DETECTING RNA REGULATORY INTERACTIONS

DETECTING RNA REGULATORY INTERACTIONS IN BACTERIAL CELLS

By

Kacper Kuryllo

B.Sc., University of Waterloo, Waterloo Ontario, 2006

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2014) McMaster University

(Biochemistry & Biomedical Sciences) Hamilton, Ontario

TITLE: Detecting RNA Regulatory Interactions in Bacterial Cells

AUTHOR: Kacper Kuryllo, B.Sc. (University of Waterloo)

SUPERVISOR: Dr. Yingfu Li

NUMBER OF PAGES: xiii 131

Abstract

Non-coding RNAs are involved in the regulation of most major cellular process in *Escherichia coli*. With current technologies, many of these molecules have been identified; however, the full scope of their regulatory interactions is still unknown. None of the techniques currently in use employ the regulatory effect of the RNAs, which is the major unifying attribute of these molecules, for their identification. This thesis presents projects involving the design of a dual-reporter plasmid and screening method in the discovery and characterization of RNA regulatory interactions

The first project details the engineering of the dual reporter plasmid. This vector utilizes one fluorescent protein to detect regulatory events and a second to normalize for off-target effects. The second project utilizes this tool in the discovery and characterization of novel regulatory responses. This is accomplished by screening a library of intergenic regions for regulatory responses against a collection of metabolite. Interesting interactions involving nitrogen abundance, iron and uracil are identified and further examined.

Finally, this thesis examines how this technology can be further expanded for the study of RNA regulatory functions. The use of the screening method for the detection of regulatory events caused by alternative minimal media composition and the potential for the dual reporter plasmid to aid in the study of riboswitches are investigated.

Acknowledgements

I would first like to thank my supervisors Dr. Yingfu Li and Dr. Eric Brown. I have benefitted greatly from their outstanding contribution to my academic development. Their careful guidance has been instrumental in shaping the direction of the research found in this document.

I would also like to thank my committee member Marie Elliot. Her depths of knowledge regarding my field of study, as well as her input and guidance towards my project design, have been a monumental help. The questions she has asked constantly encouraged me to keep reading and seek out further knowledge.

Many of my contemporaries from the Li, Brown and other labs have been a great help in not only developing my research but also in providing a wonderful environment to work in. In particular I would like to thank Casey Fowler, Simon McManus, Wendy Mok, Bobbijo Sawchyn and Zohaib Ghazi from the Li lab, Geordie Stewart and Soumaya Zlitni from the Brown lab and Marisa Azad from the Wright lab for their contributions.

Finally, I would like to thank my family who have always been there for me and provided positive support throughout my studies.

To my parents who have always been there for me.

Table of Contents

List of Tables

Supplemental

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Supplemental

List of Abbreviations

ATC	anhydrotetracycline
bp	base-pair
BFP	blue fluorescent protein
CFP	cyan fluorescent protein
DNA	2′-deoxyribonucleic acid
EV	empty vector
GFP	green fluorescent protein
IGR	intergenic region
kbp	kilo base pairs
MCS	multiple cloning site
mRNA	messenger RNA
ncRNA	non-coding RNA
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
RFP	red fluorescent protein
RNA	ribonucleic acid
UTR	untranslated region
YFP	yellow fluorescent protein

Declaration of Academic Achievement

I took the primary role in selection of all projects for the thesis, design of experiments, interpretation of the data and the writing of the manuscripts. I also performed all experimental work appearing in this document, with the exception of the western blots and q-RT-PCR in Chapter 2, and riboswitch screens and construction in Chapter 4, which I performed with the assistance of Weijia Zhu, Shahrzad Jahanshahi, Zohaib Ghazi and Michael Yan respectively.

Chapter 1

Background and Introduction

1.1 Current challenges in the field

The general concepts of biology have developed greatly in the last three decades. RNA, which was primarily viewed as an intermediary messenger between DNA and proteins, with only cursory evidence for catalytic and regulatory functions, is now widely accepted to rival proteins in these tasks. The RNA world theory, that holds that life passed through a stage where RNA was the primary molecule of life, central to catalytic, structural and genetic roles (Gilbert, 1986; Joyce, 2002), is accepted by many scientists. This paradigm shift was propelled forward by the discovery of a multitude of new functions for RNA within the cell. Many of these new roles of RNA were demonstrated to occur by mechanisms in which proteins play only accessory roles or in their absence altogether. The ability of RNA to function autonomously is best illustrated through the example of riboswitches that regulate by translational initiation, whereby allosteric RNAs respond to the binding of ligands by altering their own propensity for interaction with another RNA based machine, the ribosome (Mandal, M., Breaker, 2004).

The identification of the mechanisms forming the known set of RNA based regulatory interactions contributes greatly towards an understanding of the complexity and scope of regulatory networks within the cell. Many of the RNAs involved in these processes are capable of interacting with a multitude of targets and function as parts of

sophisticated regulatory networks, involving proteins, small molecules and other RNA based components. However, though many regulatory element-target pairs have been identified there is currently a deficit in the understanding of the biological significance of these interactions. That is to say, that although there may be a clear mechanistic understanding of how a regulatory element (RNA, protein or other) functions to modulate expression of a linked gene, the purpose of this change in expression is more often than not poorly understood. It is typically more challenging to address the relevance of a regulatory interaction within a biological context, than it is to identify or gain a mechanistic understanding of the interaction. Part of this difficulty is caused by a lack of intracellular study of these elements.

Intracellular methods can be used in conjunction with one another to provide additional insights. The development of techniques for quantifying the total metabolic component of cells has created the field of metabolomics (Patti et al., 2012). The existence of tools for studying RNA regulatory responses in high-throughput, would allow for their use alongside metabolomic methods to provide data sets for comparison. Chemical genomics is an additional tool for studying living cells that has been demonstrated to integrate well with other methods. Small molecules can be used to interrupt gene function and chart genetic interactions when combined with other techniques (Barker et al., 2010). The potential uses of chemical genomics techniques for identifying RNA genetic interactions has not been explored as of yet. The success of this approach would be highly dependent on the existence of methods for detecting RNA regulatory interactions within cells.

Tools for the intracellular study of RNA interactions can allow for observation of regulatory responses that occur only inside of cells. For instance, the examination of how a cells growth rate affects these interactions is not feasible *in vitro*. Unfortunately, protocols for intercellular observation of RNA regulatory interactions often lack sensitivity, are complex or are unsuitable for high-throughput applications. These methods in particular are impractical for identifying regulatory functions by screening, as their low throughput severely limits the quantity of potential interactions that can be tested. Development of an intercellular technique that can overcome these limitations would be of great practical benefit.

There are many aspects of RNA based regulation that have either not been addressed or are unclear. The potential for RNA regulatory elements to exist within long non-coding regions in bacterial operons is of particular interest to this study. Given that bacterial genomes greatly favor a compact arrangement of genes (Rogozin et al. 2002), the presence of these regions may be indicative of functional importance. As RNA regulatory elements tend to be concentrated in non-coding sequence, these regions are prime candidates for the discovery of RNA regulatory functions.

This thesis investigates the potential of intergenic regions (IGR) in the *Escherichia coli* genome to contain RNA regulatory elements by the use of an intracellular reporter system. This is primarily accomplished by the screening and later profiling of an IGR library against a multitude of growth conditions. The reporter system uses pairs of highly similar fluorescent proteins to detect changes in gene expression and to normalize for non-specific effects. This normalization technique alleviates many of the

limitations of fluorescent proteins caused by the slow maturation of these macromolecules, while still retaining their benefits.

The following sections serve to introduce the background information underlying the concepts described in this thesis. First, a brief history of regulatory RNAs is presented, with a focus on prokaryotic RNAs as they are of the most relevance to this study. Section 1.3 then discusses tools used to study and identify non-coding RNA (ncRNA). In section 1.4, the properties of reporter genes are examined, with a strong emphasis on fluorescent proteins. Many of the details presented are important for understanding the rationale applied during reporter selection in Chapter 2. Section 1.5 addresses how small molecules can be used as tools to study biology, specifically in the fields of metabolomics and chemical genetics. Finally, section 1.6 provides a hypothesis and an outline for the remainder of the thesis.

1.2 Bacterial Regulatory Non-coding RNA

The first non-coding RNA characterized was the alanine transfer RNA (tRNA), from *Saccharomyces cerevisiae* (Holley et al., 1965), though the existence of functional ncRNAs that could mediate translation had been previously predicted by Francis Crick in 1958 (Crick, 1958). The discovery of this tRNA demonstrated that ncRNA could possess functional roles, and was instrumental in dispelling the view that RNA was simply an intermediary conveyor of genetic information between DNA and proteins.

Since those early times, a multitude of regulatory RNAs have been discovered in diverse types of organisms. One of the first regulatory RNAs identified was the *E. coli* MicF small RNA (sRNA) (Mizuno et al., 1984), which was identified before microRNAs in eukaryotes. A large divide exists between the types of regulatory RNAs that are found in eukaryotic and prokaryotic organisms with virtually no mechanistic overlap between the two. Regulatory RNAs in eukaryotes are dominated by small interfering RNAs (siRNA) and microRNAs, while prokaryotes typically use sRNAs and riboswitches. MicroRNAs and sRNAs can be thought of as analogous systems. With both these classes, the ncRNAs bind and modulate the expression of messenger RNAs (mRNA) that are transcribed separately. However, the similarities largely end there. MicroRNAs have clearly defined lengths, maturation process and mechanisms (Ameres and Zamore, 2013). sRNAs are more dynamic, having variable lengths, structures and mechanisms (Storz et al., 2011). Likely the most unifying feature of sRNA is that many are dependent on the RNA chaperon protein Hfq for activity (Waters and Storz, 2009). Riboswitches, which do have limited occurrence in eukaryotes, are largely confined to prokaryotic organisms, and have different mechanism of regulation in these two domains of life (Wachter, 2010). Due to the lack of similarities between prokaryotic and eukaryotic regulatory RNAs, and that the primary interest of this work involves the former, from this point forwards we will exclusively concern ourselves with prokaryotic regulatory RNAs.

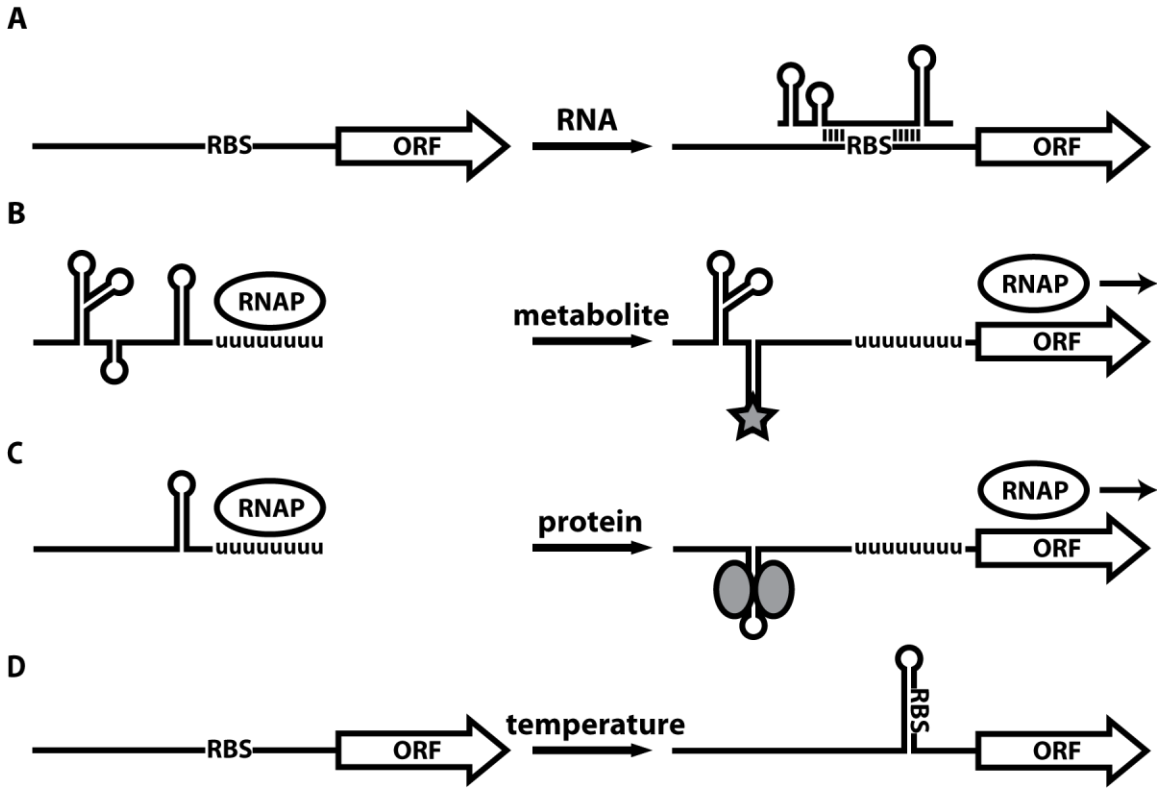

Figure 1-1 Schematic of RNA regulatory interactions in prokaryotes. (A) An RNA-RNA interaction involving the binding of RNA transcribed at one chromosomal locus to mRNA transcribed at another. The interaction blocks a ribosome binding site (RBS) and prevents translation. (B) Metabolite-RNA interaction is illustrated by a transcriptional riboswitch. Binding of a metabolite to the aptamer domain of a riboswitch prevents the formation of a transcriptional terminator. Analogous systems also exist for the detection of metal ions. (C) In this protein-RNA regulatory interaction, protein binds to a hairpin loop during transcription. This prevents the formation of a second hairpin downstream that terminates transcription. (D) This panel illustrates an RNA that responds to temperature. In this example a decrease in temperature results in the formation of a hairpin loop that prevents translation the mRNA by blocking the RBS. Though this example deals with temperature it should be noted that similar mechanisms exist to detect other parameters, such as pH.

RNA regulatory activity in prokaryotes occurs primarily by the response of mRNA elements to various types of regulatory stimuli. The stimuli almost always interact with elements located in the 5′ untranslated regions (UTR) of mRNA, and can

include other RNAs (Figure 1-1A) (Storz et al., 2011; Grundy and Henkin 1993), small molecules (Figure 1-1B) (Barrick and Breaker, 2007), metal ions (Massé et al., 2007), proteins (Figure 1-1C) (Babitzke et al., 2009), temperature (Figure 1-1D) (Kortmann and Narberhaus, 2012), and pH (Bingham et al., 1990). These interactions result in regulatory responses that alter the expression of downstream open reading frames, by affecting transcriptional termination, protein translation or by altering the mRNAs stability (Massé et al., 2003). Bacteria also are known to vary in their utilization of various regulatory strategies. *Gammaproteobacteria*, such as *Escherichia*, tend to contain many examples of sRNAs (Gottesman and Storz, 2010), while *firmicutes* and *Fusobacteria* are known to contain more riboswitches than other bacterial classes (Barrick and Breaker, 2007).

In bacteria, RNA-RNA regulatory interactions are typically mediated by sRNA-mRNA binding (Figure 1-1A), however other RNA molecules, such as tRNA, are also known to bind mRNA (Grundy and Henkin, 1993). sRNA-mRNA interactions are particularly interesting, as small RNAs often have multiple mRNA targets. The *E. coli* RyhB sRNA, is known to interact with various mRNAs involved in iron homeostasis (Massé et al., 2003). Other small RNAs regulate seemingly unrelated cellular processes. For example, the GcvB sRNA controls the expression of genes involved in amino acid transport (Sharma et al., 2011; Urbanowski et al., 2000), magnesium homeostasis (Coornaert et al., 2013) and acid resistance (Jin et al., 2009).

The most well conserved class of RNA regulatory elements among prokaryotes are riboswitches, though this observation may be skewed, as they are primarily identified

by searching for sequence and structural conservation. These Riboswitches lie in the 5′ UTRs of transcripts and regulate the expression of downstream genes (Figure 1-2B). They are typically involved in the regulation of metabolic processes and are known to respond to amino acids (Sudarsan et al., 2003), nucleic bases (Kim and Breaker, 2008), positively and negatively charged ions (Baker et al., 2012; Dann et al., 2007; Ren et al., 2012), cofactors (Nahvi et al. 2002), and signal transduction molecules (Sudarsan et al., 2008). While the vast majority of them regulate either through transcriptional attenuation or translational initiation, others function catalytically. The *glmS* (Winkler et al., 2004) and di-cyclic GMP (Sudarsan et al., 2008) riboswitches are the only two riboswitch-ribozymes to have been identified. They both function catalytically cleaving the mRNAs they are located within. These cleavage events are thought to regulate the expression of the associated open reading frames (ORF) by altering mRNA stability (Collins et al., 2007; Lee et al., 2010).

Given the vast amounts of genomic sequence data that have been produced during the last decade, there exists a wealth of sequence space for further identification of these molecules.

1.3 Overview of methods in non-coding RNA discovery

Initially, functional non-coding RNAs were discovered through the examination of discrete RNA fractions, yielding ribosomal RNA (rRNA), tRNAs and spicing associated RNAs (Storz, 2002). After these early discoveries, more advanced sequence manipulation techniques then allowed for the identification of ncRNAs by examining

discrete genomic sequences in the study of other genes. Three early examples of fortuitous discoveries include tRNA regulated transcripts (Grundy and Henkin, 1993), the MicF sRNA (Mizuno et al., 1984) and miRNA (Lee et al., 1993). A few ncRNAs were also identified by RNA isolation techniques. 4.5S and 6S RNA were discovered by ^{32}P-labelling of total RNA (Hindley, 1967) and Spot 42 (spf), 10Sa (tmRNA) and 10Sb (M1) RNA by 2D gel electrophoresis (Ikemura and Dahlberg, 1973). However, it was not until full genomic sequences were made available that techniques were invented specifically for the discovery of particular classes of ncRNA.

Some of the earliest evidence that there were yet many ncRNA still to be discovered came from the use of genome tiling arrays. These consist of sets of oligonucleotides that are used to assay differential gene expression by hybridization to mRNA pools (de Saizieu et al., 1998; Lockhart et al., 1996; Wodicka et al., 1997). Their use in the analysis of the *E. coli* genome demonstrated that virtually all DNA is transcribed, and that many non-coding regions likely contain functionally relevant small and antisense RNAs (Selinger et al., 2000).

Targeted approaches for the discovery of functional non-coding RNA typically involve either ncRNA isolation experiments or bioinformatical analysis of genomic sequence data. Early bioinformatical analysis identified ncRNAs by searching for conserved non-coding sequences, the presences of highly structured regions or arrangements of genomic elements. In 2001, studies looking at sequence conservation and genomic elements discovered 31 of the currently known *E. coli* sRNAs (Argaman et al., 2001; Rivas et al., 2001; Wassarman et al., 2001). Later these methods were

improved and used to identify sRNA genes in a wide variety of other bacterial genomes (Livny and Waldor, 2007). These techniques have evolved to better predict sRNAs (Pichon and Felden, 2008), identify their mRNAs targets and provide interactive tools for the use of the research community (Kerry, 2014; Tjaden et al., 2006; Wright et al., 2013).

Bioinformatics has been particularly useful in riboswitch discovery. Most riboswitches have been identified by comparing genomes to find non-coding regions that contain structural conservation (Barrick et al., 2004). Further refinement of this approach, and its application to genomic and metagenomic sequence data, has yielded many potential riboswitches and other structural RNAs (Weinberg 2010). Though a plethora of structural motifs have been discovered, their functions are largely unknown. To determine whether these elements are riboswitches they need to be demonstrated to interact with ligands, which is a difficult task (Breaker, 2011; Meyer et al., 2011). The creation of a technique that allows for the high-throughput screening of riboswitch-ligand interactions would prove to be of extreme importance in the study of these sequence motifs.

Experimental approaches have proven to be of use in the identification of ncRNAs (Sharma and Vogel, 2009). Shotgun cloning of total RNA fractions in the range of ~50 to ~500 nucleotides has been employed to identify novel sRNA and experimentally validate many that were computationally identified (Vogel et al., 2003). Other methods use known features of small RNAs for their discovery. Co-immunoprecipitation with Hfq has been used to identify new sRNAs, and to investigate whether previously known sRNAs bind Hfq (Zhang et al., 2003).

1.4 Reporter genes

Genes used to monitor gene expression are referred to as reporter genes. They use a multitude of mechanisms to detect gene expression and most often code for proteins, though RNA reporters exist for the study of transcription and RNA localization (Daigle and Ellenberg, 2007; Paige et al., 2007). These genes typically produce signals by colorimetery, luminescence or fluorescence. However, any gene capable of producing a tractable output could be deemed a reporter.

Reporters come with their own sets of strengths and limitations. Of particular importance, an inverse relationship exists between the sensitivity of reporters and their simplicity of use. Large gains in sensitivity are typically achieved by amplification steps that necessitate additional processing. Most often these amplification procedures are enzymatic and signal strength is increased by the production of colorimetric substrate or luminescence (Schenborn et al., 1999). The additional processing that these reporters require often limits their usefulness in live cells and high-throughput applications, or necessitates the addition of expensive substrates to growth media. Other reporters, such as fluorescent proteins, lack amplification steps and do not inherit these limitations; however they are also less sensitive.

The first fluorescence protein (GFP) was isolated from the jelly fish *Aequorea victoria* and most fluorescent protein variants in use today are derived from its gene (Kremers et al., 2011). Since then fluorescent proteins have been found in various other organisms including: mushroom corals from the *Discosoma* genus (Fradkov 2000), the sea pansy *Renilla reniformis* (Wampler et al., 1971), and the Japanese freshwater eel

Anguilla japonica (Kumagai et al., 2013). Although fluorescent proteins have multiple origins they share common structural features. Fluorescent proteins typically consist of a beta barrel structure with an internal hydrophobic core that houses a fluorescent chromophore.

The chromophore within a fluorescent protein is responsible for the observed fluorescent behavior, while the remainder of the protein provides an environment that facilitates the formation and fluorescence of the chromophore. Though the Ser-Tyr-Gly motif that forms the chromophore in the *A. victoria* wild-type GFP is commonly found throughout nature it does not react to form a chromophore outside of the protein. The mechanism for chromophore formation was conventionally thought to involve a conformational rearrangement, followed by cyclization, dehydration and oxidation reactions (Barondeau et al., 2003). However recent insights have demonstrated the oxidation of the chromophore likely occurs prior to dehydration (Figure 1-2A) (Rosenow et al., 2005; Strack et al., 2011). This enzymatic process can tolerate a large variety of chromophore motif substitutions allowing for the creation of chromophores with new spectral properties (Figure 1-2B). In general chromophore maturation can be thought of as functioning by increasing the size of the conjugated system found on a central aromatic amino acid (Gross et al., 2000). The aromatic amino acids found in fluorescent protein motifs are themselves fluorescent in the UV range of the spectra (Teale and Weber, 1957). Increasing the conjugated system shifts the fluorescence of these residues into the visible range.

There are several important exceptions to the mechanism of chromophore presented above. First, though all chromophores for fluorescent proteins mentioned in this document possess this mechanism, the sea anemone-derived dsRED chromophore undergoes a second oxidation step prior to dehydration that serves to further increase the length of the conjugated system along the protein backbone (Rosenow et al., 2005; Subach et al., 2012). Additionally, the spectral properties of a chromophore may be altered by interaction with neighboring residues. The covalent structure of the yellow fluorescent protein (YFP) chromophore is identical to that from GFP. A π-stacking interaction between the chromophore and a tyrosine residue results in a redshift in the emission spectra (Wachter et al., 1998), creating the yellow colour.

Maturation of fluorescent proteins involves both the folding of the proteins and formation of their chromophores. The process is typically rate-limited by the oxidation step in chromophore formation (Iizuka et al., 2011). The longer maturation rates of the DsRED protein (~10hours) may be explained by its possession of two oxidation steps (Bevis and Glick, 2002).

While the aforementioned structural and mechanistic features hold true for the majority of fluorescent proteins, they do not for the eel fluorescent protein (UnaG). This protein appears to be of a unique structural origin and achieves fluorescence by the binding of a heme-derivative that it uses as a chromophore (Kumagai et al., 2013). This protein has the advantage that its maturation rate is not limited by chromophore generation, however as it does not make the chromophore it can only be utilized in

environments that supply this heme-derivative. The potential applications of this new protein will likely be quite different from those of current generation fluorescent proteins.

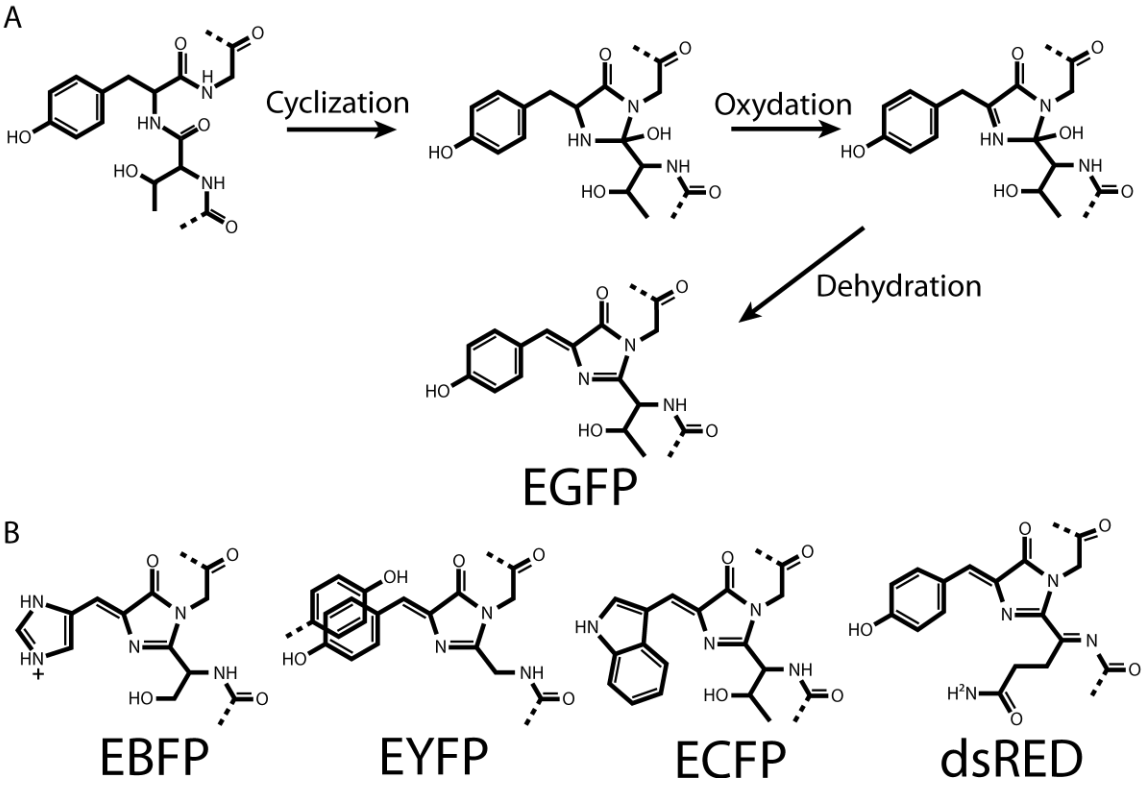

Figure 1-2 Schematic of fluorescent protein chromophores and maturation. (A) The major steps of chromophore maturation common to all fluorescent proteins examined in this thesis are presented with EGFP as an example. (B) The remaining fluorescent protein chromophores used in this study are presented. Note that formation of the dsRED chromophore involves two oxidation steps.

1.5 Small molecules in scientific discovery

The study of small molecules, their effects on cells and their response to cellular perturbance is fundamental to an understanding of life. Small molecules can be separated into two groups: metabolites and synthetic molecules. The major distinction between the two is that metabolites are naturally processed during metabolism, while synthetic molecules exist outside of this scope.

The use of small molecules to study biological systems is referred to as chemical genetics. Traditionally this has focused on the interaction of these molecules with proteins, however small molecules may also bind other targets such as RNA and DNA. These interactions may affect enzymatic function, signal transduction, gene regulation or other aspects of cellular biology (O' Connor, 2011). Two major approaches that are analogous to classic forward and reverse genetics exist within chemical genetics: forward chemical genetics and reverse chemical genetics. Forward chemical genetics involves the use of small molecules in an attempt to elicit a phenotype (Stockwell, 2000). Once a phenotype is observed, the mechanism responsible for production of that phenotype is examined. Reverse chemical genetics works by attempting to identify molecules that interact with a target protein, DNA or RNA (Stockwell, 2000). When the reverse approach is performed *in vitro,* it is constrained by the availability of purified targets and the applicability of the resulting data for in cell studies. In-cell reverse chemical genetics mitigates these issues while also allowing for the study of complex interactions such as effects on signal transduction pathways and gene regulation that may not be easily studied *in vitro*. Chemical genetic approaches have various advantages over traditional genetics

(Barker et al., 2010). While construction of mutants in essential genes requires inducible complementation of the gene, chemical genetics can target these genes relatively easily (Walsh and Chang, 2006). Small molecules are also more conducive to temporal and quantitative control (O' Connor, 2011).

It is important to note that the goal of chemical genetics is the study of biological systems and not drug discovery. This is a useful distinction, as it increases the diversity of molecules available for screening, by avoiding constraints in chemical properties and structural motifs that are typically applied during the identification of pharmaceuticals (O' Connor, 2011). Though this extensive chemical space assists in the identification of chemical genetic interactions, it is important to note one significant disadvantage of chemical genetics relative to traditional genetic approaches. While the generation of mutants involved in classical genetics is now a straight-forward process, the discovery of chemical entities capable of doing the same is typically more labor intensive with no certainty that one can be found in a library of small molecules (O' Connor, 2011).

Chemical genetics has proven to be more successful in prokaryotic organisms than eukaryotes. This is primarily due to inherent challenges with eukaryotic cell culture. However the utility of these approaches to studying more complex organisms has increased significantly. The combination of chemical genetics with other gene disruption techniques such as RNAi (Brough et al., 2011; Borisy et al., 2003), is proving to be a significant application in these organisms.

As mentioned in section 1.1 study of the metabolite composition of a cell is referred to as metabolomics. Analysis of the abundance of metabolites is performed

under various conditions to allow for the generation of data sets that correlate these metabolites to phenotypes. Untargeted metabolomics is the term used for the generation of these global data sets (Patti et al., 2012), the primary objective of which is to link various metabolites to biological pathways and advance the understanding of cellular physiology. A particular advantage of this approach is that it is possible to generate these data sets from small biological samples (Kaddurah-Daouk et al., 2008), greatly increasing its clinical applicability.

This relatively new field of study was made possible by recent technical advances. Mass spectrometry has allowed for the accumulation of metabolomic data and identified molecules that have not previously been found in databases (Patti et al., 2012). However, analysis of the resulting data sets and utilization of the involved techniques has remained a challenge.

1.6 Objectives of this thesis

This thesis works under the overlying hypothesis that novel regulatory functions exist within many non-coding genomic regions. An additional hypothesis is that a carefully designed sensor consisting of fluorescent proteins pairs can provide simple and reliable analysis of RNA regulatory function. Finally, the study proposes that chemical genetics can be used to examine how RNA regulatory regions respond within cellular environments. The aim of the projects presented in the following three chapters of this thesis (Chapters 2-4) is to address these related hypotheses and present technology that is useful to other researchers in the future. Though *E. coli* was used as a model organism

for all experimental analysis presented within this document, and thus the fluorescent sensor would likely only be directly applicable within related *Gammaproteobacteria*, the presented technology should prove useful in the design of similar sensors and methods in other prokaryotic organisms. Finally, it should be noted that while RNA regulatory interactions are our primary interest, this technology should be useful in the study of other modes of regulation, and collaborations not presented with this document are underway to examine this potential.

The first project in the thesis presents the design and characterization of a tool for in cell analysis of RNA regulatory interactions. This is accomplished with the use of a sensor consisting of two fluorescent proteins. One fluorescent protein is used to quantify regulatory responses, while the second is used to normalize for off-target effects. The sensor is designed with the explicit purpose of later testing the central hypothesis of this work. The criteria of this design is that the sensor should provide quantitative reproducible results, is capable of normalizing for off-target effects, is functional for different RNA regulatory strategies, is useful at high and low expression levels and can detect both synthetic and natural regulatory interactions.

The first three of these design criteria are examined by testing the response of known RNA regulatory targets to three distinct and well characterized RNA and protein regulatory effectors, expressed off synthetic constructs. This analysis provides highly reproducible data consistent with previously published results for all three regulatory systems. The final two criteria are tested by examining the utility of the sensor for detecting endogenously expressed regulatory effectors at lower expression levels. This

proves successful for not only replicating previously published data but also provides new insights into the function of the regulatory targets.

The second project, described in Chapter 3, is undertaken with the goal of testing the primary hypothesis of the thesis. Non-coding regions within the *E. coli MG1655* genome are examined for unknown regulatory responses, against a wide array of growth conditions. This is done by a chemical genetic approach, whereby a library of small molecules in the form of cellular metabolites are examined for their ability to elicit regulatory responses in a library of IGRs found between genes on poly-cistronic transcripts.

The IGR library is constructed by careful selection of elements for study and their insertion into the dual reporter plasmid. IGRs within operons are selected as the focus of this work as their presence may be indicative of regulatory function, due to their relative low occurrence within the compact *E. coli* genome. As we seek to primarily investigate RNA regulatory interactions, a computational method is developed for the selection of IGRs likely to contain RNA regulatory elements. This program identifies IGRs within the *E. coli MG1655* genome by examining distances between consecutive genes, selects those IGRs only found within operons and then excludes IGRs containing transcriptional promoters. The longest 96 of these IGRs, ranging from 86-560 base pairs (bp) are selected for further analysis.

