Control of gene expression in cyanobacteria

By

Gina C. Gordon

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 2/16/2018

The dissertation is approved by the following members of the Final Oral Committee:

Brian F. Pfleger, Professor, Department of Chemical and Biological Engineering

Michael G. Thomas, Professor, Department of Bacteriology

Karen M. Wassarman, Professor, Department of Bacteriology

Richard L. Gourse, Professor, Department of Bacteriology

Hiroshi Maeda, Associate Professor, Department of Botany

Abstract

Microbial cells can be manipulated to produce valuable chemicals and products. Cyanobacteria are an attractive candidate for conversion of CO_2 into useful products because of their fast generation time and ability to be cultivated on non-arable land. In other organisms as well as in cyanobacteria, metabolic engineering efforts and strategies are heavily reliant on the underlying genetic tools that facilitate changes in gene expression of both native and exogenous biochemical pathways. One relatively unexplored area in terms of manipulating gene expression is transcript stability. This thesis expands the toolset for manipulating gene expression in cyanobacteria and probes the relatively unexplored area of RNA stability. The ribonucleases that mediate RNA turnover in cyanobacteria are analyzed, and the functions of three RNase III enzymes are described. A global RNA turnover study using RNA-sequencing showed that RNA turnover was very rapid and identified preliminary links between transcript stability and sequence elements that may confer stability. Finally, genetic tools were built to control gene expression in cyanobacteria and applied to increase lactate production. These genetic tools and findings on RNA turnover in cyanobacteria provide significant advances towards reliable prediction and engineering of a synthetic transcript.

Acknowledgments

I have been fortunate to have been surrounded by a supportive family, friends, and coworkers. First, a thank you to Brian for the opportunity, support, and freedom I needed in graduate school. Two postdocs, Jeffrey Cameron and Andrew Markley, were absolutely instrumental in teaching me to be a scientist and getting me headed in a good direction. I want to acknowledge my undergraduate PI, Jake McKinlay. He taught me to be smart with experimental design: draw a circle with all possible hypotheses and design a simple and smart experiment to cut the possibilities in half. It was in his lab that I realized that I could have a rewarding and successful career in science.

Madison has been a wonderful place to work and study. I want to thank the MDTP students for being such great friends and scientific collaborators. I have yet to see a graduate student that doesn't end up in a dark place at some point during graduate school, but the inclusiveness and sense of comradery among the MDTP students gets everyone through. You are all brilliant people, I can't wait to see what you all will do.

I want to thank members of the Pfleger Lab that made going to lab every day, bearable. Matt Begemann, Daniel Mendez Perez, Jackie Rand, Chris Mehrer, Travis Korosh, Ryan Clark, Matthew Copeland, Nestor Hernandez, Austin Comer, Taylor Cook, and Dylan Courtney. I'll always remember evenings at The Library, lab parties, and fantasy football. You were a great group- always there when I needed to vent and when there was cause to celebrate.

A big thanks to my family. Visits always provided a much needed break from science, and allowed me to return re-freshed and recharged. Thank you for your endless encouragement and love. And thanks for pretending to read all of my papers I sent you. And lastly, to my two main guys. Thanks to Dixon for lots of love and laughs, and reminders of what's important (cuddling on the couch and running around in the backyard). Last but not least, I would like to thank Tony, my boyfriend, fiance, and then husband. Thank you for the endless support both scientifically and emotionally.

