$ ,  $R^2 = 0.92$ ); however, if genes appeared in the second cistron, the expression had poor correlation to that of the monocistron ( $\text{ratio}_{\text{RFP}} = 0.01 \pm 0.14$ ,  $R^2 = 0.01$ ) (**Figure 7.7.2 C**).

The second library was identical to the above except that the CRISPR hairpins were inserted at both the 5' and 3' ends to encapsulate each cistron (**Figure 7.7.2B**). Measurement of expression of these CRISPR-mediated bicistronic designs revealed a striking linear correlation between the bicistronic and monocistronic configurations: if a gene appeared as the first cistron, the measured mean protein production rate was almost the same as the monocistron for each expression unit ( $\text{ratio}_{\text{RFP}} = 1.01 \pm 0.05$ ,  $R^2 = 0.97$ ) and if a gene appeared as the second cistron, its expression was linearly correlated to that of the corresponding monocistronic circuit ( $\text{ratio}_{\text{RFP}} = 0.55 \pm 0.08$ ,  $R^2 = 0.82$ ) (**Figure 7.7.2D**). Not only RFP but also GFP showed a similar trend of change, confirming the effect is not gene-specific.



**Figure 7.7.2 RNA processing allows design of predictable synthetic operons.** (A) Six monocistrons are constructed and combined in pairs to generate eighteen bicistrons. RBS1 = Bujard RBS; RBS2 = B0030 RBS; RBS3 = Anderson RBS. (B) Synthetic operons with Csy4 cleavage elements. (C) Measured mean RFP (left) and GFP (right) production rates for the monocistrons and bicistrons. The bicistrons with RFP in the first cistron are shaded in yellow; and those with RFP in the second cistron are shaded in grey. The other cistron is shown on the top as “context”. Correlation between the bicistrons and monocistrons are shown on the bottom, with the solid line showing RFP as the first cistron and the dotted line showing RFP as the second cistron. (D) Mean RFP (left) and GFP (right) production rates for the monocistrons and bicistrons with RNA processing. Correlation between bicistrons and monocistrons are shown on the bottom.

### 7.7.3 Using synthetic operons to measure transcription polarity

Previous studies have shown that genes in operons distal from the promoter are expressed at lower levels than the proximal genes in an effect called “transcriptional polarity”<sup>28</sup>. In our synthetic operon circuits with the CRISPR hairpins, while transcript processing effectively removes post-transcriptional interactions among cistrons, the expression should be weighted by a factor determined by transcriptional polarity. After transcript cleavage (we expect the cleavage speed is very fast), we expect similar transcriptional polarity effects because distant genes have

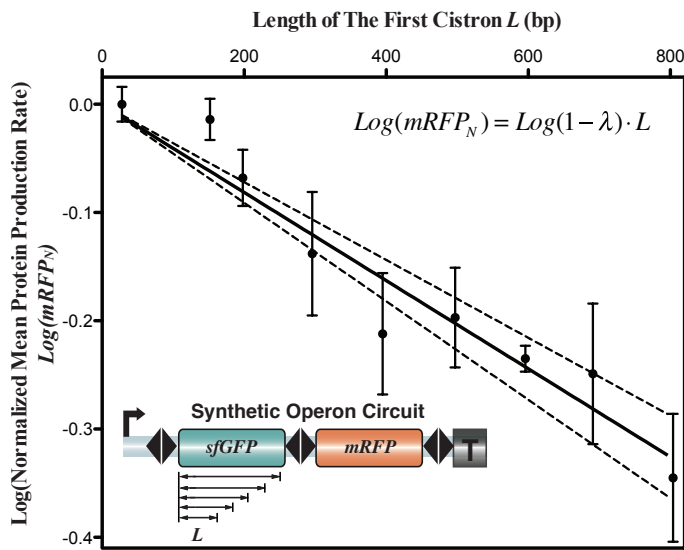
lower frequencies of RNA polymerases readthrough. To confirm our hypothesis, we shortened the length of the first cistron from the 3' end. The length of the first cistron including SD/UTR sequences and sfGFP coding sequences is 804bp in total, and we gradually shortened its length to 690bp, 594bp, 495bp, 393bp, 294bp, 198bp, 150bp, and measured expression of the second cistron including mRFP (**Figure 7.7.3**). If we assume the expression of the second cistron changes only because RNA polymerases transcribe different lengths of the first cistron, and assume this rate is  $\lambda$  (per nucleotide,  $0 < \lambda < 1$ ). Then the quantity equals to the relative expression of mRFP in the second cistron, where  $L$  is the length of the first cistron. Taking logarithm on the quantity and we expect a linear relationship to hold:

$$\text{Log}(mRFP_N) = \text{Log}(1 - \lambda) \cdot L$$

Where  $mRFP_N$  is normalized mRFP protein production rate (to that in the monocistron). By fitting measured  $mRFP_N$  to  $L$ , we obtained a linear correlation, whose slope equals to  $\text{Log}(1 - \lambda)$ . Using the 95% confidence interval (CI) obtained from the fitness, we calculated the average RNA polymerase dropoff rate was

$$\lambda = 8.3 \times 10^{-4} \sim 1.04 \times 10^{-3} / \text{nucleotide}$$

This is a relative high drop-off rate, meaning that transcriptional readthrough is dropped by 50% for transcribing every 1000bp. This probably reflects that the sequences of these fluorescent proteins (sfGFP and mRFP) aren't optimized to maximize RNA polymerase readthrough in *E. coli* Top10 cells.



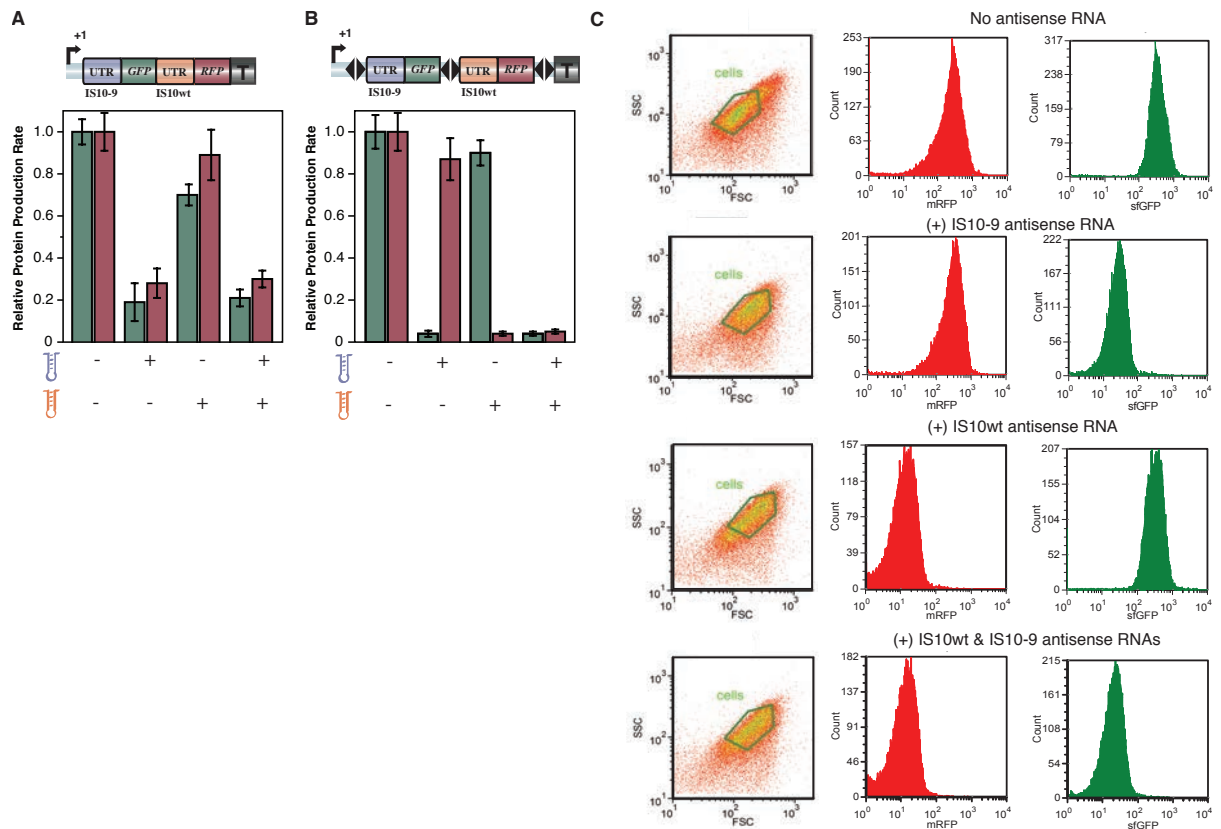
**Figure 7.7.3 Measurement of transcriptional polarity effects in the synthetic operon.** We gradually shortened the length of the first cistron from the 3' end, and measured expression of the second cistron. The plot of the normalized mRFP expression and the length of the first cistron shows a linear correlation, whose slope can be used to calculate the average dropoff rate of RNA polymerase on this construct. The dotted lines show the 95% confidence interval of fitness.

## 7.8 Engineering Independent Translational Controls Into Synthetic Operons

The ability to process a transcript into separate and independent units permits the opportunity to more easily insert intergenic regulation to modulate the all-or-nothing promoter-based control in multigene operons. For example, as noted above, effective IS10wt translation regulatory function requires a free 5' end, thus rendering it useless for regulation of promoter distal cistrons.

However, also as above, we might expect that efficacy could be restored if the cistrons containing IS10 regulator variants are cleaved apart.

To test this, we incorporated orthogonal variants of IS10-derived *cis*-acting SD/UTR elements into synthetic operons to translationally regulate individual genes by the cognate *trans*-acting antisense RNAs. We fused two orthogonal UTR sequences (IS10wt and IS10-9) separately to sfGFP and mRFP synthetic operon with and without the CRISPR hairpins. We similarly modified the pCsy4 plasmid to create three plasmids that constitutively expressed only IS10wt antisense, only IS10-9 antisense or both. Without CRISPR hairpins and cleavage, when the first antisense RNA repressed expression of the first cistron, it also knocked down that of the second gene, presumably because the antisense RNA affects not only translation but also overall transcript stability. Further, the second antisense RNA failed to repress the second UTR element (**Figure 7.8A**). In contrast, the configuration with hairpins and cleavage enabled perfect independent control - presence of one antisense RNA only repressed the target gene without affecting the other, and presence of both antisense RNAs knocked down both genes (**Figure 7.8B&C**). These results suggest that the complex global and local interactions among elements on a transcript that can be controlled through efficient transcript cleavage thereby facilitating rational design of differentially regulated multi-cistronic operons.



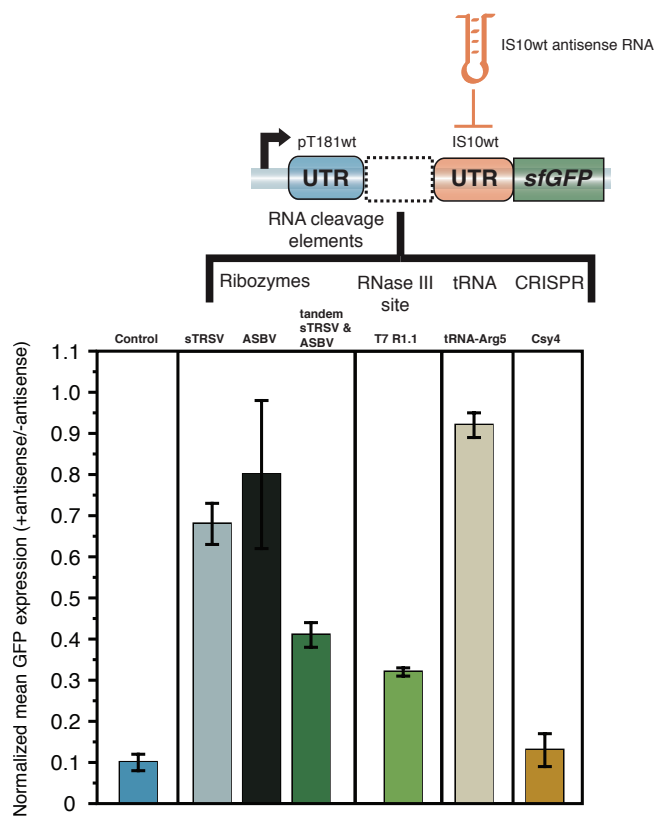
**Figure 7.8 Construction of complex UTR controls in synthetic operons.** Orthogonally acting IS10 antisense RNA-mediated elements are used to control RFP and GFP independently in the operon without (**A**) or with (**B**) RNA processing. The green bars show the relative production rates of GFP, and the red bars show that of RFP. (**C**) Flow cytometry data of the synthetic operon



regulated by two orthogonal UTR systems. From top to bottom: no antisense; only IS10-9 antisense; only IS10wt antisense; both antisense RNAs. The first column shows the forward scatter-side scatter 2D plots with the polygon gating 75% of cell populations. The second and third columns show histograms of mRFP and sfGFP respectively, with the gates used for calculating mean and standard deviation shown in red.