This library of IGRs is then screened for regulatory responses against a library of metabolites. This identifies 16 significant regulatory interactions, the majority of which are in response to nitrogen source abundance, growth rate/phase, uracil and iron. The

reproducibility of the data is demonstrated by generating kinetic plots of expression under these conditions.

The final project, comprising Chapter 4, is an exploratory chapter that presents experiments that expand on additional uses for the reporter system. While the central goal of this chapter is not to test the primary hypothesis of this thesis, many of the individual experiments do so, while also providing data for the secondary hypothesis. Experiments contained within examine our technology for detecting new regulatory interactions under metabolite limitation and from a library of known and predicted riboswitches.

The study outlined in Chapter 2 has been accepted to ChemBioChem for publication. The majority of the text in Chapter 2 can be found in the submitted publication. However, it has been modified to fit the format of the thesis. The data in Chapter 3 is considered largely ready for publication though an additional experiment to provide verification of observed regulatory interactions may be included later. Chapter 4, the final experimental chapter does not compromise a single research project for publication, but serves to further expand upon aspects of the thesis. This thesis is presented in standard thesis format. References for all sections of the thesis are located at the end of the document.

Chapter 2

A Dual Reporter System for Detecting RNA Interactions in Bacterial Cells

2.1 INTRODUCTION

There have been major efforts in recent years to develop various techniques to identify and study noncoding RNA (ncRNA) elements that are involved in gene regulation (Barrick et al., 2004; Pichon and Felden, 2008; Tjaden et al., 2006; Selinger at al., 2000; Sharma and Vogel, 2009; Vogel at al., 2003; Vogel and Sharma, 2005; Zhang et al., 2003). Although a plethora of ncRNAs have been identified, there is a lack of suitable tools for observing their interactions with various partners in cells. Such techniques would have to be simple, intracellularly functional and broadly applicable to any given ncRNA of interest.

Intracellular molecular interactions can be conveniently monitored through the use of easily tractable reporters such as fluorescent proteins, bioluminescent enzyme luciferase or colorimetric enzyme β-galactosidase. Fluorescent protein-based methods are simpler and less expensive than enzyme-based assays, as they neither need external substrates nor require cell lysis. These features render fluorescent protein more amenable to high-throughput screening: samples can be read kinetically, potentially identifying transient interactions that are often overlooked with endpoint assays. However, one major limitation with fluorescent protein-based assays is that fluorescent proteins require time to

mature before they can generate a signal (Fisher and DeLisa, 2008). Due to the lag between the expression and maturation of these proteins, cells grown at different rates will contain different proportions of maturated protein, skewing their fluorescent signals. Thus, when cells are examined under different growth conditions, typical fluorescent protein-based assays are prone to systematic errors. This is especially prevalent in faster growing prokaryotes where the doubling time of the organism is typically shorter than a fluorescent protein's half-life of maturation.

To overcome this problem, we have conceived a dual reporter system that uses a pair of fluorescent proteins to detect RNA interactions in bacteria. In our design, one protein is used to measure changes in gene expression, while the second functions to normalize the signal from the first. To minimize signal fluctuations from transcription, both fluorescent protein genes are placed under the control of a single transcriptional promoter. To limit signal divergence, the fluorescent protein pair is selected to have distinct emission spectra but otherwise very similar properties. By this design, the relative fluorescence of the protein pair becomes independent of their degree of maturation, rates of transcription and translation, and ultimately the bacterial growth conditions. In this study, we will report the investigation of utilizing such a reporter system to detect intracellular interactions of select RNA molecules with their partners in *Escherishia coli* (*E. coli*) as a model organism.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains and materials

Plasmid cloning was performed with the use of *Escherichia coli DH5αZ1*. This strain was also used for assays that involved the use of anhydrotetracyclin (ATC) as an inducer. All assays not requiring the use of ATC were performed with *Escherichia coli MG1655*. DNA oligonucleotides (Integrated DNA Technologies, Coralville, IA) were produced by standard phosphoramidite chemistry. PCR amplification was carried out with the use of Phusion DNA polymerase (New England BIOLABS Inc., Ipswich, MA). The following restriction enzymes were used in this study: KpnI, BamHI, AfeI, XmaI and SpeI (Fisher Scientific, Waltham, MA).

2.2.2 Cloning and plasmid construction

Plasmid pBnR was constructed by the insertion of genes coding for RFP, BFP and an RNA1 promoter into the pBAD30-Ara (Fowler 2011) backbone (pBAD30 with *araC* and the pBAD promoter removed). Next pCnY was created by replacing BFP with CFP and RFP with YFP. Using a process of crossover PCR, regions homologous to 1 kbp (kilo base pairs) upstream and downstream of the *lacZ* locus of *E. coli K12* were integrated into the pCnY vector to yield plasmid pCYi-amp. Replacement of the ampicillin resistance marker (*bla*) on pCYi-amp with a chloramphenicol cassette (*CAT*), followed by the predicted bidirectional *macB-cspD* terminator (Lesnik et al., 2001), yielded plasmid pCYi-cm. pCYi-fu was then produced by insertion of BamHI and NcoI restriction sites directly downstream the YFP start codon. Optimization of the pCYi-fu

CFP RBS yielded plasmid pCYi-f2, which was used to produce plasmid pCYi-f3 by optimization of the YFP RBS and multiple cloning site. Plasmid pRNY was created by the insertion of an AfeI restriction site 2 basepairs (bp) upstream of the pNYL-MCS11 transcriptional start site. gcvB-KO strain was created by the integration of kanamycin resistance cassette (*aph*) in place of the native *gcvB* gene in *Escherichia coli MG1655*. pCYi-f3, pCYi-oppA, pCYi-gltI pCYi-glmUS genomic integrants were created by linearization of their respective plasmids with SpeI, followed by recombination into the *Escherichia coli MG1655 lacZ* locus. Recombination was achieved by the use of the Warner-Datsenko method (Datsenko and Wanner, 2000). Plasmid pCYi-ter was constructed by insertion of the *dnaG-rpsU* intergenic region into vector pCYi-f3. Details of primers and plasmids used in this study can be found respectively in Supplementary Tables S2-1 and S2-2.

2.2.3 Fluorescence assays and signal normalization

E. coli K12 cell cultures were incubated at 37°C while shaking at 250 rpm. For all assays individual colonies were picked off LB Agar plates with a pipette tip to inoculate starter cultures, that were grown for 16 h in LB media. These starter cultures were then used to inoculate cultures for assay at a 1/1000 dilution. When necessary, ATC (100 ng/mL) was added to media for assay cultures prior to inoculation. This concentration of ATC was selected as the highest concentration of ATC for which no toxic growth effects were apparent. For fluorescence development experiments aliquots were read initially

every 0.5 h, then as cell growth rate retarded the time between sampling was increased to 1 h and then 2.5 h. For all other assays cultures were grown in triplicate to an optical density at 600 nm of ~0.1. Prior to reading fluorescence, samples were washed by centrifugation of 1 mL aliquots of culture for 6 minutes at 6,000 x g followed by resuspension of cell pellets in 1× PBS Calcium Magnesium (Invitrogen, Carlsbad, CA). Fluorescence was read with the use of a Tecan Safire plate reader (Tecan Group Ltd., Männedorf, Switzerland). The follow excitation/emission wavelengths were used for fluorescence reads: BFP (383 nm / 447 nm), RFP (550 nm / 585 nm), CFP (443 nm / 473 nm), YFP (512 nm / 532 nm). Normalized fluorescence was calculated by dividing RFP and YFP signals respectively by BFP and CFP signals. Detailed recipes of growth media used in this study are provided in Supplementary Table 3. In all cases we assessed fluorescence of biological replicates in triplicate and reported error bars as standard deviation.

2.2.4 Immuno-blot assay

Samples were grown at higher volume (50 mL and 200 mL for detection of proteins off plasmid and chromosome, respectively) in duplicate but otherwise treated identically to those for fluorescence assay. Bacterial cultures were concentrated to a final volume of 2 mL and lysed by sonication with ultrasonic liquid processor (Microson). Protein concentration was determined with Bradford Reagent (Sigma Aldrich, Oakville, ON) and diluted two-fold with SDS-loading buffer (SDS (4%), bromophenol blue (0.2%), glycerol (20%), DTT (125 mM)) prior to separation by SDS-PAGE (11%). Protein bands

were visualized by chemiluminescent immunoblot with anti-groEL (Abcam, Toronto, ON), anti-GFP (Life Technologies, Burlington, ON) primary antibodies and horseradish peroxidise secondary anti-body (Jackson's Labs, West Grove, PA). They were further quantified by calculation of integrated density, and GFP values were normalized to their respective GroEL loading control.

2.2.5 Quantitative RT-PCR

E. coli cultures for RNA extraction were grown identically as for fluorescence assays (note 25 ng/mL ATC was used to induce the bglGF expression while 100 ng/ ATC was used for the induced expression of other genes. Fluorescence signal was read from samples used for RNA extraction to ensure it was consistent with pervious results. Total RNA was extracted using the Aurum Total RNA Mini Kit (Bio-Rad, Mississauga, ON). The sequences of primers used in the qPCR analysis can be found in Supplementary Table 1. cDNA was synthesized with the use of SuperScript III Reverse Transcriptase (Life Technologies, Burlington, ON). qPCR was conducted with a CFX96 Real-Time PCR Detection System (Bio-Rad, Mississauga, ON) with the use of EvaGreen fluorescent dye (Biotium, Hayward, CA) and Taq DNA Polymerase (Thermo Scientific, Pittsburgh, PA). A PCR program consisting of 94°C for 240 s followed by 40 cycles of 94°C for 40 s, 50°C for 40 s and 72°C for 45 s was used. The resulting data was analyzed with the use of Bio-Rad CFX Manager software. *E. coli* cysG was amplified as a reference gene and used to normalize threshold crossing values for transcripts. CFP-coding mRNA was additionally used to normalize threshold crossing values for YFP transcripts. Relative quantization of RNA was determined using the standard curve method. All samples were

run with 3 biological replicates and 2 qPCR technical replicates. Values were reported as the average of all replicates with error as standard deviation.

2.3 RESULTS

2.3.1 Reporter characterization and selection

We first created reporter plasmids to test the signal-to-background ratio of the following five fluorescent proteins: blue (azurite),[3] cyan (SCFP3A),[4] green (SGFP),[5] yellow (SYPF2)[4] and red (dsRed(EC))[6] fluorescent protein (abbreviated as BFP, CFP, GFP, YFP and RFP, respectively). *E. coli* cells were transformed with these plasmids, along with an empty vector (EV) as a control (which was used to measure the background fluorescence at relevant excitation and emission wavelengths), and were allowed to grow in M9, a chemically defined growth medium. The signal-to-background ratio (the average of three replicates with errors as standard deviation) observed were: BFP/EV = 170 ± 6 (excitation/emission wavelengths = 383/447 nm), CFP/EV = 587 ± 16 (443/473 nm), GFP/EV = 1768 ± 80 (488/517 nm), YFP/EV = 3361 ± 47 (512/532 nm), and RFP/EV = 13349 ± 6674 (550/580 nm; note that the relatively larger error in RFP signal-to-background results from small but varying EV red signal). It is clear that the longer the emission wavelength of a fluorescent protein, the better the signal-to-background ratio.

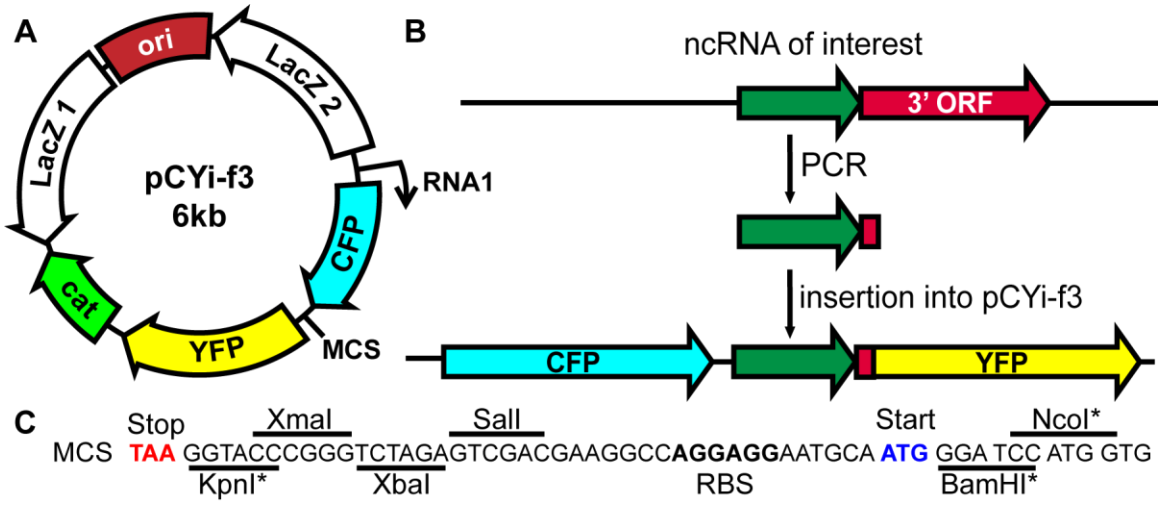

Figure 2-1 A dual reporter system. (A) Reporter plasmid layout. pCYi-f3 contains YFP and CFP coding genes under the control of the RNA1 promoter, a pACYC origin of replication (ori), a chloramphenicol resistance cassette (cat) and flanking *lacZ* homologous regions (*lacZ* 1 and 2; for possible integration at the *lacZ* locus in *E. coli* K12) and a multiple cloning site (MCS). (B) Non-coding RNA insertion scheme. The DNA fragment of interest plus a portion of the downstream 3′ ORF is amplified by PCR and inserted into pCYi-f3, leading to a translational fusion to the YFP gene. (C) MCS. Sites used for cloning in this study are marked with an *.

Two pairs of fluorescent proteins, namely BFP and RFP (BR), CFP and YFP (CY), were then selected to construct a dual reporter system. The plasmid containing CY is shown in Figure 2-1A. Each protein pair was positioned in tandem downstream of the constitutive RNA1 promoter[7] to allow for their transcription to a single RNA transcript (Figure 1-1B). Two transcriptional terminators, rrnb T1 and rrnb T2, were placed after the second fluorescent protein gene to provide efficient transcription termination. Additionally, a multiple cloning site (MCS) (Figure 2-1C) was positioned between the two fluorescent protein genes – this allows for later insertion of DNA sequence encoding RNA to be studied (Figure S2-1). The first proteins in the pairs – BFP and CFP – were to

be used for signal normalization and the second – RFP and YFP – to measure changes in expression from any inserted RNA coding elements.

To determine which fluorescent protein pair possessed the least variable normalized fluorescence under different growth conditions, E. coli cells carrying CY or BR pairs were cultured in both LB (nutrient rich) and M9 (nutrient deficient) media for 12 h. Normalized fluorescence was calculated by division of the fluorescent signals of YFP and RFP by CFP and BFP, respectively. For the CY pair, the normalized fluorescence was observed to increase within the first 4 h of cell growth and stay constant between 4-12 h (Figure 2-2A). The trend appeared to be independent of media conditions, as an almost identical pattern was observed for cells grown both in LB and M9 media. In sharp contrast, the normalized fluorescence of cells carrying the BR pair varied dramatically during the 12-h culturing period (Figure 2-2B). The CY pair was selected for further use due to its consistent signal output. The BR pair was deemed unsuitable as its signal output was strongly affected by choice of growth media, which may reflect differences in maturation properties of BFP and RFP.

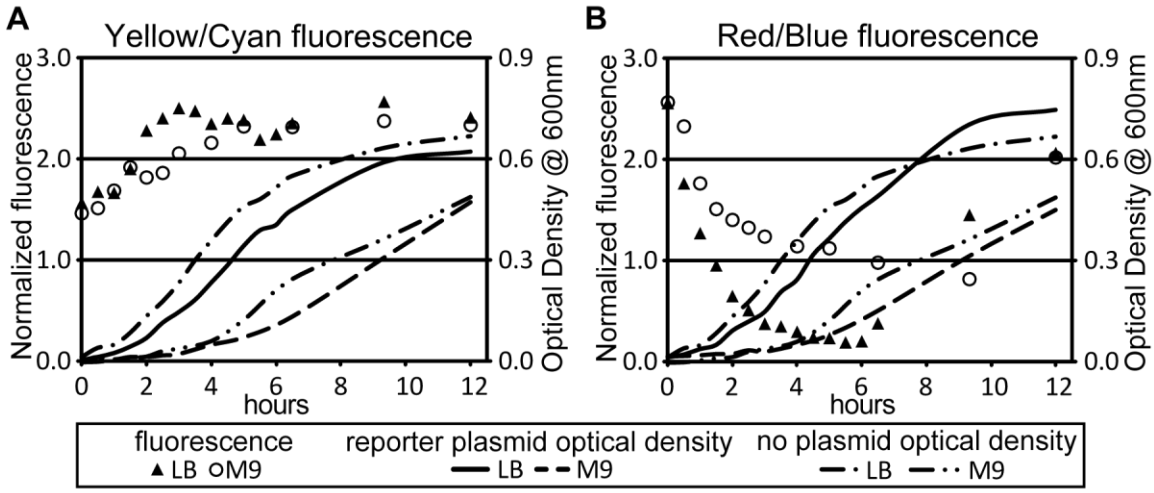

Figure 2-2 Comparison of signal development, in cells carrying (A) CFP and YFP, and (B) BFP and RFP pairs. Fluorescence intensities of *E. coli* cells were measured during 12 h culturing in LB (filled triangles) or M9 (open circles) and shown on left y-axis. The optical densities of these cultures (reporter plasmid) and wild type *E. coli* (no plasmid) are displayed for LB and M9 on the right y-axis.

2.3.2 GlmZ regulatory system

Upon establishing CY as a suitable fluorescent protein pair, we sought to investigate its utility for detecting RNA regulatory interactions in cells. We chose three distinct RNA regulatory systems based on their respective differences. The first two, GlmZ and GcvB, are two small RNA (sRNA) molecules that are known to be involved in translational regulation by RNA-RNA interactions, but diverge in that GlmZ positively regulates a bicistronic transcript though interaction with its intergenic region, while GcvB negatively regulates targeted mRNAs by binding to their 5′-UTRs. The third system, bglG/F, differs from the aforementioned ones by regulating transcription through protein-RNA interaction. In all cases, an sRNA control was provided through the use of plasmid

pRNY, which, upon induction with tetracycline, produces a 128-nucleotide RNA that does not interact with any targeted RNA molecule.

The first RNA investigated was GlmZ, a sRNA known to activate the expression of *glmS* that encodes glucosamine 6-phosphate synthase in *E. coli*. The *glmS* gene resides in the bicistronic *glmUS* operon. Upon transcription, the *glmUS* mRNA is cleaved at the *glmU* stop codon by RNase E, resulting in the monocistronic *glmU* and *glmS* mRNAs. GlmZ has been found to promote *glmS* mRNA translation by binding to the intergenic region of *glmUS* (Kalamorz et al., 2007; Urban and Vogel, 2008).

To detect the interaction between GlmZ and *glmUS* in *E. coli* cells, we designed two separate plasmids, pRNY-glmZ and pCYi-glmUS. The *glmZ* gene was placed downstream of a tetracycline-inducible promoter in pRNY-glmZ so that the production of GlmZ can be controlled by anhydrotetracycline (ATC) (Figure 2-3A). Two versions of pCYi-glmUS plasmids, pCYi-glmUS and pCYi-glmUS+7AA, were made. pCYi-glmUS contained the intergenic region of *glmUS* while pCYi-glmUS+7AA had both the intergenic region and first 7 codons of the *glmS* gene (for the consideration that some nucleotides at the beginning of the coding sequence might be involved in the GlmZ-glmUS interaction). The intergenic region was placed in front of the *yfp* gene so that the expression of *yfp* was linked to the GlmZ-*glmUS* interaction, which was under the control of ATC. A third plasmid, pCYi-f3, was used as a control, which contained two fluorescent protein genes but lacked the intergenic region insert.

As expected, the relative expression of YFP to CFP in the cells carrying pCYi-f3 showed no dependence on ATC; in contrast, the YFP signal experienced 2-fold and 4.5-fold increases in cells containing pCYi-glmUS and pCYi-glmUS+7AA, respectively (Figure 3B). The higher value for the pCYi-glmUS+7AA cell line strongly suggests that the first 7 codons play an important role in the regulatory mechanism. The 4.5-fold change with this cell line is highly consistent with a published value (~5-fold) determined using a glmS-lacZ fusion on the chromosome (Kalamorz et al., 2007).

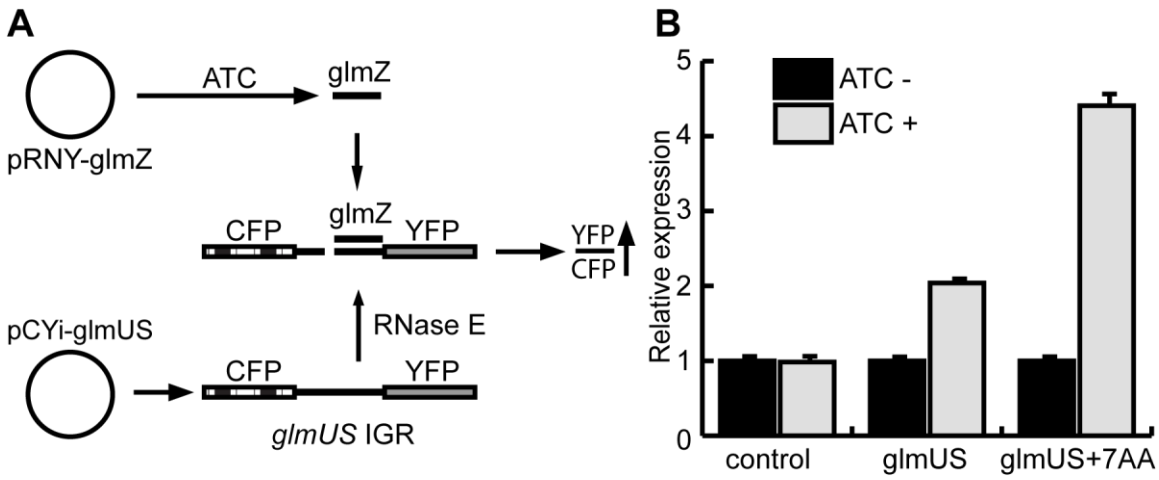

Figure 2-3 Detecting GlmZ-*glmUS* interaction. (A) The two-plasmid design. ATC induces the expression of GlmZ from pRNY-glmZ, which interacts with the IGR of the glmUS mRNA constitutively expressed from pCYi-glmUS (illustrated) or pCYi-glmUS+7AA (not shown), causing enhanced expression of YFP. (B) Relative YFP expression of *E. coli* cells carrying pRNY-glmZ along with either pCYi-f3 (a control plasmid), pCYi-glmUS or pCYi-glmUS+7AA. Samples were assayed in the absence or presence of ATC (black and gray bars, respectively).

2.3.3 GcvB regulatory system

We next applied the same approach to examine the RNA-RNA interactions involving the GcvB sRNA. GcvB has been shown to regulate a number of genes in *E. coli* and *Salmonella* species, such as *oppA* and *gltI*, that are involved in the biosynthesis and transport of amino acids and peptides (Sharma et al., 2011; Urbanowski et al., 2000). GcvB binds the *oppA* mRNA at the 5′-UTR and prevents translation by occluding ribosomes from accessing the ribosomal binding site (RBS); GcvB is known to bind the *gltI* mRNA at a putative translation enhancer located 45 nucleotides upstream of the start codon (Sharma et al., 2007).

To detect GcvB-*oppA* and GcvB-*gltI* RNA-RNA interactions in *E. coli* cells, we again exploited the two-plasmid approach. The first plasmid, pRNY-gcvB, contained the *E. coli gcvB* gene inserted downstream of the tetracycline inducible promoter. The second plasmid, pCYi-oppA or pCYi-gltI, contained the DNA sequence encoding both the 5′-UTR and the first 7 codons of the *E. coli oppA* or *gltI* gene inserted between the two reporter genes. Upon induction with ATC, GcvB should be expressed and interact with each target RNA sequence upstream of the *yfp* mRNA, thereby decreasing its expression (Figure 2-4A). This prediction was confirmed by fold changes in signal of 0.42 for pCYi-gltI and 0.59 for pCYi-oppA (Figure 2-4B). These results are comparable to published values of 0.55 (gltI) and 0.36 (oppA), reported by Sharma et al (Sharma et al., 2007) with the use of a system employing plasmid based gcvB overexpression and signal detection by Western blot.

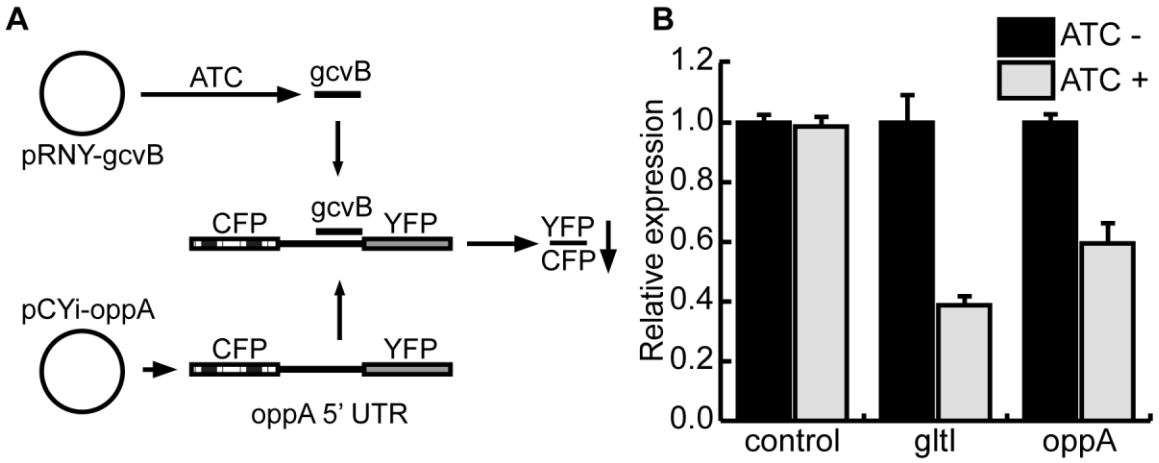

Figure 2-4 Detecting GcvB mediated RNA-RNA interactions. (A) The two-plasmid design. ATC induces expression of GcvB from pRNY-gcvB, which interacts with the 5′ UTR of the *oppA* or *gltI* mRNA expressed constitutively from pCYi-oppA (illustrated) or pCYi-gltI (not shown), causing the reduced expression of YFP. (B) Relative YFP expression of three *E. coli* strains carrying pRNY-gcvB as well as pCYi-f3 (control), pCYi-gltI (*gltI*) or pCYi-oppA (*oppA*), in the absence or presence of ATC (black and gray bars, respectively).

2.3.4 BglGF regulatory system

In the bglGFB operon, there exists an RNA regulatory region within the bglGF-intergenic region (Amster-Choder et al., 1992, 2005). Here, the RNA forms a hairpin that prematurely terminates transcription of the *bglBFG* operon. However, when bound by the *bglG* gene product (BglG protein) the ability of this structure to function as a transcription terminator is diminished, increasing expression of genes within the operon. BglG itself exists in two states: an active dephosphorylated state that binds to the mRNA and an inactive phosphorylated state that does not. The equilibrium between active/inactive BglG is dependent on the import of β-glucosides into the cell, with uptake favoring the active form of the protein. These sugar analogs are transported into the cell and metabolised by

the other two gene products of the operon, BglF and BglB, respectively (Schnetz et al., 1987). Thus, β-glucosides activate expression of genes within the bglGFB operon, which are otherwise repressed by the protein binding RNA element.

The *bglGFB* operon is considered cryptic: it is never naturally expressed in *E. coli* (Defez and De Felice, 1981; Prasad and Schaefler, 1974, Reynolds et al., 1981). Thus, to enable expression of BglG and BglF in order to study its regulatory mechanism, we isolated *bglG* and *bglF* genes and inserted them downstream of the tetracycline-inducible promoter in pRNY, creating the plasmid pRNY-bglGF. Next we amplified the *bglGF* intergenic region and inserted it into the CY reporter system to create plasmid pCYi-bglGF. The pCYi-bglGF vector was then co-transformed with either pRNY-bglGF or pRNY in *E. coli MG1655*.

Upon induction with ATC, the strain containing pCYi-bglGF and pRNY-bglGF should respond to the presence of β-glucosides by a decrease in transcriptional termination downstream of the bglGF intergenic region, resulting in higher YFP expression (Figure 2-5A). This was confirmed by observing an 8-fold change in YFP signal with the addition of both ATC and salicin, consistent with a previous study that has shown 90% reduction in transcript termination under these conditions. (Amster-Choder and Wright, 1992). A small 2-fold change was observed with salicin but without the addition of ATC, and no significant change was observed for bacteria when plasmids pRNY and pCYi-f3 were used as negative controls (Figure 2-5B).

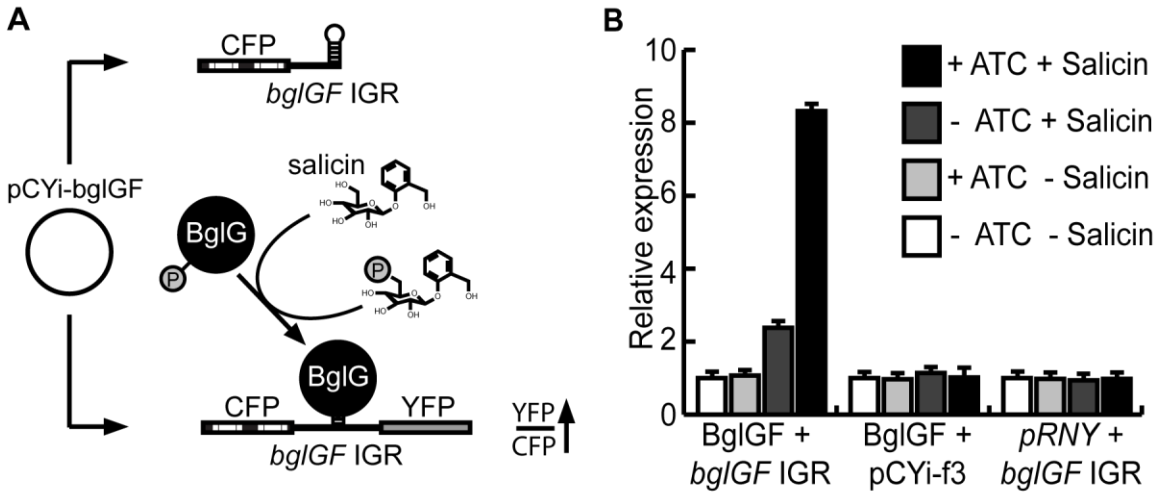

Figure 2-5 *bglGF* regulatory system. (A) Reporter construct containing *bglGF* intergenic region (IGR) expressed from plasmid pCYi-bglGF. In the presence of salicin, BglG is dephosphorylated and blocks transcriptional termination at bglGF. (B) Relative expression for plasmids pCYi-bglGF and pRNY-bglGF. bglGF-IGR and BglGF correspond to presence of plasmids pCYi-bglGF and pRNY-bglGF, respectively.

2.3.5 Endogenous GcvB response

The feasibility of using the reporter construct to monitor low level endogenous expression was examined using the GcvB system. Vectors pCYi-oppA, pCYi-gltI and pCYi-f3 (as a control) were integrated into the E. coli chromosome, using the Warner-Datsenko method (Datsenko and Wanner, 2000). This was done to decrease the copy number, as the copy number of endogenous regulatory targets is likely to be lower than those expressed from plasmid constructs. After integration, the ratio of fluorescence of both reporters was maintained (Figure S2-2A). Fluorescence from the integrated control vector was 20-fold over background for CFP and 203-fold over background for YFP (Figure S2-2B).

Fluorescence was monitored in cultures grown in M9, M9 plus glycine, M9 plus all supplements and LB. It was observed that media complexity was inversely proportional to YFP signal (Figure 2-6). LB showed the greatest decrease in accordance to previous results (Pulvermacher et al., 2009). We were able to observe effects for both *oppA* and *gltI* translational fusions. Though the change in the *oppA* fusion was rather small at 0.66-fold, the *gltI* one was more significant at 0.5-fold.

To further investigate the relationship between the GcvB sRNA and the observed molecular events, we created a *gcvB* deletion in the pCYi-gltI integrated strain with the use of the previously mentioned Warner-Datsenko method. If GcvB were entirely responsible for the observed relationship between growth media and expression from the fluorescent construct, we would expect the *gcvB* deletion to produce equal expression with all media. Interestingly, while the response of this strain to growth media was significantly decreased, relative expression in LB and rich defined media were still noticeably lower than that in M9 (Figure 2-6). We suspect this reduced level of signal results from the presence of additional regulatory effector(s) within the *gltI* 5′-UTR.

We found that the values obtained when monitoring low level endogenous expression are highly dependent on the stage of cell growth they are read at. To ensure consistent and reproducible results, samples must be read during early exponential growth. This phase of growth corresponded to OD_{600} of less than 0.1 for M9 and M9 plus glycine samples, and less than 0.4 for LB and rich media.