Table of Contents

Page

Abstract **i**

Acknowledgments **ii**

Table of Contents **iii**

List of Figures **v**

List of Tables **viii**

1 Chapter 1: Genetic tools for controlling gene expression in cyanobacteria **1**

1.1 Motivations for controlling gene expression 2

1.2 Thesis overview 3

1.3 Review of regulatory tools for controlling gene expression in cyanobacteria 6

2 Chapter 2: Identifying RNase III targets in *Escherichia coli* **46**

2.1 Introduction 47

2.2 Identifying RNase III Sites using RNA-sequencing 50

2.3 Materials and Methods 77

3 Chapter 3: RNases present in cyanobacteria **87**

3.1 Introduction 88

3.2 Genetic and genomic analysis of RNases in model cyanobacteria 90

3.3 Materials and Methods 112

4 Chapter 4: Function of three homologs of RNase III in cyanobacteria **116**

4.1 Introduction 117

4.2 Role of three homologs of RNase III 118
4.3 Discussion 135
4.4 Materials and Methods 143

5 Chapter 5: Global RNA decay in cyanobacteria 149

5.1 Introduction 150
5.2 Global RNA degradation in PCC 7002 151
5.3 Discussion 156
5.4 Materials and Methods 162

6 Chapter 6: Tools to control gene expression in cyanobacteria 166

6.1 Introduction 167
6.2 CRISPR interference in PCC 7002: physiology and chemical production 167
6.3 Promoter characterization 187
6.4 RBS characterization 189
6.5 5' hairpin stabilizes transcripts in PCC 7002 190
6.6 Materials and Methods 196

7 Chapter 7: Conclusions 205

7.1 Summary of thesis research 206
7.2 Future Trends 211

References 215

Appendices 255

RNA spike-in protocol 256
RNA-sequencing supplementary files 263

List of Figures

Page

Figure 1.1 Synthetic biology tools built to control gene expression 4
Figure 1.2 Overview of key factors governing gene expression 10
Figure 1.3 Fold induction ratios from inducible systems in cyanobacteria 19
Figure 1.4 Organization and transcripts of the *cpc* operon 33
Figure 2.1 Identification of RNase III cleavages sites by RNA-sequencing 51
Figure 2.2 RNA-seq coverage surrounding known *pnp* cleavage site 51
Figure 2.3 Degradation surrounding known RNase III cleavage sites 52
Figure 2.4 RNase III cleavage of in vitro transcripts 61
Figure 2.5 Lack of RNase III cleavage of in vitro transcripts 62
Figure 2.6 Predicted secondary structure of novel RNase III sites 63
Figure 2.7 Read coverage and decay near novel *aceEF* cleavage site 65
Figure 2.8 Read coverage and decay surrounding novel RNase III cleavage sites 66
Figure 2.9 Read coverage and decay of *ndh* transcript 69
Figure 2.10 Read coverage and decay near novel *pflBA* cleavage site 70
Figure 2.11 Read coverage and decay of *yajQ* and *panE* transcripts 71
Figure 2.12 RNase III cleavage of *hisL* leader transcript 71
Figure 2.13 RNase III cleavage of *pheM* leader transcript 72
Figure 2.14 Pyruvate dehydrogenase activity assays of WT and *rnc*- 73
Figure 2.15 Predicted secondary structure of cleaved transcripts 75
Figure 2.16 Influence of RNase activity on central metabolism 78
Figure 2.17 Purification of *E. coli* RNase III 81
Figure 3.1 Schematic diagram of RNases in bacteria. 89

Figure 3.2 Possible essential RNases in PCC 7002. 94
Figure 3.3 Diversity and distribution of RNase III homologs in cyanobacteria 100
Figure 3.4 Deletion and segregation analysis of RNase III homologs 101
Figure 3.5 Growth of RNase III deletion mutants in liquid and on solid medium 103
Figure 3.6 Growth of ΔA0384 strains in varying CO_2 concentrations 104
Figure 3.7 Genetic complementation of the RNase III triple mutant 105
Figure 3.8 Generation and physiological characterization of a Δ1543 mutant 108
Figure 3.9 Growth and physiology of ΔA1543 strains 109
Figure 3.10 Light-dependent oxygen evolution activity of RNase III mutants 111
Figure 4.1 BioAnalyzer traces from RNase III mutants 120
Figure 4.2 Role of RNase III in rRNA processing 121
Figure 4.3 Proposed secondary structure of the rRNA operon 122
Figure 4.4 Differential expression patterns in RNase III mutants 125
Figure 4.5 Increased expression of pAQ3 genes without A2542 127
Figure 4.6 A2542 regulates copy number control through ctRNA cleavage 129
Figure 4.7 *In vitro* cleavage assays of pAQ3 replication RNA and ctRNA 130
Figure 4.8 Differential expression between single and triple mutant strains 132
Figure 4.9 Differentially expressed genes in all samples 134
Figure 4.10 Cyanobacterial RNase III's cleave *E. coli* RNase III targets 136
Figure 4.11 Oligomerization of purified RNase III enzymes 137
Figure 4.12 Comparison of *E. coli* and cyanobacteria RNase III datasets 140
Figure 5.1 Distribution of gene half-lives in PCC 7002 152
Figure 5.2 Photosynthesis genes have long half-lives 153
Figure 5.3 High expression of photosynthesis genes 153
Figure 5.4 Expression level within energy metabolism subgroup. 154
Figure 5.5 L-shaped rho-independent terminators enriched in stable genes 155