## 7.9 Comparing CRISPR to Other RNA Cleavage Elements

To compare the efficacy of Cys4-mediated cleavage to other RNA cleavage elements, we inserted either hammerhead ribozyme, tandem ribozymes or RNase III cleavage sites<sup>197</sup> into the same circuit between the two UTR elements. However, none of these elements were as effective as the CRISPR-based system. Use of the most effective alternative element, RNase III T7R1.1, allowed recovery of  $67\% \pm 3\%$  of the full IS10wt ON/OFF dynamic range. In contrast, the use of the CRISPR element recovered  $97\% \pm 5\%$  (**Figure 7.9**), suggesting that the CRISPR cleavage targets may be more robust in different sequence contexts. This is might be a requirement given the variable adjacent sequences in which they are embedded naturally.

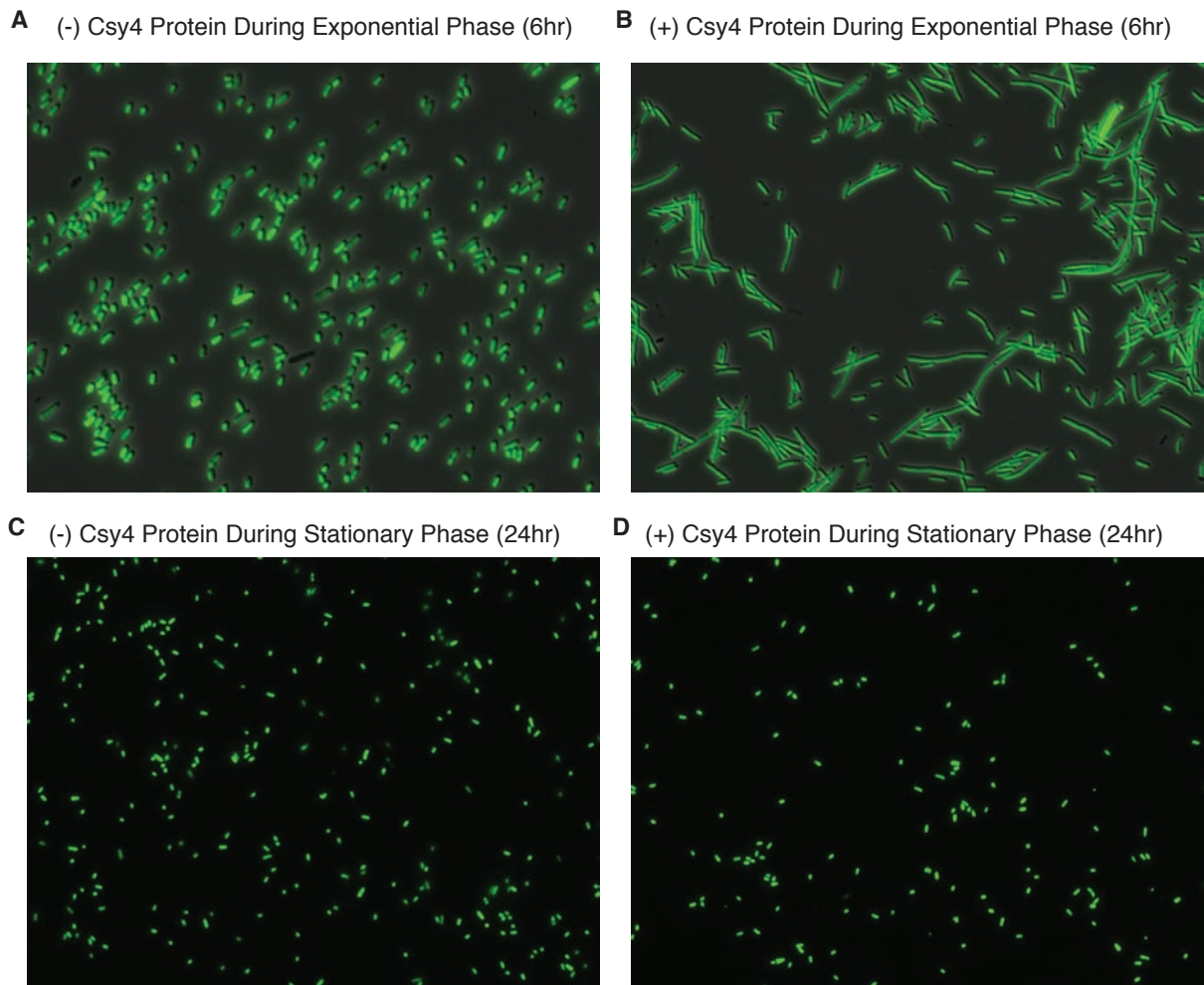


**Figure 7.9 Comparison of the efficacy of RNA cleavage elements.** Different RNA cleavage elements were inserted into the tandem UTR construct (PT181wt and IS10wt) as shown in Fig. 4d. Fluorescence of each construct in the presence of IS10wt antisense RNA was measured with flow cytometry, averaged over three biological replicates, and normalized to that without IS10wt antisense. The control shows the repression of IS10wt antisense RNA on IS10wt UTR that is not fused in tandem to PT181wt UTR. sTRSV is short for small Tobacco RingSpot Virus

hammerhead ribozyme<sup>118</sup>. ASBV is short for Avacado SunBlotch Virus ribozyme<sup>202</sup>. T7 1.1 is the RNase III recognitve site derived from T7 phage<sup>197</sup>. tRNA-Arg5 is obtained from<sup>203</sup>. Among all constructs, CRISPR enabled maximal repression and exhibited the highest efficacy of restoring riboregulator-UTR function.

### 7.10 Growth Effects of Csy4 Expression

We noticed slower growth of *E. coli* cells during the exponential phase when inducing the expression Csy4 protein expression (normally the growth rate is 0.3~0.4 h<sup>-1</sup> for cells without Csy4 expression, and the growth rate is 0.15~0.25 h<sup>-1</sup> for cells with Csy4 expression). The cells were visualized under the Zeiss Axio Observer D1 microscope (**Figure 7.10A-D**). The cells that express Csy4 protein phase grow much larger than the cells without Csy4 protein expression during the exponential. However, after the cells entered the stationary phase, no size difference was noticed for both cells with and without Csy4 protein expression. We suspect the *E. coli* cells that express Csy4 protein triggered some type of SOS response, which needs to be further investigated.



**Figure 7.10 Visualization of cells (-) or (+) Csy4 expression under the microscope. (A)** The *E. coli* Top10 cells (-) Csy4 expression during the exponential phase. **(B)** Cells (+) Csy4

expression during the exponential phase. (C) Cells (-) Csy4 expression during the stationary phase. (D) Cells (+) Csy4 expression during the stationary phase.

### 7.11 Discussion on Utilities of RNA processing

In this work, we have developed the CRISPR-derived transcript processing system to achieve predictable control of differentially regulated multigene operons. The segmentation of transcripts by controlled cleavage alleviates post-transcriptional interactions among transcript components and suggests that standard expression control parts, properly bounded by CRISPR hairpins, provide for rational and reliable engineering of gene expression. Indeed, we showed that using the CRISPR transcript processing system, we could characterize the activity of individual components and compose them to form predictable gene expression units. These data demonstrate that CRISPR-mediated insulation has advantages over other RNA cleavage elements such as ribozymes or RNase III cleavage sites. Due to its robust behavior in different contexts, we anticipate that this system is transferrable to other prokaryotes and across kingdoms. Cys4 belongs to a family of RNA cleavage enzymes that each recognizes their individual target hairpins<sup>178</sup>. This exquisite specificity suggests that it is unlikely that these proteins will cleave off-target substrates. Further, since the target hairpins have presumably evolved to maintain structure when flanked by random phage sequences, it is expected that they will maintain their structure despite variable sequence context. Indeed, the existence of orthogonal cleavage proteins and targets opens the possibility of using combinations of these components in the same cell. Differential processing of RNA into defined segments in a regulated fashion might allow another layer of complex control of expression function and might also, conceivably, be used to create different RNA fragments to generate dynamic RNA scaffolds<sup>198</sup>.

More generally, the approach of enforcing standard junctions among well-characterized biological parts is a defining feature of the synthetic biology approach. The CRISPR parts described here are an example of specific elements designed to remove “parasitic” interactions among regulatory regions by physically uncoupling them while maintaining their functional composition. Context-controlling parts are one emerging theme in synthetic biology and have spanned efforts such as the development of orthogonal ribosomes designed, in part, to decouple from host physiology and the creation of biological microcompartments that encapsulate possibly toxic biosynthetic products away from the cytoplasmic milieu<sup>199-201</sup>. In our case, creating such insulation at the post-transcriptional level revealed clear the strong transcriptional polarity that comes part-and-parcel with multicistronic operons. Whether this sort of context-dependence can be removed or must remain a possibly predictable property of multigene operons remains to be seen.

Finally, the CRISPR system as a whole provides a number of other potentially useful functions for the engineering of cellular systems. For example, the somewhat mysterious process of acquisition in which sequences from infecting phage are packaged and inserted into the genome at a precise location between a promoter and the last inserted sequence might be harnessed for a number of interesting tasks ranging from design of cellular stack memory and regulation, to *in vivo* sequence synthesis. In any case, this relatively recently elucidated bacterial system is a reminder that there is a vast repository of useful biological function in earth’s pan-genome beyond the classically sought biosynthetic enzymes.

## Chapter 8 Conclusions

### 8.1 Summary of Contributions

This dissertation focuses on the scalable programming of synthetic RNA-based genetic and cellular systems. In this work, we have established the bacterial noncoding RNAs as an intriguing engineering substrate for programming cells. They are highly amenable to rational design following simple rules of Watson-Crick base pairing. They can perform a wide range of diverse functions, which can be relatively easily mapped to their sequences and structures that can be rapidly resolved using cutting-edge techniques. We provide a set of engineering principles for using these synthetic RNA elements as predictable controllers to sense environment signals and form complex genetic systems. Combining with the synthetic RNA processing system, these synthetic RNA elements can recapitulate the functions of natural ncRNAs, and provide a versatile and powerful engineering platform for higher-order cellular information processing and computation useful for applications ranging from chemical production to therapeutics.

A first step in programming cells is to create a large number of versatile, predictable, and interoperable synthetic regulatory elements that can be combined together for complex functions. To create synthetic RNA regulatory elements, we started with two primary types of ncRNA-mediated systems. Both systems modulate the mRNA-level regulatory signals encoded in the 5' untranslated region, and are inhibited by distantly encoded ncRNAs. In the first system (the *S. aureus* pT181 system) the ncRNA controls transcription elongation, whereas in the second system (the *E. coli* IS10 system) the ncRNA controls translation initiation<sup>132</sup>. To create new orthogonal RNA elements that act independently, we modified the RNA-RNA interaction of the natural systems. We hypothesize that the loop motif on the ncRNA determines its interaction with the mRNA target. We systematically mutated this motif and the target region on the mRNA, and measured a large library of ncRNA-mRNA interactions using fluorescent protein reporter systems. Our characterization results in families of new RNA elements that are orthogonal with each other for both transcription and translation controls. Furthermore, we developed mathematical thermodynamic models to predict new RNA elements *in silico*. In the case of translational RNA elements, the predictions of our mathematical model highly correlated with the experiment data.

The designed synthetic RNA elements can act as wires of genetic circuits, connecting one gene to the other. To engineer these RNAs to directly sense and integrate cellular and external signals, we designed a novel type of allosteric RNA chimera molecule by fusing ncRNAs to RNA aptamers. We demonstrate the design principles for creating such allosteric RNA molecules that can sense either proteins (such as MS2 coat protein) or small molecules (such as theophylline) and control transcription or translation. We show that our design strategy is highly modular between the ligand-sensing motif on the RNA aptamer and the target-recognition motif on the ncRNA, which allows us to reconfigure these two motifs in combination to engineer families of orthogonal RNA chimeras that respond to different ligands and regulate different gene targets. We further show that multiple RNA chimeras allow logical integration of molecular signals in the same cell, which can be used to perform complex cellular information processing.