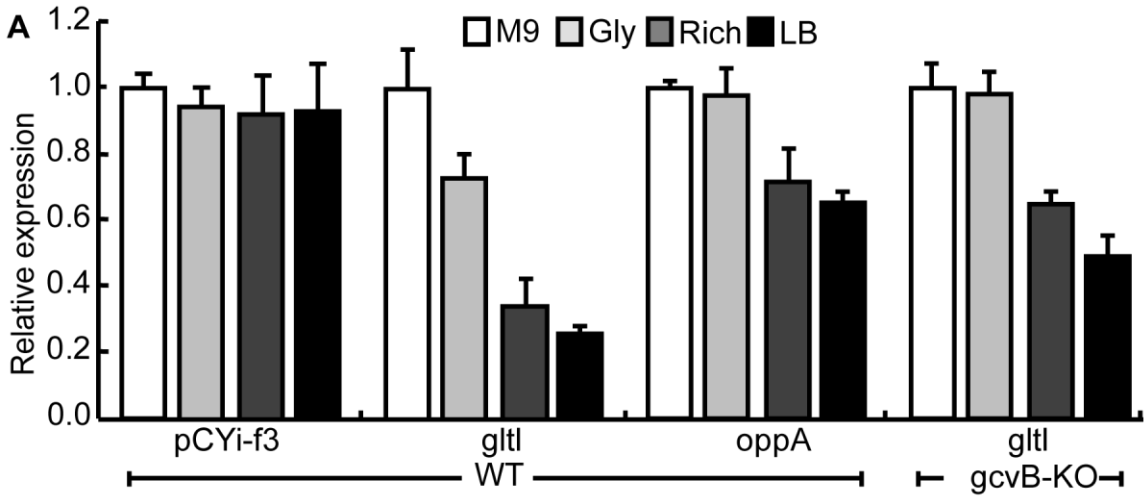

Figure 2-6 Monitoring endogenous regulatory response. Fluorescence was read with the following media: M9, M9 plus 0.2% glycine (Gly), M9 plus all supplements (Rich) and LB. (A) Response of genomic integrants to various culture media. oppA, gltI and pCYi-f3 are respectively pCYi-f3, pCYi-oppA, pCYi-gltI constructs integrated into *E. coli* chromosome at lacZ. (B) WT: *E.* coli MG1655 chromosome; gcvB-KO: *MG1655* with a gcvB deletion.

2.3.5 Comparison of normalization methods

When comparing normalization by CFP to that by optical density we observed the following. Our fluorescent protein approach produced significantly lower average standard deviations, 5% vs 15%, in two of our experiments (Figure S2-3). Additionally in two other cases, normalization by optical density resulted in significant artefacts that were not present with our approach (Figures S2-4 and S2-5). These artefacts were a result of cell densities falling below the linear range of optical density. This highlights an advantage of fluorescence normalization over optical density when working with cells at low cell counts.

2.3.7 Western Blot analysis

Because the cell densities of select samples in our bglGF experiments were too low to allow for normalization by optical density, we sought to verify the results we observed with the reporter system through protein detection via Western blot. Samples with the addition of both salicin and ATC had greatly reduced growth rates, presumably due to high level expression of BglG and BglF. The cell densities of these samples at the time of reading were below our target OD_{600} range of 0.05 to 1.0. This resulted in artefacts when YFP fluorescence was normalized by optical density (Figure S2-3).

Western Blot analysis presented a unique challenge with our system, due to the shared origin of CFP and YFP they are indistinguishable by immune-blot. To overcome this problem, we established a baseline for CFP expression with the use of control plasmid pCYi-ter, which does not produce any YFP fluorescence because there is a transcriptional terminator upstream of YFP. Upon induction with ATC and salicin, we observed a statistically significant increase in total fluorescent protein (Figure S2-6A) in cells containing both pCYI-bglGF and pRNY-bglGF but not in cells containing either pCYi-ter or pCYi-f3 control plasmids. This is consistent with our observations made with fluorescent protein reporters.

We also wished to verify the observed signals for chromosomally integrated pCYi-gltI, as these represent novel results. To this end we assayed our cell line with pCYi-gltI integrated into the chromosome when grown in LB, M9 and M9 plus glycine media. We used a cell line with pCYi-f3 integrated into the chromosome as a negative

control. We were able to detect a significant difference when pCYi-f3 grown in M9 and LB media (Figure S2-6B), consistent with our previous results made with the fluorescent protein reporter system. It should be noted that while we observed a decrease when this cell line was grown with the addition of glycine to M9 media, we do not consider it significant due to the high error in these values. No significant difference was observed with the pCYi-f3 control.

2.3.8 Quantitative RT-PCR analysis

For all regulatory systems examined in this study, quantitative RT-PCR analysis was performed to measure changes in the level of RNA expression.

With the addition of ATC, expression of GlmZ, GcvB and bglGF RNAs from plasmids pRNY-GlmZ, pRNY-GcvB and pRNY-bglGF were found to change by 382 ± 99, 211 ± 21 and 15.8 ± 0.9 fold, respectively (Figure S2-7A). The lower fold expression of bglGF can be attributed to the use of less ATC for its induction (25 ng/ml for bglGF vs. 100 ng/ml for GlmZ and GcvB; note: reducing the ATC concentration in the case of bglGF was necessary in order to minimize the negative effect of RNA expression on cell growth so that cell densities suitable for RNA extraction could be achieved).

The level of YFP mRNA from pCYi-glmUS+7AA was observed to change by 1.9 ± 0.3 fold (Figure S2-7B) when GlmZ was induced from plasmid pRNY-glmZ. This observed change is consistent with previous studies showing stabilization of the *glmS* portion of the *glmUS* transcript (which resides within the YFP mRNA) by GlmZ binding

(Kalamorz et al., 2007; Urban and Vogel, 2008). This value is about 2.5-fold lower than the 4.5-fold change in the YFP expression observed by fluorescence measurement (Figure 3B), suggesting that the remainder of the observed increases in the YFP expression occurs at the translational level. No significant change was observed when the control plasmid pCYi-f3 was used in place of pCYi-glmUS.

Cells containing the pCYi-gltI and pRNY-gcvB plasmids did not produce a significant change (1.1 ± 0.1 fold) in the level of the YFP mRNA (Figure S2-7B) upon GcvB induction with ATC. This is expected because the regulation of *gltI* by GcvB occurs translationally (Sharma et al., 2007).

Induction of bglGF with ATC and salicin resulted in a 13.5 ± 3.9-fold change in the level of the YFP mRNA from cells containing the plasmids pCYi-bglGF and pRNY-bglGF (Figure S7B). This value is consistent (within the margin of error) with the 8-fold change in the YFP expression observed by fluorescence measurement (Figure 2-5B). When pRNY-bglGF was substituted with control plasmid pRNY, no significant change in expression was observed.

2.4 DISCUSSION

We have demonstrated a strategy where a dual fluorescent reporter is used to monitor intracellular RNA-RNA and protein-RNA interactions in bacterial cells. This sensor can minimize signal fluctuations unrelated to the targeted molecular interactions through the

use of a pair of highly similar fluorescent proteins. We have shown that the sensor is useful for the study of RNA targets that are located within 5'-UTRs or intergenic regions within operons. Our sensor was demonstrated to be functional when binding partners of RNA targets were induced either by artificial constructs or through the endogenous response of cells to metabolites.

Our two-reporter strategy should be easily expanded to the study of RNA-small molecule interactions and is compatible with high-throughput experiments. We intend to pursue the use of our sensor for these purposes in the future.

Chapter 3

Probing regulatory RNA responses to metabolites

3.1 INTRODUCTION

Many RNA regulatory interactions have been identified, however, reported studies have focused primarily on identification of binding partners for candidate RNAs and their regulatory mechanisms. How these interactions respond within the cell to the multitude of possible growth conditions is typically not examined. Only after a potential condition is established as a likely regulatory effector, is it examined for the ability to induce a regulatory response in live cells. This approach ignores the potential for an RNA sequence to respond to multiple effectors or alternatively, for growth conditions to have indirect effects on gene expression. While the growth conditions that effect gene expression may take various forms, most often they occur as the presence or absence of cellular metabolites. Examining the response of RNA sequences to a variety of metabolites has the potential to identify new regulatory interactions in addition to providing a more detailed understanding of how these RNAs function within the cell. In this study, we present the first intracellular high-throughput screen of a library of IGRs designed for ncRNA expression against a collection of metabolites.

RNA regulatory interactions are primarily elucidated through the use of reverse genetics, whereby regulatory ncRNAs are first identified, and then attempts are made to discern the regulatory interactions they are involved in. This approach is limited in bacterial systems as many regulatory ncRNAs lack universal features to assist in their discovery. Instead, methodologies for their identification attempt to utilize non-universal aspects of these RNAs. Bioinformatical analyses of genomic data are well suited for identifying functional ncRNA sequences based on the conservation of their structure and sequence (Livny and Waldor, 2007), while the isolation and sequencing of cellular RNA fragments focus on their size and stability (Sharma and Vogel, 2009). However, even with a combination of these approaches, ncRNAs lacking these characteristics will be overlooked. Additionally, these and other *in silico* and *in vitro* protocols that are typically used for RNA identification and mechanistic study may potentially omit details otherwise gained through intracellular observation.

We have sought a forwards genetics approach, whereby RNA regulatory interactions may be identified by observation of their regulatory outcomes; specifically the effects that regulatory elements within 5′ untranslated regions have on the expression of downstream ORFs intracellularly. In our model organism *E. coli* and bacteria in general, translational regulation typically occurs through the interaction of elements contained within the 5′ UTRs of mRNA. To monitor these regulatory responses we decided to use fluorescent proteins as reporters since they are easily amenable to high-throughput study and may be monitored in live cells. Although fluorescent proteins often have signal issues caused by their maturation rates, we previously showed (in Chapter 2)

that normalization by a second reporter is a viable method to control signal fluctuations when examining ncRNAs.

When constructing our library of ncRNA, we focused on those regions found within polycistronic operons. We hypothesised, with the genetic organization of the *E. coli* genome strongly favoring a compact arrangement of genes (Rogozin et al. 2002), that a set of non-coding IGRs from between genes within operons would be enriched for potential regulatory function. Use of IGRs has the added benefit of closely mirroring the transcriptional organisation of our reporter vector, which expresses any RNA inserted into it within a polycistronic operon. Additionally, as IGRs within operons are part of continuous transcriptional units, we ensure that only transcribed sequences are inserted during cloning. Non-transcribed sequences directly upstream of 5′ UTRs contain promoters that may register as false positives in our analyses. While promoters may also exist in transcribed sequences, their occurrence is not guaranteed. Moreover, we eliminated those IGRs that contain known promoters.

Clones in the IGR library differed in their expression in M9 media by up to ~7000 +/- 1000 fold. This signifies that the 5′ UTRs within these regions act to set basal levels of expression even if they do not function as regulatory elements. Additionally, the large dynamic range covered by these signals, signifies that our method should be capable of studying and identifying regulatory interactions that function at both high and low levels of expression. The importance of these large differences in expression is something that warrants further examination.

The IGR library was screened against a metabolite array (Zliti et al. 2013) consisting of amino acids, nucleobase compounds, enzymatic cofactors and metabolic intermediates. This metabolite collection was designed to target a number of major metabolic pathways with redundancy to assist in the elimination of false positives. To this set we made the addition of metals ions, (Fe^{2+} and Cu^{2}) and Kreb cycle intermediates (2-oxogultarate, malate, and citrate). The screen identified 16 5′ UTRs, three of which are known to regulated downstream genes - *glmS, iscS* and *pyrE* - by previously identified interactions, serving as internal controls in the validation of our method.

The regulatory hits were then further examined, producing kinetic plots of expression under growth conditions that displayed large regulatory effects. This analysis replicated observations from our initial regulatory hits and it also provided evidence for growth rate and growth phase affects on many of expression profiles. In particular the IGR for *nirC* displayed a 15 fold decrease in rich defined media over the course of the assay. Conversely, other IGRs such as that for *pyrE*, showed no significant changes in relation to growth phase. Further investigations will be necessary to address the mechanisms behind these observed effects.

3.2 MATERIALS AND METHODS

3.2.1 Strains and plasmid cloning

All cell culture for molecular cloning was performed with the use of LB (Luria Bertani**)** Broth (Sigma Aldrich, St. Louis, MO, USA). IGRs were inserted into plasmid pCYi-f3 (Figure 2-1A) intergenically between CFP and YFP ORFs (Figure 2-1B), arranged as an artificial operon that is expressed from a single constitutive RNA1 promoter. Cloning of IGRs into plasmid pCYi-f3 was carried out in a 96 well format by the following protocol. DNA oligonucleotide primers (Integrated DNA Technologies, Coralville, IA, USA) were used to amplify the IGRs with Phusion DNA polymerase (New England BIOLABS Inc., Ipswich, MA, USA). PCR products were purified with the use of a PCR filter plate (Millipore, Billerica, MA, USA). The purified inserts and pCYi-f3 plasmid were digested with KpnI and BamHI Fastdigest Enzymes (Fisher Scientific, Waltham, MA, USA), and digestion reactions were inactivated by incubation at 85°C for 15 minutes. The inserts were then ligated to pCYi-f3 with T4 DNA Ligase (Fisher Scientific, Waltham, MA, USA) and transformed into *E. coli DH5αZ1*. For transformations not yielding colonies the procedure was repeated individually for their respective IGRs, with PCR product purified by agarose gel electrophoresis. All clones were verified by Sanger DNA sequencing (MOBIX Lab, McMaster University). To construct three replicate libraries for fluorescence assays, IGR clones in pCYi-f3 were transformed by electroporation into *E. coli MG1655*. All replicate library clones were verified by colony PCR. Fluorescence of the triplicate IGR clones was examined for inconsistencies, and any outliers found were retransformed.

3.2.2 Fluorescence assays and signal normalization

For fluorescence assays, cell cultures were grown for 16 hours in M9 media and then used to inoculate M9 secondary cultures at 1/50 concentration. The secondary cultures were grown to an OD600 of ~0.20 and then used to inoculate assay media at 1/200 concentration in 96 well black untreated microplates with lids (Corning Inc., Corning, NY). The culture plates were incubated at 37°C without shaking at high humidity. For screening of the IGR library at a single time point, the cell cultures were incubated for six hours and fluorescence readings were then repeated every two hours until cyan fluorescence was at least ~10 fold greater than background for the majority of samples. We had previously determined that this occurred at an optical density of ~0.1 (data not shown). Background levels of YFP and CFP fluorescence were taken as their fluorescence signal at a time point of zero. For kinetic assays of fluorescence development, readings were taken every ten minutes for three hours, when cells were grown in M9-All media and every ten minutes for ten hours with all other media. Fluorescence readings were taken with a Tecan Safire plate reader (Tecan Group Ltd., Männedorf, Switzerland), at excitation/emission wavelengths of 443nm/473nm for CFP and 512/532nm for YFP.

With all fluorescence assays, YFP signal was normalized by dividing it by the CFP signal read from the same sample. Reported data for samples in non-kinetic assays are from the earliest reliable reading in log phase growth (a cyan value closest to 7 fold over background). Additionally, readings with a CFP or YFP signals less than 3 fold over background were discarded. Background was not subtracted from reported fluorescent

values, as it was observed to be non-additive. All readings were measured from biological replicates in duplicates with error values represented as average deviation.

3.2.3 Growth media

Media for the metabolite array was prepared in accordance with protocols described by Zliti et al. (Zliti et al. 2013). No compounds were included for which transport mechanisms are not thought to exist. Fe^{2+}, Cu^{2}, 2-oxogultarate, malate and citrate additions to the metabolite array were supplied by the following compounds at the indicated concentrations: $FeCl_2$ 1 mM, $CuCl_2$ 1 µM, 2-oxoglutaric acid 1 mg/mL, malic acid 0.5 mg/mL and citric acid 0.5 mg/mL. Chelex media was prepared by treating M9 media with Chelex 100 resin (Biorad) (Falconer et al., 2014). All chemicals were obtained from Sigma-Aldrich (Oakville, ON).

3.3 RESULTS

3.3.1 Selection of non-coding elements for study

Two data sets of *E. coli MG1655* operons were combined. These lists included experimentally verified operons from the DOOR database (Mao et al., 2009) and computationally predicted operons by Dam and Xu (Dam et al., 2007). Next, to restrict focus to the subset of IGRs found between genes in operons, all monocistronic operons were excluded (Figure 3-1). The resulting list included 3268 genes in 900 operons. From this list, a set of IGRs was generated from non-coding regions between successive genes.

A minimum size cut-off of 30 bp was applied to narrow the list of non-coding regions to those likely to contain regulatory function. This size limit corresponds to approximately twice the typical distance from the beginning of a ribosomal binding site to a start codon. A total of 687 non-coding regions greater than 30 bp were identified. The set of IGRs (Table S3-1) was then annotated for experimentally identified transcription start sites. This included data from the regulonDB database (Salgado et al., 2013) and from Cho and Palsson's examination of transcriptional unit architecture (Cho et al., 2009). Of this list the longest 96 IGRs with a downstream protein coding gene and without evidence for an internal promoter were selected for further study.

Next an algorithm was created to generate primer sequences for the cloning of these IGRs into plasmid pCYi-f3. Primers were designed to amplify the non-coding region plus an additional 9 codons (27 bp) on either end. The length of the primers was initially set so as to have a predicted annealing temperature of 54°C. Annealing temperatures were estimated by summation of 2°C for A and T residues, and 4°C for G and C resides, in their IGR complementary regions. Primer lengths were then adjusted, by extension into or retraction away from the non-coding region, so that residues at their 3′ ends would not be followed by a run of identical residues. A sequence -CTAGTAGGTACC- containing a KpnI restriction site plus an additional 6 nucleotides was appended to the 5′ end of the forward primer, while a similar sequence -ACCATGGATCC- with a BamHI restriction site was appended the 3′ end of the reverse primer. No IGR sequences contained internal BamHI or KpnI restriction sites.

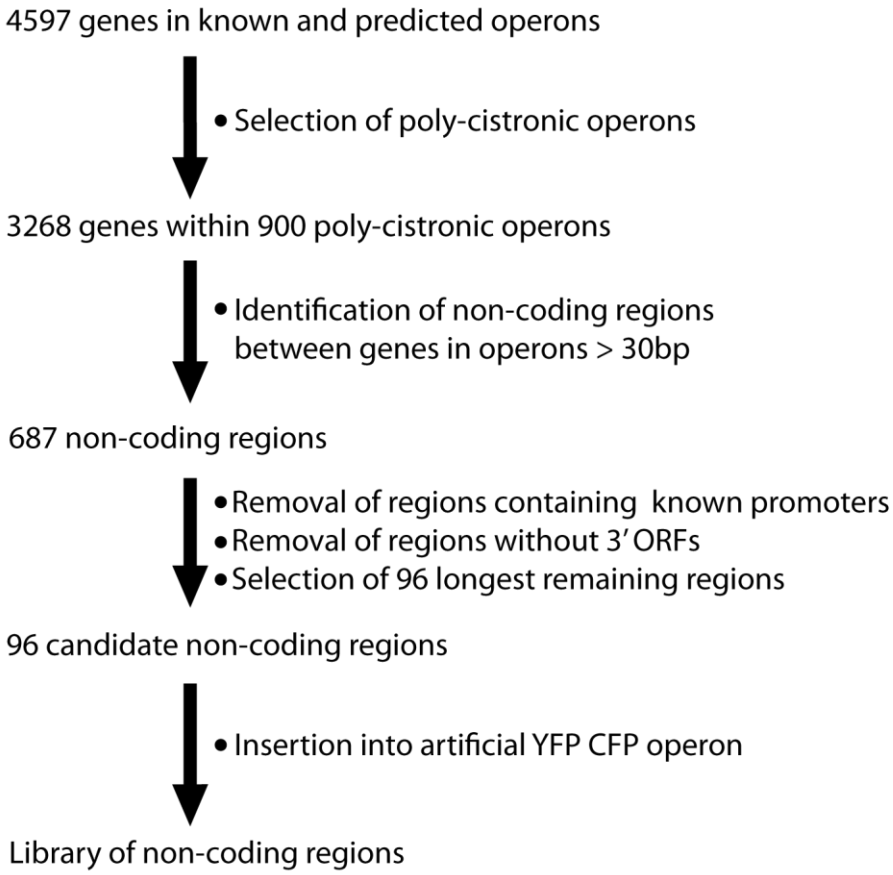

Figure 3-1 IGR selection schematic. This flow-through of the IGR selection process begins with all 4597 genes in the *E. coli MG1655* genome and ends with creation of the IGR library by insertion into plasmid pCYi-f3 that contains an artificial CFP-YFP operon.

3.3.2 Basal expression levels

The effect of the IGRs on the expression of both the CFP and YFP was assessed by examining the fluorescent signal of the individual reporters for the entire IGR library. Fluorescence was read at an OD600 of 0.1 in M9 growth media. Relative expression was plotted against percentage average deviation (Figure 3-2). The clones did not vary significantly with respect to CFP expression. However, YFP fluorescence was observed

to vary from clone to clone up to ~7000 +/- 1000 fold (Figure 3-2); with the lowest expressing clone indistinguishable from cells lacking YFP. Low expression clones were also observed to have higher average deviation than higher expression clones (Figure 3-2). With low basal levels of expression, it would be harder to discriminate between the effects of metabolites. Therefore, to reduce the possibility of false positives during screening, the following 16 low expression clones were omitted from all further analysis: *bglF*, *degS*, *der*, *dnaG*, *frlR*, *ilvX*, *leuA*, *lhr*, *lipB*, *pheA*, *pheS*, *potG*, *prpC*, *speB*, *yfiM* and *ygcS*.

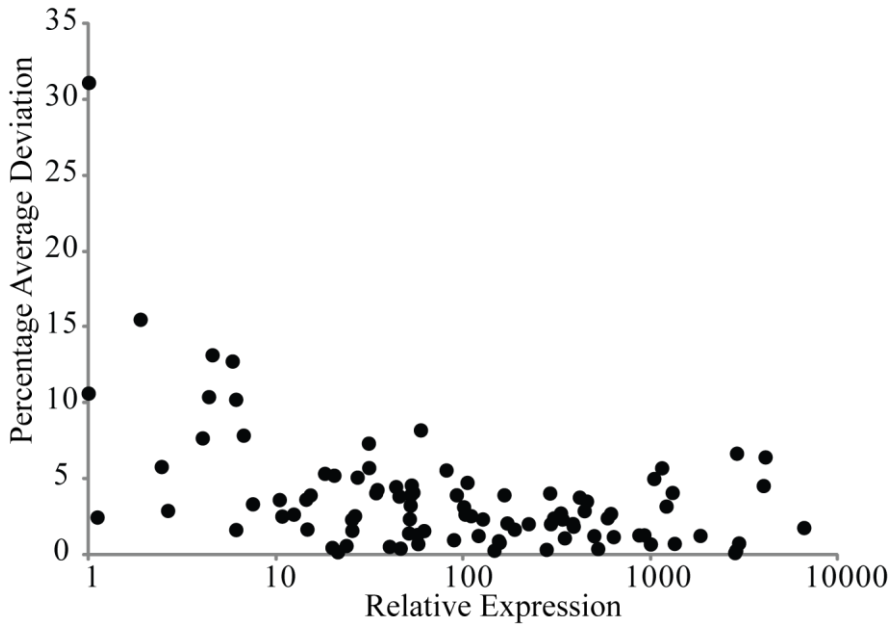

Figure 3-2 Basal expression levels of IGR elements. Expression levels of all IGR library clones in M9 media are presented as the average of two replicates. Expression is shown relative to lowest expressing IGR. Percentage average deviation of the two replicates is presented on the Y-axis.

3.3.3 Metabolite Array Screen

The library of IGRs was screened for regulatory interactions by comparing expression in M9 media to M9 media with the addition of metabolites. Heatmaps of fold change in expression during log growth phase were created for IGRs in the library and used to identify regulatory interactions (Figure 3-3). A minimum cut-off of 2 fold increase/decrease was chosen for samples with metabolite combinations while a 1.5 fold increase/decrease was chosen for single metabolites. A higher cut-off was selected for pools of metabolites, to compensate for greater variation in expression with these samples. The following 16 IGRs were identified as robust regulatory hits: *ahpF*, *chbF*, *clpX*, *dnaJ*, *dppB*, *flgG*, *glmS*, *gltF*, *iscS*, *nirC*, *pstC*, *pyrE*, *rpsA*, *wcaA*, *yafX*, *yfeX*. Three of these IGR contain 5′ UTRs for which previously published RNA based regulatory mechanisms are known (*glmS*, *iscS*, *and pyrE*). All of the hits responded with either increase or decrease in expression to pools of amino acids, with the exception of *gltF*, *iscS*, *pyrE*, *yafX* and *yfeX*. The *gltF* IGR responded with an increase in expression to the addition of many amino acids individually and α-ketoglutarate. The *iscS* IGR responded to the addition of Fe^{2+} or citrate with an increase in expression. Additionally *pyrE*, *yafX* and *yfeX* responded to uracil, with a decrease for *pyrE* and *yafX* and increase for *yfeX*.

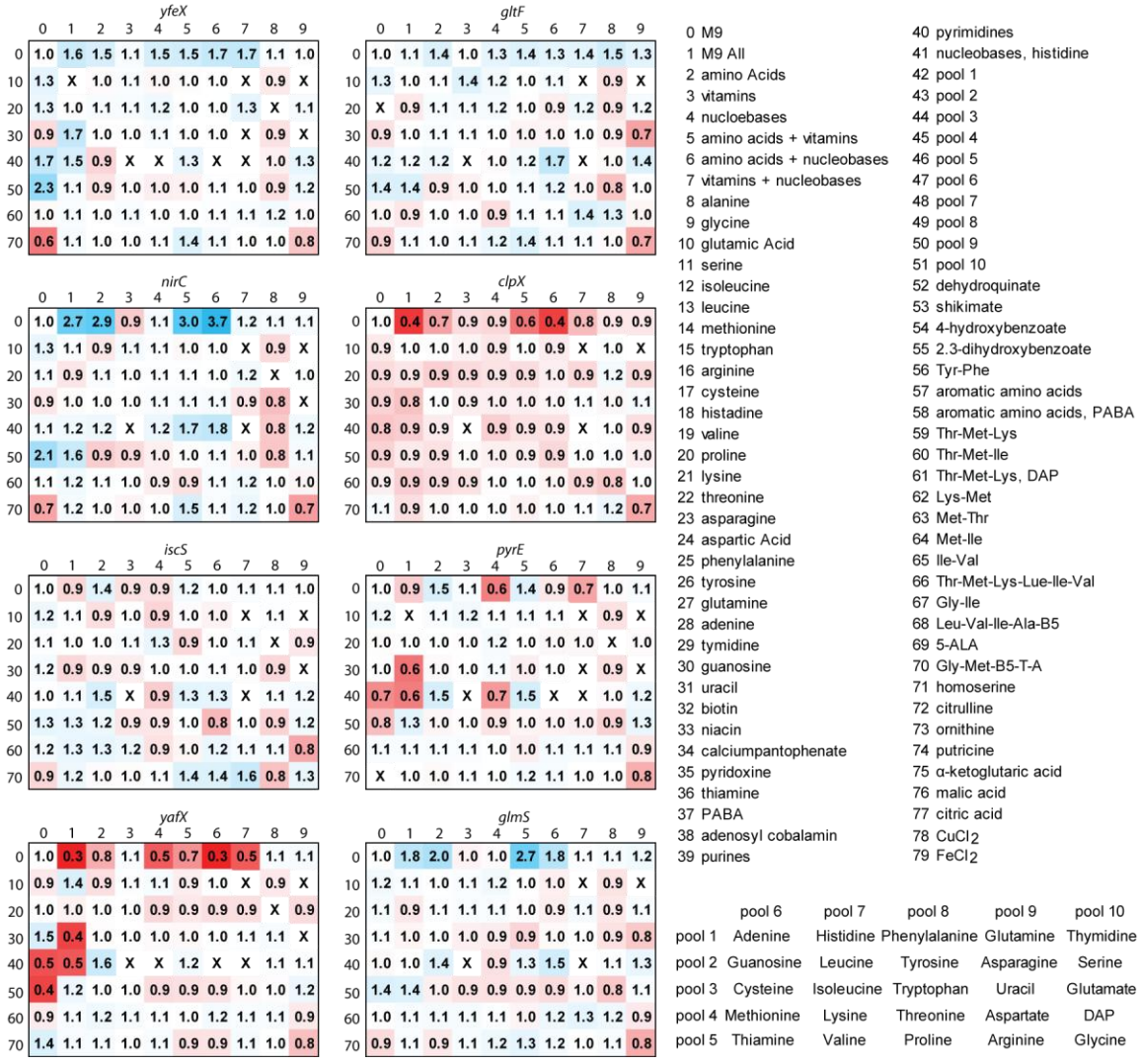

Figure 3-3 Heatmaps of select intergenic region. The name of the gene normally found downstream of each intergenic region is displayed above each heatmap. A list of the media conditions used in the screen is provided to the right of the heat maps, with the composition of metabolite pools 1 to 10 provided below it. The media conditions in individual samples can be found by cross-referencing a summation of heatmap column and row labels to the media condition list. All values are provided relative to M9 media.

3.3.4 Kinetic expression analysis

To demonstrate reproducibility of hits from the primary screen and further examine expression, kinetic expression data were obtained for all cloned 5′ UTRs under the three media conditions that displayed the greatest responses. Expression profiles were obtained for M9 media (M9), M9 with the addition of uracil (M9+Uracil) or M9 all media (M9+All) (Figure 3-4). Additionally, M9 media plus Fe^{2+} was compared to chelex resin treated M9 media to examine the effect of iron depletion on expression (Figure 3-5). Profiles from 5′ UTRs previously identified as low expression clones were excluded from analysis as they had also been for the primary screen.

When comparing M9, M9+Uracil and M9+All, similar fold changes as obtained in the primary screen were observed during early-log phase. For elements showing changes with the addition of M9+All, the following fold expressions were observed in early log phase for M9+All compared to those from the primary screen: *ahpF* 2.3 vs. 1.8, *chbF* 0.25 vs. 0.38, *clpX* 0.65 vs. 0.4, *dnaJ* 2.35 vs. 1.62, *dppB* 2.27 vs. 1.73, *flgG* 1.82 vs. 1.71, *glmS* 3.10 vs. 2.71, *nirC* 8.49 vs. 3.7, *pstC* 1.89 vs. 1.77, *rpsA* 1.94 vs. 1.68 and *wcaA* 0.40 vs. 0.39. While the magnitude of these values differed significantly for certain 5′ UTRs (*nirC* in particular) they were all consistent in either decreasing or increasing in expression. For the hits with M9+uracil, the fold changes during log phase in the kinetic curve also mirror those from the primary screen: *pyrE* 0.78 vs. 0.6, *yafX* 0.40 vs. 0.43, *yfeX* 2.16 vs. 1.72. Although the *pyrE* hit appeared to have less of a change in early log phase, this difference increased later in the growth curve to 0.65. Many expression

profiles demonstrated strong growth phase dependence. In particular, *ahpF*, *dnaJ*, *dppB*, *flgG*, *glmS*, *gltF*, *nirC*, *pstC*, *rpsA* and *yfeX* had profiles where fold changes in expression were highest in early log phase then decreased with time, with expression curves converging to the M9 values (Figure 3-4A). Conversely, *chbF*, *clpX*, *pyrE*, *wcaA* and *yafX* had no or little convergence to M9 values, displaying similar fold changes in expression through the growth profile (Figure 3-4B).

M9 media plus iron was compared to chelex resin treated M9 media to further examine the effect iron may have upon these 5′ UTRs. Although Chelex resin removes most of the essential divalent metal ions, cell growth will still occur typically at a much lower rate, likely due to residual metals. As expected, growth was very slow in the Chelex treated media, with many cultures requiring ~50 hours of growth to reach densities sufficient for fluorescent reads. Readings were taken in duplicates; however, were not averaged due to inconsistent growth rate with duplicates. Relative expression was plotted against CFP signal to normalize for growth phase (Figure 4-5). A minimum cut-off of two standard deviations in relative expression during early log phase was used to identify the following hits, with values reported as fold increase upon the addition of iron: *yafX* 2.67, *cysM* 2.15, *hycB* 2.12, *caiD* 2.09, *glmS* 1.96, *iscS* 1.78, *emrA* 1.68 and *pyrE* 1.65. Three of these 5′ UTRs – *cysM*, *hycB* and *iscS*- are located in front of known iron-sulfur cluster proteins. Kinetic expression curves of these elements show that many of these hits would have higher fold increase if the comparisons between iron and Chelex media were made later in the growth phase; the 1.78 fold increase of *iscS* observed in early log phase increased to 22.6 fold in late log/stationary phase (Figure 3-5).