Figure 5.6 Half-life increases along transcript 161
Figure 5.7 RNA-seq read counts from RNA spike-ins 164
Figure 6.1 Repression of YFP reporter with CRISPRi 171
Figure 6.2 Induction does not cause repression without both dCas9 and sgRNA 172
Figure 6.3 Altering dCas9 promoter does not relieve repression 173
Figure 6.4 aTc inducible promoters improve dynamic range of repression 174
Figure 6.5 Extra nucleotides at 5' end of sgRNA do not affect repression 174
Figure 6.6 Titration of repression by decreasing dCas9 RBS strength 176
Figure 6.7 Truncation of sgRNA 176
Figure 6.8 sgRNA targeting different locations within YFP 177
Figure 6.9 CRISPRi repression can be reversed 179
Figure 6.10 Tunable repression of the phycobilisome operon 180
Figure 6.11 Repression of the carboxysome operon 182
Figure 6.12 Schematic of proposed mechanism for increased lactate production 184
Figure 6.13 Lack of phycobilisome and chlorophyll a degradation with *glnA* sgRNA . . 185
Figure 6.14 Repression of glutamine synthetase improves lactate production 186
Figure 6.15 Comparison of YFP reporter libraries in *E. coli* and PCC 7002 188
Figure 6.16 Inducible expression from *ccmR* promoter 190
Figure 6.17 A RBS library of a synthetic YFP transcript. 191
Figure 6.18 Comparison of RBS calculator predictions and experimental results 192
Figure 6.19 Diagram of constructs used to test effects of 5' hairpin 194
Figure 6.20 5' hairpins stabilize transcripts 195
Figure 6.21 Fluorescence levels increase in RBS library with a 5' hairpin 196
Figure A1.1 Relationship between read count and spike-in addition 261

List of Tables

Page

Table 1.1 Cyanobacteria strains and abbreviations 8

Table 2.1 Known and putative RNase III cleavage sites 58

Table 2.2 Oligonucleotides used to verify RNase III cleavage sites 82

Table 3.1 Ribonuclease homologs in commonly studied cyanobacteria 92

Table 3.2 Summary of PCC 7002 RNase deletion strains 93

Table 3.3 Analysis of RNase III-like proteins in diverse cyanobacterial genomes . . . 96

Table 3.4 Oligonucleotides used to make RNase mutants and verify segregation . . . 114

Table 3.5 Plasmids used to delete RNases in PCC 7002. 115

Table 4.1 RNase III transcript levels remain constant across mutant strains 124

Table 4.2 Primers used to study role of RNase III homologs in cyanobacteria 148

Table 4.3 Plasmids used to over express RNase III enzymes 148

Table 6.1 Plasmids used to create CRISPRi strains 198

Table 6.2 *Synechococcus* sp. strain PCC 7002 CRISPRi strains 199

Table 6.3 Strains and plasmids used to measure effects of 5' hairpin 201

Table 6.4 Strains and plasmids used to build RBS libraries 202

Table A1.1 Primers used to generate RNA spike-in DNA templates 258

Table A1.2 Copy number of spike-ins 259

Table A1.3 qPCR primers used for RNA spike-in detection 260

Table A2.4 Summary of access to RNA-sequencing files 264

Table A2.5 Top 100 putative RNase III sites in *E. coli* 264

Table A2.6 Top 40 differentially expressed genes in ΔA0061 268

Table A2.7 Top 40 differentially expressed genes in ΔA2542 269

Table A2.8 Top 40 differentially expressed genes in ΔA0384 270
Table A2.9 Top 40 differentially expressed genes in ΔA0061ΔA2542 271
Table A2.10 Top 40 differentially expressed genes in ΔA0061ΔA0384 272
Table A2.11 Top 40 differentially expressed genes in ΔA2542ΔA0384 273
Table A2.12 Top 40 differentially expressed genes in ΔA0061ΔA2542ΔA0384 274
Table A2.13 Top 80 most stable PCC 7002 genes 275
Table A2.14 Top 80 least stable PCC 7002 genes 276

1 Chapter 1: Genetic tools for controlling gene expression in cyanobacteria

Authors and Contributors:

Gina Gordon: Organized material, made figures, and wrote text.