We assembled synthetic RNA elements to build fundamental regulatory network motifs. Using two orthogonal pairs of transcriptional synthetic RNA elements and their targets, we

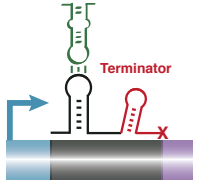
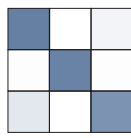
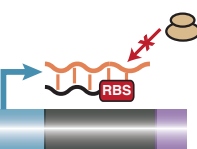
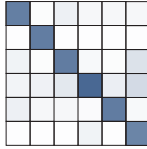
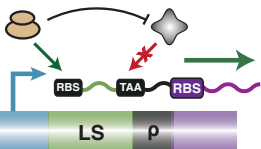
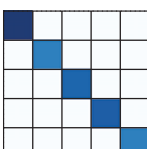
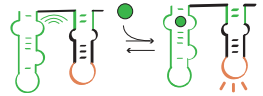
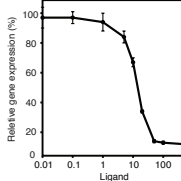
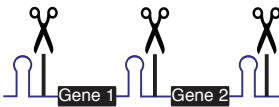

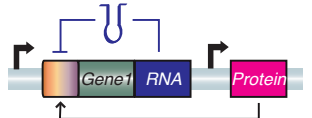
created three basic motifs. The first motif is the independent control of multiple genes in the same cell. The second motif is the logic integration of multiple ncRNA signals by combining tandem ncRNA targets on the same transcript. The third motif is the RNA-only transcriptional cascade. In this motif, multiple ncRNAs are chained together to propagate genetic signals along the cascade. To note, this is also the first reported network that is mediated solely by RNAs. We characterized the properties of these RNA circuits and measured the time response, signal sensitivity, and noise across cell populations. Our results show that these RNA circuits are highly dynamic and can respond to stimuli very rapidly (~minutes), which is likely due to the fast production, degradation, and folding of RNA molecules. On the other hand, they exhibit lower sensitivity and are noisier than equivalent protein circuits. Our results suggest the utilities of RNA circuits in controlling stress responses (time response), filtering high-frequency stimuli (sensitivity), and balancing expression over a wide range in unpredictable environments (noise).

To further improve the scalability of the designed genetic parts, we created a method for converting translational controllers to transcriptional controllers. This is motivated by the fact that although larger families of orthogonal translational ncRNA regulators can be more easily predicted and engineered, it is harder to combine them into complex genetic circuits. To convert translation to transcription, we utilized the *E. coli* leader sequence-mediated regulatory system, the *tna* system. We fused multiple orthogonal versions of IS10 UTR targets to control the translation of the leader sequence. Our results showed that this successfully created five mutually orthogonal transcriptional controllers. Furthermore, the system exhibited a much bigger ON/OFF range, implying this type of chimeric system behaved synergistically. Combining with the RNA processing system, we created an expandable multi-input logic gate.

A critical challenge in genetic engineering and genetic circuit design is that complex interactions between molecular components such as DNA, RNA, and proteins often lead to highly variable behaviors across different cellular contexts. To engineer predictable biological systems on the RNA level, we developed a synthetic RNA processing platform from the bacterial CRISPR genetic immune pathway. The synthetic RNA processing system can efficiently and specifically cleave desired precursor mRNAs at designed loci. Using this system, we show that transcript cleavage enables quantitative programming of gene expression by modular assembly of promoters, ribosome binding sites, *cis* regulatory elements, and riboregulators. These basic components can be grouped into multi-gene synthetic operons that behave predictably only after RNA processing. Physical separation of otherwise linked elements within biological assemblies (mRNAs) allows design of sophisticated RNA-level regulatory systems that are not possible without it. Thus, our results exemplify a crucial design principle based on controllable RNA processing for improving the modularity and reliability of genetic systems.

A summary of designed genetic parts and synthetic circuits is listed in **Figure 8.1**.

## New RNA-Level Genetic Elements

	Name	Mechanism	Performance	Notes
Current	<b>Orthogonal Switches</b>			
	PT181			<ul style="list-style-type: none"> <li>* Transcriptional control</li> <li>* 3 mutually orthogonal pairs</li> <li>* 7~10 fold repression</li> </ul>
	IS10			<ul style="list-style-type: none"> <li>* Translational control</li> <li>* 6 mutually orthogonal pairs</li> <li>* 10~20 fold repression</li> </ul>
	Lead peptide			<ul style="list-style-type: none"> <li>* Converts translation control to transcript control</li> <li>* Retains orthogonality</li> <li>* High dynamic range: &gt;500 fold</li> </ul>
	<b>RNA Sensors</b>	Theophylline/ MS2 coat protein/ Q $\beta$ coat protein		
Future	<b>RNA Processing Systems</b>			
				<ul style="list-style-type: none"> <li>* Sequence-specific cleavage</li> <li>* Highly efficient</li> <li>* Controllable</li> </ul>
	<b>Applications</b>			
	Metabolic pathway			<ul style="list-style-type: none"> <li>* Differential control of operonic genes inside an operon</li> <li>* Tune the relative expression for rapid optimization</li> </ul>
	Dynamic circuits e.g. adaptation			<ul style="list-style-type: none"> <li>* RNA degradation is fast</li> <li>* Possibly enables dynamic control</li> </ul>

**Figure 8.1 Summary of genetic parts in this dissertation and their applications.**

## 8.2 Future Directions

We propose a few future directions based on the current work.

The first direction is to further expand the orthogonal libraries for both transcriptional and translational regulators. We obtained libraries of orthogonal RNA parts, but the number of mutually orthogonal parts is still limited (3 for transcriptional controllers, and 6 for translational controllers). It is important to understand what is the upper limit on the number of mutually

orthogonal parts. Furthermore, the performance of the genetic parts needs to be improved, especially the ON/OFF dynamic range.

Second, we need to address to major drawbacks of the ncRNA-mediated systems. The first drawback is that ncRNA molecules always need to be overexpressed for full repression on the target. Unless there is an effective method for ncRNA amplification, their uses in constructing circuits could be limited. The second drawback is that most ncRNA-mediated systems lack cooperativity. Since it has been shown that many useful functions require a high sigmoidality<sup>20</sup>, lack of cooperativity is likely to restrict their uses. However, this might be partially compensated by the used of positive feedbacks<sup>204</sup>.

Third, to understand the unique regulatory abilities of ncRNA and proteins in depth, we need to characterize the genetic circuits built from these two types of regulators. When ncRNA instead of protein should be used? What is the difference between transcriptional feedbacks and translational feedbacks? Answering these questions will establish a set of design principles for building complex biological systems by choosing the best genetic parts.

Fourth, the ncRNA-based systems should be applied to applications including balancing enzyme productions, refactoring complex genetic systems, and constructing higher-order genetic circuits especially the dynamic circuits such as adaptation and counters.

Finally, we need a powerful computation platform to forward design parts, combine them into basic regulatory motifs, and assemble these motifs into large-scale genetic networks. This type of CAD platform, combining with increasing DNA sequencing and synthesis capabilities, should provide a powerful approach for programming cells that can be widely used to application in therapeutics, medicine, environment, agriculture, energy, and chemical production.

## Reference

1. Kay, L. *The molecular vision of life: Caltech, the Rockefeller Foundation, and the rise of the new biology*. (1996).
2. Watson, J.D. & Crick F.H.C. A Structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
3. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
4. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
5. Rothberg, J. The development and impact of 454 sequencing. *Nat. Biotechnol.* **26**, 1117-1124 (2008).
6. Wetterstrand, K. *DNA sequencing costs: data from the NHGRI large-scale genome sequencing program*. (2011).
7. Carlson, R. The changing economics of DNA synthesis. *Nat. Biotechnol.* **27**, 1091–1094 (2009).
8. Gibson, D.G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
9. Purnick, P.E.M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
10. Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
11. Arkin, A. Setting the standard in synthetic biology. *Nat. Biotechnol.* **26**, 771–774 (2008).
12. Collins, J. Synthetic Biology: Bits and pieces come to life. *Nature* **483**, S8–10 (2012).
13. Voigt, C.A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17**, 548–557 (2006).
14. Lucks, J.B., Qi, L., Whitaker, W.R. & Arkin, A.P. Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* **11**, 567–573 (2008).
15. Nandagopal, N. & Elowitz, M.B. Synthetic biology: integrated gene circuits. *Science* **333**, 1244–1248 (2011).
16. Li, G.-W., Oh, E. & Weissman, J.S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* (2012). doi:10.1038/nature10965
17. Sørensen, M.A., Kurland, C.G. & Pedersen, S. Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* **207**, 365–377 (1989).
18. Holm, L. Codon usage and gene expression. *Nucleic Acids Res.* (1986).
19. Buck, M., Miller, S. & Drummond, M. Upstream activator sequences are present in the promoters of nitrogen fixation genes. *Nature* **320**, 374-378 (1986).
20. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
21. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
22. You, L., Cox, R.S., Weiss, R. & Arnold, F.H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
23. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
24. Haldimann, A., Daniels, L.L. & Wanner, B.L. Use of new methods for construction of



- tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. *J. Bacteriol.* **180**, 1277–1286 (1998).
25. Whiteley, M. & Greenberg, E.P. Promoter Specificity Elements in Pseudomonas aeruginosa Quorum-Sensing-Controlled Genes. *J. Bacteriol.* **183**, 5529–5534 (2001).
  26. Seila, A.C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
  27. Toulmé, F., Mosrin-Huaman, C., Artsimovitch, I. & Rahmouni, A.R. Transcriptional pausing in vivo: a nascent RNA hairpin restricts lateral movements of RNA polymerase in both forward and reverse directions. *J. Mol. Biol.* **351**, 39–51 (2005).
  28. Wek, R.C., Sameshima, J.H. & Hatfield, G.W. Rho-dependent transcriptional polarity in the ilvGMEDA operon of wild-type Escherichia coli K12. *J. Biol. Chem.* **262**, 15256–15261 (1987).
  29. Skordalakes, E. Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell* **127**, 553–564 (2006).
  30. Cardinale, C., Washburn, R. & Tadigotla, V. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* **320**, 935–938 (2008).
  31. Lesnik, E.A. *et al.* Prediction of rho-independent transcriptional terminators in Escherichia coli. *Nucleic Acids Res.* **29**, 3583–3594 (2001).
  32. Larson, M.H., Greenleaf, W.J., Landick, R. & Block, S.M. Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**, 971–982 (2008).
  33. Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7**, 481 (2011).
  34. Hall, M., Gabay, J. & Débarbouillé, M. A role for mRNA secondary structure in the control of translation initiation. *Nature* **295**, 616–618 (1982).
  35. Chen, H., Bjercknes, M. & Kumar, R. Determination of the optimal aligned spacing between the Shine–Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res.* **22**, 4953–4957 (1994).
  36. Novick, R.P., Iordanescu, S., Projan, S.J., Kornblum, J. & Edelman, I. pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell* **59**, 395–404 (1989).
  37. Gottesman, S. & Storz, G. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb. Perspect. Biol.* **1**, 1–16 (2010).
  38. Gottesman, S. The small RNA regulators of Escherichia coli: roles and mechanisms. *Annu. Rev. Microbiol.* **58**, 303–328 (2004).
  39. Peabody, D.S. Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein-RNA interaction. *J. Biol. Chem.* **265**, 5684–5689 (1990).
  40. Hattman, S., Newman, L., Murthy, H.M. & Nagaraja, V. Com, the phage Mu mom translational activator, is a zinc-binding protein that binds specifically to its cognate mRNA. *Proc. Natl Acad. Sci. USA* **88**, 10027–10031 (1991).
  41. de Smit, M.H. & van Duin, J. Translational initiation at the coat-protein gene of phage MS2: native upstream RNA relieves inhibition by local secondary structure. *Mol. Microbiol.* **9**, 1079–1088 (1993).
  42. Mandal, M. & Breaker, R.R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463 (2004).
  43. Cech, T. Ribozyme engineering. *Curr. Opin. Struct. Biol.* **2**, 605–609 (1992).

44. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
45. Rauhut, R. & Klug, G. mRNA degradation in bacteria. *FEMS Microbiol. Rev.* **23**, 353–370 (1999).
46. Kazantsev, A.V. & Pace, N.R. Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol.* **4**, 729–740 (2006).
47. Régnier, P. & Arraiano, C.M. Degradation of mRNA in bacteria: emergence of ubiquitous features. *Bioessays* **22**, 235–244 (2000).
48. Beguiristain, N., Robertson, H.D. & Gómez, J. RNase III cleavage demonstrates a long range RNA: RNA duplex element flanking the hepatitis C virus internal ribosome entry site. *Nucleic Acids Res.* **33**, 5250–5261 (2005).
49. Lundblad, E.W. & Altman, S. Inhibition of gene expression by RNase P. *Nat. Biotechnol.* **27**, 212–221 (2010).
50. Keasling, J. Controlling messenger RNA stability in bacteria: strategies for engineering gene expression. *Biotechnol. Prog.* **13**, 699–708 (1997).
51. Keasling, J. Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol. Prog.* **15**, 58–64 (1999).
52. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl Acad. Sci. USA* **104**, 14283–14288 (2007).
53. Wang, G.-Z., Lercher, M.J. & Hurst, L.D. Transcriptional coupling of neighboring genes and gene expression noise: evidence that gene orientation and noncoding transcripts are modulators of noise. *Genome Biol. Evol.* **3**, 320–331 (2011).
54. Schümperli, D., McKenney, K., Sobieski, D.A. & Rosenberg, M. Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell* **30**, 865–871 (1982).
55. Couzin, J. MicroRNAs make big impression in disease after disease. *Science* **319**, 1782–1784 (2008).
56. Tsai, M.-C., Spitale, R.C. & Chang, H.Y. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* **71**, 3–7 (2011).
57. Hobert, O. Gene regulation by transcription factors and microRNAs. *Science* **319**, 1785–1786 (2008).
58. Halic, M. & Moazed, D. Transposon silencing by piRNAs. *Cell* **138**, 1058–1060 (2009).
59. Kim, D. & Rossi, J. RNAi mechanisms and applications. *Biotechniques* **44**, 613–616 (2008).
60. Bunka, D.H.J. & Stockley, P.G. Aptamers come of age - at last. *Nat. Rev. Microbiol.* **4**, 588–596 (2006).
61. Mehta, P., Goyal, S. & Wingreen, N.S. A quantitative comparison of sRNA-based and protein-based gene regulation. *Mol. Syst. Biol.* **4**, 221 (2008).
62. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
63. Low, J.T. & Weeks, K.M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–158 (2010).
64. Lucks, J.B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl Acad. Sci. USA* **108**, 11063–11068 (2011).

65. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
66. Reuter, J.S. & Mathews, D.H. *BMC Bioinformatics* **11**, 129 (2010).
67. Hofacker, I.L. & Stadler, P.F. Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics* **22**, 1172–1176 (2006).
68. Zadeh, J.N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
69. Xayaphoummine, A., Bucher, T. & Isambert, H. Kinifold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* **33**, W605–10 (2005).
70. Gardner, D.P., Ren, P., Ozer, S. & Gutell, R.R. Statistical potentials for hairpin and internal loops improve the accuracy of the predicted RNA structure. *J. Mol. Biol.* **413**, 473–483 (2011).
71. Dixon, R., Eady, R., Espin, G., Hill, S. & Iaccarino, M. Analysis of regulation of *Klebsiella pneumoniae* nitrogen fixation (*nif*) gene cluster with gene fusions. *Nature* **286**, 128–132 (1980).
72. Zhan, J. *et al.* Develop reusable and combinable designs for transcriptional logic gates. *Mol. Syst. Biol.* **6**, 388 (2010).
73. Gordon, G., Sitnikov, D., Webb, C. & Teleman, A. Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. *Cell* **90**, 1113–1121 (1997).
74. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102**, 3581–3586 (2005).
75. Tabor, J., Bayer, T., Simpson, Z. & Levy, M. Engineering stochasticity in gene expression. *Mol. BioSyst.* **4**, 754–761 (2008).
76. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
77. Blake, W.J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).
78. Glick, B. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–261 (1995).
79. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26**, 787–793 (2008).
80. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
81. Barrick, D. *et al.* Quantitative analysis of ribosome binding sites in *E. coli*. *Nucleic Acids Res.* **22**, 1287–1295 (1994).
82. Dueber, J.E., Yeh, B.J., Chak, K. & Lim, W.A. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **301**, 1904–1908 (2003).
83. Voigt, C., Wolf, D., Arkin A.P. The *Bacillus subtilis* *sin* Operon. *Genetics* **169**, 1187–1202 (2005).
84. Reina, J. *et al.* Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol.* **9**, 621–627 (2002).
85. Rackham, O. & Chin, J.W. A network of orthogonal ribosome x mRNA pairs. *Nat. Chem. Biol.* **1**, 159–166 (2005).
86. Isaacs, F.J. *et al.* Engineered riboregulators enable post-transcriptional control of gene

- expression. *Nat. Biotechnol.* **22**, 841–847 (2004).
87. Beisel, C.L. & Smolke, C.D. Design principles for riboswitch function. *PLoS Comput. Biol.* **5**, e1000363 (2009).
  88. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25**, 795–801 (2007).
  89. Temme, K. *et al.* Induction and relaxation dynamics of the regulatory network controlling the type III secretion system encoded within Salmonella pathogenicity island 1. *J. Mol. Biol.* **377**, 47–61 (2008).
  90. Sambrook, J. *Molecular cloning: a laboratory manual.* (2001).
  91. Shetty, R.P., Endy, D. & Knight, T.F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).
  92. Huang, S.H. Inverse polymerase chain reaction. An efficient approach to cloning cDNA ends. *Mol. Biotechnol.* **2**, 15–22 (1994).
  93. Johannsen, M.W., Veedu, R.N., Madsen, A.S. & Wengel, J. Enzymatic polymerisation involving 2'-amino-LNA nucleotides. *Bioorg. Med. Chem. Lett.* (2012). doi:10.1016/j.bmcl.2012.03.073
  94. Qin, Y. *et al.* Improved method for rapid and efficient determination of genome replication and protein expression of clinical hepatitis B virus isolates. *J. Clin. Microbiol.* **49**, 1226–1233 (2011).
  95. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6**, 242–251 (2011).
  96. Li, M.Z. & Elledge, S.J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
  97. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
  98. Pédrelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
  99. Campbell, R.E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).
  100. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909 (2005).
  101. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616 (2006).
  102. Steen, K.-A., Malhotra, A. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease. *J. Am. Chem. Soc.* **132**, 9940–9943 (2010).
  103. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* **4**, 1413–1421 (2009).
  104. Leveau, J.H. & Lindow, S.E. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J. Bacteriol.* **183**, 6752–6762 (2001).
  105. Heidrich, N. & Brantl, S. Antisense RNA-mediated transcriptional attenuation in plasmid pIP501: the simultaneous interaction between two complementary loop pairs is required for efficient inhibition by the antisense RNA. *Microbiology* **153**, 420–427 (2007).

106. Brantl, S. & Wagner, E.G. Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. *Mol. Microbiol.* **35**, 1469–1482 (2000).
107. Brantl, S. & Wagner, E.G.H. An antisense RNA-mediated transcriptional attenuation mechanism functions in *Escherichia coli*. *J. Bacteriol.* **184**, 2740–2747 (2002).
108. Nordström, K. & Wagner, E.G. Kinetic aspects of control of plasmid replication by antisense RNA. *Trends Biochem. Sci.* **19**, 294–300 (1994).
109. Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D. & Arkin, A.P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl Acad. Sci. USA* **108**, 8617–8622 (2011).
110. Qi, L., Lucks, J.B., Liu, C.C., Mutalik, V.K. & Arkin, A.P. Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Res.* (2012).doi:10.1093/nar/gks168
111. Kolb, F.A. *et al.* Progression of a loop-loop complex to a four-way junction is crucial for the activity of a regulatory antisense RNA. *EMBO J.* **19**, 5905–5915 (2000).
112. Kolb, F.A. *et al.* Four-way junctions in antisense RNA-mRNA complexes involved in plasmid replication control: a common theme? *J. Mol. Biol.* **309**, 605–614 (2001).
113. Franch, T., Petersen, M., Wagner, E.G., Jacobsen, J.P. & Gerdes, K. Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.* **294**, 1115–1125 (1999).
114. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
115. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
116. Breaker, R.R. Complex riboswitches. *Science* **319**, 1795–1797 (2008).
117. Tamsir, A., Tabor, J.J. & Voigt, C.A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212–215 (2011).
118. Hampel, A. & Tritz, R. RNA catalytic properties of the minimum (-)sTRSV sequence. *Biochemistry* **28**, 4929–4933 (1989).
119. Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S.D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10**, 708–712 (2003).
120. Rosenfeld, N. & Alon, U. Response delays and the structure of transcription networks. *J. Mol. Biol.* **329**, 645–654 (2003).
121. Behrouzi, R., Roh, J.H., Kilburn, D., Briber, R.M. & Woodson, S.A. Cooperative tertiary interaction network guides RNA folding. *Cell* **149**, 348–357 (2012).
122. Brantl, S. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* **10**, 102–109 (2007).
123. Simons, R.W. & Kleckner, N. Translational control of IS10 transposition. *Cell* **34**, 683–691 (1983).
124. Jain, C. IS10 antisense control in vivo is affected by mutations throughout the region of complementarity between the interacting RNAs. *J. Mol. Biol.* **246**, 585–594 (1995).
125. Kittle, J.D., Simons, R.W., Lee, J. & Kleckner, N. 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 (1989).
126. Case, C.C., Simons, E.L. & Simons, R.W. The IS10 transposase mRNA is destabilized during antisense RNA control. *EMBO J.* **9**, 1259–1266 (1990).

127. Ma, C. & Simons, R.W. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *EMBO J.* **9**, 1267–1274 (1990).
128. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
129. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
130. Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab.* **58**, 109-130 (2001).
131. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).
132. Mutalik, V.K., Qi, L., Guimaraes, J.C., Lucks, J.B. & Arkin, A.P. Rationally designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* **8**, 447–454 (2012).
133. Woodson, S.A. Compact intermediates in RNA folding. *Annu. Rev. Biophys.* **39**, 61–77 (2010).
134. Roth, A. & Breaker, R.R. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78**, 305–334 (2009).
135. Winkler, W.C. & Breaker, R.R. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* **59**, 487–517 (2005).
136. Jenison, R.D., Gill, S.C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).
137. Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P. & Pardi, A. Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat. Struct. Biol.* **4**, 644–649 (1997).
138. Romaniuk, P.J., Lowary, P., Wu, H.N., Stormo, G. & Uhlenbeck, O.C. RNA binding site of R17 coat protein. *Biochemistry* **26**, 1563–1568 (1987).
139. Gesnel, M.-C., Del Gatto-Konczak, F. & Breathnach, R. Combined use of MS2 and PP7 coat fusions shows that TIA-1 dominates hnRNP A1 for K-SAM exon splicing control. *J. Biomed. Biotechnol.* **2009**, 104853 (2009).
140. Peabody, D.S. & Lim, F. Complementation of RNA binding site mutations in MS2 coat protein heterodimers. *Nucleic Acids Res.* **24**, 2352–2359 (1996).
141. Culler, S.J., Hoff, K.G. & Smolke, C.D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **330**, 1251–1255 (2010).
142. Convery, M.A. *et al.* Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat. Struct. Biol.* **5**, 133–139 (1998).
143. Carothers, J.M., Goler, J.A., Kapoor, Y., Lara, L. & Keasling, J.D. Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic Acids Res.* **38**, 2736–2747 (2010).
144. Spingola, M. & Peabody, D.S. MS2 coat protein mutants which bind Qbeta RNA. *Nucleic Acids Res.* **25**, 2808–2815 (1997).
145. Sinha, J., Reyes, S.J. & Gallivan, J.P. Reprogramming bacteria to seek and destroy an herbicide. *Nat. Chem. Biol.* **6**, 464–470 (2010).
146. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).

147. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.* **23**, 337–343 (2005).
148. Lee, J.F., Hesselberth, J.R., Meyers, L.A. & Ellington, A.D. Aptamer database. *Nucleic Acids Res.* **32**, D95–100 (2004).
149. Hermann, T. & Patel, D.J. Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820–825 (2000).
150. Vicens, Q., Mondragón, E. & Batey, R.T. Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic Acids Res.* **39**, 8586–8598 (2011).
151. Mannironi, C., Di Nardo, A., Fruscoloni, P. & Tocchini-Valentini, G.P. In vitro selection of dopamine RNA ligands. *Biochemistry* **36**, 9726–9734 (1997).
152. Steber, M., Arora, A., Hofmann, J., Brutschy, B. & Suess, B. Mechanistic basis for RNA aptamer-based induction of TetR. *ChemBioChem* **12**, 2608–2614 (2011).
153. Sussman, D., Nix, J.C. & Wilson, C. The structural basis for molecular recognition by the vitamin B 12 RNA aptamer. *Nat. Struct. Biol.* **7**, 53–57 (2000).
154. Watts, J.M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
155. Sivaprakasam, K., Pagán, O.R. & Hess, G.P. Minimal RNA aptamer sequences that can inhibit or alleviate noncompetitive inhibition of the muscle-type nicotinic acetylcholine receptor. *J. Membr. Biol.* **233**, 1–12 (2010).
156. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab. Eng.* **9**, 258–267 (2007).
157. Topp, S. & Gallivan, J.P. Guiding bacteria with small molecules and RNA. *J. Am. Chem. Soc.* **129**, 6807–6811 (2007).
158. Kortmann, J. & Narberhaus, F. Bacterial RNA thermometers: molecular zippers and switches. *Nat. Rev. Microbiol.* **10**, 255–265 (2012).
159. Liu, C.C. & Qi, L. *et al.* A converter from translation to transcription. *Submitted*.
160. Gong, F. & Yanofsky, C. Instruction of translating ribosome by nascent peptide. *Science* **297**, 1864–1867 (2002).
161. Seidelt, B. *et al.* Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412–1415 (2009).
162. Lunde, B.M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
163. Babitzke, P., Baker, C.S. & Romeo, T. Regulation of translation initiation by RNA binding proteins. *Annu. Rev. Microbiol.* **63**, 27–44 (2009).
164. Babitzke, P. Regulation of transcription attenuation and translation initiation by allosteric control of an RNA-binding protein: the *Bacillus subtilis* TRAP protein. *Curr. Opin. Microbiol.* **7**, 132–139 (2004).
165. Lim, F. & Peabody, D.S. RNA recognition site of PP7 coat protein. *Nucleic Acids Res.* **30**, 4138–4144 (2002).
166. Lim, F., Spingola, M. & Peabody, D.S. The RNA-binding site of bacteriophage Qbeta coat protein. *J. Biol. Chem.* **271**, 31839–31845 (1996).
167. Witherell, G.W., Wu, H.N. & Uhlenbeck, O.C. Cooperative binding of R17 coat protein to RNA. *Biochemistry* **29**, 11051–11057 (1990).
168. Romaniuk, P.J., Lowary, P., Wu, H.N., Stormo, G. & Uhlenbeck, O.C. RNA binding site of R17 coat protein. *Biochemistry* **26**, 1563–1568 (1987).