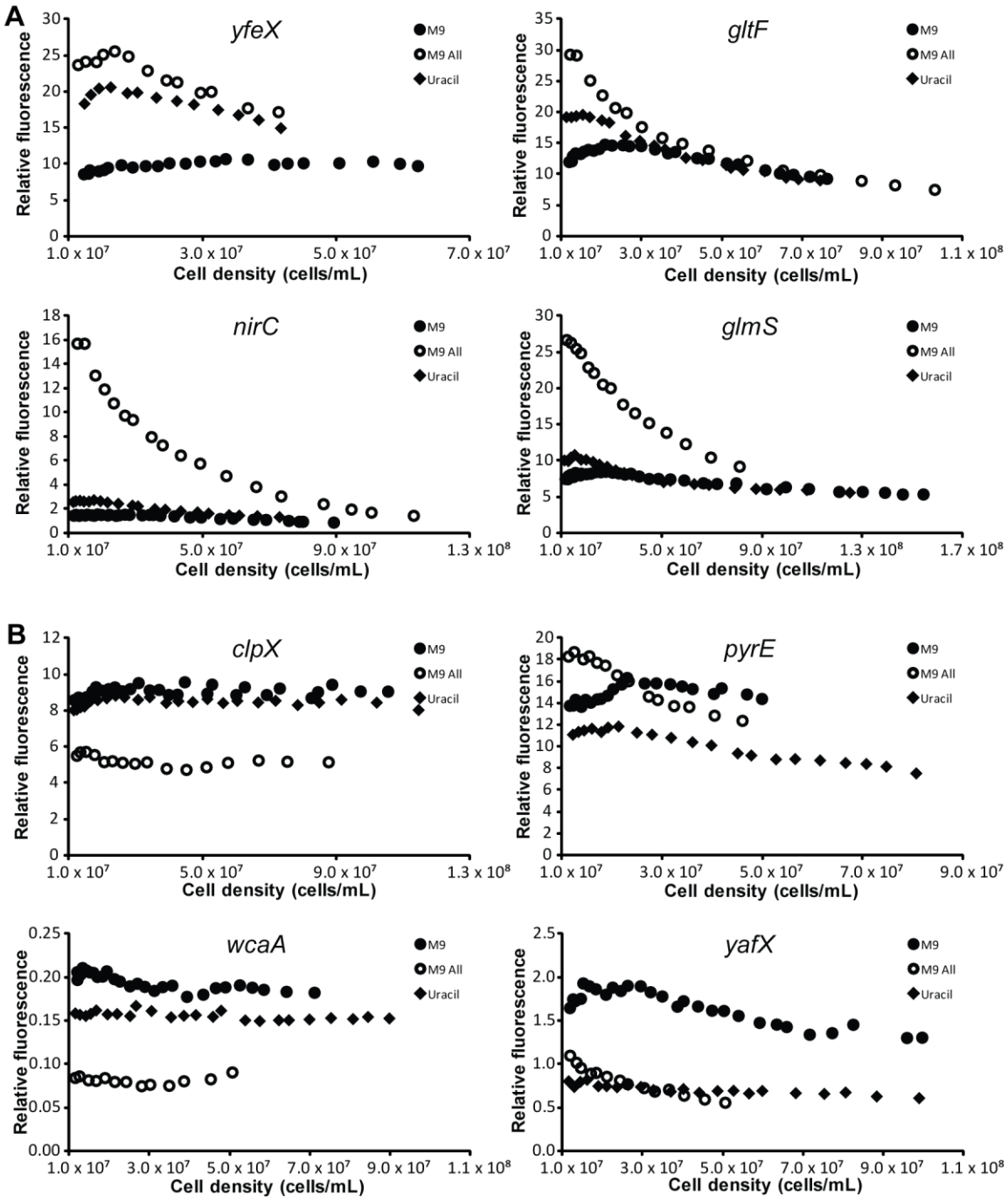

Figure 3-4 Kinetic plots of select intergenic region. The name of the gene normally found downstream of each intergenic region is displayed above each heatmap. Cell densities were calculated from cyan fluorescence signal with the conversion ratio of 1 unit cyan fluorescence signal = ~ 1.6 x 10^4 cells/mL. (A) Kinetic plots in which signal decreases with time. (B) Kinetic plots in which signal is stable with time.

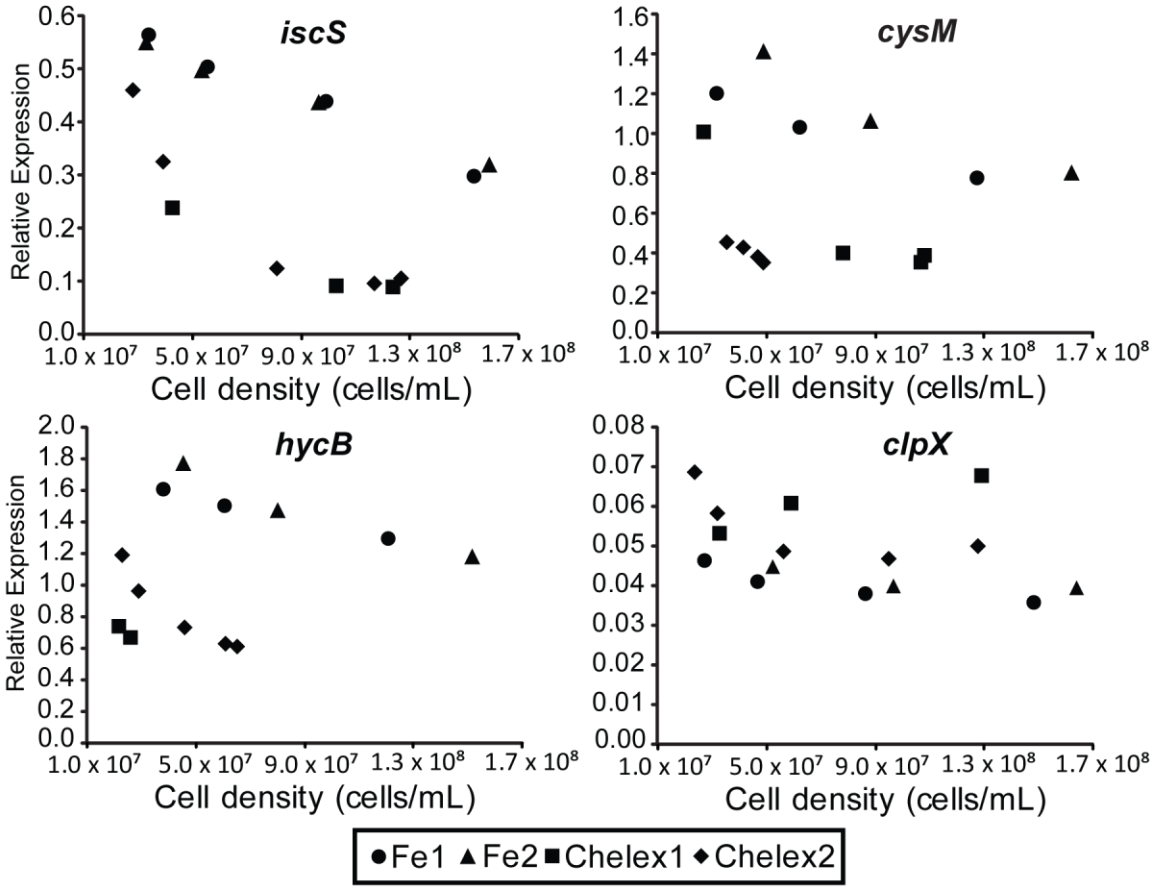

Figure 3-5 Chelex kinetic assays. Kinetic fluorescence reads were performed on *iscS*, *cysM*, *hycB* and *clpX* intergenic regions. M9 media for Fe1 and Fe2 samples had the addition of 1 mM Fe^{2+}, and was treated with Chelex resin for Chelex1 and Chelex2 samples. Fe1, Fe2 and Chelex1 and Chelex2 represent duplicates that were not combined due to inconsistent growth rates in chelex treated media.

3.4 DISCUSSION

With this study, we have been able to demonstrate that many IGRs possess regulatory function. Out of 80 elements, 16 were identified as hits in a multi-metabolite primary screen and eight in an iron depletion screen giving hit rates of 20% and 10%, respectively. Our hits can be divided into specific hits and non-specific hits: specific hits

occurring when growth conditions (in our case addition of small molecules) cause responses by mechanisms that exist specifically for their detection. Conversely, non-specific hits arise when growth conditions alter states of the cell and regulatory responses occur indirectly as a consequence of those states. For example, a growth condition causing a regulatory response in a ncRNA by altering the growth rate of the cell would be considered non-specific.

The majority of our hits fall into the non-specific category and are one of the most useful aspects of our research, providing evidence for previously uncharacterized regulatory functions. These hits may further be examined in later studies to determine their underlying specific interactions. For example, we detected regulation of *glmS* as a non-specific interaction dependent on either nitrogen abundance or growth rate. This IGR is known to respond to the presence of glucosamine-6-phosphate (Glc-6-P) to regulate this gene (Kalamorz et al., 2007; Urban and Vogel, 2008). Thus, we would hypothesize that Glc-6-P abundance is sensitive to nitrogen abundance and/or growth rate. These non-specific regulatory interactions highlight the complexity of regulatory networks and homeostasis within the cell, demonstrating that while study of more specific interactions are invaluable in gaining mechanistic insights, they only encapsulate a fraction of the regulatory responses that occur within cells.

The majority of the observed non-specific hits appear to result from changes in growth rate, nitrogen abundance or growth phase. As these states are not mutually exclusive, it can be difficult to ascertain which state results in the observed regulatory signal. The addition of a nitrogen source will typically increase growth rate; cell growth

rates also changes when entering different phases in a growth curve. Although we attempted to control for growth phase by taking all readings during early log phase (where the majority of elements displayed steady signal), examination of kinetic expression profiles for *gltF*, *nirC*, *glmS* show little if any steady phase signal, as signals vary between different growth phases. Data obtained at single time points for these elements would be highly dependent upon the selection of time points. Of all the IGRs tested, *nirC* displayed the greatest variance in signal, with expression in M9+All over M9 decreasing from 16 to 2 fold between early log and stationary phases. Conversely, other elements possess regulatory profiles that show little dependence on growth phase, making single time point readings for these elements more relevant.

We were also able to detect specific hits for iron and uracil regulatory elements that have both been previously identified and are unique to this screen. The *pyrE* 5′ UTR for which we observed a decrease in expression in the presence of uracil is known to respond to Uridine-5′-triphosphate (UTP) by transcription attenuation (Poulsen et al., 1984). Also, the *iscS* 5′ UTR that showed a decrease in signal in the presence of iron and citrate is known to be a regulatory target of the RyhB sRNA, which functions to decrease its expression under iron limiting conditions (Desnoyers et al., 2009). These hits made for good internal controls for the validation of our methodology. We hypothesized that two additional hits in the iron assay - the *cysM* and *hycB* 5′ UTRs - are regulated in a similar fashion to *pyrE* as these non-coding regions are also upstream of iron-sulfur cluster proteins (Flint et al., 1996; Sauter et al., 1992). Further investigation will be necessary to determine the nature of these effects.

While we were able to identify various hits, our screen did contain a reproducible artifact. Many of the kinetic profiles displayed a gradual decrease in signal during later growth stages. We had determined that this was due to the accumulation of autofluorescence contributing to our CFP signal. For single time point readings this extra signal can be removed by washing the cells prior to readings; however, this additionally processing is not feasible in experiments generating high-throughput kinetic data. A future improvement on our method could involve the replacement of CFP with a reporter not prone to this interference. In section 2.5.1, we demonstrated that fluorescent proteins with longer emission wavelengths have higher fluorescence signal relative to background levels. Substitution of CFP by a fluorescent protein that emits at longer wavelengths may be one wave to solve this problem.

It is possible that certain regulatory interactions were overlooked if expression of our ncRNA was outside the effective range for interaction with a binding partner. This could result from use of a multicopy vector. This can be investigated in the future by screening at a lower copy number, possibly to be achieved by chromosomal integration of our IGR library. However, any attempt to do so would need to be carried out with care to ensure that signalling from the reporters is maintained significantly above background levels.

In this study, the first ever screen of a library of putative regulatory ncRNA elements against a metabolite array was performed. It successfully identified multiple regulatory interactions. A subset of these corresponds to known RNA regulatory mechanisms, while the majority of them are novel and may serve as a starting point for

future studies. The screening methodology presented here provides a simple, flexible and cost effective tool for the discovery of new regulatory interactions.

Chapter 4

Riboswitch Sensors and the Study of Nutrient Substitutions

4.1 INTRODUCTION

In this chapter, additional applications for the technology discussed in Chapters 2 and 3 are explored. First, a screening approach using the IGR library to identify regulatory interactions caused by nutrient substitutions is examined. Previously, the effect of metabolite additions to M9 minimal media on gene expression had been investigated. However, as *E. coli* prioritizes the metabolism of certain nutrients by turning off gene expression for alternative metabolic pathways (Chubukov et al., 2014; Liu et al., 2005), it is conceivable that in the presence of high priority nutrients, genes regulating response to lower priority metabolites will be suppressed. However, this should not be an issue when a metabolite exists as the sole source of nutrition it can provide. For instance when glucose is supplied alongside acetate, *E. coli* preferentially utilizes glucose, shutting down genes involved in the transport and metabolism of acetate (Cozzone, 1998; Oh et al., 2002).

This examination was performed by screening *E. coli* cells containing the IGR library, in media featuring alternative carbon sources, nitrogen sources and varied pH. The screen has identified multiple regulatory responses, including many from the screen discussed in Chapter 3. However in general, this screen identifies larger regulatory responses than were observed in the previous screen, suggesting that many regulatory effects are suppressed by the presence of high priority nutrients in M9 media.

The *gltF* IGR was selected for additional follow-up as it showed the largest regulatory response when comparing any two growth conditions. This element was examined using combinations of different nitrogen and carbon sources. When carbon and nitrogen sources glucose and glutmate were substituted for α-ketoglutarate and ammonia, a ~15 fold increase in expression was observed. This regulatory effect was further examined in *gltD* and *gltF* genomic knockouts. The *gltD* gene was found to be necessary for the regulatory increase resulting from the substitution of α-ketoglutarate for glucose, but not for any other metabolite substitutions.

This chapter also examines the use of the dual reporter plasmid to monitor regulatory interactions from riboswitches. Though many sequences have been identified as potential riboswitches, their study is often limited due to difficulties in identifying and testing ligand interactions (Breaker, 2011; Meyer et al., 2011). A screening methodology enabling the high-throughput identification of riboswitch-ligand interactions would prove invaluable to the study of riboswitches. For this purpose, a sensor in the form of plasmid pCYi-f4 was created by modifying pCYi-f3 to allow for the expression of riboswitches at the 5′ termini of mRNA. This new sensor was able to identify known interactions for *E. coli* lysine (Caron et al., 2012) and vitamin B12 (Nahvi et al., 2004) riboswitches. However when candidate riboswitch sequences were tested the screen failed to identify any novel riboswitch-ligand interactions.

The two components of this chapter seek to expand the uses for methods developed within this thesis. Our approach is shown to be effective in the identification and examination of diverse regulatory interactions. As such, these tools should prove

useful for the study of these regulatory interactions and other areas of gene expression.

4.2 MATERIALS AND METHODS

4.2.1 Library assay procedures

All riboswitch assays were performed using single time point assay procedures with the metabolic suppression array found in section 3.4.2 of this thesis. The alternative metabolism screen of the IGR library was performed with the use of this procedure as well, substituting the appropriate alternative metabolism media. Hits were identified as samples with signal greater than three standard deviations relative to M9. Standard deviations were calculated using fold change values relative to M9, for all samples in that particular media.

The gltF follow-up assay was performed by transforming the gltF clone from the IGR library into *E. coli MG1655, E. coli MG1655 ΔgltD* and *E. coli MG1655 ΔgltF*. These strains were inoculated into growth media at 1/100 concentration, and grown to an optical density of 0.1. Cell pellets were obtained from 1 mL of culture and resuspended in phosphate buffered saline. Fluorescence intensity was read at excitation and emission wavelengths of 443/473nm for CFP and 512/532nm for YFP. Yellow fluorescence was normalized by dividing it by cyan fluorescence. Error bars were reported as standard deviation and all samples were run in triplicate.

4.2.2 Plasmid construction

First, primers CFP-f4-fw and CFP-f4-rv were used to amplify *cfp* and the RNA1

promoter from plasmid pCYi-f3, creating DNA fragment CFPrev. CFPrev was extended with use of primer CFP-extend to insert an additional RNA1 promoter and new MCS, resulting in DNA fragment CFPrev2prom. Next, primers lacZ2-fw and alt-f4-fw were used to linearize the pCYi-f3 plasmid and insert a KpnI restriction site, creating the fragment backboneKpnI. CFPrev2prom and backboneKpnI were digested with KpnI and XbaI restriction enzymes (Fisher Scientific, Waltham, MA) and the digested fragments were ligated together to produce plasmid pCYi-f4. See Table S4-1 for a list of primers used in the creation of pCYi-f4.

4.2.3 Cloning and generation of genomic knockouts

E. coli MG1655 ΔgltD and *E. coli MG1655 ΔgltF* genomic knockouts were constructed with the Warner-Datsenko method (Datsenko and Wanner, 2000) by inserting a kanamycin resistance cassette in place of *gltD* and *gltF* respectively.

Clones for putative and known riboswitches were created by inserting DNA fragments coding for the RNA motifs into pCYi-f4 (Figure 4-1A). The riboswitches were cloned such that transcription would begin from their native transcriptional start sites (Figure 4-1B). DNA fragments were obtained by performing PCR on the *E. coli MG1655* genome for *btuB*, *lysC* and *mini-ykkC* elements. DNA fragments for *plf*, *ykoY*, *crcB*, *ykxD*, *manA*, *glnA* and *ydaO* RNA motifs were obtained by PCR-based fusion of smaller oligonucleotides (Shevchuk et al., 2004). Sequences for these elements were obtained from information on putative riboswitches identified by the Breaker lab (Weinberg et al., 2010). All DNA oligonucleotides (Integrated DNA Technologies, Coralville, IA) were produced by standard phosphoramidite chemistry.

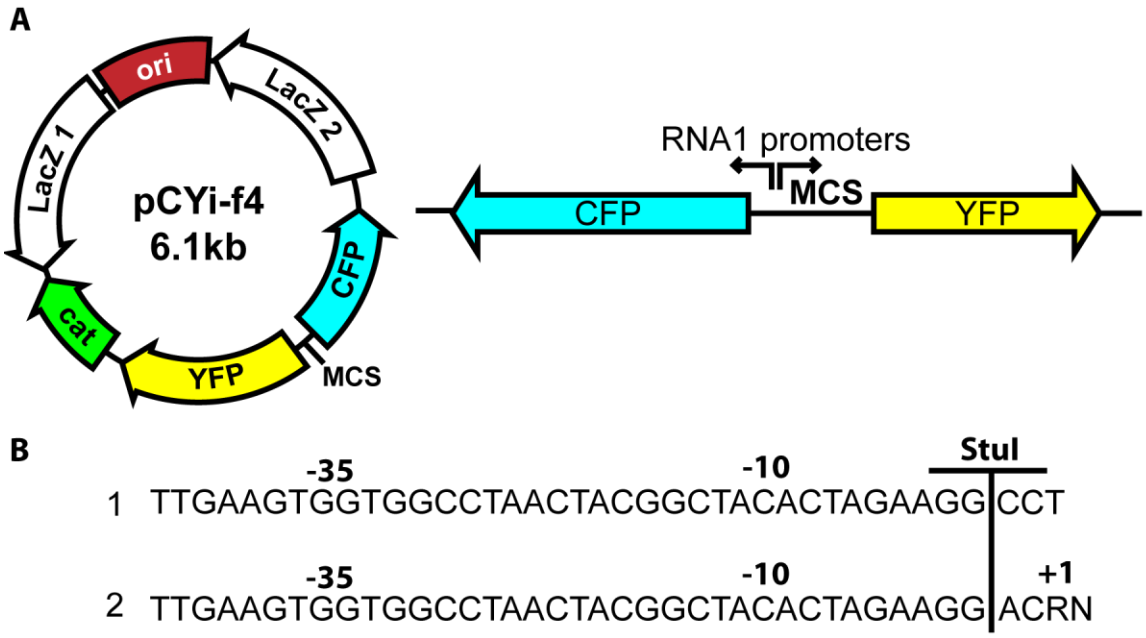

Figure 4-1 pCYi-f4 plasmid diagram and cloning schematic. (A) pCYi-f4 plasmid contains a pACYC origin of replication (*ori*), YFP and CFP coding genes under the control of the RNA1 promoter, a chloramphenicol resistance cassette (*cat*) and lacZ homologous regions (lacZ 1 and 2) and a multiple cloning site (MCS). (B) Non-coding RNA insertion scheme. The RNA1 promoter upstream of YFP is incomplete due to the presence of a StuI restriction site. To insert sequence for transcription downstream of this promoter the sequence must begin with ACR. This will complete the promoter and result in transcription at the indicated transcriptional start site (+1).

4.2.4 Growth media

Metabolite media for riboswitch screens was prepared as described in Section 3.4.3 of this thesis. Recipes for media used in nutrients substitution assays can be found in Table S4-2. All molecules were obtained, unless otherwise stated, from Sigma-Aldrich (Oakville, ON).

4.3 RESULTS

4.3.1 Response of IGRs to alternative carbon and nitrogen sources

The IGR library was screened for regulatory response against alternative metabolism media which features substitutions to basic components of M9 media. α-Ketoglutarate, Malate and mannose were supplied as alternative carbon sources in place of glucose. Glutamate, glycine, arginine and glutamine were used in place of ammonia as nitrogen sources. Finally, CMM 5.5 media that substitutes a citric acid buffer in place of phosphate buffer was used to test gene expression at a pH of 5.5 vs. the typical M9 pH of 7.4. The data was analyzed to produce a heat plot of fold change in regulatory response (Figure 4-2). The eight strongest hits, accompanied here by their respective fold regulatory change values, were as follows: *iscS* 7.8 ± 0.2, *nirC* 6.5 ± 0.3 and *yfjQ* in 6.1 ± 0.2 in CMM5.5, *yadI* 4.1 ± 0.1, *ygcS* 4.1 ± 0.1, *speB* 3.7 ± 0.1 and *psi*F 2.6 ± 0.1 in glutamate, and *gltF* 5.5 ± 0.1 in α-ketoglutarate. The greatest fold change observed between any two growth conditions was for *gltF*, where glycine relative to α-ketoglutarate produced a 15 fold change in expression. All other hits yielded only increases or decreases in expression relative to M9. We chose to focus on *gltF* for follow up experiments as it produced a unique pattern of expression.

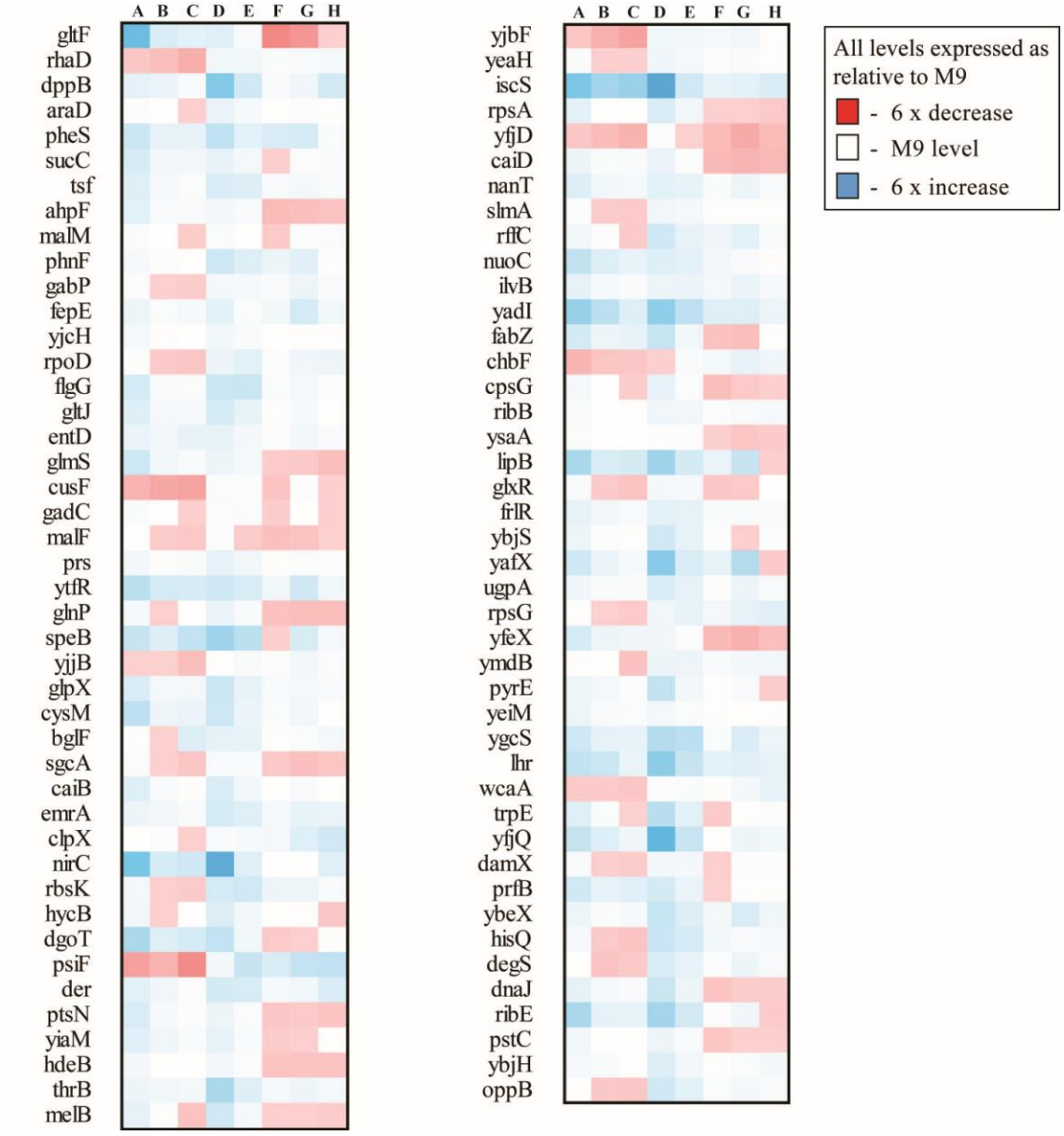

Figure 4-2 Heatmap of alternative metabolism screen. Media labels are as follows: **A** - α-ketoglutarate, **B** – Malate, **C** – Mannose, **D** - CMM pH 5.5, **E** – glutamate, **F** – glycine, **G** – Arginine, **H** – Putricine. All samples are displayed as relative to M9 media. Alternative nitrogen source samples (glutamate, glycine, arginine and putricine) contain no ammonia and 0.4% glucose. Alternative carbon source samples (α-ketoglutarate, mannose, malate) contain 10 mM ammonia and no glucose. CMM pH 5.5 contains 10 mM ammonia and 0.4% glucose in pH 5.5 media.

4.3.2 *gltF* follow-up assay

The *gltDF* IGR was transformed into *E. coli MG1655* wildtype, *ΔgltD* and *ΔgltF* strains. These strains were examined for regulatory action under the following carbon and nitrogen source combinations: glucose and ammonia, α-ketoglutarate and ammonia, glucose and glutamine, glucose and glycine, glucose and glutamate and α-ketoglutarate and glycine (Figure 4-3). An 8.9 ± 0.1 increase in fold expression was observed when carbon and nitrogen sources were respectively changed from glucose and glutamate to α-ketoglutarate and ammonia. The substitution of α-ketoglutarate for glucose as a carbon source resulted in increases in expression in all instances, except in a *ΔgltD* genetic background (Figure 4-3). The deletion of *gltF* had no effect on expression. Glutamine was the only nitrogen source that when substituted for ammonia did not result in a ~7 fold decrease in expression. This was observed for all instances of glutamine substitution (Figure 4-3).

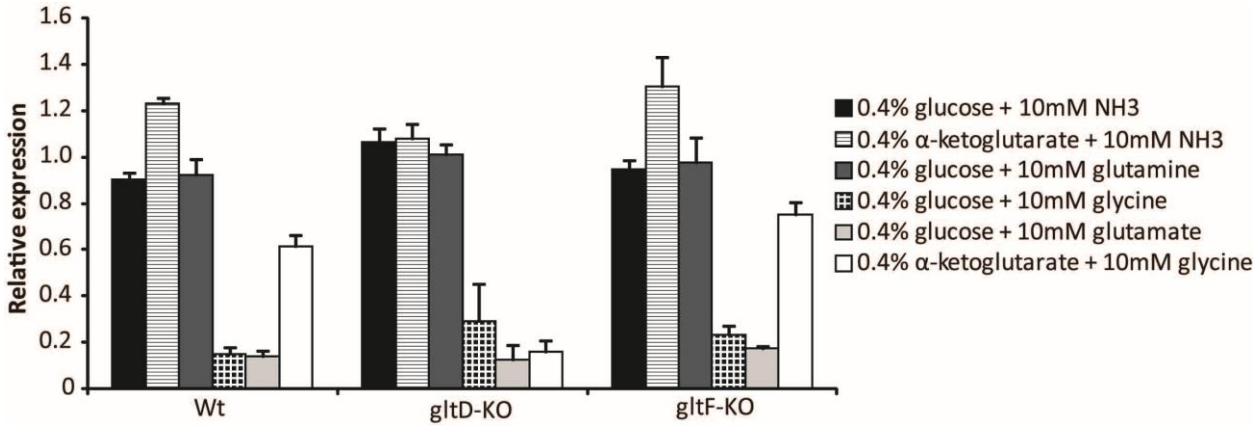

Figure 4-3 *gltF* expression assay. All samples contain the *gltI* clones from the IGR library. Wt corresponds to *E. coli MG155* and gltD-KO and gltF-KO correspond to that same strain with genomic deletions in *gltD* and *gltF* respectively.

4.3.3 Screening known and putative riboswitches

Known and putative riboswitch motifs, *btuB, lysC, plf, ykoY, crcB, ykxD, manA, glnA* and *ydaO* were introduced to plasmid pCYi-f4 and subsequently transformed into *E. coli MG1655*. Each motif was then screened against the metabolite library. The elements *crcB*, *ykxD*, and *glnA* were excluded from further analysis due to low signal output. In addition, *manA* and *ydaO* were also excluded due to the slow growth of strains harboring these plasmids. The poor growth in strains harboring these two elements was a result of extremely high expression of YFP. As the *lysC* riboswitch is known to bind lysine and decrease the expression of downstream genes (Sudarsan et al., 2003), a decrease in YFP signal was expected in the presence of lysine. This was confirmed by an ~3 fold decrease in YFP expression for lysine alone, as well as similar decreases for all other samples containing lysine (Figure 4-4). Similarly, the addition of adenosylcobalamin (vitamin B12) was expected to reduce YFP expression for the *btuB* riboswitch, which has previously been demonstrated to respond to this ligand (Nahvi et al., 2002, 2004). This was confirmed for all samples containing vitamin B12 (Figure 4-4). The sample containing vitamine B12 as the sole metabolite addition resulted in an ~5.7 fold change in expression.

The *ykoY* RNA motif responded to pools of amino acids and nucleobases, demonstrating a ~2-4 fold increase in YFP expression (Figure 4-4). The pfl RNA motif gave rise to a ~2-3 fold decrease in expression in response to purine nucleobases As these conditions are likely to have strong effects on growth rate it is likely that the observed regulatory interactions are non-specific. The only single metabolite addition

that had a strong effect on these two elements is adenine, which is known to be toxic to *E. coli* (Levine and Taylor, 1982).

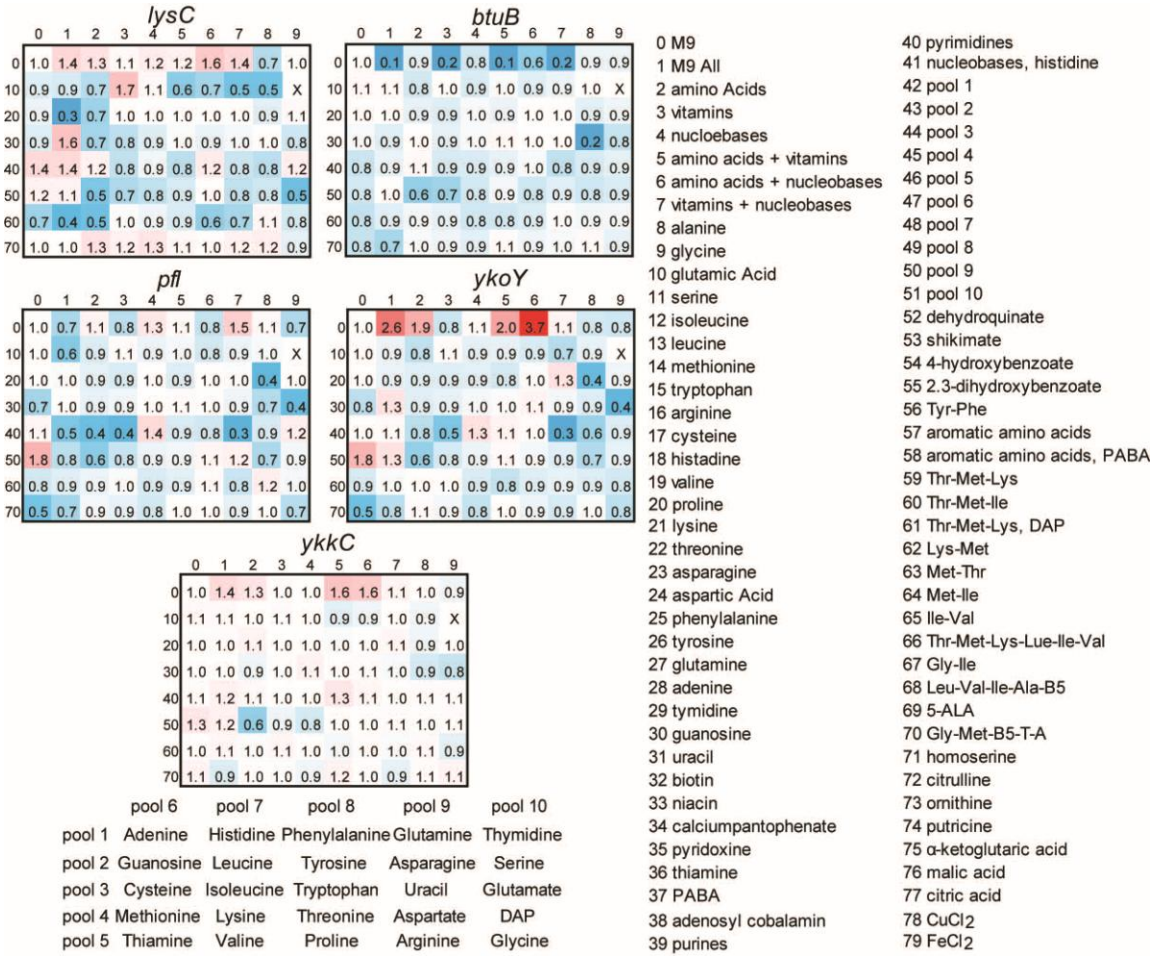

Figure 4-4 Heatmaps of known and predicted riboswitch. Labels above each heatmap correspond to the name of the riboswitch. The media condition in a cell can be found by summation of the row and column labels and cross-referencing this value to the list on the right of the figure. The metabolite composition of pools 1-10 are given in the diagram at the bottom left of the figure. All values in heat map are presented relative to M9 media. An "X" indicates data that was omitted due to low signal.

4.4 DISCUSSION

The investigation of metabolite substitutions for screening purposes, outlined above, provides a potential strategy for examining gene expression. The observation that metabolite substitutions produced regulatory effects of greater magnitudes than metabolite additions is consistent with knowledge of how *E. coli* prioritizes gene expression in response to nutrient availability. Though individual regulatory circuits for metabolite utilization in *E. coli* typically function by positive feedback, they are subject to catabolite repression, with the bacterium strongly favoring particular nutrient sources (Chubukov et al., 2014). In the case of carbon source metabolism catabolite repression often decreases expression of transporters for low priority carbon sources (Görke and Stülke, 2008; Deutscher et al., 2006). The observation that the *gltDF* IGR produced a regulatory response to α-ketoglutarate and glycine only in the respective absence of glucose and ammonia may be explained by this mechanism.

The response of the *gltDF* IGR to nitrogen and carbon sources is particularly unique. Out of all the nitrogen sources tested glutamine was the only one to not repress expression from this IGR when substituted for ammonia. Both glutamine and α-ketoglutarate have been observed to regulate nitrogen source metabolism in *E. coli* by action on GlnB and the NtrB/NtrC two-component system (Maheswaran and Forchhammer, 2003). It is conceivable that *gltF* regulation occurs either through an internal promoter controlled by this two-component system, or by interaction with an element under the control of this regulatory system. Although the initial mapping of the *gltBDF* operon failed to detect the presence of a promoter within the *gltDF* IGR (Castaño

et al., 1988), a *lacZ* promoter assay later provided evidence that this region may contain an extremely weak promoter (Grass et al., 1999). However, a promoter of this strength would not be expected to alter the expression of YFP significantly enough to be detected above the moderately strong RNA1 promoter (Liang et al., 1999). As such, a promoter is unlikely to account for the observed regulatory changes.

The GltF polypeptide codes for a periplasmic protein of unknown function (Grass et al., 1999). Though it was originally suspected that GltF is necessary for induction of genes by the NtrB/NtrC two-component system (Castaño et al., 1992), this hypothesis was later demonstrated to be incorrect (Grass et al., 1999; Goss et al., 2001). A *gltF* genomic knockout had no effect on the regulatory changes observed for this IGR (Figure 4-3). Thus, if the observed regulatory effect results from the function of the two-component system it is independent of GltF. Determining the mechanism responsible for these regulatory responses may provide useful evidence for understanding the function of GltF.

The development of the sensor for use with riboswitches presented unique technical challenges. Earlier attempts in our lab to regulate the expression of fluorescent proteins by riboswitches often proved unsuccessful; particularly when riboswitches were to be transcribed from non-native promoters (data not shown). It is likely that the inclusion of non-native RNA sequence upstream of a riboswitch could interfere with its functionality. Vector pCYi-f4 was created to circumvent this issue, enabling the expression of riboswitches without additional upstream sequence. The plasmid was able to detect regulatory interactions for *btu*B and *lysC* riboswitches in the presence of their

respective ligands, adenosylcobalamin and lysine. The *btuB* riboswitch was previously used to construct a sensor for studying aspects of vitamin B12 metabolism (Fowler et al., 2010). Similarly, the *lysC* riboswitch sensor created within this work may be useful in studies of lysine metabolism.

The observed responses of RNA motifs *pfl* and *ykoY* to nucleotide containing compounds provide evidence for regulatory function. However, the observed regulatory changes may occur indirectly as a response to the growth rate of the organism. This is particularly true for the *Ykoy* motif, as it also responded to amino acids, and recently evidence from the Storz lab points to it being a manganese sensor (Dambach et al., 2014). The *pfl* motif, on the other hand, often resides ahead of genes involved in purine biosynthesis (Meyer et al., 2011), suggesting that the regulatory response may stem from a more direct interaction. Further investigation will be necessary to understand these observed regulatory changes.

The experiments presented within this chapter have expanded upon aspects of the technology developed in previous chapters. These strategies have not only demonstrated new methodologies, but have also provided evidence of novel regulatory functions within intergenic regions that may serve as the basis of future study. Due to the simplicity, flexibility and low cost of the presented techniques, they should find practical use in future studies of regulatory RNAs.

Chapter 5

Discussion and Future Directions

This thesis involves the investigation of non-coding regions for regulatory function and the development of technology necessary for their examination. The projects presented in this study examine a novel approach for monitoring regulatory interactions in *E. coli*, the first high-throughput screen of an IGR library against an array of metabolites and the expansion of these methodologies. The results of these projects validate the primary hypothesis that many non-coding regions contain unknown regulatory functions, and provide data that can serve as the starting point for future scientific inquiry. This chapter serves to address the outcomes and issues that arose with each of the three research projects. Additionally future directions to expand upon the research in these projects are discussed.

The project presented in Chapter 2 involves the creation of a sensor for the intercellular detection of RNA regulatory interactions in *E. coli*. The functionality of this sensor was tested by examining 3 distinct regulatory systems expressed from synthetic constructs. For these systems, the magnitudes of the observed regulatory interactions were consistent with previous studies that used similar experimental setups. However, for the GlmZ and GcvB systems, these interactions were lower than those observed when studied *in vitro* (Kalamorz et al., 2007; Sharma et al., 2007). The reason that *in vitro*

setups display greater responses was not investigated; however, it is an area worthy of further examination.

Two major limitations of this project exist. First, the utility of the sensor for measuring kinetic responses and studying regulation in stationary growth phase were not investigated. The first of these limitations was examined in the study presented in Chapter 3; however the second was not addressed. The use of a constitutive promoter on the dual reporter plasmid results in the accumulation of fluorescent protein. During cellular replication, the protein pool is continuously diluted with newly expressed protein, providing some semblance of turnover. However this effect is minimal in stationary phase, due to slow or non-existent growth of the organism. It should be noted, that stationary phase does not necessarily indicate a lack of growth, but that the rate of growth is identical to the rate of death (Navarro-Llorens and Martínez-García, 2010). Thus it may be possible to achieve accurate quantitative measurements in stationary phase, however it would be dependent both on the rate of cell turnover and the length of the grow phase. Growth phases may be increased by use of chemostats in continuous microbial culture (Hoskisson, and Hobbs, 2005; Novick and Szilard, 1950). However, this involves extensive instrumental setup and is not compatible with high-throughput applications. An alternative solution would be to use an inducible promoter on the dual reporter vector. The bicistronic mRNA containing the fluorescent proteins and regulator region could then be induced upon entering stationary phase. While this approach is technically simpler, the data it produces could be less consistent if it were affected by the timing of induction.

Secondly, while the presented method is successful when applied to the GcvB system to detect low level expression, other regulatory systems with lower expression likely exist that it would be unsuitable for. Evidence for this is presented in Chapter 2, where a subset of IGR library clones has fluorescent signals that are too low for detection. The fluorescent signal quality of the dual reporter sensor is limited by three factors: the dynamic range of the detector used to read fluorescence, the amplitude of fluorescence produced by the reporter and the level of background fluorescence. When examining yellow fluorescence, non-fluorescent media (M9) and cells not expressing a fluorescent protein are indistinguishable once cells are washed. This signifies that detection limits from the fluorometer (Tecan Safire M100) are encountered at the low end of its dynamic range. Use of a more sensitive detector would likely improve the yellow fluorescence signal at low expressed levels.

No upper limit to the detection of yellow fluorescence was observed, which could not be corrected for by lowering the sensitivity of the detector. Thus, selection of brighter fluorescent proteins should improve the lower detection limit, while not negatively affecting the upper one. Unfortunately, no YFPs currently available are significantly brighter than the one used in this study, SYFP2 (Kremers et al., 2006). The RFP, dsRED-(Pfleger et al., 2005), is particularly bright however due to its slow maturation it was found to be unsuitable for use in this study. Alternative RFPs, like those developed in the Tsien lab, may be more suitable for the task. In particular, tdTomato has brightness of ~1.6 times that of dsRED and an ~10 fold faster maturation rate (Shaner et al. 2004).

This maturation rate is very similar to that of the CFP that was used (CFP3A) (Kremers et al., 2006) and thus would likely make an excellent substitution.

While background in fluorescence signal was not a significant issue with the YFP, it was in the CFP channel. Though the fluorescent signal from CFP was significantly above background even when expressed at single copy, a brighter CFP would allow for fluorescence to be read at lower cell densities, and thus earlier within exponential growth. Possible substitutions for CFP3A include mTurquoise2, which is a variant of CFP3A that is 1.76 fold brighter with similar maturation properties (Goedhart et al., 2012). Additionally, a green fluorescent protein could be used in place of the cyan if a red protein was substituted for the yellow. The yellow to red substitution would be necessary, as the excitation/emission spectra of green and yellow fluorescent proteins overlap; however those of green and red fluorescent proteins are significantly distinct and can be used together. Signal over empty vector was ~3 fold higher for GFP (SGFP) than it was for CFP, thus substitution of this GFP for CFP would be expected to produce a significant improvement.

The third chapter explores the use of chemical genetic screens for the identification of RNA regulatory function. In this study, the first ever screen of a library of potential RNA regulatory targets against a metabolite library is performed. The ultimate goal of this is to test the primary hypothesis of this thesis: many non-coding regions contain unknown regulatory function. This was successful, as 16 IGRs were found to have regulatory response. Additionally, the screen was demonstrated to be suitable for identifying RNA regulator responses, as three of the IGRs had RNA

regulatory interaction that are known to be induced by the growth conditions they were identified with. However, the project did not directly address the mechanistic nature of the remaining 13 unknown interactions, nor did it verify these interactions with an alternative method.

When constructing the IGR library, certain constraints were applied to enrich it for RNA sequences that would likely contain regulator elements. IGRs were selected that are located between genes found on operons, and any containing known promoters were excluded. In compact prokaryotic genomes, non-coding regions are thought to contain mostly regulatory elements (Rogozin et al. 2002). As the occurrence of promoters is expected to be lower in IGRs found between genes on operons than those at the ends of operons, it would be expected that regions lacking promoters contain other regulatory function. However, this is largely an assumption, as little is known about the genome wide distribution of RNA regulatory elements in bacteria. The data obtained in this work should prove valuable in testing this hypothesis in the future, once a more complete understanding of the full regulatory complement *of E. coli* is achieved.

While the majority of elements in the IGR library had robust signals, 16% of them had signals that were too low to allow for their analysis. The reasons for this are not entirely clear; however, half have known transcriptional terminators within them. The *bglF* IGR in particular is known to contain a transcriptional terminator that functions as an RNA regulatory element (Amster-Choder and Wright, 1992; Amster-Choder 2005). This element is described in detail in Chapter 2 of this thesis. Other IGRs known to contain terminators include *degS* (Waller and Sauer, 1996), *dnaG* (Yajnik and Godson,

1993), *ilvX* (Lawther, and Hatfield, 1980), *leuA* (Wessler and Calvo, 1981), *pheS* (Fayat et al., 1983) and potentially *pheA* (Gurvich et al., 2003). Out of these the *ilvX*, *leuA*, *pheS* and *phe*A IGRs are part of leader-attenuator regulatory regions. As the open reading frames for the leader peptides are not present, their translation cannot suppress transcriptional termination in the IGRs, rendering these regulatory mechanisms incomplete (stuck in an off-state). This is similar to the *bglF* IGR, as the proteins involved in its regulatory mechanism are cryptic. Thus, the cause of low signal in most of these clones was due to incomplete protein-RNA regulatory mechanisms that were stuck in off-states. As the median size of these IGRs is within the shortest 20 percentile of the library, regulatory function is likely to exist within IGRs below our size cut-off. Future studies should likely include this sequence space.

A major challenge in the study presented in Chapter 3 occurred as a result of the accumulation of autofluorescence in the cyan channel. When yellow fluorescence was normalized by cyan, this resulted in a gradual decrease in yellow fluorescent signal over time. Cellular autofluorescence is most often caused by the presence of metabolites and structural components such as flavins, lipofuscins, and NAD(P)H (Billinton and Knight, 2001). The NAD(P)H emission spectra features a broad peak with maximal emission at 470 nm (Lakowicz et al., 1992) that coincides with our CFP emissions at 473 nm. The close spectral overlap strongly suggests that NAD(P)H is the source of the observed autofluorescence. Washing cells in phosphate buffered saline was found to be the most effective method of decreasing autofluorescence. Whether this occurred by the removal of autofluorescence along with growth media, or by inducing a change in the intercellular

environment, is unclear. As NAD(P)H is far less fluorescent in the oxidized state (Billinton and Knight, 2001), the decrease in autofluorescence may have occurred due to its oxidation, that resulted from the replacement of growth media by PBS. Development of a method that can correct for the progressive signal decrease, would greatly improve the use of the dual reporter sensor for the production of kinetic data.

In Chapter 4, additional applications of the technology in the preceding chapters are presented. First, the screening methodology from Chapter 3 is examined in the context of growth media substitutions, as opposed to additions that were tested in Chapter 3, primarily with focus on the use of alternative carbon and nitrogen sources. By this approach, the *gltF* 5′ UTR was identified as a strong regulatory hit under multiple media substitutions. Additionally, other hits from the screen in Chapter 3 - *nirC* and *iscS* - were re-identified. The amplitudes of these interactions were ~5 fold greater in this screen, signifying that screening media substitutions is superior to media additions. The second aspect examined in this chapter involves the use of the dual reporter plasmid from Chapter 2 to monitor regulatory interactions from riboswitches. Although multiple regulatory RNA systems were examined in the previous studies, none of these involved riboswitches. To show our system works with riboswitches, the dual reporter plasmid was first modified to express riboswitches at the 5′ termini of mRNA transcripts, without any additional upstream sequence. A screen of 10 known and putative riboswitches against the metabolite library was able to identify known lysine and B12 riboswitches, but was unable to identify ligands for the putative riboswitches.

One of the most significant finding presented in Chapter 4 is that nutrient substitutions appear to produce stronger regulatory interactions than nutrient additions. When the *gltF* IGR was screened for regulatory function with the addition of compounds from the metabolite array, it displayed 1.4 ± 0.06 fold induction with α-ketoglutarate, compared to 5.6 ± 0.36 fold with the substitution of α-ketoglutarate for glucose as a carbon source. This can be explained by how *E. coli* alters its gene expression based on carbon source availability. The bacterium prioritizes carbon sources by deactivating genes associated with lower priority sources when presented with more preferential ones (Liu et al., 2005). As glucose is typically the preferred carbon source of these bacteria (Deutscher 2008), it is not surprising that the inclusion of alternative sources has little effect. Additionally, when supplied as additions, insignificant fold changes of 1.3 ± 0.06 for glycine and 1.1 ± 0.01 for arginine were observed, compared to respective values of 2.69 ± 0.03 and 2.32 ± 0.03 when these nutrients were supplied as the sole nitrogen source in place of ammonia. It is not surprising that these amino acids failed to produce an effect when supplied as additions to media containing ammonia, as ammonia is viewed as *E. coli's* preferred nitrogen source (Reitzer, 2003).

A major limitation with the use of the dual reporter sensor for the discovery of riboswitch-ligand interactions is that there is no way to tell if an absence of activity results from a non-functional construct or from a failure to raise the intercellular ligand concentration. There is no guarantee that addition of ligands to media will result in their transport into cells. For instance, although *E. coli* contains a flavin mononucleotide

(FMN) riboswitch (Vitreschak et al., 2002) it does not possesses a transporter for FMN (Abbas and Sibirny, 2011).

The method presented here allows the possibility of testing riboswitches outside their native organisms, where they may or may not be functional. As of yet the cross-species activity of riboswitches has not been thoroughly investigated, however, the tools presented here should be useful for this purpose.

The detection of regulatory outcomes is a useful strategy in the identification of novel regulatory interactions. Previous work in this area has focused primarily on promoter libraries. Zaslaver *et al.* constructed an extensive library of 2000 promoter fusions to GFP for the purpose of promoter identification and transcription network mapping (Zaslaver et al., 2006). The construction of an ncRNA library of this extent, that includes only those 5′ UTRs without internal promoter, would allow for greatly expanded screening of RNA regulatory interactions. However, identifying all promoters and mapping their transcriptional start sites would be very challenging. This undertaking could potentially be achieved through an iterative process, whereby first a 5′ UTR library is constructed using the positions of known and predicted promoters, then upon detection of a regulatory interaction, promoter mapping could be performed by 5′ RACE (Frohman et al., 1988) to test for transcription start site. If an internal promoter were identified, that particular clone would be remade. The pCYi-f4 plasmid would be an excellent vector for cloning of the 5′ UTRs, as it expresses RNAs without any additional upstream sequence that may interfere with their functionality. Ultimately, this library of transcribed non-

coding regions could be assessed purely for RNA based gene regulation. This would be particularly useful, as it would allow for quick classification of observed regulator interactions.

In conclusion, this thesis describes the development and characterization of a sensor and screening methodology for the identification and analysis of RNA regulatory interactions. The sensor was used to identify multiple regulatory interactions by performing the first ever screen of an IGR library of against a metabolite collection. A subset of the identified interactions corresponded to IGRs containing known RNA based regulatory mechanisms, while the majority demonstrated novel functions. Of equal importance, the sensor produced highly reproducibly data consistent with previous publications, when examining known regulatory systems. The ease of use, simplicity and flexibility of the screening methodology and dual reporter sensor, and the many potential applications that they have, suggests that they will contribute greatly to future scientific endeavors.

APPENDIX

CHAPTER 2 SUPPLEMENTARY MATERIALS:

Table S2-1 DNA oligonucleotides used in this study

Name	Sequence	Purpose
pNYL+1-fw	tcgagtccctatcagtgatagagattgacatccctatcagtgatagagatactg agcgctggtaccgggcccgtcgacat	pRNY construction
pNYL+1-rv	cgatgtcgacgggcccggtaccagcgctcagtatctctatcactgatagggа tgtcaatctctatcactgatagggac	pRNY construction
CnY-fw	cacacaatctaagctgtcg	pCYi-f3 MCS sequencing primer
CnY-rv	ctgcacgccatagccaagt	pCYi-f3 MCS sequencing primer
bglt2-fw	ctagtagggtacccgcgtgcgcaaagaacactg	pCYi-bglGF construction
bglt2-fu	caccatggatccttttctggctaactccgtcataa	pCYi-bglGF construction
bglGF-fw	acgaacctggatgttcgttataaaaac	pRNY-bglGF construction
bglGF-rv	ctagtagcccgggtcgttagcgaatgatggataaca	pRNY-bglGF construction
gcvB-NY-fw	acacttcctgagccggaacgaaaag	pRNY-gcvB construction
gcvB-rv	ctactagcccgggaaaaaaaagcaccgcaattaggcggtg	pRNY-gcvB construction
oppA-fw	ctagtagggtaccaatcgacataaggtgatcgtct	pCYi-gcvB construction
oppA-7aa	acctttagacaccatggatcctctcttggtgatgttggtcattg	pCYi-gcvB construction
gltI-fw	ctagtagggtaccacaacactgcacaataaagttgc	pCYi-gltI construction
gltI-7aa	acctttagacaccatggatccggcaggtttacgtaattgcatat	pCYi-gltI construction
gcvB-KO-fw	ttagttttttaatctgagccattataaattgtccgttgagcttctaccagcaaata cctatagtggcggcggaattgccagctggggcg	gcvB knockout generation
gcvB-KO-rv	caaacagcagatcaaccgtactgacgtgaaagagatggtcgaactggatca gtaattcgcgatcgcaaggttcgaaccccagagtcccgc	gcvB knockout generation
glmUS-fw	ctagtagggtacctggcgtcgtccggtaaagaaaaag	pCYi-glmUS-(7AA/ATG) construction
glmUS-7AA	atgtgtggaattgttggcgcgggat*ccatgg*tgtctaaaggt	pCYi-glmUS-7AA construction
glmUS-ATG	acctttagacaccatggatcccatagtttttgattccgatttatatc	pCYi-glmUS-ATG construction
glmZ-NY-fw	acgtagatgctcattccatctcttat	pRNY-glmZ construction
glmZ-rv	ctactagcccgggaaaaaaacgcctgctcttattacggag	pRNY-glmZ construction
lacZ1-check	catatggaaaccgtcgatattca	Verification of integration at lacZ
lacZ2-check	tggccgattcattaatgcagct	Verification of integration at lacZ
YFPfwd	ttatcctatcaatcaaaactgagc	YFP qPCR
YFPrev	aatcttctctcatccgcc	YFP qPCR
CFPfwd	ttgtagtcgtacctcgtatgc	CFP qPCR
CFPrev	aacaccccaagtcagagtagtg	CFP qPCR
GlmZfwd	gtagatgctcattccatctct	GlmZ qPCR
GlmZrev	aaaacaggtctgtatgacaacaag	GlmZ qPCR
GcvBfwd	acttcctgagccggaac	GcvB qPCR
GcvBrev	attaatcactatggacagacaggg	GcvB qPCR
bglGFfwd	atgaacatgcaaatcacc	bglGF qPCR
bglGFrev	aaggcatactctttttctattcc	bglGF qPCR
cysGfwd	acttgaatcactgctgcc	cysG qPCR
cysGrev	accacatctgcctgctga	cysG qPCR

Table S2-2 Plasmids used in this study

Name	Purpose	Resistance
pBAD30	pBAD30 construction	ampicillan
pBAD30-Ara	pBnR construction	ampicillan
pBnR	BFP RFP fluorescence development pCnY construction	ampicillan
pCnY	CFP RFP fluorescence development pCYi-amp construction	ampicillan
pCYi-amp	pCYi-cm construction	ampicillan
pCYi-cm	pCYi-fu construction	chloramphenicol
pCYi-fu	pCYi-f2 construction	chloramphenicol
pCYi-f2	pCYi-f3 construction	chloramphenicol
pCYi-f3	Empty vector control in regulatory assays	chloramphenicol
pRNY	Expression of RNA binding regulatory elements	kanamycin
pCYi-glmUS-ATG	glmZ regulatory assay	chloramphenicol
pCYi-glmUS-7AA	glmZ regulatory assay	chloramphenicol
pRNY-GlmZ	glmZ regulatory assay	kanamycin
pCYi-oppA	gcvB regulatory assay	chloramphenicol
pCYi-gltI	gcvB regulatory assay	chloramphenicol
pRNY-gcvB	gcvB regulatory assay	kanamycin
pCYi-bglGF	bglGF regulatory assay	chloramphenicol
pRNY-bglGF	bglGF regulatory assay	kanamycin
pCYi-ter	Western blot no YFP control	chloramphenicol

Table S2-3 Growth Media used in this study

Media/Supplement	Base Media	Components
LB	N/A	LB Broth Miller (BioShop, Burlington, ON)
M9	N/A	64 g/L $Na_2HPO_4 \cdot 7H_20$ 15 g/L KH_2PO_4 2.5 g/L NaCl 5.0 g/L NH_4Cl 0.4% glucose 2 mM $MgSO_4$ 0.1 mM $CaCl_2$
Gly	M9	2 mg/mL glycine
Salicin	M9	2 mg/mL salicin

Rich	M9	100 μg/mL DL-alanine 22 μg/mL L-arginine 100 μg/mL L-asparagine 100 μg/mL L-aspartic acid 22 μg/mL L-cysteine 100 μg/mL glycine 100 μg/mL L-glutamic acid 100 μg/mL L-glutamine 22 μg/mL L-histidine 20 μg/mL L-leucine 44 μg/mL L-lysine 20 μg/mL L-isoleucine 10 μg/mL L-methionine 20 μg/mL L-phenylalanine 30 μg/mL L-proline 50 μg/mL L-serine 40 μg/mL L-threonine 20 μg/mL L-tyrosine 20 μg/mL L-tryptophan 20 μg/mL L-valine 40 μg/mL adenine 40 μg/mL thymidine 40 μg/mL thymine 40 μg/mL uracil 0.5 μg/mL biotin 1 μg/mL calcium pantothenate 1 μg/mL niacin 1 μg/mL pyridoxine HCl 1 μg/mL thiamine HCl

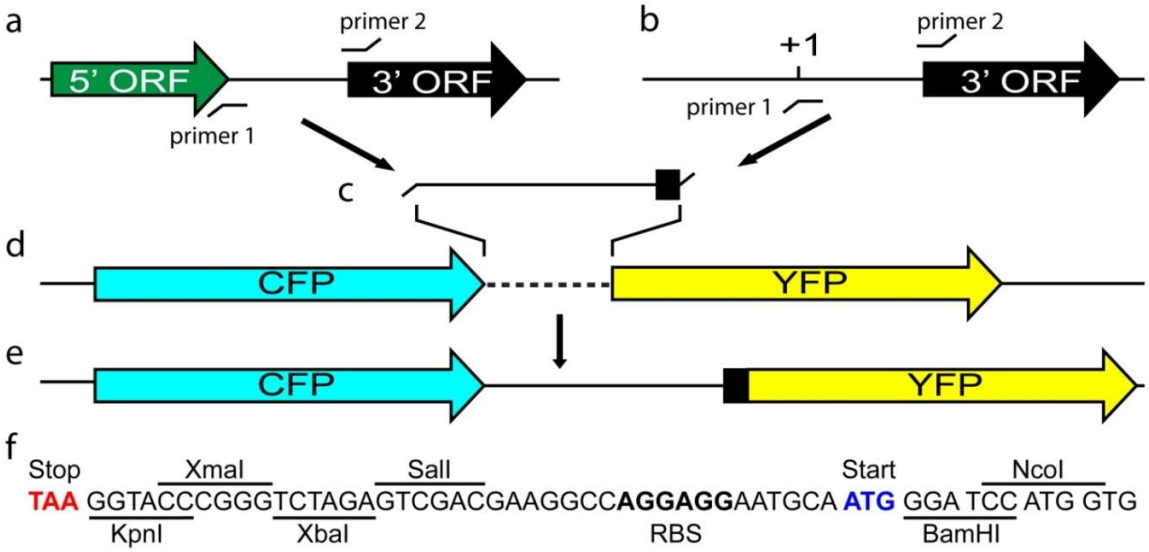

Figure S2-1 Insertion of RNA-coding DNA elements into pCYi-f3 vector. (a) Encompassing the region between genes or (b) the 5′ UTR of an operon, DNA fragments corresponding to RNA of interest are amplified by PCR. (c) PCR products include an additional 27 base pairs upstream and downstream of their non-coding region. (d) Ligation of digested DNA fragments and pCYi-f3 results in an operon with Cyan Fluorescent protein (CFP) upstream of the non-coding region and yellow fluorescent protein (YFP) downstream of non-coding region. (e) The resulting plasmid has YFP translational fused to 9 amino acids downstream of the non-coding region. CFP has no translational fusion due to a stop codon in the (f) multiple cloning site.

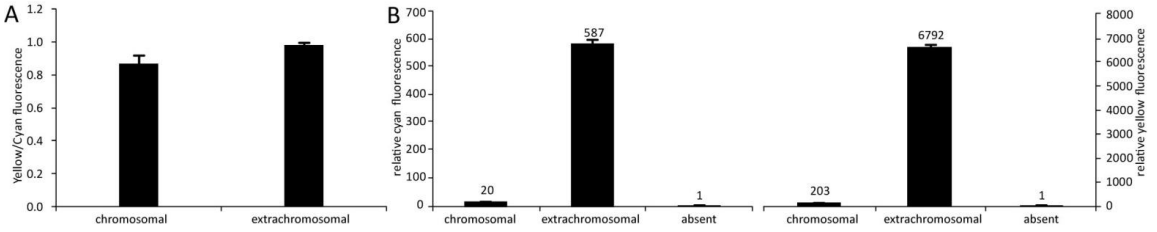

Figure S2-2 pCYi-f3 fluorescence intensity for extrachromosomal and chromosomally integrated pCYi-f3. Values are reported as an average of 6 replicates. Cells were grown in M9 to an optical density at 600 nm of 1.0. Chromosomal, extrachromosomal and absent respectively refer to pCYi-f3 integrated into the *Escherichia coli MG1655* chromosome at *lacZ*, present as an extrachromosomal vector and not present in the cell line. (A) Ratio of Yellow/Cyan fluorescence to demonstrate the maintenance of fluorescence ratio upon chromosomal integration. (B) Yellow and

Cyan fluorescence scaled to that from cells lacking *cfp* and *yfp* (absent). Exact values are given above bars to provide for easier comparison.

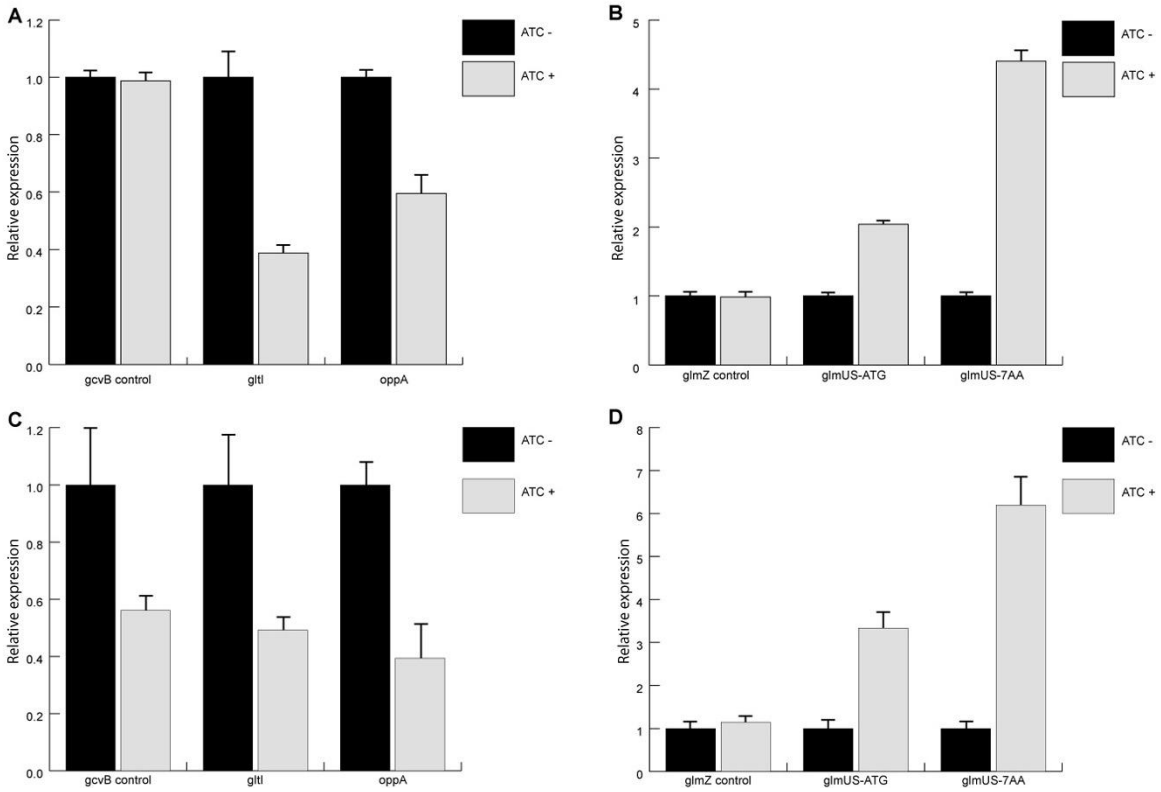

Figure S2-3 Comparison of normalization methods for data obtained with *gcvB* and *glmZ* plasmid constructs. The YFP fluorescence from *gcvB* and *glmZ* plasmids was normalized by the CFP fluorescence (A and B) or by optical density at 600 nm (C and D). Average standard deviations when normalize by fluorescence and optical density were 5% and 15% respectively.

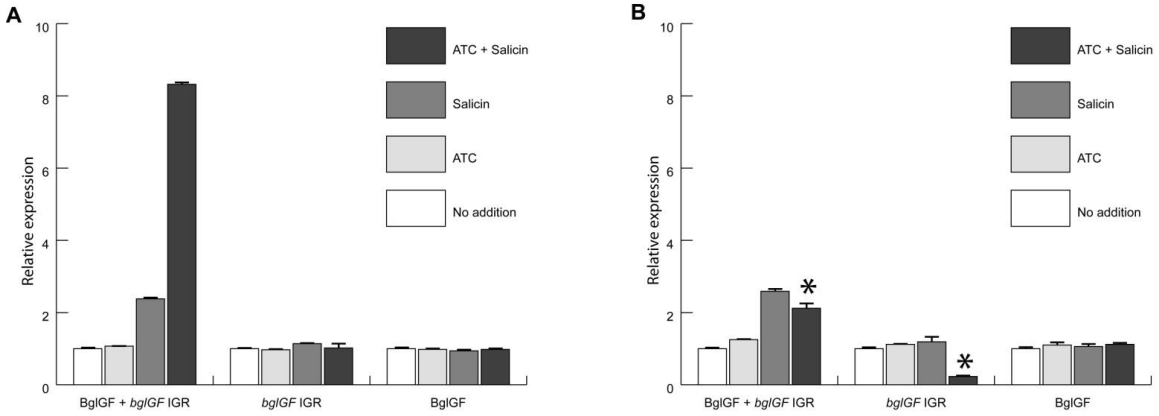

Figure S2-4 Comparison of normalization methods for data obtained with *bglGF* itergenic region. The YFP fluorescence was normalized by the CFP fluorescence (A) or by optical density at 600 nm (B). Average standard deviations when normalized by fluorescence and optical density were 2.6% and 4.6%, respectively. * indicates that the normalization by optical density resulted in a low YFP signal.