169. Ni, C.Z. *et al.* Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* **3**, 255–263 (1995).
170. Morgan, G.J., Hatfull, G.F., Casjens, S. & Hendrix, R.W. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. *J. Mol. Biol.* **317**, 337–359 (2002).
171. Burrill, D.R. & Silver, P.A. Making cellular memories. *Cell* **140**, 13–18 (2010).
172. Haseltine, E.L. & Arnold, F.H. Synthetic gene circuits: design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 1–19 (2007).
173. Mukherji, S. & van Oudenaarden, A. Synthetic biology: understanding biological design from synthetic circuits. *Nat. Rev. Genet.* **10**, 859–871 (2009).
174. Drubin, D.A., Way, J.C. & Silver, P.A. Designing biological systems. *Genes Dev.* **21**, 242–254 (2007).
175. Gibcus, J.H. & Dekker, J. The context of gene expression regulation. *F1000 Biol. Rep.* **4**, 8 (2012).
176. Makeyev, E.V. & Maniatis, T. Multilevel regulation of gene expression by microRNAs. *Science* **319**, 1789–1790 (2008).
177. Qi, L., Haurwitz, R.E., Shao W., Doudna, J.A., Arkin, A.P. RNA processing enables predictable programming of prokaryotic gene expression. *Submitted*.
178. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
179. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
180. Al-Attar, S., Westra, E.R., van der Oost, J. & Brouns, S.J.J. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392**, 277–289 (2011).
181. Karginov, F.V. & Hannon, G.J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37**, 7–19 (2010).
182. Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M. & Macmillan, A.M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **18**, 688–692 (2011).
183. Lintner, N.G. *et al.* Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286**, 21643–21656 (2011).
184. Brouns, S.J.J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
185. Sashital, D.G., Jinek, M. & Doudna, J.A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **18**, 680–687 (2011).
186. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 (2010).
187. Wilkie, G.S., Dickson, K.S. & Gray, N.K. Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends Biochem. Sci.* **28**, 182–188 (2003).
188. Mignone, F., Gissi, C. & Liuni, S. Untranslated regions of mRNAs. *Genome Biol.* **3**, reviews0004.1-reviews0004.10 (2002).
189. Gama-Castro, S. *et al.* RegulonDB version 7.0: transcriptional regulation of Escherichia



- coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* **39**, D98–105 (2011).
190. Rhodius, V.A. & Mutalik, V.K. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigmaE. *Proc. Natl Acad. Sci. USA* **107**, 2854–2859 (2010).
191. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl Acad. Sci. USA* **102**, 12678–12683 (2005).
192. Vivek, M. *et al.* Composition and quality of irregular transcription and translation genetic elements. *Submitted*
193. Kelly, J.R. *et al.* Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* **3**, 4 (2009).
194. Rocha, E.P.C. The organization of the bacterial genome. *Annu. Rev. Genet.* **42**, 211–233 (2008).
195. Pflieger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24**, 1027–1032 (2006).
196. Lim, H.N., Lee, Y. & Hussein, R. Fundamental relationship between operon organization and gene expression. *Proc. Natl Acad. Sci. USA* **108**, 10626–10631 (2011).
197. Dunn, J.J. & Studier, F.W. T7 early RNAs and *Escherichia coli* ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease 3. *Proc. Natl Acad. Sci. USA* **70**, 3296–3300 (1973).
198. Delebecque, C.J., Lindner, A.B., Silver, P.A. & Aldaye, F.A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).
199. Dominak, L.M., Gundermann, E.L. & Keating, C.D. Microcompartmentation in artificial cells: pH-induced conformational changes alter protein localization. *Langmuir* **26**, 5697–5705 (2010).
200. Retterer, S.T. & Simpson, M.L. Microscale and nanoscale compartments for biotechnology. *Curr. Opin. Biotechnol.* (2012).doi:10.1016/j.copbio.2012.01.002
201. Bonacci, W. *et al.* Modularity of a carbon-fixing protein organelle. *Proc. Natl Acad. Sci. USA* **109**, 478–483 (2012).
202. Daròs, J.A., Marcos, J.F., Hernández, C. & Flores, R. Replication of avocado sunblotch viroid: evidence for a symmetric pathway with two rolling circles and hammerhead ribozyme processing. *Proc. Natl Acad. Sci. USA* **91**, 12813–12817 (1994).
203. Espéli, O., Moulin, L. & Boccard, F. Transcription attenuation associated with bacterial repetitive extragenic BIME elements1. *J. Mol. Biol.* **314**, 375–386 (2001).
204. Mitrophanov, A.Y. & Groisman, E.A. Positive feedback in cellular control systems. *Bioessays* **30**, 542–555 (2008).

## Appendix A Definitions of Concepts Used In The Study

A **ribosome binding site (RBS)** is defined as a sequence on mRNA that is bound by the ribosome during translation initiation. An RBS usually contains the SD sequence as well as nucleotides that bracket the SD. In our opinion, the concept of the RBS is not well defined because it is hard to characterize the exact boundary experimentally.

The **Shine-Dalgarno (SD) Sequence** is defined as the nucleotides in the untranslated region that are complementary to the 16S rRNA in the ribosome, generally located 8 basepairs upstream of the start codon.

**Untranslated Region (UTR)** is referred to either of two sections (or part of them) on each side of a coding sequence on a strand of mRNA.

**5' UTR** is defined as the **whole** section of untranslated region found on the 5' side of the coding sequence.

An **operon** is a set of two or more genes and their associated regulatory elements that are transcribed as a single element.

A **cistron** is a gene, which is usually a segment within an operon that exhibits its own translation.

## Appendix B Plasmids Used In The Study

### *Chapter 3 Plasmid table*

<b>Name</b>	<b>Description</b>	<b>Origin</b>	<b>Resistance</b>
pAPA1509	Blank vector	p15A	Cam
pAPA1260	Blank vector	ColE1	Amp
pAPA1506	P <sub>J23119</sub> – Attenuator(WT,T2) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1256	P <sub>J23119</sub> – Antisense(WT) – T	ColE1	Amp
pAPA1507	P <sub>J23119</sub> – Attenuator(WT,T3) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1272	P <sub>J23119</sub> – Attenuator(WT,T4) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1273	P <sub>J23119</sub> – Attenuator(LS,T4) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1257	P <sub>J23119</sub> – Antisense(LS) – T	ColE1	Amp
pAPA1557	P <sub>J23119</sub> – Attenuator(LS2,T4) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1556	P <sub>J23119</sub> – Antisense(LS2) – T	ColE1	Amp
pAPA1719	P <sub>J23119</sub> – T – P <sub>C</sub> – T	ColE1	Amp
pAPA1721	P <sub>J23119</sub> – Antisense(WT) – T – P <sub>C</sub> – T	ColE1	Amp
pAPA1726	P <sub>J23119</sub> – T – P <sub>J23119</sub> – Antisense(LS) – T	ColE1	Amp
pAPA1423	P <sub>J23119</sub> – Antisense(WT) – T – P <sub>C</sub> – Antisense(LS) – T	ColE1	Amp
pAPA1464	P <sub>J23119</sub> – Attenuator(WT,T4) – Attenuator(WT,T4) – RBS(S) – sfGFP – T	p15A	Cam
pAPA1736	P <sub>J23119</sub> – T	ColE1	Amp
pAPA1737	P <sub>J23119</sub> – Antisense(LS) – T	ColE1	Amp
pAPA1450	P <sub>J23119</sub> – Attenuator(LS,T4) – Antisense(WT) – T	ColE1	Amp
pAPA1448	P <sub>J23119</sub> – Attenuator(LS,T4) – Insulator(sTRSV) - Antisense(WT) – T	ColE1	Amp
pAPA1347	P <sub>J23119</sub> – Attenuator(LS,T4) – (Insulator(sTRSV) - Antisense(WT),2) – T	ColE1	Amp
pAPA1728	P <sub>J23119</sub> – Attenuator(LS,T4) – Insulator(sTRSVmut1) - Antisense(WT) – T	ColE1	Amp
pAPA1741	P <sub>J23119</sub> – Attenuator(LS,T4) – Insulator(sTRSV) – (Antisense(WT),2) – T	ColE1	Amp
pAPA1742	P <sub>J23119</sub> – Attenuator(LS,T4) – Antisense(WT) - Insulator(sTRSV) - Antisense(WT) – T	ColE1	Amp
pAPA1743	P <sub>J23119</sub> – Attenuator(LS,T4) – (Antisense(WT),2) – T	ColE1	Amp
pAPA1744	P <sub>J23119</sub> – Attenuator(LS,T4) – Antisense(WT) – T	ColE1	Amp
pAPA1745	P <sub>J23119</sub> – Attenuator(LS,T4) – Insulator(sTRSV) - Antisense(WT) – T	ColE1	Amp
pAPA1749	P <sub>J23119</sub> – Attenuator(LS,T4) – Antisense(WT) – T	ColE1	Amp
pAPA1739	P <sub>J23119</sub> – T – P <sub>J23119</sub> – T	ColE1	Amp
pAPA1740	P <sub>J23119</sub> – Antisense(WT) – T – P <sub>J23119</sub> – T	ColE1	Amp
pAPA1445	P <sub>J23119</sub> – Attenuator(LS,T4) – (Insulator(sTRSV) – Antisense(WT),2) – T - P <sub>J23119</sub> – T	ColE1	Amp
pAPA1438	P <sub>J23119</sub> – Attenuator(LS,T4) – (Insulator(sTRSV) – Antisense(WT),2) – T - P <sub>J23119</sub> – Antisense(LS) – T	ColE1	Amp
pAPA1701	P <sub>J23119</sub> – Attenuator(LS,T4) – (Insulator(sTRSV) – Antisense(WT),2) – T - P <sub>J23119</sub> – Insulator(sTRSV) - Antisense(LS) – T	ColE1	Amp
pAPA1702	P <sub>J23119</sub> – Attenuator(LS,T4) – (Insulator(sTRSV) –	ColE1	Amp

	Antisense(WT),2) – T - P <sub>J23119</sub> – Antisense(LS) - Insulator(sTRSV) - Antisense(LS) – T		
pAPA1703	P <sub>J23119</sub> – Attenuator(LS,T4) – (Insulator(sTRSV) – Antisense(WT),2) – T - P <sub>J23119</sub> – (Insulator(sTRSV) - Antisense(LS),2) – T	ColE1	Amp
pAPA1469	P <sub>J23119</sub> – Attenuator(LS,T4) – Attenuator(LS,T4) – RBS(S) – sfGFP – T	p15A	Cam
pAPA1465	P <sub>J23119</sub> – Attenuator(WT,T4) – Attenuator(LS,T4) – RBS(S) – sfGFP – T	p15A	Cam
pAPA1472	P <sub>l</sub> lacO-1 – T	ColE1	Amp
pAPA1473	P <sub>l</sub> lacO-1 – Antisense(WT) – T	ColE1	Amp
pAPA1217	P <sub>l</sub> lacO-1 – Antisense(LS) – T	ColE1	Amp
pAPA1280	P <sub>l</sub> lacO-1 – Attenuator(LS,T4) – RBS <sub>medium</sub> – sfGFP – T	p15A	Cam
pAPA1282	P <sub>J23119</sub> – Attenuator(LS,T4) – RBS <sub>weak</sub> – sfGFP – T	p15A	Cam
pAPA1291	P <sub>J23119</sub> – Attenuator(LS,T4) – RBS <sub>medium</sub> – mRFP – T	p15A	Cam
pAPA1502	P <sub>J23119</sub> – Attenuator(L,T2) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1501	P <sub>J23119</sub> – Antisense(L) – T	ColE1	Amp
pAPA1504	P <sub>J23119</sub> – Attenuator(S,T2) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1503	P <sub>J23119</sub> – Antisense(S) – T	ColE1	Amp
pAPA1508	P <sub>J23119</sub> – Attenuator(LS,T2) – RBS <sub>str</sub> – sfGFP – T	p15A	Cam
pAPA1729	P <sub>J23119</sub> – Insulator(sTRSV) – Antisense(WT) – T	ColE1	Amp
pAPA1460	P <sub>J23119</sub> – Insulator(sTRSV) – Antisense(LS) – T	ColE1	Amp