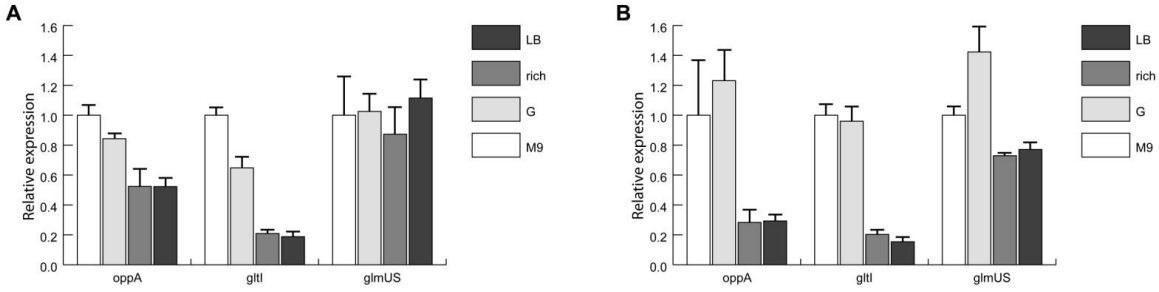

Figure S2-5 Comparison of normalization methods for data obtained with *gcvB* targets placed on *E. coli* chromosome. The YFP fluorescence was normalized by the CFP fluorescence (A) and optical density at 600 nm (B). Average standard deviations were ~13.5% for both normalization methods. Normalization by optical density was unable to yield an observable decrease in *OppA* and *gltI* samples with the addition of glycine. Additionally when normalized by optical density *glmUS-7AA* controls had unexpected changes in expression.

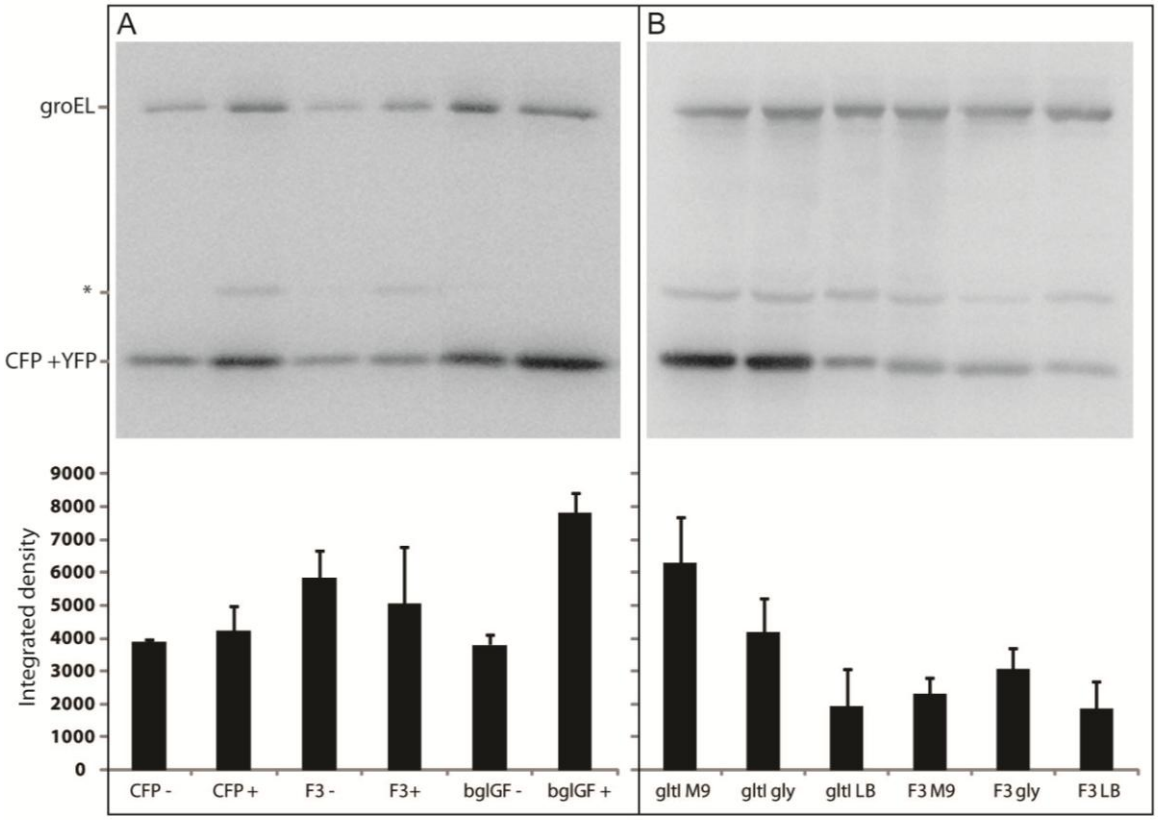

Figure S2-6 Western blot analysis of fluorescent protein levels. Antibodies were used to detect groEL (57 KDa) and CFP/YFP (27 KDa). *A non-specific band is visible at 33 KDa. (A) Blot of two plasmid experiments with (+) and without (-) 100 μg/ml ATC. Labels correspond to pairs of plasmids as follows. CFP: pCYi-rpsU + pRNY; F3: pCYI-f3 + pRNY; gcvB: pCYi-gltI + pRNY-gcvB; glmZ: pCYi-glmUS + pRNY-glmZ; bglGF: pCYi-bglGF + pRNY-bglGF. bglGF samples also contained 0.2% salicin. (B) Blot of endogenous gcvB expression experiments with M9 media (M9), M9 media plus 0.2% glycine and LB media (LB). gltI and F3 correspond respectively to pCYI-gltI and pCYi-f3 integrated into the *E. Coli MG1655* chromosome.

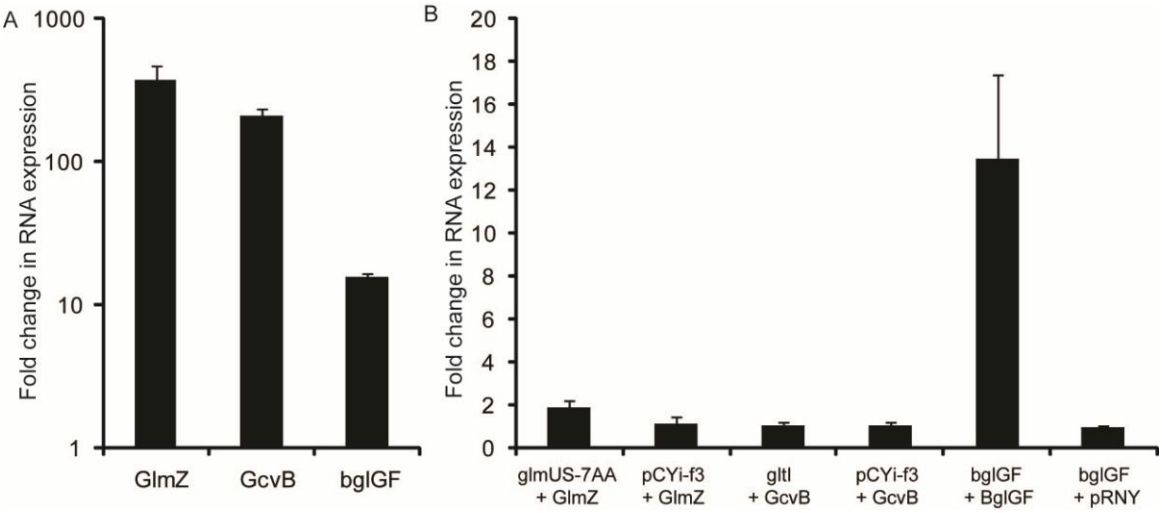

Figure S2-7. **Results of qRT-PCR experiments**. (A) The fold changes in the expression of RNAs from pRNY variants upon the addition of ATC. GlmZ and GcvB were induced from plasmids pRNY-glmZ and pRNY-gcvB with 100 ng/ml ATC. bglGF was induced from plasmid pRNY-bglGF with 25 ng/ml. (B) The fold changes in the expression of the YFP mRNA from plasmids pCYi-glmUS-7AA (glmUS-7AA), pCYi-f3, pCYi-gltI (gltI), and pCYi-bglGF (bglGF) in the presence of the RNA expressed from pRNY-glmZ (GlmZ), pRNY-gcvB (GcvB) or pRNY-bglGF (BglGF). Each YFP mRNA level was normalized to the relevant CFP mRNA level.

CHAPTER 3 SUPPLEMENTARY MATERIALS:

Table S3-1 Intergenic regions within operons. IGRs are listed in the order that they appear in the *E. coli MG1655* genome. For all boolean values: 0 = false, 1 = true.

Gene1 and Gene2	Genes flanking IGR, appearing in genome order.
b#1 and b#1	b numbers for Gene1 and Gene2 respectively.
Start and Stope	Nucleotide positions of beginning and end of IGR.
Size	The IGRs length.
Strand	Genome strand the IGR is located on.
TSS1 and TSS2	Presence of transcriptional start sites determined by studies.
P1 and P2	Presence of promoters determined by studies.
Exp	Whether operon has been experimental verified.

Gene1	Gene2	b#1	b#2	Start	Stop	Size	Strand	TSS1	TSS2	P1	P2	Exp
thrL	thrA	b0001	b0002	255	337	82	forward	0	0	0	1	1
talB	mog	b0008	b0009	9191	9306	115	forward	0	1	0	1	0
dnaK	dnaJ	b0014	b0015	14079	14168	89	forward	0	0	0	1	1
nhaA	nhaR	b0019	b0020	18655	18715	60	forward	0	0	0	1	1
ribF	ileS	b0025	b0026	22348	22391	43	forward	0	0	0	1	1
lspA	fkpB	b0027	b0028	25701	25826	125	forward	0	1	0	1	1
caiD	caiC	b0036	b0037	36162	36271	109	reverse	0	0	0	1	1
caiC	caiB	b0037	b0038	37824	37898	74	reverse	0	0	0	1	1
caiB	caiA	b0038	b0039	39115	39244	129	reverse	0	0	0	1	1
caiA	caiT	b0039	b0040	40386	40417	31	reverse	0	0	0	1	1
fixB	fixC	b0042	b0043	44129	44180	51	forward	0	0	0	1	1
fixX	yaaU	b0044	b0045	45750	45807	57	forward	0	0	0	1	0
surA	lptD	b0053	b0054	54702	54755	53	reverse	0	0	0	1	1
araD	araA	b0061	b0062	66550	66835	285	reverse	0	0	0	0	1
araC	yabI	b0064	b0065	71265	71351	86	forward	1	0	0	1	0
tbpA	sgrR	b0068	b0069	75480	75644	164	reverse	0	0	0	1	1
sgrT	setA		b0070	77519	77621	102	forward	0	0	0	0	1
leuA	leuL	b0074	b0075	83529	83622	93	reverse	0	0	0	1	1
murG	murC	b0090	b0091	100711	100765	54	forward	0	1	0	1	1
ftsA	ftsZ	b0094	b0095	105244	105305	61	forward	0	0	0	1	1
ftsZ	lpxC	b0095	b0096	106456	106557	101	forward	1	1	1	1	1
secM	secA	b0097	b0098	108217	108279	62	forward	0	0	0	1	1
secA	mutT	b0098	b0099	110984	111044	60	forward	0	0	0	1	1
pdhR	aceE	b0113	b0114	122856	123017	161	forward	0	0	0	0	1
aceF	lpd	b0115	b0116	127587	127912	325	forward	0	1	1	1	1
speE	yacC	b0121	b0122	136464	136570	106	reverse	0	1	0	1	1
yadH	yadI	b0128	b0129	144472	144577	105	forward	0	0	0	1	0
yadC	yadK	b0135	b0136	150953	151003	50	reverse	0	0	0	1	1

htrE	ecpD	b0139	b0140	155426	155461	35	reverse	0	0	0	1	1
pcnB	yadB	b0143	b0144	159126	159186	60	reverse	0	0	1	1	0
yadB	dksA	b0144	b0145	160112	160149	37	reverse	0	0	0	1	1
dksA	sfsA	b0145	b0146	160604	160782	178	reverse	0	1	0	1	1
fhuA	fhuC	b0150	b0151	169727	169778	51	forward	0	0	0	1	1
clcA	erpA	b0155	b0156	176528	176610	82	forward	0	1	0	1	0
yadS	btuF	b0157	b0158	177624	177662	38	reverse	0	0	0	1	1
dapD	glnD	b0166	b0167	185947	185978	31	reverse	0	0	1	1	1
glnD	map	b0167	b0168	188650	188712	62	reverse	0	0	0	1	1
rpsB	tsf	b0169	b0170	190599	190857	258	forward	0	0	0	1	1
rseP	bamA	b0176	b0177	197898	197928	30	forward	0	1	0	1	0
bamA	hlpA	b0177	b0178	200360	200482	122	forward	0	0	1	0	1
lpxD	fabZ	b0179	b0180	201996	202101	105	forward	0	0	0	1	1
rnhB	dnaE	b0183	b0184	205089	205126	37	forward	0	0	0	1	1
ldcC	yaeR	b0186	b0187	211820	211877	57	forward	0	0	0	1	0
yaeR	tilS	b0187	b0188	212266	212331	65	forward	0	0	0	1	0
metQ	metI	b0197	b0198	220928	220968	40	reverse	0	0	0	1	1
rrsH	ileV	b0201	b0202	225312	225381	69	forward	0	0	0	1	1
ileV	alaV	b0202	b0203	225457	225500	43	forward	0	0	0	1	1
alaV	rrlH	b0203	b0204	225575	225759	184	forward	0	0	0	1	1
rrlH	rrfH	b0204	b0205	228662	228756	94	forward	0	0	0	1	1
rrfH	aspU	b0205	b0206	228875	228928	53	forward	0	0	0	1	0
mltD	gloB	b0211	b0212	233955	234027	72	reverse	0	0	0	1	0
dinB	yafN	b0231	b0232	251953	252005	52	forward	0	0	0	1	0
frsA	crl	b0239	b0240	257771	257829	58	forward	1	1	0	1	0
proA	thrW	b0243	b0244	261980	262095	115	forward	0	1	1	1	0
yafX	ykfF	b0248	b0249	264430	264528	98	reverse	0	0	0	1	1
ykfF	ykfB	b0249	b0250	264767	264844	77	reverse	0	0	0	1	1
yafZ	ykfA	b0252	b0253	267229	267321	92	reverse	1	0	0	0	1

insI-1	insO-1	b0256	b0257	270978	271054	76	forward	0	0	0	0	1
yagB	yagA	b0266	b0267	279959	280053	94	reverse	0	1	0	0	1
matC	matB	b0292	b0293	309250	309308	58	reverse	1	0	0	1	0
matB	matA	b0293	b0294	309895	309970	75	reverse	0	0	0	1	0
eaeH	insE-1		b0298	314468	314506	38	forward	0	0	0	0	1
yahD	yahE	b0318	b0319	335109	335149	40	forward	0	0	0	1	1
prpB	prpC	b0331	b0333	348796	349236	440	forward	0	0	0	0	1
prpC	prpD	b0333	b0334	350405	350439	34	forward	0	0	0	1	1
prpD	prpE	b0334	b0335	351890	351930	40	forward	0	0	0	1	1
cynT	cynS	b0339	b0340	358682	358713	31	forward	0	0	0	1	1
cynS	cynX	b0340	b0341	359183	359216	33	forward	0	0	0	1	1
lacA	lacY	b0342	b0343	361084	361150	66	reverse	0	0	0	1	1
lacY	lacZ	b0343	b0344	362403	362455	52	reverse	0	0	0	1	1
lacI	mhpR	b0345	b0346	366734	366811	77	reverse	0	0	0	1	1
frmB	frmA	b0355	b0356	377592	377686	94	reverse	0	1	0	0	1
frmA	frmR	b0356	b0357	378795	378830	35	reverse	0	0	0	1	1
phoA	psiF	b0383	b0384	402386	402505	119	forward	0	0	0	0	1
aroL	yaiA	b0388	b0389	406153	406203	50	forward	0	1	0	1	1
yaiA	aroM	b0389	b0390	406394	406652	258	forward	0	1	0	1	1
aroM	yaiE	b0390	b0391	407329	407401	72	forward	1	1	0	1	0
phoB	phoR	b0399	b0400	417055	417113	58	forward	0	0	0	1	1
brnQ	proY	b0401	b0402	420134	420210	76	forward	0	0	0	1	1
queA	tgt	b0405	b0406	425305	425361	56	forward	0	0	0	1	0
ribD	ribE	b0414	b0415	433782	433871	89	forward	0	0	0	1	1
nusB	thiL	b0416	b0417	434780	434858	78	forward	0	0	0	1	1
yajO	dxs	b0419	b0420	437359	437539	180	reverse	0	0	0	0	1
ampG	yajG	b0433	b0434	452769	452813	44	reverse	0	0	0	1	0
clpP	clpX	b0437	b0438	456524	456650	126	forward	0	0	0	1	1
clpX	lon	b0438	b0439	457924	458112	188	forward	0	1	2	1	1

queC	ybaE	b0444	b0445	464771	464836	65	reverse	0	1	0	1	0
ybaO	mdlA	b0447	b0448	468065	468095	30	forward	0	0	0	1	0
glnK	amtB	b0450	b0451	472160	472190	30	forward	0	0	0	1	1
dnaX	ybaB	b0470	b0471	493247	493300	53	forward	0	0	0	1	0
ybbN	ybbO	b0492	b0493	517503	517564	61	reverse	0	0	0	1	0
rhsD	ybbC	b0497	b0498	526765	526805	40	forward	0	0	0	0	1
ybbB	allS	b0503	b0504	530450	530519	69	reverse	0	0	0	1	0
allA	allR	b0505	b0506	532157	532235	78	forward	0	1	0	1	0
hyi	glxR	b0508	b0509	535710	535810	100	forward	0	0	0	1	1
glxR	ybbV	b0509	b0510	536688	536720	32	forward	0	0	0	0	1
ybbW	allB	b0511	b0512	538311	538371	60	forward	0	0	0	1	1
allB	ybbY	b0512	b0513	539732	539789	57	forward	0	0	0	1	1
purE	lpxH	b0523	b0524	552323	552441	118	reverse	0	0	1	1	0
sfmC	sfmD	b0531	b0532	558889	558920	31	forward	0	0	0	1	1
sfmD	sfmH	b0532	b0533	561523	561565	42	forward	0	0	0	1	0
renD	emrE	b0542	b0543	567470	567538	68	forward	0	0	0	0	1
ompT	envY	b0565	b0566	584856	585370	514	reverse	0	0	0	0	1
cusC	cusF	b0572	b0573	596196	596354	158	forward	0	0	0	0	1
ybdF	ybdJ	b0579	b0580	605109	605174	65	reverse	0	0	0	1	0
ybdJ	ybdK	b0580	b0581	605422	605488	66	reverse	0	0	0	1	0
entD	fepA	b0583	b0584	609311	609477	166	reverse	0	0	0	0	1
entF	fepE	b0586	b0587	617261	617477	216	forward	0	0	0	1	1
ahpC	ahpF	b0605	b0606	638731	638976	245	forward	0	0	0	0	1
lipB	ybeD	b0630	b0631	661501	661602	101	reverse	0	0	0	1	1
ybeD	dacA	b0631	b0632	661865	661975	110	reverse	0	1	2	1	0
mrdA	rlmH	b0635	b0636	667440	667471	31	reverse	0	0	0	1	1
gltJ	gltI	b0654	b0655	685892	686062	170	reverse	0	1	0	1	1
ybeX	ybeY	b0658	b0659	691007	691097	90	reverse	0	0	0	1	1
glnX	glnV	b0664	b0665	695727	695765	38	reverse	0	0	0	1	1

glnV	metU	b0665	b0666	695839	695887	48	reverse	0	0	0	1	1
glnW	glnU	b0668	b0670	696053	696088	35	reverse	0	1	0	1	1
nagD	nagC	b0675	b0676	699549	699597	48	reverse	0	0	0	1	1
nagA	nagB	b0677	b0678	701974	702034	60	reverse	0	0	0	1	1
ybfM	ybfN	b0681	b0682	708963	709013	50	forward	0	0	0	1	0
uof	fldA	b4637	b0684	709948	710158	210	reverse	0	1	0	1	1
ybfG	ybfH			715580	715611	31	reverse	0	0	0	0	1
ybgL	nei	b0713	b0714	745122	745158	36	forward	0	0	0	1	1
ybgQ	ybgD	b0718	b0719	751392	751452	60	reverse	0	0	0	1	0
sdhB	sucA	b0724	b0726	757628	757929	301	forward	0	1	1	1	1
sucB	sucC	b0727	b0728	761962	762237	275	forward	0	0	0	1	1
tolR	tolA	b0738	b0739	775500	775565	65	forward	0	0	0	1	1
tolB	pal	b0740	b0741	778255	778290	35	forward	0	0	0	1	1
lysT	valT	b0743	b0744	779852	779988	136	forward	0	0	0	0	1
nadA	pnuC	b0750	b0751	782351	782389	38	forward	0	0	0	1	1
modF	modE	b0760	b0761	793011	793079	68	reverse	0	0	0	1	1
ybhH	ybhI	b0769	b0770	801034	801110	76	forward	0	0	0	1	0
ybhB	bioA	b0773	b0774	807132	807191	59	reverse	0	0	0	1	0
ybiI	ybiX	b0803	b0804	837679	837753	74	reverse	0	0	0	1	0
ybiX	fiu	b0804	b0805	838430	838472	42	reverse	0	0	0	1	0
glnP	glnH	b0810	b0811	846342	846481	139	reverse	0	0	0	1	1
deoR	ybjG	b0840	b0841	881957	882015	58	reverse	0	0	0	1	0
ybjH	ybjI	b0843	b0844	884453	884539	86	reverse	0	0	0	1	0
nfsA	rimK	b0851	b0852	891129	891190	61	forward	0	0	0	1	1
rimK	ybjN	b0852	b0853	892092	892180	88	forward	0	0	0	1	1
potF	potG	b0854	b0855	894119	894214	95	forward	0	0	0	1	1
potI	ybjO	b0857	b0858	897152	897212	60	forward	0	0	0	1	0
ybjO	rlmC	b0858	b0859	897700	897741	41	forward	0	0	0	1	1
ybjS	ybjT	b0868	b0869	905976	906075	99	reverse	0	0	0	1	0

ltaE	poxB	b0870	b0871	908517	908554	37	reverse	1	0	0	1	1
clpS	clpA	b0881	b0882	922456	922487	31	forward	0	0	0	1	0
aat	cydC	b0885	b0886	926655	926697	42	reverse	0	0	0	1	0
rarA	serS	b0892	b0893	938560	938651	91	forward	1	1	0	1	0
pflB	focA	b0903	b0904	952777	952832	55	reverse	0	0	0	1	1
serC	aroA	b0907	b0908	957964	958035	71	forward	0	0	0	1	1
cmk	rpsA	b0910	b0911	961107	961218	111	forward	0	0	0	1	1
rpsA	ihfB	b0911	b0912	962891	963051	160	forward	0	1	1	1	1
ycaI	msbA	b0913	b0914	965807	965844	37	forward	0	0	0	1	1
lpxK	ycaQ	b0915	b0916	968575	968612	37	forward	0	0	0	1	1
ycaQ	ycaR	b0916	b0917	969844	969896	52	forward	0	0	1	1	0
fabA	ycbZ	b0954	b0955	1015693	1015762	69	reverse	0	0	1	1	0
hspQ	rlmI	b0966	b0967	1027944	1028002	58	reverse	0	1	0	1	0
yccB	appA	b4592	b0980	1039760	1039840	80	forward	0	0	0	1	1
gfcE	gfcD	b0983	b0984	1045026	1045072	46	reverse	0	0	0	1	0
ghrA	ycdX	b1033	b1034	1098047	1098102	55	forward	0	0	0	1	0
csgB	csgA	b1041	b1042	1103629	1103670	41	forward	0	0	0	1	1
csgA	csgC	b1042	b1043	1104125	1104184	59	forward	0	0	0	1	1
ymdA	ymdB	b1044	b1045	1104948	1105043	95	forward	0	0	0	1	0
grxB	mdtH	b1064	b1065	1123277	1123341	64	reverse	0	0	0	1	0
flgM	flgA	b1071	b1072	1129351	1129427	76	reverse	0	0	1	1	1
flgF	flgG	b1077	b1078	1133780	1133952	172	forward	0	0	0	1	1
flgG	flgH	b1078	b1079	1134734	1134787	53	forward	0	0	0	1	1
psrD	yceD	b4418	b1088	1145980	1146017	37	forward	0	0	0	1	0
yceD	rpmF	b1088	b1089	1146538	1146590	52	forward	1	0	0	1	1
rpmF	plsX	b1089	b1090	1146763	1146844	81	forward	0	1	0	1	1
plsX	fabH	b1090	b1091	1147914	1147982	68	forward	0	0	0	1	1
fabG	acpP	b1093	b1094	1150627	1150838	211	forward	0	1	0	1	1
acpP	fabF	b1094	b1095	1151074	1151162	88	forward	0	1	0	0	1

ymfA	potD	b1122	b1123	1180948	1181006	58	reverse	0	0	0	1	0
hflD	mnmA	b1132	b1133	1191854	1191890	36	reverse	0	0	0	1	0
mnmA	nudJ	b1133	b1134	1192996	1193050	54	reverse	0	0	0	1	0
xisE	ymfH	b1141	b1142	1200255	1200292	37	reverse	0	0	0	1	1
ymfT	ymfL	b1146	b1147	1202447	1202479	32	forward	0	0	0	0	1
ymfO	ymfP	b1151	b1152	1205427	1205483	56	forward	0	0	0	0	1
ymgG	ymgI	b1172	b4593	1222142	1222213	71	reverse	0	1	0	1	0
ycgL	ycgM	b1179	b1180	1227230	1227302	72	forward	0	0	0	1	0
ychF	pth	b1203	b1204	1257035	1257152	117	reverse	0	0	1	0	1
prs	ispE	b1207	b1208	1261098	1261249	151	reverse	0	0	0	0	1
hemA	prfA	b1210	b1211	1264193	1264235	42	forward	0	0	0	1	1
ychA	kdsA	b1214	b1215	1267352	1267388	36	forward	0	0	0	1	1
chaB	chaC	b1217	b1218	1271572	1271709	137	forward	0	0	1	0	1
tpr	tyrV	b1229	b1230	1286411	1286467	56	reverse	0	0	0	0	1
tyrV	tyrT	b1230	b1231	1286551	1286761	210	reverse	0	0	0	0	1
purU	ychJ	b1232	b1233	1287847	1287897	50	reverse	0	0	0	1	0
rssA	rssB	b1234	b1235	1289373	1289465	92	forward	0	0	1	0	1
oppA	oppB	b1243	b1244	1300837	1300923	86	forward	0	0	0	0	1
yciU	cls	b1248	b1249	1305174	1305209	35	reverse	0	0	0	1	1
yciA	yciB	b1253	b1254	1310270	1310375	105	reverse	0	1	0	1	0
yciB	yciC	b1254	b1255	1310914	1310944	30	reverse	0	0	0	1	1
yciE	yciF	b1257	b1258	1313248	1313294	46	reverse	0	0	0	1	1
yciF	yciG	b1258	b1259	1313794	1313880	86	reverse	0	0	0	1	1
trpE	trpL	b1264	b1265	1320970	1321062	92	reverse	0	0	0	1	1
yciM	pyrF	b1280	b1281	1339751	1339945	194	forward	0	1	1	1	1
yciT	yciZ	b1284	b4596	1342370	1342460	90	reverse	0	1	0	0	1
rnb	yciW	b1286	b1287	1346936	1347004	68	reverse	0	0	1	1	0
ycjD	sapF	b1289	b1290	1349784	1349852	68	reverse	0	0	0	1	0
puuB	puuE	b1301	b1302	1363536	1363574	38	forward	0	0	0	1	1

pspA	pspB	b1304	b1305	1366771	1366825	54	forward	0	0	0	1	1
pspD	pspE	b1307	b1308	1367638	1367713	75	forward	0	1	0	1	1
ycjP	ycjQ	b1312	b1313	1372956	1372987	31	forward	0	0	0	1	1
ycjF	tyrR	b1322	b1323	1384596	1384744	148	forward	0	1	1	1	1
abgT	abgB	b1336	b1337	1399797	1399834	37	reverse	0	0	0	1	1
C0343	dbpA		b1343	1407461	1407535	74	forward	0	0	0	0	1
ttcA	intR	b1344	b1345	1409972	1410024	52	reverse	0	0	0	1	0
ydaQ	ydaC	b1346	b1347	1411476	1411555	79	reverse	0	0	0	1	1
lar	recT	b1348	b1349	1411951	1412008	57	reverse	0	0	0	1	1
racC	ydaE	b1351	b4526	1415787	1415862	75	reverse	0	0	0	1	0
ydaW	rzpR	b1361	b1362	1421336	1421369	33	forward	0	0	0	0	1
lomR_2	stfR	b1371	b1372	1427008	1427073	65	forward	0	0	0	0	1
ydcX	ydcY	b1445	b1446	1515586	1515672	86	forward	0	1	0	1	0
yncG	yncH	b1454	b1455	1524888	1524964	76	forward	0	0	0	1	0
yddE	narV	b1464	b1465	1533882	1533961	79	reverse	0	0	0	1	0
narZ	narU	b1468	b1469	1540614	1540696	82	reverse	0	0	0	1	0
yddJ	yddK	b1470	b1471	1542743	1542782	39	reverse	0	0	0	0	1
sra	bdm	b1480	b1481	1553987	1554089	102	reverse	0	1	1	1	1
gadC	gadB	b1492	b1493	1568513	1568669	156	reverse	0	0	0	0	1
pqqL	yddB	b1494	b1495	1573226	1573271	45	reverse	0	0	0	1	0
yddB	yddA	b1495	b1496	1575643	1575681	38	reverse	0	0	0	1	1
ydeO	yneN	b1499	b1500	1581711	1581786	75	reverse	0	0	0	0	1
ydeU	ydeK	b1509	b1510	1592089	1592133	44	reverse	0	0	0	1	0
lsrK	lsrR	b1511	b1512	1598233	1598312	79	reverse	0	0	0	0	1
lsrG	tam	b1518	b1519	1605313	1605370	57	forward	0	0	0	1	1
yneH	sad	b1524	b1525	1611275	1611339	64	reverse	0	1	0	1	0
marA	marB	b1531	b1532	1617981	1618013	32	forward	0	0	0	1	1
ynfO	ydfO	b4533	b1549	1635013	1635071	58	forward	0	0	0	1	1
hokD	relE	b1562	b1563	1643298	1643370	72	reverse	0	0	0	1	1

ydfB	rzpQ	b1572	b1573	1646817	1646847	30	forward	0	0	0	1	1
ydfD	ydfE	b1576	b1577	1648009	1648102	93	forward	0	0	0	1	1
ynfB	speG	b1583	b1584	1654173	1654208	35	forward	0	0	0	1	1
ynfE	ynfF	b1587	b1588	1658519	1658580	61	forward	0	0	0	1	1
ynfH	dmsD	b1590	b1591	1662487	1662530	43	forward	0	0	0	1	1
ynfK	dgsA	b1593	b1594	1665243	1665368	125	reverse	0	1	0	0	1
ydgI	folM	b1605	b1606	1678963	1679000	37	forward	0	0	0	1	1
rstB	tus	b1609	b1610	1682207	1682283	76	forward	0	0	0	1	0
uidC	uidB	b1615	b1616	1690875	1690914	39	reverse	0	0	0	0	1
ydgK	rsxA	b1626	b1627	1703714	1703791	77	forward	0	1	0	0	1
pdxY	tyrS	b1636	b1637	1713913	1713972	59	reverse	0	1	0	1	1
tyrS	pdxH	b1637	b1638	1715246	1715375	129	reverse	1	1	0	1	1
sodC	ydhF	b1646	b1647	1722679	1722760	81	reverse	1	1	1	1	0
ydhF	ydhL	b1647	b1648	1723656	1723705	49	reverse	0	1	0	1	0
nemR	nemA	b1649	b1650	1724646	1724683	37	forward	0	0	0	1	1
nemA	gloA	b1650	b1651	1725780	1725861	81	forward	0	1	1	1	0
rnt	lhr	b1652	b1653	1727018	1727111	93	forward	0	0	0	0	1
ydiB	aroD	b1692	b1693	1772679	1772710	31	forward	0	0	0	1	1
ydiR	ydiS	b1698	b1699	1779363	1779419	56	forward	0	0	0	1	1
ydiU	ydiV	b1706	b1707	1789268	1789331	63	reverse	0	0	0	1	0
nlpC	btuD	b1708	b1709	1790755	1790833	78	reverse	0	1	0	1	0
btuE	btuC	b1710	b1711	1792133	1792196	63	reverse	0	0	0	1	1
pheS	pheM	b1714	b1715	1796966	1797250	284	reverse	0	0	0	1	1
pheM	rplT	b1715	b1716	1797294	1797417	123	reverse	0	1	0	0	1
rplT	rpmI	b1716	b1717	1797773	1797826	53	reverse	0	0	0	0	1
rpmI	infC	b1717	b1718	1798023	1798120	97	reverse	0	1	0	1	1
chbF	chbR	b1734	b1735	1816524	1816629	105	reverse	0	0	0	1	1
chbA	chbC	b1736	b1737	1817829	1817880	51	reverse	0	0	0	1	1
chbC	chbB	b1737	b1738	1819238	1819323	85	reverse	0	0	0	1	1

selD	ydjA	b1764	b1765	1846032	1846149	117	reverse	0	0	0	1	1
yeaC	msrB	b1777	b1778	1859998	1860040	42	reverse	0	0	0	1	0
gapA	yeaD	b1779	b1780	1861790	1861874	84	forward	0	1	0	0	1
yeaG	yeaH	b1783	b1784	1866866	1866979	113	forward	0	0	0	1	1
yeaJ	yeaK	b1786	b1787	1871555	1871598	43	forward	0	1	0	1	0
yeaN	yeaO	b1791	b1792	1874878	1874933	55	forward	0	1	0	1	0
yeaV	yeaW	b1801	b1802	1882657	1882689	32	forward	0	0	0	1	1
yeaY	yeaZ	b1806	b1807	1888556	1888596	40	reverse	0	0	1	1	1
yeaZ	yoaA	b1807	b1808	1889291	1889349	58	reverse	0	0	0	1	0
manX	manY	b1817	b1818	1901043	1901106	63	forward	0	0	0	1	1
manZ	yobD	b1819	b1820	1902770	1902825	55	forward	0	0	0	1	0
yebV	yebW	b1836	b1837	1920040	1920145	105	forward	0	1	0	1	0
yebF	yebG	b1847	b1848	1928414	1928481	67	reverse	0	1	0	1	0
eda	edd	b1850	b1851	1930780	1930817	37	reverse	0	0	0	1	1
ruvC	yebC	b1863	b1864	1945400	1945435	35	reverse	0	0	0	1	1
yecE	yecN	b1868	b1869	1950237	1950290	53	forward	0	1	0	1	0
yecN	cmoA	b1869	b1870	1950685	1950726	41	forward	0	0	0	0	1
cutC	yecM	b1874	b1875	1956984	1957304	320	reverse	0	0	1	1	0
tap	tar	b1885	b1886	1969008	1969054	46	reverse	0	0	0	1	1
araG	araF	b1900	b1901	1983093	1983163	70	reverse	0	0	0	1	1
cysT	glyW	b1910	b1911	1990011	1990066	55	reverse	0	0	0	1	1
pgsA	uvrC	b1912	b1913	1990841	1990898	57	reverse	0	1	0	1	0
fliY	fliZ	b1920	b1921	1998409	1998497	88	reverse	0	1	0	0	1
fliZ	fliA	b1921	b1922	1999048	1999094	46	reverse	0	0	0	1	1
fliT	amyA	b1926	b1927	2004102	2004180	78	forward	0	1	1	1	0
yodC	yedI	b1957	b1958	2026394	2026473	79	reverse	0	0	0	1	0
dcm	yedJ	b1961	b1962	2030341	2030408	67	reverse	0	1	0	1	0
yedJ	yedR	b1962	b1963	2031103	2031143	40	reverse	0	0	0	1	1
mtfA	asnT	b1976	b1977	2042472	2042573	101	forward	0	0	1	1	0

asnW	yeeO	b1984	b1985	2056126	2056227	101	reverse	0	1	0	1	0
yeeS	yeeT	b2002	b2003	2074778	2074841	63	forward	0	0	0	1	1
yeeT	yeeU	b2003	b2004	2075062	2075136	74	forward	0	0	0	1	1
yeeU	yeeV	b2004	b2005	2075504	2075593	89	forward	0	0	0	1	1
yeeY	yeeZ	b2015	b2016	2086282	2086328	46	reverse	0	0	0	1	0
hisL	hisG	b2018	b2019	2088070	2088216	146	forward	0	1	0	0	1
rfbA	rfbD	b2039	b2040	2109043	2109101	58	reverse	0	0	0	1	1
wcaJ	cpsG	b2047	b2048	2119578	2119633	55	reverse	0	0	0	1	1
cpsG	cpsB	b2048	b2049	2121003	2121108	105	reverse	0	0	0	1	1
wcaA	wzc	b2059	b2060	2131421	2131514	93	reverse	0	0	0	1	1
dcd	udk	b2065	b2066	2140239	2140331	92	reverse	0	0	0	1	1
yegZ	yegR		b2085	2165844	2166013	169	reverse	0	0	0	0	1
gatD	gatC	b2091	b2092	2170897	2170945	48	reverse	0	0	0	0	1
gatB	gatA	b2093	b2094	2172588	2172619	31	reverse	0	0	0	1	1
yegW	yegX	b2101	b2102	2180803	2180855	52	reverse	0	0	0	1	0
yegX	thiD	b2102	b2103	2181673	2181738	65	reverse	0	0	0	1	0
yehC	yehD	b2110	b2111	2189667	2189702	35	reverse	0	0	0	1	1
yehI	yehK	b2118	b4541	2201933	2201994	61	forward	0	0	0	1	0
yehS	yehT	b2124	b2125	2210218	2210265	47	reverse	0	0	0	1	0
mlrA	yohO	b2127	b4542	2213619	2213679	60	forward	0	0	0	1	0
mglA	mglB	b2149	b2150	2237311	2237372	61	reverse	0	0	0	1	1
yeiH	nfo	b2158	b2159	2248788	2248862	74	forward	0	0	1	1	0
yeiM	pscG	b2164	b2165	2255357	2255451	94	reverse	0	0	0	1	0
yeiR	lpxT	b2173	b2174	2266837	2266876	39	forward	0	0	0	1	1
rtn	yejA	b2176	b2177	2270304	2270386	82	forward	0	0	1	1	0
yejM	proL	b2188	b2189	2284158	2284233	75	forward	0	1	1	1	0
ccmC	ccmB	b2199	b2200	2294342	2294384	42	reverse	0	0	0	1	1
yojI	alkB	b2211	b2212	2306637	2306713	76	reverse	0	0	1	1	0
atoE	atoB	b2223	b2224	2324100	2324131	31	forward	0	0	0	1	1

nrdA	nrdB	b2234	b2235	2345172	2345406	234	forward	0	1	0	0	1
yfaV	yfaW	b2246	b2247	2358174	2358231	57	reverse	0	0	0	1	1
menD	menF	b2264	b2265	2377281	2377370	89	reverse	0	0	0	1	1
elaB	elaA	b2266	b2267	2379049	2379104	55	reverse	0	1	0	1	0
nuoM	nuoL	b2277	b2278	2391063	2391227	164	reverse	0	0	0	1	1
nuoG	nuoF	b2283	b2284	2398187	2398240	53	reverse	0	0	0	1	1
nuoC	nuoB	b2286	b2287	2401867	2401973	106	reverse	0	0	0	1	1
yfbQ	yfbR	b2290	b2291	2406800	2406884	84	forward	0	1	0	1	0
ackA	pta	b2296	b2297	2412694	2412769	75	forward	0	1	0	1	1
yfcD	yfcE	b2299	b2300	2417198	2417256	58	reverse	0	0	0	1	0
yfcG	folX	b2302	b2303	2419290	2419347	57	forward	0	0	0	1	0
hisQ	hisJ	b2308	b2309	2423938	2424028	90	reverse	0	0	0	1	1
hisJ	argT	b2309	b2310	2424810	2425031	221	reverse	0	1	1	1	1
ubiX	purF	b2311	b2312	2426648	2426743	95	reverse	0	0	0	1	1
purF	cvpA	b2312	b2313	2428260	2428297	37	reverse	0	0	0	1	1
folC	accD	b2315	b2316	2430964	2431034	70	reverse	0	0	0	1	0
dedA	truA	b2317	b2318	2432763	2432846	83	reverse	0	0	0	1	1
usg	pdxB	b2319	b2320	2434671	2434737	66	reverse	0	0	0	1	1
yfcL	yfcM	b2325	b2326	2442191	2442225	34	reverse	0	0	0	1	1
aroC	prmB	b2329	b2330	2445495	2445530	35	reverse	0	0	0	1	1
yfdC	argW	b2347	b2348	2464255	2464331	76	forward	0	1	1	1	0
yfdL	yfdM	b2355	b2356	2470084	2470134	50	reverse	0	0	0	0	1
yfdP	yfdQ	b2359	b2360	2471988	2472054	66	forward	0	0	0	0	1
yfdV	oxc	b2372	b2373	2488208	2488278	70	reverse	0	1	0	1	0
oxc	frc	b2373	b2374	2489972	2490026	54	reverse	0	0	0	1	0
alaX	alaW	b2396	b2397	2516138	2516178	40	reverse	0	0	0	1	1
valU	valX	b2401	b2402	2519028	2519073	45	forward	0	0	0	1	1
valX	valY	b2402	b2403	2519148	2519195	47	forward	0	0	0	1	1
xapB	xapA	b2406	b2407	2522007	2522067	60	reverse	0	0	0	1	1

ligA	zipA	b2411	b2412	2528198	2528269	71	reverse	1	1	0	1	0
ptsH	ptsI	b2415	b2416	2532043	2532088	45	forward	0	0	0	1	1
ptsI	crr	b2416	b2417	2533815	2533856	41	forward	0	0	0	1	1
yfeK	yfeS	b2419	b2420	2535738	2535771	33	forward	0	0	0	1	1
cysM	cysA	b2421	b2422	2537605	2537739	134	reverse	0	0	0	0	1
yfeX	yfeY	b2431	b2432	2548567	2548663	96	reverse	0	0	0	1	1
eutR	eutK	b2437	b2438	2553204	2553250	46	reverse	0	0	0	1	1
eutM	eutD	b2457	b2458	2570472	2570511	39	reverse	0	0	0	1	1
nudK	aegA	b2467	b2468	2581500	2581568	68	reverse	0	0	0	1	0
ypfH	tmcA	b2473	b2474	2591792	2591866	74	reverse	0	0	0	1	0
hyfJ	hyfR	b2490	b2491	2609892	2609922	30	forward	0	0	0	1	1
hda	uraA	b2496	b2497	2616843	2616893	50	reverse	0	0	0	1	0
uraA	upp	b2497	b2498	2618182	2618268	86	reverse	0	0	0	0	1
guaA	guaB	b2507	b2508	2630557	2630626	69	reverse	0	0	0	1	1
yfgJ	der	b2510	b2511	2633836	2633906	70	reverse	0	1	0	1	0
der	bamB	b2511	b2512	2635378	2635496	118	reverse	0	0	0	0	1
hscB	iscA	b2527	b2528	2657489	2657585	96	reverse	0	0	1	1	0
iscS	iscR	b2530	b2531	2659553	2659665	112	reverse	0	0	0	1	1
yphC	yphD	b2545	b2546	2673783	2673849	66	reverse	0	0	0	1	0
glnB	yfhA	b2553	b2554	2685430	2685491	61	reverse	0	0	1	1	0
yfhK	glmY	b2556	b4441	2689120	2689179	59	reverse	0	0	0	1	0
tadA	yfhB	b2559	b2560	2695879	2695937	58	reverse	0	0	0	1	1
yfhH	yfhL	b2561	b2562	2697629	2697685	56	forward	0	0	0	1	0
rseA	rpoE	b2572	b2573	2707426	2707459	33	reverse	0	1	0	1	1
yfiP	yfiQ	b2583	b2584	2717943	2717975	32	forward	0	0	0	1	1
yfiQ	pssA	b2584	b2585	2720635	2720749	114	forward	0	1	0	1	0
pssA	yfiM	b2585	b2586	2722104	2722201	97	forward	0	0	0	1	0
rrfG	rrlG	b2588	b2589	2724210	2724303	93	reverse	0	1	0	1	1
rrlG	gltW	b2589	b2590	2727206	2727391	185	reverse	0	0	0	1	1

gltW	rrsG	b2590	b2591	2727466	2727638	172	reverse	0	0	0	1	1
pheL	pheA	b2598	b2599	2735668	2735767	99	forward	0	0	0	0	1
rplS	trmD	b2606	b2607	2742552	2742594	42	reverse	0	0	0	1	1
trmD	rimM	b2607	b2608	2743361	2743392	31	reverse	0	0	0	1	1
ypjD	yfjD	b2611	b4461	2746775	2746886	111	forward	0	0	0	1	0
yfjP	yfjQ	b2632	b2633	2766595	2766687	92	forward	0	0	0	0	1
yfjR	ypjK	b2634	b2635	2768426	2768467	41	forward	0	0	0	1	1
ypjJ	yfjZ	b4548	b2645	2775099	2775137	38	forward	0	0	0	1	1
gabT	gabP	b2662	b2663	2792037	2792275	238	forward	0	0	0	1	1
ygaU	yqaE	b2665	b2666	2794808	2794892	84	reverse	0	1	0	1	0
proW	proX	b2678	b2679	2805096	2805154	58	forward	0	0	0	1	1
ygaX	ygaY	b2680	b2681	2806604	2806634	30	forward	0	0	0	0	1
mprA	emrA	b2684	b2685	2809322	2809449	127	forward	0	0	0	0	1
gshA	yqaA	b2688	b2689	2814461	2814534	73	reverse	0	0	0	1	0
argQ	argZ	b2691	b2692	2815882	2816081	199	reverse	0	0	0	0	1
argZ	argY	b2692	b2693	2816157	2816220	63	reverse	0	0	0	1	1
argY	argV	b2693	b2694	2816296	2816495	199	reverse	0	0	0	1	1
recX	recA	b2698	b2699	2820661	2820730	69	reverse	0	1	0	0	1
srlD	gutM	b2705	b2706	2826538	2826643	105	forward	0	1	0	1	1
gutM	srlR	b2706	b2707	2827002	2827069	67	forward	0	0	0	1	1
hypF	hydN	b2712	b2713	2835447	2835600	153	reverse	0	1	0	0	1
hycB	hycA	b2724	b2725	2847871	2847996	125	reverse	0	0	0	1	1
hypD	hypE	b2729	b2730	2851279	2851318	39	forward	0	0	0	1	1
hypE	fhlA	b2730	b2731	2852286	2852360	74	forward	0	1	0	1	1
rpoS	nlpD	b2741	b2742	2865573	2865636	63	reverse	1	0	0	1	1
ygbE	cysC	b2749	b2750	2871359	2871409	50	reverse	0	1	0	1	0
cysH	cysI	b2762	b2763	2886334	2886409	75	reverse	0	0	0	1	1
queD	ygcN	b2765	b2766	2890601	2890679	78	forward	0	0	0	1	0
ygcS	ygcU	b2771	b4463	2895892	2895986	94	reverse	0	0	0	1	0

eno	pyrG	b2779	b2780	2905963	2906051	88	reverse	1	0	0	1	1
mazG	chpA	b2781	b2782	2908707	2908778	71	reverse	0	0	0	1	1
chpR	relA	b2783	b2784	2909361	2909439	78	reverse	0	0	0	1	1
relA	rlmD	b2784	b2785	2911673	2911721	48	reverse	0	0	0	1	0
sdaC	sdaB	b2796	b2797	2927540	2927598	58	forward	0	0	0	1	1
fucP	fucI	b2801	b2802	2933573	2933606	33	forward	0	0	0	1	1
fucI	fucK	b2802	b2803	2935381	2935460	79	forward	0	0	0	1	1
fucU	fucR	b2804	b2805	2937332	2937390	58	forward	0	0	0	1	1
metZ	metW	b2814	b2815	2945485	2945519	34	forward	0	1	0	1	1
metW	metV	b2815	b2816	2945595	2945629	34	forward	0	0	0	1	1
kduD	kduI	b2842	b2843	2981280	2981310	30	reverse	0	0	0	1	0
ygeK	ygeL	b2855	b2856	2992925	2992959	34	reverse	0	0	0	0	1
glyU	ygeR	b2864	b2865	2997079	2997158	79	reverse	0	0	1	1	0
ygeW	ygeX	b2870	b2871	3005474	3005532	58	forward	0	0	0	1	0
ygeX	ygeY	b2871	b2872	3006728	3006786	58	forward	0	0	0	1	0
ygeY	hyuA	b2872	b2873	3007997	3008050	53	forward	0	0	0	1	0
hyuA	yqeA	b2873	b2874	3009435	3009483	48	forward	0	0	0	1	0
yqeB	yqeC	b2875	b2876	3012261	3012309	48	reverse	0	0	0	1	0
ssnA	ygfM	b2879	b2880	3018511	3018562	51	forward	0	0	0	1	0
guaD	ygfQ	b2883	b4464	3025107	3025143	36	forward	0	0	0	1	1
prfB	recJ	b2891	b2892	3034304	3034395	91	reverse	0	0	0	1	1
gcvP	gcvH	b2903	b2904	3047063	3047182	119	reverse	0	0	0	1	1
zapA	ssrS	b2910	b2911	3053963	3054005	42	forward	0	0	1	1	0
ssrS	ygfA	b2911	b2912	3054187	3054263	76	forward	0	0	0	1	1
ygfA	sibC	b2912	b4446	3054811	3054873	62	forward	0	0	0	1	0
yggE	argO	b2922	b2923	3066102	3066195	93	reverse	0	1	1	1	0
fbaA	pgk	b2925	b2926	3069266	3069481	215	reverse	0	1	0	0	1
pgk	epd	b2926	b2927	3070644	3070694	50	reverse	0	0	0	1	1
speB	speA	b2937	b2938	3081819	3081957	138	reverse	0	0	0	0	1

galP	yggI	b2943	b2944	3087700	3087777	77	forward	0	0	0	1	0
yggI	endA	b2944	b2945	3088274	3088369	95	forward	0	0	0	1	0
yggX	mltC	b2962	b2963	3102390	3102455	65	forward	0	0	0	1	1
mltC	nupG	b2963	b2964	3103534	3103736	202	forward	0	1	1	1	1
yghG	pppA	b2971	b2972	3111499	3111565	66	reverse	0	0	0	1	0
glcA	glcB	b2975	b2976	3119301	3119656	355	reverse	0	1	0	0	1
yghR	yghS	b2984	b2985	3131234	3131266	32	reverse	0	0	0	1	1
yqhD	dkgA	b3011	b3012	3154540	3154645	105	forward	1	0	0	1	0
yqhG	yqhH	b3013	b3014	3156598	3156649	51	forward	0	0	0	1	0
ftsP	plsC	b3017	b3018	3160691	3160766	75	reverse	0	0	0	1	0
ygiV	ygiW	b3023	b3024	3167253	3167306	53	reverse	0	0	0	1	0
mdaB	ygiN	b3028	b3029	3171133	3171164	31	forward	0	1	0	1	0
ribB	sroG	b3041		3182488	3182592	104	reverse	0	0	0	0	1
rfaE	glnE	b3052	b3053	3194775	3194823	48	reverse	0	0	0	1	0
htrG	cca	b3055	b3056	3199849	3199913	64	forward	0	0	0	1	1
ttdB	ttdT	b3062	b3063	3205998	3206046	48	forward	0	0	0	1	1
rpsU	dnaG	b3065	b3066	3209018	3209129	111	forward	0	0	0	1	1
dnaG	rpoD	b3066	b3067	3210874	3211069	195	forward	0	0	0	1	1
ebgC	ygjI	b3077	b3078	3224193	3224256	63	forward	0	0	0	1	0
ygjP	ygjQ	b3085	b3086	3234485	3234562	77	forward	0	0	0	1	0
ygjQ	ygjR	b3086	b3087	3235254	3235315	61	forward	1	0	0	1	0
yqjC	yqjD	b3097	b3098	3247359	3247397	38	forward	0	0	0	1	1
tdcG	tdcF	b4471	b3113	3257671	3257743	72	reverse	0	0	0	1	1
tdcE	tdcD	b3114	b3115	3260440	3260474	34	reverse	0	0	0	1	1
tdcB	tdcA	b3117	b3118	3264050	3264149	99	reverse	0	0	0	1	1
rnpB	garK	b3123	b3124	3268614	3268647	33	reverse	0	0	0	1	1
garR	garL	b3125	b3126	3270779	3270809	30	reverse	0	0	0	1	1
kbaY	agaB	b3137	b3138	3282025	3282192	167	forward	0	0	0	1	1
agaB	agaC	b3138	b3139	3282668	3282707	39	forward	0	0	0	1	1

yhbV	yhbW	b3159	b3160	3301389	3301470	81	forward	0	1	0	1	0
nlpI	pnp	b3163	b3164	3306946	3307055	109	reverse	0	1	0	1	0
pnp	rpsO	b3164	b3165	3309190	3309437	247	reverse	0	1	1	0	1
rpsO	truB	b3165	b3166	3309706	3309855	149	reverse	0	1	0	0	1
rbfA	infB	b3167	b3168	3311200	3311364	164	reverse	0	1	0	1	1
rimP	metY	b3170	b3171	3316028	3316235	207	reverse	0	1	1	1	1
hflB	rlmE	b3178	b3179	3324957	3325057	100	reverse	0	1	0	0	1
murA	yrbA	b3189	b3190	3334516	3334571	55	reverse	0	0	0	1	0
lptB	rpoN	b3201	b3202	3342691	3342739	48	forward	0	1	0	1	1
hpf	ptsN	b3203	b3204	3344482	3344600	118	forward	0	0	0	1	1
ptsN	yhbJ	b3204	b3205	3345091	3345137	46	forward	0	0	0	1	1
arcB	yhcC	b3210	b3211	3351047	3351143	96	reverse	1	1	0	1	0
gltD	gltF	b3213	b3214	3358638	3359198	560	forward	0	0	0	1	1
nanE	nanT	b3223	b3224	3369058	3369106	48	reverse	0	0	0	1	1
nanT	nanA	b3224	b3225	3370596	3370705	109	reverse	0	0	0	1	1
degQ	degS	b3234	b3235	3380132	3380222	90	forward	0	0	0	0	1
mreC	mreB	b3250	b3251	3398000	3398066	66	reverse	0	0	0	1	1
yhdJ	yhdU	b3262	b3263	3410559	3410643	84	forward	0	0	0	1	0
yhdW	yhdX	b3268	b3269	3418088	3418156	68	forward	0	0	0	0	1
rrfF	thrV	b3272	b3273	3421564	3421602	38	reverse	0	0	0	1	1
rrfD	rrlD	b3274	b3275	3421809	3421902	93	reverse	0	0	0	1	1
rrlD	alaU	b3275	b3276	3424805	3424980	175	reverse	0	0	0	1	1
alaU	ileU	b3276	b3277	3425055	3425098	43	reverse	0	0	0	1	1
ileU	rrsD	b3277	b3278	3425174	3425243	69	reverse	0	0	0	1	1
fmt	rsmB	b3288	b3289	3433183	3433229	46	forward	0	0	0	1	0
rplQ	rpoA	b3294	b3295	3438021	3438062	41	reverse	0	0	0	1	1
rpsD	rpsK	b3296	b3297	3439697	3439731	34	reverse	0	0	0	1	1
rpmJ	secY	b3299	b3300	3440756	3440788	32	reverse	0	0	0	1	1
rpsH	rpsN	b3306	b3307	3444567	3444601	34	reverse	0	0	0	1	1

rplC	rpsJ	b3320	b3321	3450948	3450981	33	reverse	0	0	0	1	1
bfr	bfd	b3336	b3337	3464747	3464819	72	reverse	0	1	0	0	1
tufA	fusA	b3339	b3340	3469351	3469422	71	reverse	0	0	0	1	1
rpsG	rpsL	b3341	b3342	3472103	3472200	97	reverse	0	0	0	1	1
yheU	prkB	b3354	b3355	3482458	3482512	54	forward	0	0	0	1	0
crp	yhfK	b3357	b3358	3484774	3484813	39	forward	0	0	0	1	0
pabA	fic	b3360	b3361	3488851	3488883	32	reverse	0	0	1	1	1
nirD	nirC	b3366	b3367	3494899	3495025	126	forward	0	0	0	1	1
frlB	frlC	b3371	b4474	3500312	3500362	50	forward	0	0	0	1	1
frlD	frlR	b3374	b3375	3501974	3502074	100	forward	0	0	0	1	1
yhfX	yhfY	b3381	b3382	3508998	3509082	84	reverse	0	0	0	1	0
dam	damX	b3387	b3388	3513935	3514042	107	reverse	0	0	1	1	1
damX	aroB	b3388	b3389	3515328	3515420	92	reverse	0	0	0	1	1
aroB	aroK	b3389	b3390	3516508	3516565	57	reverse	0	0	0	1	1
yrfF	yrfG	b3398	b3399	3526626	3526691	65	forward	0	0	0	1	0
gntX	nfuA	b3413	b3414	3543587	3543646	59	forward	0	0	0	1	1
glpG	glpE	b3424	b3425	3559475	3559520	45	reverse	0	0	0	1	1
gntK	gntR	b3437	b3438	3575615	3575754	139	reverse	0	1	1	1	1
ugpA	ugpB	b3452	b3453	3588934	3589032	98	reverse	0	0	0	1	1
livH	livK	b3457	b3458	3594426	3594474	48	reverse	0	0	0	1	1
yhhN	zntA	b3468	b3469	3604400	3604474	74	forward	0	1	1	1	0
yhhQ	dcrB	b3471	b3472	3607905	3607978	73	forward	0	0	0	1	0
yhhT	acpT	b3474	b3475	3610937	3610992	55	forward	0	0	0	1	0
yhiR	gor	b3499	b3500	3644250	3644322	72	forward	0	0	0	1	0
arsR	arsB	b3501	b3502	3646904	3646958	54	forward	0	0	0	1	1
slp	dctR	b3506	b3507	3652550	3652706	156	forward	0	1	0	0	1
yhiD	hdeB	b3508	b3509	3653925	3653989	64	reverse	0	0	0	0	1
hdeB	hdeA	b3509	b3510	3654315	3654431	116	reverse	0	0	0	0	1
gadE	mdtE	b3512	b3513	3656916	3657255	339	forward	0	1	0	0	1

gadW	gadX	b3515	b3516	3662641	3663009	368	reverse	0	1	0	0	1
gadX	gadA	b3516	b3517	3663833	3664203	370	reverse	0	0	1	0	1
yhjC	yhjD	b3521	b3522	3671336	3671385	49	forward	0	0	0	1	0
yhjG	yhjH	b3524	b3525	3676388	3676443	55	reverse	0	0	0	1	0
dppB	dppA	b3543	b3544	3703813	3704121	308	reverse	0	0	0	1	1
proK	eptB	b3545	b3546	3706715	3706807	92	reverse	0	0	1	1	0
yiaD	ghrB	b3552	b3553	3715229	3715333	104	forward	0	1	0	1	0
xylB	xylA	b3564	b3565	3727394	3727466	72	reverse	0	0	0	0	1
xylF	xylG	b3566	b3567	3730146	3730224	78	forward	0	0	0	1	1
xylH	xylR	b3568	b3569	3732924	3733002	78	forward	0	0	0	1	1
ysaA	yiaJ	b3573	b3574	3739605	3739707	102	reverse	0	0	0	1	0
yiaL	yiaM	b3576	b3577	3742233	3742351	118	forward	0	0	0	1	1
selA	yibF	b3591	b3592	3759272	3759370	98	reverse	0	1	0	1	0
mtlA	mtlD	b3599	b3600	3772217	3772447	230	forward	0	1	0	0	1
cysE	gpsA	b3607	b3608	3780585	3780665	80	reverse	0	1	0	1	0
secB	grxC	b3609	b3610	3782151	3782214	63	reverse	0	1	0	1	0
gpmM	envC	b3612	b3613	3784827	3784861	34	forward	0	0	0	1	1
waaU	rfaZ	b3623	b3624	3797335	3797368	33	reverse	0	0	0	1	1
rfaZ	rfaY	b3624	b3625	3798219	3798290	71	reverse	0	0	0	1	1
rfaJ	rfaI	b3626	b3627	3800022	3800062	40	reverse	0	0	0	1	1
rfaS	rfaP	b3629	b3630	3803139	3803176	37	reverse	0	0	0	1	1
mutM	rpmG	b3635	b3636	3809175	3809273	98	reverse	0	1	1	0	1
rpmB	yicR	b3637	b3638	3809697	3809914	217	reverse	0	1	0	1	1
dut	slmA	b3640	b3641	3812410	3812517	107	forward	0	0	0	0	1
pyrE	rph	b3642	b3643	3813791	3813886	95	reverse	0	0	0	0	1
gmk	rpoZ	b3648	b3649	3820074	3820129	55	forward	0	1	1	1	0
yicN	yicO	b3663	b3664	3840425	3840478	53	reverse	0	0	0	1	0
ilvB	ivbL	b3671	b3672	3850807	3850913	106	reverse	0	0	0	1	1
ibpB	ibpA	b3686	b3687	3864920	3865032	112	reverse	0	1	1	0	1

dgoT	dgoD	b3691	b4478	3869753	3869873	120	reverse	0	0	0	0	1
tnaC	tnaA	b3707	b3708	3886532	3886753	221	forward	0	1	0	0	1
tnaA	tnaB	b3708	b3709	3888168	3888259	91	forward	0	1	0	0	1
yieH	cbrB	b3715	b3716	3895462	3895529	67	forward	0	0	0	1	0
cbrB	cbrC	b3716	b3717	3895996	3896045	49	forward	0	0	0	1	0
bglH	bglB	b3720	b3721	3900243	3900312	69	reverse	0	0	0	1	0
bglF	bglG	b3722	b3723	3903620	3903754	134	reverse	0	0	0	1	1
pstB	pstA	b3725	b3726	3906389	3906572	183	reverse	0	0	1	1	1
pstC	pstS	b3727	b3728	3908421	3908508	87	reverse	0	0	0	1	1
glmS	glmU	b3729	b3730	3911691	3911853	162	reverse	0	0	0	0	1
atpG	atpA	b3733	b3734	3916288	3916339	51	reverse	0	0	0	1	1
atpF	atpE	b3736	b3737	3918911	3918973	62	reverse	0	0	0	1	1
atpE	atpB	b3737	b3738	3919212	3919259	47	reverse	0	0	0	1	1
rsmG	mnmG	b3740	b3741	3921703	3921767	64	reverse	0	0	0	1	1
mnmG	mioC	b3741	b3742	3923656	3924035	379	reverse	0	1	0	1	1
mioC	asnC	b3742	b3743	3924478	3924568	90	reverse	1	1	0	0	1
rbsB	rbsK	b3751	b3752	3935191	3935317	126	forward	0	0	0	1	1
rrsC	gltU	b3756	b3757	3941372	3941458	86	forward	0	0	0	1	1
gltU	rrlC	b3757	b3758	3941533	3941727	194	forward	0	0	0	1	1
rrlC	rrfC	b3758	b3759	3944630	3944723	93	forward	0	0	0	1	1
rrfC	aspT	b3759	b3760	3944842	3944895	53	forward	0	0	0	1	0
ilvL	ilvX	b3766	b4669	3948443	3948530	87	forward	0	0	0	0	1
ilvG_1	ilvG_2			3949566	3949646	80	forward	0	0	0	0	1
ilvE	ilvD	b3770	b3771	3951436	3951501	65	forward	0	0	0	1	1
rhoL	rho	b3782	b3783	3964355	3964440	85	forward	0	0	0	0	1
wzzE	rffE	b3785	b3786	3968100	3968156	56	forward	0	1	0	1	1
rffH	rffC	b3789	b3790	3972512	3972619	107	forward	0	0	0	1	1
yifK	argX	b3795	b3796	3980295	3980398	103	forward	0	1	1	1	0
argX	hisR	b3796	b3797	3980474	3980532	58	forward	0	0	0	1	1

leuT	proM	b3798	b3799	3980715	3980758	43	forward	0	0	0	1	1
yifL	dapF	b4558	b3809	3992748	3992785	37	forward	0	0	0	1	1
rarD	yigI	b3819	b3820	4002201	4002253	52	reverse	0	0	0	1	0
recQ	rhtC	b3822	b3823	4005716	4005780	64	forward	0	0	0	1	0
yigL	yigM	b3826	b3827	4009023	4009099	76	forward	0	0	0	1	0
tatC	tatD	b3839	b4483	4021535	4021577	42	forward	0	0	0	1	1
ubiD	fre	b3843	b3844	4024504	4024550	46	forward	1	0	0	1	0
yigZ	trkH	b3848	b3849	4031129	4031168	39	forward	0	0	0	1	1
rrsA	ileT	b3851	b3852	4035095	4035164	69	forward	0	0	0	1	1
ileT	alaT	b3852	b3853	4035240	4035283	43	forward	0	0	0	1	1
alaT	rrlA	b3853	b3854	4035358	4035542	184	forward	0	0	0	1	1
rrlA	rrfA	b3854	b3855	4038446	4038540	94	forward	0	0	0	1	1
yihD	rdoA	b3858	b3859	4040361	4040438	77	forward	0	1	1	1	0
glnL	glnA	b3869	b3870	4054362	4054648	286	reverse	0	0	1	1	1
ompL	yihO	b3875	b3876	4062318	4062386	68	reverse	0	0	0	1	0
yihO	yihP	b3876	b3877	4063789	4063832	43	reverse	0	0	0	1	1
yihV	yihW	b3883	b3884	4072658	4072692	34	forward	0	0	0	0	1
rhaD	rhaA	b3902	b3903	4092295	4092746	451	reverse	0	0	0	0	1
rhaS	rhaR	b3905	b3906	4096595	4096669	74	forward	0	0	0	1	1
glpX	glpK	b3925	b3926	4113602	4113737	135	reverse	0	0	0	0	1
rraA	menA	b3929	b3930	4117353	4117446	93	reverse	1	0	1	0	1
menA	hslU	b3930	b3931	4118372	4118439	67	reverse	1	1	1	1	0
hslV	ftsN	b3932	b3933	4120310	4120403	93	reverse	0	0	1	1	0
katG	yijE	b3942	b3943	4134038	4134131	93	forward	0	1	0	1	0
frwB	pflD	b3950	b3951	4141967	4142018	51	forward	0	0	0	1	0
argB	argH	b3959	b3960	4154812	4154873	61	forward	0	0	0	1	1
rrsB	gltT	b3968	b3969	4166223	4166395	172	forward	0	0	0	1	1
gltT	rrlB	b3969	b3970	4166470	4166664	194	forward	0	0	0	1	1
rrlB	rrfB	b3970	b3971	4169567	4169660	93	forward	0	0	0	1	1