#### Chapter 4 Plasmid table

Plasmid	Decription	Origin	Resistance
pVKM84	A vector that has been used for the construction of sense plasmid library. This vector backbone has PJ23119 promoter-RBS region-sfGFP reporter gene replacing the LacI repressor gene, PLLaco1 promoter and <i>mRFP1</i> gene from the pBbS6c vector. RBS region is replaced with RNA-IN region in the sense library translationally fused to the second codon of sfGFP reporter.	pSC101	Cam
pVKM85	A vector that has been used for the construction of sense plasmid library. This vector backbone has PJ23119 promoter-RBS region-mRFP1 reporter gene replacing the LacI repressor gene and PLLaco1 promoter from the pBbS6c vector. RBS region is replaced with RNA-IN region in the sense library translationally fused to the second codon of mRFP1 reporter.	pSC101	Cam
pVKM87	A vector that has been used for the construction of antisense plasmid library. The variants of antisense RNA-OUT region were cloned between promoter PLLaco1 and terminator in the antisense plasmids	ColE1	Amp
pVKM86	P <sub>J23119</sub> –RNA-IN mutant S1-sfGFP	pSC101	Cam
pVKM113	P <sub>J23119</sub> –RNA-IN mutant S3-sfGFP	pSC101	Cam
pVKM114	P <sub>J23119</sub> –RNA-IN mutant S4-sfGFP	pSC101	Cam
pVKM115	P <sub>J23119</sub> –RNA-IN mutant S5-sfGFP	pSC101	Cam
pVKM116	P <sub>J23119</sub> –RNA-IN mutant S6-sfGFP	pSC101	Cam
pVKM118	P <sub>J23119</sub> –RNA-IN mutant S7-sfGFP	pSC101	Cam
pVKM117	P <sub>J23119</sub> –RNA-IN mutant S8-sfGFP	pSC101	Cam
pVKM129	P <sub>J23119</sub> –RNA-IN mutant S17-sfGFP	pSC101	Cam

pVKM132	P <sub>J23119</sub> –RNA-IN mutant S19-sfGFP	pSC101	Cam
pVKM131	P <sub>J23119</sub> –RNA-IN mutant S21-sfGFP	pSC101	Cam
pVKM130	P <sub>J23119</sub> –RNA-IN mutant S22-sfGFP	pSC101	Cam
pVKM134	P <sub>J23119</sub> –RNA-IN mutant S23-sfGFP	pSC101	Cam
pVKM133	P <sub>J23119</sub> –RNA-IN mutant S26-sfGFP	pSC101	Cam
pVKM125	P <sub>J23119</sub> –RNA-IN mutant S27-sfGFP	pSC101	Cam
pVKM126	P <sub>J23119</sub> –RNA-IN mutant S28-sfGFP	pSC101	Cam
pVKM127	P <sub>J23119</sub> –RNA-IN mutant S29-sfGFP	pSC101	Cam
pVKM128	P <sub>J23119</sub> –RNA-IN mutant S30-sfGFP	pSC101	Cam
pVKM124	P <sub>J23119</sub> –RNA-IN mutant S31-sfGFP	pSC101	Cam
pVKM123	P <sub>J23119</sub> –RNA-IN mutant S32-sfGFP	pSC101	Cam
pVKM135	P <sub>J23119</sub> –RNA-IN mutant S34-sfGFP	pSC101	Cam
pVKM136	P <sub>J23119</sub> –RNA-IN mutant S43-sfGFP	pSC101	Cam
pVKM138	P <sub>J23119</sub> –RNA-IN mutant S49-sfGFP	pSC101	Cam
pVKM137	P <sub>J23119</sub> –RNA-IN mutant S52-sfGFP	pSC101	Cam
pVKM182	P <sub>J23119</sub> –RNA-IN mutant S13-sfGFP	pSC101	Cam
pVKM184	P <sub>J23119</sub> –RNA-IN mutant S40-sfGFP	pSC101	Cam
pVKM22	Blank vector	ColE1	Amp
pVKM40	P <sub>LlacO-1</sub> –RNA-OUT mutant A1	ColE1	Amp
pVKM100	P <sub>LlacO-1</sub> –RNA-OUT mutant A3	ColE1	Amp
pVKM101	P <sub>LlacO-1</sub> –RNA-OUT mutant A4	ColE1	Amp
pVKM102	P <sub>LlacO-1</sub> –RNA-OUT mutant A5	ColE1	Amp
pVKM103	P <sub>LlacO-1</sub> –RNA-OUT mutant A6	ColE1	Amp
pVKM105	P <sub>LlacO-1</sub> –RNA-OUT mutant A7	ColE1	Amp
pVKM104	P <sub>LlacO-1</sub> –RNA-OUT mutant A8	ColE1	Amp
pVKM147	P <sub>LlacO-1</sub> –RNA-OUT mutant A17	ColE1	Amp
pVKM150	P <sub>LlacO-1</sub> –RNA-OUT mutant A19	ColE1	Amp
pVKM149	P <sub>LlacO-1</sub> –RNA-OUT mutant A21	ColE1	Amp
pVKM148	P <sub>LlacO-1</sub> –RNA-OUT mutant A22	ColE1	Amp
pVKM152	P <sub>LlacO-1</sub> –RNA-OUT mutant A23	ColE1	Amp
pVKM151	P <sub>LlacO-1</sub> –RNA-OUT mutant A26	ColE1	Amp
pVKM143	P <sub>LlacO-1</sub> –RNA-OUT mutant A27	ColE1	Amp
pVKM144	P <sub>LlacO-1</sub> –RNA-OUT mutant A28	ColE1	Amp
pVKM145	P <sub>LlacO-1</sub> –RNA-OUT mutant A29	ColE1	Amp
pVKM146	P <sub>LlacO-1</sub> –RNA-OUT mutant A30	ColE1	Amp
pVKM142	P <sub>LlacO-1</sub> –RNA-OUT mutant A31	ColE1	Amp
pVKM141	P <sub>LlacO-1</sub> –RNA-OUT mutant A32	ColE1	Amp
pVKM153	P <sub>LlacO-1</sub> –RNA-OUT mutant A34	ColE1	Amp
pVKM154	P <sub>LlacO-1</sub> –RNA-OUT mutant A43	ColE1	Amp
pVKM156	P <sub>LlacO-1</sub> –RNA-OUT mutant A49	ColE1	Amp
pVKM155	P <sub>LlacO-1</sub> –RNA-OUT mutant A52	ColE1	Amp
pVKM183	P <sub>LlacO-1</sub> –RNA-OUT mutant A13	ColE1	Amp
pVKM185	P <sub>LlacO-1</sub> –RNA-OUT mutant A40	ColE1	Amp
pSLQ221	Two RNA-IN mutant constructs S34-SFGFP, S5-mRFP1	pSC101	Cam
pSLQ222	Two RNA-IN mutant constructs S34-SFGFP, S31-mRFP1	pSC101	Cam
pSLQ223	Two RNA-IN mutant constructs S34-SFGFP, S49-mRFP1	pSC101	Cam
pSLQ224	Two RNA-IN mutant constructs S34-SFGFP, S4-mRFP1	pSC101	Cam
pSLQ225	Two RNA-IN mutant constructs S5-SFGFP, S31-mRFP1	pSC101	Cam
pSLQ226	Two RNA-IN mutant constructs S5-SFGFP, S49-mRFP1	pSC101	Cam

pSLQ227	Two RNA-IN mutant constructs S4-SFGFP, S5 -mRFP1	pSC101	Cam
pSLQ228	Two RNA-IN mutant constructs S4-SFGFP, S31-mRFP1	pSC101	Cam
pSLQ229	Two RNA-IN mutant constructs S4-SFGFP, S49-mRFP1	pSC101	Cam
pSLQ230	Two RNA-IN mutant constructs S31-SFGFP, S49-mRFP1	pSC101	Cam
pSLQ231	Two RNA-OUT mutant constructs A34 and null	ColE1	Amp
pSLQ232	Two RNA-OUT mutant constructs A34 and A5	ColE1	Amp
pSLQ233	Two RNA-OUT mutant constructs A34 and A31	ColE1	Amp
pSLQ234	Two RNA-OUT mutant constructs A34 and A49	ColE1	Amp
pSLQ235	Two RNA-OUT mutant constructs A5 and null	ColE1	Amp
pSLQ236	Two RNA-OUT mutant constructs A5 and A31	ColE1	Amp
pSLQ237	Two RNA-OUT mutant constructs A5 and A49	ColE1	Amp
pSLQ238	Two RNA-OUT mutant constructs A31 and null	ColE1	Amp
pSLQ239	Two RNA-OUT mutant constructs A31 and A49	ColE1	Amp
pSLQ240	Two RNA-OUT mutant constructs null and A49	ColE1	Amp
pSLQ241	Two RNA-OUT mutant constructs null and null	ColE1	Amp
pSLQ242	Two RNA-OUT mutant constructs A4 and null	ColE1	Amp
pSLQ243	Two RNA-OUT mutant constructs A4 and A5	ColE1	Amp
pSLQ244	Two RNA-OUT mutant constructs A4 and A31	ColE1	Amp
pSLQ245	Two RNA-OUT mutant constructs A4 and A49	ColE1	Amp
pSLQ254	Two RNA-OUT mutant constructs A34 and A4	ColE1	Amp

**Chapter 5 Plasmid table**

<b>Plasmid #</b>	<b>Description</b>	<b>Origin</b>	<b>Resistance</b>
pAPA1272	WT pT181 reporter plasmid	p15A	Cam
pAPA1301	WT IS10 reporter plasmid	pSC101	Cam
pAPA1273	MT pT181 reporter plasmid	p15A	Cam
pAPA1509	GFP knockout plasmid (no fluorescent protein expression)	p15A	Cam
pAPA1307	Theo-P-IS10 ncRNA fusion plasmid	ColE1	Amp
pAPA1310	WT IS10 ncRNA knockout plasmid (positive control plasmid for IS10)	ColE1	Amp
pAPA1309	WT IS10 ncRNA expression plasmid (negative control plasmid for IS10)	ColE1	Amp
pAPA1305	Theo-P-pT181ncRNA fusion plasmid	ColE1	Amp
pAPA1260	WT pT181 ncRNA knockout plasmid (positive control plasmid for pT181)	ColE1	Amp
pAPA1308	WT pT181 ncRNA expression plasmid (negative control plasmid for pT181)	ColE1	Amp
pAPA1304	Theo-SE-pT181WT ncRNA fusion plasmid	ColE1	Amp
pAPA1306	Theo-SE-pT181MT ncRNA fusion plasmid	ColE1	Amp
pAPA1302	MS2-SE-pT181WT ncRNA fusion plasmid which contains MS2 coat protein cassette	ColE1	Amp
pAPA1303	MS2-SE-pT181WT ncRNA fusion plasmid which contains MS2-RFP protein fusion cassette	ColE1	Amp
pAPA1315	MS2-SE-pT181MT fusion plasmid which contains MS2-RFP protein fusion cassette	ColE1	Amp
pAPA1311	Theo-SE-pT181 MT-1 fusion plasmid	ColE1	Amp
pAPA1312	Theo-SE-pT181 MT-3 fusion plasmid	ColE1	Amp

pAPA1313	Theo-SE-pT181 MT-2 fusion plasmid	ColE1	Amp
pAPA1314	Theo-SE-pT181WT and MS2-SE-pT181WT expression plasmid	ColE1	Amp

### Chapter 6 Plasmid table

Plasmid #	Description	Origin	Resistance
pSLQ439	P <sub>J23119</sub> -RNA IN(3)-tna- sfGFP -Term	pSC101	Amp
pSLQ440	P <sub>J23119</sub> -RNA IN(4)-tna- sfGFP -Term	pSC101	Amp
pSLQ441	P <sub>J23119</sub> -RNA IN(9)-tna- sfGFP -Term	pSC101	Amp
pSLQ442	P <sub>J23119</sub> -RNA IN(20)-tna- sfGFP -Term	pSC101	Amp
pSLQ443	P <sub>J23119</sub> -RNA IN(23)-tna-sfGFP-Term	pSC101	Amp
pSLQ444	P <sub>J23119</sub> -RNA OUT(3)-Term	ColE1	Kan
pSLQ445	P <sub>J23119</sub> -RNA OUT(4)-Term	ColE1	Kan
pSLQ446	P <sub>J23119</sub> -RNA OUT(9)-Term	ColE1	Kan
pSLQ447	P <sub>J23119</sub> -RNA OUT(20)-Term	ColE1	Kan
pSLQ448	P <sub>J23119</sub> -RNA OUT(23)-Term	ColE1	Kan
pSLQ449	P <sub>L</sub> lacO-1-PP7-Term- P <sub>L</sub> tetO-1-RBS <sub>PP7</sub> -sfGFP-Term	p15A	Cam
pSLQ450	P <sub>L</sub> lacO-1-MS2-Term- P <sub>L</sub> tetO-1-RBS <sub>MS2</sub> -sfGFP -Term	p15A	Cam
pSLQ451	P <sub>L</sub> lacO-1-Com-Term- P <sub>L</sub> tetO-1-RBS <sub>Com</sub> -sfGFP -Term	p15A	Cam
pSLQ452	P <sub>L</sub> lacO-1-Term- P <sub>L</sub> tetO-1-RBS <sub>PP7</sub> -sfGFP -Term	p15A	Cam
pSLQ453	P <sub>L</sub> lacO-1-Term- P <sub>L</sub> tetO-1-RBS <sub>MS2</sub> -sfGFP -Term	p15A	Cam
pSLQ454	P <sub>L</sub> lacO-1-Term- P <sub>L</sub> tetO-1-RBS <sub>Com</sub> -sfGFP -Term	p15A	Cam
pVKM72	P <sub>L</sub> lacO-1-PP7-Term	ColE1	Amp
pVKM112	P <sub>L</sub> tetO-1-RBS <sub>pp7</sub> -sfGFP -Term	p15A	Cam
pVKM74	P <sub>L</sub> lacO-1-MS2-Term	ColE1	Amp
pVKM108	P <sub>L</sub> tetO-1-RBS <sub>MS2</sub> -sfGFP -Term	p15A	Cam
pVKM54	P <sub>L</sub> lacO-1-Com-Term	ColE1	Amp
pVKM107	P <sub>L</sub> tetO-1-RBS <sub>Com</sub> -sfGFP -Term	p15A	Cam
pVKM73	P <sub>L</sub> lacO-1-Qβ-Term	ColE1	Amp
pVKM111	P <sub>L</sub> tetO-1-RBS <sub>Qβ</sub> -sfGFP -Term	p15A	Cam
pSLQ560	P <sub>L</sub> tetO-1-RBS <sub>PP7</sub> -PP7::mRFP fusion-Term	p15A	Cam
pSLQ567	P <sub>L</sub> tetO-1-RBS <sub>PP7</sub> -PP7(R54M)::mRFP fusion-Term	p15A	Cam
pSLQ607	P <sub>L</sub> tetO-1-RBS <sub>COM</sub> -COM::sfGFP-Term	p15A	Cam
pSLQ548	P <sub>L</sub> lacO-1-COM(-26U)-Term	ColE1	Amp

### Chapter 7 Plasmid table

Plasmid #	Description	Origin	Resistance
pSLQ679	P <sub>L</sub> tetO-1-Csy4-Term	ColE1	Amp
pSLQ636	J23119(BglII)-Repeat site-RNA-IRNA N (9)-mRFP-Repeat site-RNA IN(wt)-Repeat site-sfGFP-Term	pSC101	Cam
pSLQ676	P <sub>L</sub> tetO-1-Term	ColE1	Amp
pSLQ677	P <sub>L</sub> tetO-1-Csy4-Term- P <sub>J23119</sub> -IN(9)	ColE1	Amp
pSLQ678	P <sub>L</sub> tetO-1-Csy4-Term-P <sub>J23119</sub> -IN(WT)-Term- P <sub>J23119</sub> -IN(9)-Term	ColE1	Amp
pSLQ505	P <sub>L</sub> tetO-1-RBS <sub>str</sub> -Csy4-RNA OUT(WT)	pSC101	Cam
pSLQ515	P <sub>J23119</sub> -pT181(Sense)-Repeat-RNA IN(WT)-Term	p15A	Cam

## Appendix C Sequences For Important Genetic Elements

### Chapter 3 - genetic elements

WT	AACAAAATAAAAAGGAGTCGCTCACGCCCTGACCAAAGTTTGTGAACGACATCATTCAAA	60
L	.....CG.....	60
S	.....TGT.....GCA.....	60
LS	.....TGT.....CG.....GCA.....	60
LS2	.....GTA.....CTG.....AAC.....	60
T3	.....	60
T4	.....	60
WT	GAAAAAACACTGAGTTGTTTTTATAATCTTGTATATTTAGATATTAACGATATTTAAA	120
L	.....	120
S	.....	120
LS	.....	120
LS2	.....	120
T3	.....	120
T4	.....	120
WT	TATACATAAAGATATATATTTGGGTGAGCGATTTCCTTAAACGAAATTGAGATTAAGGAGT	180
L	.....	180
S	.....	180
LS	.....	180
LS2	.....	180
T3	.....	180
T4	.....	180
WT	CGATTTTTTT--	189
L	.....--	189
S	.....--	189
LS	.....--	189
LS2	.....--	189
T3	..C.....--	189
T4	..C.C....TT	191

### Chapter 4 - genetic elements

#	Sequence
S1	<b>GCG</b> AAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S2	CCGAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S3	GGGAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S4	GCCAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S5	GCGTAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S6	GCGATAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S7	CGGAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S8	CCCAAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S9	CCGTAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S10	CCGATAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S11	GGCAAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S12	GGGTAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S13	GGGATAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S14	GCCTAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S15	GCCATAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S16	GCGTTAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S17	CGCAAAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat



S18	CGGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S19	CGGATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S20	CCCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S21	CCCATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S22	CCGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S23	GGCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S24	GGCATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S25	GGGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S26	GCCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S27	CGCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S28	CGCATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S29	CGGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S30	CCCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S31	GGCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S32	CGCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S33	CCGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S34	GGGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S35	GCCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S36	CGGTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S37	CCCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S38	GGCTAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S39	CCGATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S40	GGGATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S41	GCCATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S42	CGGATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S43	CCCATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S44	GGCATAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S45	CCGCGAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S46	GGGCGAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S47	GCCCCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S48	CGGCGAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S49	CCCCGAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S50	GGCCGAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S51	CCGGCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S52	GGGGCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S53	GCCGCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S54	CGGGCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S55	CCCGCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
S56	GGCGCAAAATCAATAAGGAGACAACAAGATGTGCGAAActcgat
A1	TCGCACATCTTGTTGTCTGATTATTGATTTT <b>TCGCG</b> GAAACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A2	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TCGGG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A3	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TCCGG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A4	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TGGCG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A5	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TACGCG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A6	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TATCGCG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A7	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TCCGG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A8	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TGGGG</b> AACCATTTGATCATATGACAAGATGTGTATCC ACCTTAACTTAATGATTTTTTACC AAAATCATTAGGGGATTCATCAG
A9	TCGCACATCTTGTTGTCTGATTATTGATTTTT <b>TACGGG</b> AACCATTTGATCATATGACAAGATGTGTATCC



A38	TCGCACATCTTGTTGTCTGATTATTGATTTTTTAGCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A39	TCGCACATCTTGTTGTCTGATTATTGATTTTTATCGGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A40	TCGCACATCTTGTTGTCTGATTATTGATTTTTATCCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A41	TCGCACATCTTGTTGTCTGATTATTGATTTTTATGGCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A42	TCGCACATCTTGTTGTCTGATTATTGATTTTTATCCGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A43	TCGCACATCTTGTTGTCTGATTATTGATTTTTATGGGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A44	TCGCACATCTTGTTGTCTGATTATTGATTTTTATGCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A45	TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGCGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A46	TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A47	TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGGGCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A48	TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A49	TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGGGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A50	TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGGCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A51	TCGCACATCTTGTTGTCTGATTATTGATTTTTTGCCGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A52	TCGCACATCTTGTTGTCTGATTATTGATTTTTTGCCCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A53	TCGCACATCTTGTTGTCTGATTATTGATTTTTTGCGGCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A54	TCGCACATCTTGTTGTCTGATTATTGATTTTTTGCCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A55	TCGCACATCTTGTTGTCTGATTATTGATTTTTTGCGGGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG
A56	TCGCACATCTTGTTGTCTGATTATTGATTTTTTGCGCCGAAACCATTTGATCATATGACAAGATGTGTAT CCACCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG

### *Chapter 5 - genetic elements*

DNA sequences of IS10 and pT181 systems

(A) Sequence of 115bp WT IS10 ncRNA

TCGCACATCTTGTTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAAGATGTGTATCCA  
CCTTAACTTAATGATTTTTTACCAAAATCATTAGGGGATTCATCAG

(B) Sequence of 40bp WT IS10 ncRNA

GCGAAAAATCAATAAGGAGACAACAAGATGTGCGAA

(C) Sequence of 58bp WT pT181 ncRNA

TCTTTGAATGATGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTTATTT

(D) Sequence of 287bp WT pT181 ncRNA target

AACAAAATAAAAAGGAGTCGCTCACGCCCTGACCAAAGTTTGTGAACGACATCATTCAAAGAAAAAACA  
CTGAGTTGTTTTTATAATCTTGTATATTTAGATATTAACGATATTTAAATATACATAAAGATATATATT  
TGGGTGAGCGATTCCTTAAACGAAATTGAGATTAAGGAGTCGCTCTTTTTTATGTATAAAAACAATCATG  
CAAATCATTCAAATCATTTGGAAAATCACGATTTAGACAATTTTTCTAAAACCGGCTACTCTAATAGCCG  
GTTGTAA

(E) Sequence of 58bp MT pT181 ncRNA

TCTTTGAATGATGTCGTTCAAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT

(F) Sequence of 287bp MT pT181 ncRNA target

AACAAAATAAAAAGGAGTCGCTCACGCCCTGACCAAAGTTTGTGAACGACATCATTCAAAGAAAAAACA  
CTGAGTTGTTTTTATAATCTTGTATATTTAGATATTAACGATATTTAAATATACATAAAGATATATATT  
TGGGTGAGCGATTCCTTAAACGAAATTGAGATTAAGGAGTCGCTCTTTTTTATGTATAAAAACAATCATG  
CAAATCATTCAAATCATTTGGAAAATCACGATTTAGACAATTTTTCTAAAACCGGCTACTCTAATAGCCG  
GTTGTAA

DNA sequence of aptamer-ncRNA fusion molecules

(A) Sequence of theo-P-IS10 ncRNA fusion

GGTGATACCAGATTTTCGCGAAAAATCCCTTGGCAGCACCTCGCACATCTTGTTGTCTGATTATTGATTTT  
TCGCGAAACCATTTGATCATATGACAAGATTGAG

(B) Sequence of theo-P-pT181 ncRNA fusion

GGTGATACCAGCTGACCAAAGGCCCTTGGCAGCACCTCTTTGAATGATGTCGTTCAAAACTTTGGTCA  
GGGCGTGAGCGACTCCTTTTTATTT

(C) Sequence of theo-SE-pT181WT ncRNA fusion

GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATGGTGCTGCCACAAACTTTGGTC  
AGGGCGTGGGCGACTCCTTTTTATTT

(D) Sequence of theo-SE-pT181MT ncRNA fusion

GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATGGTGCTGCCCTGCAACTTTGGCG  
AGGGACAGGGCGACTCCTTTTTATTT

(E) Sequence of MS2-SE-pT181WT ncRNA fusion

AAACATGAGGACCACCCATGTTCTTTGAATGGTGTGGTCCACAAACTTTGGTCAGGGCGTGAGCCACTCC  
TTTTTATTT

(F) Sequence of MS2-SE-pT181MT ncRNA fusion

AAACATGAGGACCACCCATGTTCTTTGAATGGTGTGGTCCCTGCAACTTTGGCGAGGGACAGAGCCACTCC  
TTTTTATTT

(G) Sequence of theo-SE-pT181 MT-1

GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATGGTGCTGCCACAAACTTTGGTC  
AGGGCGTGGGCGACTCCTTTTTATTT

(H) Sequence of theo-SE-pT181 MT-2

GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATGGTGCTGCCACAAACTTTGGTC  
AGGGCGTGAGCGGCTCCTTTTTATTT

(I) Sequence of theo-SE-pT181 MT-3

GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATGGTGCTGCTCACAAACTTTGGTC  
AGGGCGTGAGCGGCTCCTTTTTATTT

Mutants of theophylline-sensing IS10 ncRNA fusions based on the pseudoknot design