tyrU	glyT	b3977	b3978	4173579	4173696	117	forward	0	0	0	1	1
thrT	tufB	b3979	b3980	4173852	4173967	115	forward	0	0	0	1	1
rplA	rplJ	b3984	b3985	4177606	4178019	413	forward	0	0	0	1	1
rplJ	rplL	b3985	b3986	4178516	4178583	67	forward	0	0	0	1	1
rplL	rpoB	b3986	b3987	4178948	4179268	320	forward	0	1	0	1	1
rpoB	rpoC	b3987	b3988	4183296	4183373	77	forward	0	0	0	1	1
nudC	hemE	b3996	b3997	4195699	4195739	40	forward	0	0	0	1	1
nfi	yjaG	b3998	b3999	4197484	4197527	43	forward	0	0	0	1	0
rrsE	gltV	b4007	b4008	4207711	4207797	86	forward	0	0	0	1	1
gltV	rrlE	b4008	b4009	4207872	4208066	194	forward	0	0	0	1	1
rrlE	rrfE	b4009	b4010	4210969	4211063	94	forward	0	0	0	1	1
aceB	aceA	b4014	b4015	4215102	4215132	30	forward	0	0	0	1	1
aceA	aceK	b4015	b4016	4216436	4216619	183	forward	0	1	0	1	1
yjbE	yjbF	b4026	b4027	4234171	4234285	114	forward	0	0	0	0	1
malF	malE	b4033	b4034	4243098	4243252	154	reverse	0	0	0	1	1
malK	lamB	b4035	b4036	4245922	4245994	72	forward	0	0	0	1	1
lamB	malM	b4036	b4037	4247334	4247577	243	forward	0	0	0	1	1
dnaB	alr	b4052	b4053	4263752	4263805	53	forward	0	0	0	1	0
yjcH	acs	b4068	b4069	4283236	4283436	200	reverse	0	0	0	1	1
nrfA	nrfB	b4070	b4071	4287223	4287268	45	forward	0	0	0	1	1
nrfD	nrfE	b4073	b4074	4289455	4289535	80	forward	0	0	0	1	1
alsA	alsB	b4087	b4088	4309003	4309130	127	reverse	0	1	0	1	1
alsB	rpiR	b4088	b4089	4310065	4310124	59	reverse	0	0	0	1	1
phnL	phnK	b4096	b4097	4315918	4316029	111	reverse	0	0	0	1	1
phnF	phnE	b4102	b4104	4320445	4320684	239	reverse	0	0	0	1	1
phnE	phnD	b4104	b4105	4321304	4321359	55	reverse	0	0	0	0	1
phnE	phnD		b4105	4321304	4321359	55	reverse	0	0	0	0	1
melA	melB	b4119	b4120	4341289	4341404	115	forward	0	0	0	0	1
fumB	dcuB	b4122	b4123	4345349	4345427	78	reverse	0	0	1	1	1

cadA	cadB	b4131	b4132	4356640	4356720	80	reverse	0	0	0	0	1
pheU	yjdC	b4134	b4135	4360649	4360756	107	reverse	0	0	1	1	0
yjdC	dipZ	b4135	b4136	4361331	4361368	37	reverse	1	0	0	1	1
dcuA	aspA	b4138	b4139	4364796	4364914	118	reverse	0	0	0	0	1
groS	groL	b4142	b4143	4369004	4369048	44	forward	0	0	0	1	1
efp	ecnA	b4147	b4410	4374288	4374340	52	forward	0	1	0	1	0
ecnA	ecnB	b4410	b4411	4374465	4374576	111	forward	0	1	1	0	1
blc	ampC	b4149	b4150	4375745	4375834	89	reverse	0	0	1	1	0
ampC	frdD	b4150	b4151	4376967	4377030	63	reverse	0	0	1	1	0
yjeM	yjeN	b4156	b4157	4383364	4383416	52	forward	0	0	0	1	0
glyV	glyX	b4163	b4164	4390458	4390495	37	forward	0	0	0	1	1
glyX	glyY	b4164	b4165	4390570	4390606	36	forward	0	0	0	1	1
miaA	hfq	b4171	b4172	4398225	4398311	86	forward	0	1	1	1	1
hfq	hflX	b4172	b4173	4398619	4398695	76	forward	0	0	0	0	1
hflX	hflK	b4173	b4174	4399975	4400061	86	forward	0	0	0	1	1
nsrR	rnr	b4178	b4179	4404638	4404677	39	forward	0	0	0	1	1
rnr	rlmB	b4179	b4180	4407118	4407298	180	forward	0	1	0	1	1
rlmB	yjfI	b4180	b4181	4408029	4408156	127	forward	0	0	0	0	1
yjfJ	yjfK	b4182	b4183	4409274	4409325	51	forward	0	0	0	1	0
yjfC	aidB	b4186	b4187	4412214	4412298	84	forward	0	0	2	1	0
rpsR	rplI	b4202	b4203	4424089	4424131	42	forward	0	0	0	1	1
ytfF	ytfG	b4210	b4211	4431088	4431187	99	reverse	0	1	0	1	0
ytfQ	ytfR	b4227	b4485	4448941	4449081	140	forward	0	0	0	0	1
nrdG	nrdD	b4237	b4238	4458387	4458545	158	reverse	0	0	0	0	1
treC	treB	b4239	b4240	4462732	4462782	50	reverse	0	0	0	1	1
yjgR	idnR	b4263	b4264	4488086	4488164	78	reverse	0	0	0	1	0
idnR	idnT	b4264	b4265	4489162	4489229	67	reverse	0	0	0	1	1
idnT	idnO	b4265	b4266	4490548	4490610	62	reverse	0	0	0	1	1
yjhC	ythA	b4280	b4655	4504428	4504471	43	forward	0	0	0	1	0

fecB	fecA	b4290	b4291	4512331	4512376	45	reverse	0	0	0	1	1
sgcA	sgcQ	b4302	b4303	4526003	4526134	131	reverse	0	0	0	1	1
yjhS	nanM	b4309	b4310	4535617	4535682	65	reverse	0	0	0	1	0
fimA	fimI	b4314	b4315	4541686	4541751	65	forward	0	0	0	1	1
fimI	fimC	b4315	b4316	4542290	4542327	37	forward	0	0	0	1	1
fimC	fimD	b4316	b4317	4543052	4543119	67	forward	0	0	0	1	1
uxuA	uxuB	b4322	b4323	4550843	4550924	81	forward	0	0	0	1	1
yjiE	iadA	b4327	b4328	4556312	4556377	65	reverse	0	0	0	1	0
yjiN	mdtM	b4336	b4337	4565269	4565310	41	reverse	0	1	0	1	1
yjiP	yjiQ	b4338	b4339	4567332	4567381	49	forward	0	0	0	0	1
hsdM	hsdR	b4349	b4350	4581071	4581272	201	reverse	0	1	0	1	1
yjiX	yjiY	b4353	b4354	4587102	4587152	50	reverse	0	0	0	1	0
yjjA	dnaC	b4360	b4361	4598212	4598261	49	reverse	0	0	0	1	1
dnaT	yjjB	b4362	b4363	4599540	4599647	107	reverse	0	0	0	1	1
yjjB	yjjP	b4363	b4364	4599973	4600111	138	reverse	0	0	0	1	0
leuV	leuP	b4368	b4369	4604188	4604223	35	reverse	0	0	0	1	1
deoC	deoA	b4381	b4382	4616125	4616252	127	forward	0	0	0	0	1
deoA	deoB	b4382	b4383	4617574	4617626	52	forward	0	0	0	1	1
deoB	deoD	b4383	b4384	4618849	4618906	57	forward	0	0	0	1	1
serB	radA	b4388	b4389	4623886	4623935	49	forward	0	0	0	1	1
slt	trpR	b4392	b4393	4630693	4630783	90	forward	0	0	1	1	0

CHAPTER 4 SUPPLEMENTARY MATERIALS:

Table S4-1 Primers for pCYi-f4 creation

Primer	Sequence
CFP-f4-fw	CATACAAGCTACGGACCATCGATTACAGAGTTCTTGAAG
CFP-f4-rv	TGCAGGAGGGTACCAAAAAAAAACCCCGCTTCGGCGGGGTTTTTTTATTTGTATAGTTCA

CFP-extend	GTCGACTCTAGACCCGGGAGGCCTTCTAGTGTAGCCGTAGTT AGGCCACCACTTCAAGAACTCTGTACATACAAGCTACGGACC ATCGAT
lacZ2-fw	GTTTTTTTTTGGTACCCTCCTGCACTGGATGGTGG
alt-f4-fw	CCCGGGTCTAGAGTCGACG

Table 4-2 Media used in study

Media/Supplement	Base Media	Components
LB	N/A	LB Broth Miller (BioShop, Burlington, ON)
M9	N/A	64 g/L $Na_2HPO_4 \cdot 7H_20$ 15 g/L KH_2PO_4 2.5 g/L NaCl 5.0 g/L NH_4Cl 0.4% glucose 2 mM $MgSO_4$ 0.1 mM $CaCl_2$
M9-limited	N/A	64 g/L $Na_2HPO_4 \cdot 7H_20$ 15 g/L KH_2PO_4 2.5 g/L NaCl 2 mM $MgSO_4$ 0.1 mM $CaCl_2$
Glucose	M9-limited	0.4% glucose
α-Ketoglutarate	M9-limited	0.44% α-ketoglutarate
Ammonia	M9-limited	10 mM NH_3 0.4% glucose
Glycine	M9-limited	10 mM Glycine 0.4% glucose
Arginine	M9-limited	10 mM Arginine 0.4% glucose
Glutamate	M9-limited	10 mM Glutamate 0.4% glucose
Glutamine	M9-limited	10 mM Glutamine 0.4% glucose
CMM 5.5	N/A	52.5 g/L citric acid 15 g/L KH_2PO_4 2.5 g/L NaCl 10 mM NH_3 0.4% glucose 2 mM $MgSO_4$ 0.1 mM $CaCl_2$
Mannose	M9-limited	0.4% Putricine

		10 mM NH_3
Malate	M9-limited	0.4% Malate 10 mM NH_3
Putricine	M9-limited	10 mM Putricine 10 mM NH_3

References

Abbas, C.A., and Sibirny, A.A., (2011). Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. Microbiol. Mol. Biol. Rev. *75*, 321-360.

Ameres, S.L., and Zamore, P.D., (2013). Diversifying microRNA sequence and function. Nat. Rev. Mol. Cell Biol. *14*, 475-488,

Amster-Choder, O., and Wright, A., (1992). Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation. Science *257*, 1395-1398.

Amster-Choder, O., (2005). The bgl sensory system: a transmembrane signaling pathway controlling transcriptional antitermination. Curr. Opin. Microbiol. *8*, 127-34.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. Curr. Biol. *11*, 941-50.

Babitzke, P., Baker, C.S., Romeo, T., (2009). Regulation of translation initiation by RNA binding proteins. Annu. Rev. Microbiol. *63*, 27-44.

Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., Breaker, R.R., (2012). Widespread genetic switches and toxicity resistance proteins for fluoride. Science *335*, 233-235.

Barondeau, D.P., Putnam, C.D., Kassmann, C.J., Tainer, J.A., Getzoff, E.D., (2003). Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. Proc. Natl. Acad. Sci. U. S. A. *100*, 12111-12116.

Barker, C.A., Farha, M.A., Brown, E.D., (2010). Chemical genomic approaches to study model microbes. Chem. Biol. *17*, 624-632.

Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J.K., Breaker, R.R., (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc. Natl. Acad. Sci. U. S. A. *101*, 6421-6426.

Barrick, J.E., and Breaker, R.R., (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol. *8*, R239.

Bevis, B.J., Glick, B.S., (2002). Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *20*, 83-87.

Billinton, N., and Knight, A.W., (2001). Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. Anal. Biochem. *291*, 175-197.

Bingham, R.J., Hall, K.S., Slonczewski, J.L., (1990). Alkaline induction of a novel gene locus, alx, in Escherichia coli. J. Bacteriol. *172*, 2184-2186.

Borisy, A.A., Elliott, P.J., Hurst, N.W., Lee, M.S., Lehar, J., Price, E.R., Serbedzija, G., Zimmermann, G.R., Foley, M.A., Stockwell, B.R., Keith, C.T., (2003). Systematic discovery of multicomponent therapeutics. Proc. Natl. Acad. Sci. U. S. A. *100*, 7977-7982.

Breaker, R.R., (2011). Prospects for Riboswitch Discovery and Analysis. Mol. Cell *43*, 867-879.

Brough, R., Frankum, J.R., Costa-Cabral, S., Lord, C.J., Ashworth, A., (2011). Searching for synthetic leathality in cancer. Curr. Opin. Genet. Dev. *21*, 34-41.

Castaño, I., Bastarrachea, F., Covarrubias, A.A., (1988). gltBDF operon of Escherichia coli. J. Bacteriol. *170*, 821-827.

Castaño, I., Flores, N., Valle, F., Covarrubias, A.A., Bolivar, F., (1992). gltF, a member of the gltBDF operon of Escherichia coli, is involved in nitrogen-regulated gene expression. Mol. Microbiol. *6*, 2733-2741.

Caron, M.P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., Lafontaine, D.A., (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. Proc. Natl. Acad. Sci. U. S. A. *109*, E3444-3453

Cho, B.K., Zengler, K., Qiu, Y., Park, Y.S., Knight, E.M., Barrett, C.L., Gao, Y., Palsson, B.Ø., (2009). The transcription unit architecture of the Escherichia coli genome. Nat. Biotechnol. *27*, 1043-1049.

Chubukov, V., Gerosa, L., Kochanowski, K., Sauer, U., (2014). Coordination of microbial metabolism. Nat. Rev. Microbiol. *12*, 327-340.

Collins, J.A., Irnov, I., Baker, S., Winkler, W.C., (2007). Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev. *21*, 3356-3368.

Coornaert, A., Chiaruttini, C., Springer, M., Guillier, M., (2013). Post-transcriptional control of the Escherichia coli PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. PLoS Genet. *9*, e1003156.

Cozzone, A.J., (1998). Regulation of acetate metabolism by protein phosphorylation in enteric bacteria. Annu. Rev. Microbiol. *52*, 127-64.

Crick F.H., (1958). On protein synthesis. Symp. Soc. Exp. Biol. *12*, 138–63.

Daigle, N., and Ellenberg, J., (2007), LambdaN-GFP: an RNA reporter system for live-cell imaging. Nat. Methods *4*, 633-636.

Dam, P., Olman, V., Harris, K., Su, Z., Xu, Y., (2007). Operon prediction using both genome-specific and general genomic information. Nucleic Acids Res. *25*, 288-298.

Dambach, M., Sandoval, M., Waters, L., Storz, G., (2014). The Yybp/ Ykoy Orphan Riboswitch Is A Manganese Responsive Genetic Element. ASM 2014, Conference presentation.

Dann, C.E. 3rd, Wakeman, C.A., Sieling, C.L., Baker, S.C., Irnov, I., Winkler, W.C., (2007). Structure and mechanism of a metal-sensing regulatory RNA. Cell *130*, 878-892.

Defez, R., and De Felice, M., (1981). Cryptic operon for beta-glucoside metabolism in Escherichia coli K12: genetic evidence for a regulatory protein. Genetics *97*, 11-25.

Desnoyers, G., Morissette, A., Prévost, K., Massé, E., (2009). Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. EMBO J. *28*, 1551-1561.

Datsenko, K.A., and Wanner, B.L., (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. *97*, 6640-6645.

de Saizieu, A., Certa, U., Warrington, J., Gray, C., Keck, W., Mous, J., (1998). Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. Nat. Biotechnol. *16*, 45-48.

Deutscher, J., Francke, C., Postma, P.W., (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol Mol. Biol. Rev. *70*, 939-1031.

Deutscher J., (2008). The mechanisms of carbon catabolite repression in bacteria. Curr. Opin. Microbiol. *11*, 87-93.

Falconer, S.B., Wang, W., Gehrke, S.S., Cuneo, J.D., Britten, J.F., Wright, G.D., Brown, E.D. (2014). Metal-induced isomerization yields an intracellular chelator that disrupts bacterial iron homeostasis. Chem. Biol. *21*, 136-145.

Fayat, G., Mayaux, J.F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M., Blanquet, S., (1983). Escherichia coli phenylalanyl-tRNA synthetase operon region. Evidence for an attenuation mechanism. Identification of the gene for the ribosomal protein L20. J. Mol. Biol. *171*, 239-261.

Fisher, A.C., and DeLisa, M.P., (2008). Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. PLoS One *3*, e2351.

Flint, D.H., Tuminello, J.F., Miller, T.J., (1996). Studies on the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase in escherichia coli crude extract. Isolation of O-

acetylserine sulfhydrylases A and B and beta-cystathionase based on their ability to mobilize sulfur from cysteine and to participate in Fe-S cluster synthesis. J. Biol Chem. *271*, 16053-16067.

Fowler, C.C., Brown, E.D., Li, Y., (2010), Using a riboswitch sensor to examine coenzyme B(12) metabolism and transport in E. coli. Chem. Biol. *17*, 756-765.

Fowler, C. C., (2011). PhD Thesis, McMaster University.

Fradkov, A.F., Chen, Y., Ding, L., Barsova, E.V., Matz, M.V., Lukyanov, S.A., (2000). Novel fluorescent protein from Discosoma coral and its mutants possesses a unique far-red fluorescence. FEBS Lett. *479*, 127-130.

Frohman, M.A., Dush, M.K., Martin, G.R., (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. U. S. A. *85*, 8998-9002.

Gilbert, W., (1986). Origins of life: The RNA world. Nature *319*, 618.

Goedhart, J., von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M.A., van Weeren, L., Gadella, T.W. Jr, Royant, A., (2012). Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat. Commun. *3*, 751.

Görke, B., and Stülke, J., (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat. Rev. Microbiol. *6*, 613-624.

Goss, T.J., Perez-Matos, A., Bender, R.A., (2001). Roles of glutamate synthase, gltBD, and gltF in nitrogen metabolism of Escherichia coli and Klebsiella aerogenes. J. Bacteriol. 183, 6607-6619.

Grass, G., Bufe, B., Müller, B., Rösel, M., Kleiner, D., (1999). Characterization of the gltF gene product of Escherichia coli. FEMS Microbiol Lett. *179*, 79-84.

Gross, L.A., Baird, G.S., Hoffman, R.C., Baldridge, K.K., Tsien, R.Y., (2000). The structure of the chromophore within DsRed, a red fluorescent protein from coral. Proc. Natl. Acad. Sci. U. S. A. *97*, 11990-11995.

Grundy, F.J. and Henkin, T.M. (1993). tRNA as a positive regulator of transcripton antitermination in *B. subtilis.* Cell *74*, 475–482.

Gurvich, O.L., Baranov, P.V., Zhou, J., Hammer, A.W., Gesteland, R.F., Atkins, J.F., (2003). Sequences that direct significant levels of frameshifting are frequent in coding regions of Escherichia coli. EMBO J. *22*, 5941-5950.

Huang, H., (2007). Design and Characterization of Artificial Transcriptional Terminators. Doctoral Thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA.

Herbert, D., Elsworth, R., Telling, R.C., (1956). The continuous culture of bacteria; a theoretical and experimental study. J. Gen. Microbiol. *14*, 601-622.

Hindley, J., (1967). Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. J. Mol. Biol. *30*, 125-136.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., Zamir, A., (1965). Structure of a ribonucleic acid. Science *147*, 1462-1456.

Hoskisson, P.A., and Hobbs, G., (2005). Continuous culture--making a comeback?. Microbiology *151*, 3153-3159.

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

Ikemura, T., and Dahlberg, J.E., (1973). Small ribonucleic acids of *Escherichia coli. I.* Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. J. Biol. Chem. *248*, 5024-32.

Jin, Y., Watt, R.M., Danchin, A., Huang, J.D., (2009). Small noncoding RNA GcvB is a novel regulator of acid resistance in Escherichia coli. BMC Genomics *10*, 165

Joyce, G.F., (2002). The antiquity of RNA-based evolution. Nature *418*, 214-221.

Kaddurah-Daouk, R., Kristal, B.S., Weinshilboum, R.M., (2008). Metabolomics: a global biochemical approach to drug response and disease. Annu. Rev. Pharmacol. Toxicol. *48*, 653-683.

Kalamorz, F., Reichenbach, B., März, W., Rak, B., Görke, B., (2007). Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Mol. Microbiol. *65*, 1518-33.

Kery, M.B., Feldman, M., Livny, J., Tjaden, B., (2014). TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res. *42*, W124-9

Kim, J.N., and Breaker, R.R., (2008). Purine sensing by riboswitches. Biol. Cell 100, 1-11.

Kortmann, J., Narberhaus, F., (2012). Bacterial RNA thermometers: molecular zippers and switches. Nat. Rev. Microbiol. *10*, 255-265.

Kremers, G.J., Goedhart, J., van Munster, E.B., Gadella, T.W. Jr. (2006). Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. Biochemistry *45*, 6570-6580.

Kremers, G.J., Goedhart, J., van den Heuvel, D.J., Gerritsen, H.C., Gadella, T.W. Jr., (2007). Improved green and blue fluorescent proteins for expression in bacteria and mammalian cells. Biochemistry *46*, 3775-3783.

Kremers, G.J., Gilbert, S.G., Cranfill, P.J., Davidson, M.W., Piston, D.W., (2011). Fluorescent proteins at a glance. J. Cell. Sci. *124*, 157-160.

Kumagai, A., Ando, R., Miyatake, H., Greimel, P., Kobayashi, T., Hirabayashi, Y., Shimogori, T., Miyawaki, A., (2013). A bilirubin-inducible fluorescent protein from eel muscle. Cell *153*, 1602-1611.

Lakowicz, J.R., Szmacinski, H., Nowaczyk, K., Johnson, M.L., (1992). Fluorescence lifetime imaging of free and protein-bound NADH. *89*, 1271-1275.

Lawther, R.P., and Hatfield, G.W., (1980). Multivalent translational control of transcription termination at attenuator of ilvGEDA operon of Escherichia coli K-12. Proc. Natl. Acad. Sci. U. S. A. *77*, 1862-1866.

Lee, R. C., Feinbaum, R. L. & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843–854.

Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., Breaker, R.R., (2010). An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science *329*, 845-848.

Lesnik, E.A., Sampath, R., Levene, H.B., Henderson, T.J., McNeil, J.A., Ecker, D.J., (2001). Prediction of rho-independent transcriptional terminators in Escherichia coli. Nucleic Acids Res. *29*, 3583-3594.

Levine, R.A., and Taylor, M.W., (1982). Mechanism of adenine toxicity in Escherichia coli. J. Bacteriol. *149*, 923-930.

Liang, S., Bipatnath, M., Xu, Y., Chen, S., Dennis, P., Ehrenberg, M., Bremer, H., (1999). Activities of constitutive promoters in Escherichia coli. J. Mol. Biol. *292*, 19-37.

Liu, M., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J., Blattner, F.R., (2005). Global transcriptional programs reveal a carbon source foraging strategy by Escherichia coli. J. Biol. Chem. *280*, 15921-15927.

Livny, J. and Waldor, M.K., (2007). Identification of small RNAs in diverse bacterial species. Curr. Opin. Microbio. *10*, 96-101.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E.L. (1996).

Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*, 1675-80.

Maheswaran, M, and Forchhammer, K., (2003). Carbon-source-dependent nitrogen regulation in Escherichia coli is mediated through glutamine-dependent GlnB signalling. Microbiology *149*, 2163-2172.

Mao, F., Dam, P., Chou, J., Olman, V., Xu, Y., (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Res. *36*, D459-463.

Mandal, M., and Breaker, R.R. (2004). Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol. *5*, 451-463.

Massé, E., Escorcia, F.E., Gottesman, S., (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. Genes Dev. *17*, 1374-2383.

Massé, E., Salvail, H., Desnoyers, G., Arguin, M., (2007). Small RNAs controlling iron metabolism. Curr. Opin. Microbiol. *10*, 140-145.

Mena, M.A., Treynor, T.P., Mayo, S.L., Daugherty, P.S., (2006). Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. Nat. Biotechnol. *24*, 1569-1571.

Meyer, M.M., Hammond, M.C., Salinas, Y., Roth, A., Sudarsan, N., Breaker, R.R. (2011). Challenges of ligand identification for riboswitch candidates. RNA Biol. *8*, 5-10.

Mizuno, T., Chou, M.Y., Inouye, M., (1984). A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc. Natl. Acad. Sci. U. S. A. *81*, 1966-1970.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., Breaker, R.R., (2002). Genetic control by a metabolite binding mRNA. Chem. Biol. *9*, 1043-1049.

Nahvi, A., Barrick, J.E., Breaker, R.R., (2004). Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res. *32*, 143-150.

Navarro Llorens, J.M., Tormo, A., Martínez-García, E., (2010). Stationary phase in gram-negative bacteria. FEMS Microbiol. Rev. *34*, 476-495.

Nicholson, J.K., and Lindon, J.C. (2008). Systems biology: Metabonomics. Nature *455*, 1054–1056.

Novick, A., and Szilard, L., (1950). Description of the chemostat. Science *112*, 715-716.

O' Connor, C.O., Laraia, L., Spring, D.R., (2011). Chemical genetics. Chem. Soc. Rev. *40*, 4332-4345.

Oh, M.K., Rohlin, L., Kao, K.C., Liao, J.C., (2002). Global expression profiling of acetate-grown Escherichia coli. J. Biol. Chem. *277*, 13175-13183.

Paige, J.S., Wu, K.Y., Jaffrey, S.R., (2011). RNA mimics of green fluorescent protein. Science *333*, 642-646.

Patti, G.J., Yanes, O., Siuzdak, G., (2012). Innovation: Metabolomics: the apogee of the omics trilogy. Nat. Rev. Mol. Cell Biol. *13*, 263-269.

Pfleger, B.F., Fawzi, N.J., Keasling, J.D., (2005). Optimization of DsRed production in Escherichia coli: effect of ribosome binding site sequestration on translation efficiency. Biotechnol. Bioeng. *92*, 553-558.

Pichon, C., and Felden, B., (2008). Small RNA gene identification and mRNA target predictions in bacteria. Bioinformatics *24*, 2807-2813.

Poulsen, P., Bonekamp, F., Jensen, K.F., (1984). Structure of the Escherichia coli pyrE operon and control of pyrE expression by a UTP modulated intercistronic attentuation. EMBO J. *3*, 1783-1790.

Prasad, I., and Schaefler, S., (1974). Regulation of the beta-glucoside system in Escherchia coli K-12. J. Bacteriol. *120*, 638-50.

Pulvermacher, S.C., Stauffer, L.T., Stauffer, G.V., (2009). The small RNA GcvB regulates sstT mRNA expression in Escherichia coli. J. Bacteriol. *191*, 238-248.

Reitzer, L., (2003). Nitrogen assimilation and global regulation in Escherichia coli. Annu Rev. Microbiol. *57*, 155-176.

Ren, A., Rajashankar, K.R., Patel, D.J., (2012). Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. Nature *486*, 85-89.

Reynolds, A.E., Felton, J., Wright, A., (1981). Insertion of DNA activates the cryptic bgl operon in E. coli K12. Nature *293*, 625-629.

Rivas, E., Klein, R.J., Jones, T.A., Eddy, S.R., (2001). Computational identification of noncoding RNAs in E. coli by comparative genomics. Curr. Biol. *11*, 1369-1373.

Rogozin, I.B., Makarova, K.S., Natale, D.A., Spiridonov, A.N., Tatusov, R.L., Wolf, Y.I., Yin, J., Koonin, E.V., (2002). Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. Nucleic Acids Res. *30*, 4264-4271.

Rosenow, M.A., Patel, H.N., Wachter, R.M., (2005). Oxidative chemistry in the GFP active site leads to covalent cross-linking of a modified leucine side chain with a histidine imidazole: implications for the mechanism of chromophore formation. Biochemistry *44*, 8303-8311.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernández, S., Alquicira-Hernández, K., López-Fuentes, A., Porrón-Sotelo, L., Huerta, A.M., Bonavides-Martínez, C., Balderas-Martínez, Y.I., Pannier, L., Olvera, M., Labastida, A., Jiménez-Jacinto, V., Vega-Alvarado, L., Del Moral-Chávez, V., Hernández-Alvarez, A., Morett, E., Collado-Vides, J., (2003). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. *41*, D203-213.

Sauter, M., Böhm, R., Böck, A., (1992). Mutational analysis of the operon (hyc) determining hydrogenase 3 formation in Escherichia coli. Mol. Microbiol. *11*, 1523-1532.

Schenborn, E., and Groskreutz, D., (1999). Reporter gene vectors and assays. Mol. Biotechnol. *13*, 29-44.

Schnetz, K., Toloczyki, C., Rak, B., (1987). Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two Bacillus subtilis genes. J. Bacteriol. *169*, 2579-2590.

Selinger, D.W., Cheung K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J., and Church, G.M., (2000). RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. Nat. Biotechnol. *18*, 1262-1268.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., Tsien, R.Y., (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. *22*, 1567-1572.

Shaner, N.C., Patterson, G.H., Davidson, M.W., (2007). Advances in fluorescent protein technology. J. Cell Sci. *120*, 4247-4260.

Sharma, C.M., Darfeuille, F., Plantinga, T.H., Vogel, J., (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev. *21*, 2804-2817.

Sharma, C.M., and Vogel, J., (2009). Experimental approaches for the discovery and characterization of regulatory small RNA. Curr. Opin. Microbiol. *12*, 536-546.

Sharma, C.M., Papenfort, K., Pernitzsch, S.R., Mollenkopf, H.J., Hinton, J.C., Vogel, J., (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol. Microbiol. *81*, 1144-1165.

Shevchuk, N.A., Bryksin, A.V., Nusinovich, Y.A., Cabello, F.C., Sutherland, M., Ladisch, S., (2004). Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res. *32*, e19.

Stockwell, B.R., (2000). Chemical genetics: ligand-based discovery of gene function. Nat. Rev. Genet. *1*, 116-125.

Storz, G., (2002). An expanding universe of noncoding RNAs. Science *296*, 1260-1263.

Storz, G., Vogel, J., Wassarman, K.M., (2011). Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell. *43*, 880-891.

Strack, R.L., Strongin, D. E., Mets, L., Glick, B.S., Keenan, R.J., (2011). Chromophore formation in DsRed occurs by a branched pathway. J. Am. Chem. Soc. *132*, 8496-8505.

Subach, F.V., and Verkhusha, V.V., (2012). Chromophore transformations in red fluorescent proteins. Chem. Rev. *112*, 4308-4327.

Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., Breaker, R.R., (2003). An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev. *17*, 1688-2697.

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., Breaker, R.R., (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science *321*, 411-413.

Teale, F.W., and Weber, G., (1957). Ultraviolet fluorescence of the aromatic amino acids. Biochem. J. *65*, 376-482.

Tjaden, B., Goodwin, S.S., Opdyke, J.A., Guillier, M., Fu, D.X., Gottesman, S., Storz, G., (2006), Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res. *34*, 2791-2802.

Urban, J.H., and Vogel, J., (2008). Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. PLoS Biol. *6*, e64.

Urbanowski, M.L., Stauffer, L.T., Stauffer, G.V., (2000). The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in Escherichia coli. Mol. Microbiol. *37*, 856-868.

Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., Gelfand, M.S., (2002). Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. Nucleic Acids Res. *30*, 3141-3151.

Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jäger, J.G., Hüttenhofer, A., Wagner, E.G. (2003). RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res. *31*, 6435-6443.

Vogel, J., and Sharma, C.M., (2005). How to find small non-coding RNAs in bacteria. Biol. Chem. *386*, 1219-1238.

Wachter, R.M., Elsliger, M.A., Kallio, K., Hanson, G.T., Remington, S.J., (1998). Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. Structure *6*, 1267-1277.

Wachter, A., (2010). Riboswitch-mediated control of gene expression in eukaryotes. RNA Biol. *7*, 67-76.

Waller, P.R., and Sauer, R.T., (1996). Characterization of degQ and degS, Escherichia coli genes encoding homologs of the DegP protease. J. Bacteriol. *178*, 1146-1153.

Walsh, D.P., and Chang, Y.T., (2006). Chemical genetics. Chem. Rev. *106*, 2476-2530.

Wampler, J.E., Hori, K., Lee, J.W., Cormier, M.J., (1971). Structured bioluminescence. Two emitters during both the in vitro and the in vivo bioluminescence of the sea pansy, Renilla. Biochemistry *10*, 2903-2909.

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., Gottesman, S., (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. *15*, 1637-1651.

Waters, L.S., and Storz, G., (2009). Regulatory RNAs in bacteria. Cell *136*, 615-628.

Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H., Breaker, R.R. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol. *11*, R31.

Wessler, S.R., and Calvo, J.M., (1981). Control of leu operon expression in Escherichia coli by a transcription attenuation mechanism. J. Mol. Biol. *149*, 579-597.

Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., Breaker, R.R., (2004). Control of gene expression by a natural metabolite-responsive ribozyme. Nature *428*, 281-286.

Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., Lockhart, D.J., (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. Nat. Biotechnol. *15*, 1359-1367.

Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., Vogel, J., Hess, W.R., Backofen, R., Georg, J., (2013). Comparative genomics boosts target prediction for bacterial small RNAs. Proc. Natl. Acad. Sci. U. S. A. *110*, E3487-3496.

Yajnik, V., and Godson, G.N., (1993), Selective decay of Escherichia coli dnaG messenger RNA is initiated by RNase E. *268*, 13253-13260.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., Alon, U., (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nat. Methods. *3*, 623-628.

Zhang A1, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol. Microbiol. *50*, 1111-1124.

Zlitni, S., Ferruccio, L.F., Brown, E.D. (2013). Metabolic suppression identifies new antibacterial inhibitors under nutrient limitation. Nat. Chem. Biol. *9*, 796-804.