Theo-IS10 MUTANT #	SEQUENCE
1	GGTGATACCAGGTTTTCGCGAAAAATCCCTTGGCAGCACCTCGCACATCTG TTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAAGATGT GTATCC
2	GGTGATACCAGGTTTTCGCGAAAAATCCCTTGGCAGCACCTCGCACATCT TGTTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAAGAT GTGTATCC
3	GGTGATACCAGGTTTTCGCGAAAAATGCCCTTGGCAGCACCTCGCACATCT TGTTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAAGAT GTGTATCC
4	GGTGATACCAGGTTTTCGCGAAAAATGCCCTTGGCAGCACCTCGCACATCT TGTTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAAG
5	GGTGATACCAGTTTTTCGCGAAAAACCTTGGCAGCACCTCGCACATCTTGT GTCTGATTATTGATTTTTTCGCGAAACCATTTGATCAGACAACAA
6	GGTGATACCAGATTTTCGCGAAAAATCCCTTGGCAGCACCTCGCACATCTG TTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAAGATTG AG
7	GGTGATACCAGATTTTCGCGAAAAATCCCTTGGCAGCACCTCGCACATCTG TTGTCTGATTATTGATTTTTTCGCGAAAAATCCCTTGGCAGCACCTCGCACA TCTTGTGTCTGATTATTGATTTTTTCGCGAAACCATTTGATCATATGACAA GAT

Mutants of theophylline-sensing pT181 ncRNA fusions based on the pseudoknot design

Theo-P-pT181 MUTANT #	SEQUENCE
1	GGTGATACCAGCCCTGACCAAAGTCCCTTGGCAGCACCTCTTTGAATGA TGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
2	GGTGATACCAGCCCTGACCAAAGTCCCTTGGCAGCACCTCTTTGAATGA TGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
3	GGTGATACCAGCCCTGACCAAAGGGCCCTTGGCAGCACCTCTTTGAATGA TGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
4	GGTGATACCAGCCTGACCAAAGGCCCTTGGCAGCACCTCTTTGAATGATG TCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
5	GGTGATACCAGCCTGACCACAGGCCCTTGGCAGCACCTCTTTGAATGATG TCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
6	GGTGATACCAGCCTTACCAAAGGCCCTTGGCAGCACCTCTTTGAATGATG TCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
7	GGTGATACCAGGCCCTGACCAAAGTGTCCCTTGGCAGCACCTCTTTGAAT

	GATGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
8	GGTGATACCAGGCCCTGACCACAGTTTCCCTTGGCAGCACCTCTTTGAAT GATGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT

Mutants of theophylline-sensing pT181 ncRNA fusions based on the strand exchange design

Theo-SE-pT181 MUTANT #	SEQUENCE
1	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG ATGCTGCTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
2	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGTTGCTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
3	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCCGCTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
4	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCTGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
5	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGTCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
6	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG ATGCCGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
7	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG ATGTTGTTACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
8	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG ATGTCGCTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
9	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGTCGCTCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
10	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGTCGCCCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
11	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCTGCCACAAACTTTGGTCAGGGCGTGAGCGACTCCTTTTTATTT
12	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCTGCTCACAAACTTTGGTCAGGGCGTGAGCGGCTCCTTTTTATTT
13	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCTGCTCACAAACTTTGGTCAGGGCGTGAGCGACACCTTTTTATTT
14	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCTGCCACAAACTTTGGTCAGGGCGTGGGCGACTCCTTTTTATTT
15	GGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACCTCTTTGAATG GTGCTGCCACAAACTTTGGTCAGGGCGTGAGCGGCTCCTTTTTATTT

Mutants of MS2 coat protein-sensing pT181 ncRNA fusions based on the strand exchange design

MS2-SE-pT181 MUTANT #	SEQUENCE
1	AAACATGAGGACCACCCATGTTCTTTGAATGGGGTGGTCCACAAACTT TGGTCAGGGCGTGAGCCACTCCTTTTTATTT
2	AAACATGAGGACCACCCATGTTCTTTGAATGGTGTGGTCCACAAACTT TGGTCAGGGCGTGAGCCACTCCTTTTTATTT
3	AAACATGAGGACCACCCATGTTCTTTGAATGGTGTGCTCCACAAACTT TGGTCAGGGCGTGAGCGACTCCTTTTTATTT
4	AAACATGAGGACCACCCATGTTCTTTGAATGGGGTGGTCCACAAACTT

	TGGTCAGGGCGTGAGCGACTCCTTTTTATTT
5	AAACATGAGGACGACCCATGTTCTTTGAATGGTGTGTCGCCACAACTT TGGTCAGGGCGTGAGCGACTCCTTTTTATTT

## Chapter 6 - genetic elements

### The tna element

CTCGATATATGTGTGACCTCAAAATGGTTCAATATTGACAACAAAATTGTCGATCACCGCCCTTGATTTG  
CCCTTCTGTAGCCATCACCAGAGCCAAACCGATTAGATTCAATGTGATCTATTTGTTTGCATATATCTTAA  
TTTTGCCTTTTGC AAAGGTCATCTCTCGTTTATTTACTTGTTTTAGTAAATGATGGTGCCTTGCATATATA  
TCTGGCGAATTAATCGGTATAGCAGATGTAATATTCACAGGGATCACTGTAATTAATAAATAAATGAAGGAT  
TATGTAATGGAAAACTTTAAACATCTCCCTGAACCG

### PP7 binding site

TAAGGAGTTTATATGGAAACCCTTAGTAGCGCT

### MS2 binding site

AAACATGAGGATTACCCATGTGCAAGACAACA

### Com binding site

TGCTGAATGCCTGCGAGCATCCCACGGAGAAACATGTGGGAAAAGAGAAAAATCACGCATTCTG

### Q $\beta$ binding site

TAAGGATTAATTGCATGTCTAAGACAGCATCT

### PP7 coding sequence

ATGTCCAAAACCATCGTTCTTTCGGTCGGCGAGGCTACTCGCACTCTGACTGAGATCCAGTCCACCGCAG  
ACCGTCAGATCTTGAAGAGAAGGTCGGGCTCTGGTGGGTGCGCTGCGCCTCACGGCTTCGCTCCGTCA  
AAACGGAGCCAAGACCGCGTATCGCGTCAACCTAAAACCTGGATCAGGCGGACGTCGTTGATTCCGGACTT  
CCGAAAGTGCCTACACTCAGGTATGGTGCACGACGTGACAATCGTTGCGAATAGCACCGAGGCCTCGC  
GCAAATCGTTGTACGATTTGACCAAGTCCCTCGTCGCGACCTCGCAGGTGCAAGATCTTGTCTGCAACCT  
TGTGCCGCTGGGCCGTTGATAA

### MS2 coding sequence

ATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTGCCCCAA  
GCAACTTCGCTAACGGGGTTCGCTGAATGGATCAGTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTG  
TAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACC  
CAGACTGTTGGTGGTGTAGAGCTTCCCTGTAGCCGCATGGCGTTCGTAATAAATATGGAACCTAACCATTC  
CAATTTTCGCCACGAATTCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAA  
CCCGATTCCCTCGGCAATCCCAGCAAACCTCCGGCATCTACTAA

### Com coding sequence

ATGAAATCAATTTCGCTGTAAAACTGCAACAAACTGTTATTTAAGGCGGATAGTTTTGATCACATTGAAA  
TCAGGTGTCCGCGTTGCAAACGTCACATCATAATGCTGAATGCCCTGCGAGCATCCCACGGAGAAACATTG  
TGGGAAAAGAGAAAAAATCACGCATTCTGACGAAACCGTTCGCTTATTG

### Q $\beta$ coding sequence

ATGGCAAATTAGAGACTGTTACTTTAGGTAACATCGGGAAAGATGGAAAACAACTCTGGTCTCAATC  
CGCGTGGGGTAAATCCCCTAACGGCGTTGCCTCGCTTTCACAAGCGGGTGCAGTTCCTGCGCTGGAGAA  
GCGTGTACCCTTTCGGTATCTCAGCCTTCTCGCAATCGTAAGAACTACAAGGTCCAGGTTAAGATCCAG  
AACCCGACCGCTTGCCTGCAACGGTCTTGTGACCCATCCGTTACTCGCCAGGCATATGCTGACGTGA

CCTTTTCGTTACGCAGTATAGTACCGATGAGGAACGAGCTTTTGTTCGTACAGAGCTTGCTGCTCTGCT  
CGCTAGTCCTCTGCTGATCGATGCTATTGATCAGCTGAACCCAGCGTATTGA

## Chapter 7 - genetic elements

### Synthetic UTR sequences

“Bujard SD/UTR” <sup>3</sup>	gaattcattaaag <b>aggagaa</b> aggtacc
“B0030 SD/UTR” ( <a href="http://partsregistry.org/Part:BBa_B0030">http://partsregistry.org/Part:BBa_B0030</a> )	tttaaga <b>aggag</b> atatacat
“Weiss SD/UTR” <sup>9</sup>	attaaag <b>aggagaa</b> attaagc
“Anderson SD/UTR” ( <a href="http://partsregistry.org/Part:BBa_J61100">http://partsregistry.org/Part:BBa_J61100</a> )	tctagagaaag <b>ggggaca</b> aaactagt

### Genomic UTR sequences

UTR	Length	Sequence
lacZp1	18	AATTGTGAGCGGATAACA
serB	19	GTTATTTTTCCCTGCTTCGA
ChiA	23	GTAGGACTTTTTGTTTTGCAGTTT
lacY	31	TAATAACCGGGCAGGCCATGTCTGCCCGTAT
sodA	31	ACTGCTTACGCGGCATTAACAATCGGCCGCC
ompRp3	35	GCTTTTTTAAGAATACACGCTTACAAATTGTTGCG
trpR	36	AGCGAGTACAACCGGGGGAGGCATTTTGCTTCCCCC
glpA	44	AAATCAAACAATTCATGTTTTTACTATGGCTAAATGGTAAAAAA
rhoL	45	GACTTCGTATTAACATACCTTATTAAGTTTGAATCTTGTAATTTCCAA
CRISPRI	53	GTGGGTTGTTTTTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGA
fixA	58	TCAAAATTAAGGGCGTGATATCTGTAATTAACACCACCGATATGAACGACGTTTCCT
lldP	90	GTCATTATCCCTACACAACAATGGCAGTGCCACTTTTACACAACGTGTGACAAGGA GATGAGCAACAGACTCATTACACGATGTGCG

### Promoter sequences

Promoter	Constitutive or Inducible	Sequence ( <b><u>bold underline</u></b> - transcriptional start site; <b><i>bold italic</i></b> – operator sites)
J23119	Constitutive	ttgacagctagctcagtcctaggtataatagatc <b><u>t</u></b>
J23101	Constitutive	tttacagctagctcagtcctaggtattatagatc <b><u>t</u></b>
J23105	Constitutive	tttacggctagctcagtcctaggtactatagatc <b><u>t</u></b>



J23110	Constitutive	ttacggctagctcagtcctaggtacaatagatct
P <sub>T7A1</sub>	Constitutive	cgaggccaacttaagagacttaaagattaatttaaatttatcaaaaagagtattgactt aaagttaacctataggatacttacagccatcgagagggga
P <sub>L</sub> lacO-1	Inducible by IPTG	<b>ataaatgtgagcggataaacattgacattgtgagcggataacaagatactgagcacatc</b> agcaggacgcactgacc
P <sub>A1</sub> lacO-1	Inducible by IPTG	aaaatttatcaaaaagagtgtgactt <b>gtgagcggataacaatgatacttagattcaattg</b> <b>tgagcggataacaatttcacaca</b>

#### IS10 mutant sequences

Description	Sequence ( <b>bold</b> - Shine-Dalgarno sequences; <b>bold italic</b> – specificity sites)
IS10wt UTR	<b>gcgaaaaatcaataaggagacaacaag</b>
IS10-9 UTR	<b>ggcttaaatcaataaggagacaacaag</b>
IS10wt antisense RNA	tgcacatctgtgtctgattattg <b>at</b> <b>tttcgc</b> gaaaccatttgatcatatgacaagatgtgtatccaccttaa cttaatgattttacaaaatcattaggggattcatcag
IS10-9 antisense RNA	tgcacatctgtgtctgattattg <b>at</b> <b>tttaugc</b> gaaaccatttgatcatatgacaagatgtgtatccaccttaa acttaatgattttacaaaatcattaggggattcatcag

#### PT181 sequences

Description	Sequence
pT181 UTR	aacaaaataaaaaggagtcgctcacgccctgaccaaagtttgaacgacatcattcaagaaaaaacact gagtgTTTTATAATcttgatatattagatataaacgatattaaatatacataaagatatatttgggtgagcgatt ccttaaacgaaattgagattaaggagtcgctctTTTTATgataaaaacaatcatgcaaatcattcaaatcattgg aaaatcacgatttagacaattttctaaaaccggctactctaatagccggtgtaa
pT181 antisense RNA	atacaagattataaaaacaactcagtgTTTTTctttgaatgatgtcgttcacaaaacttggcagggcgtgagcg actcctttttattt