Dr. Brian Pfleger: Provided guidance, edited, and reviewed the text.

Portions of this chapter were submitted for publication as a book chapter. Regulatory tools for controlling gene expression in cyanobacteria. *Synthetic Biology of Cyanobacteria*. Springer/Nature Press.

1.1 Motivations for controlling gene expression

Bacteria can be engineered to produce valuable chemicals and products that can be used in many industries including transportation, agriculture, health, and the environment. These products can be novel to the host requiring introduction of a foreign biochemical pathway. Alternatively, the yield of native compounds can be increased by manipulating flux through existing pathways. Both of these approaches require fine-tuning of protein levels to achieve high yields.

Different bacterial hosts can be used for production due to specific advantages of their metabolism, growth rate, tolerance of the product, or required feedstock. There has been significant effort to capture CO_2 from the atmosphere and produce high value biofuels (Kumar, Sundaram, et al. 2018). Using bacteria for CO_2 sequestration and chemical production has many benefits compared with using plants including increased genetic tractability, fast generation times, and the ability to grow on non-arable land. Cyanobacteria are diverse photosynthetic microbes with the ability to convert CO_2 into useful products. However, metabolic engineering of cyanobacteria remains challenging because of the limited resources for modifying the expression of endogenous and exogenous biochemical pathways. Fine-tuned control of protein production will be critical to optimize the biological conversion of CO_2 into desirable molecules.

Genetic tools can operate on many different levels to alter protein and enzyme levels (**Figure 1.1**). Some genetic tools change transcription levels: examples include changing promoters or placing sequence elements that in combination with protein partners facilitate the blocking of transcriptional machinery. Other genetic tools operate by allowing or blocking translation; the ribosome binding sequence can be altered or conditionally obscured. Protein levels can also be altered by changing the rate of protein degradation by adding a protein degradation tag. One area that remains relatively unexplored in terms of manipulating gene expression is the area of RNA stability.

Messenger RNA (mRNA) are labile intermediates that play critical roles in determining the translation rate and steady state protein concentrations in the cell. There are very few examples of

using transcript stability to change gene expression, and the few examples there are have been tested on single genes in a single organism (Carrier et al. 1997; Smolke et al. 2000). There is a significant need to advance our understanding of mRNA turnover to enable accurate predictions of transcript stability. Ideally, the stability of a given transcript could be predicted based on its sequence. It would be necessary to know precisely which RNases will be involved in the degradation of the message. Unfortunately, there is no current method to predict the stability of a transcript or even determine if it's half-life will be short, medium, or long.

The majority of studies on mRNA turnover have focused on the model heterotrophic bacteria *Escherichia coli* and *Bacillus subtilis*. These studies have elucidated many RNA modifying and processing enzymes and have highlighted the differences between these Gram-negative and Gram-positive bacteria. In contrast, much less is known about mRNA turnover in cyanobacteria. Expanding our knowledge of transcript degradation in an organism with a unique combination of enzymes will advance our understanding of this essential, fundamental process as well as enable the engineering of synthetic transcripts.

1.2 Thesis Overview

This thesis explores the native control of gene expression in cyanobacteria, specifically how transcripts are turned over. There is an in-depth analysis of the role of a specific ribonuclease, RNase III, and a study of global RNA turnover in cyanobacteria. Additionally, genetic tools were built to manipulate gene expression in model cyanobacteria *Synechococcus* sp. strain PCC 7002.

Chapter 1 reviews the current genetic and regulatory tools used to control cyanobacteria gene expression. Several cyanobacteria-specific aspects of cellular physiology are described that could contribute to the observed differences in gene expression between cyanobacteria and other microbial species.

Chapter 2 contains a preliminary study using RNA-sequencing as a method to examine ribonuclease activity in *E. coli*. We found that RNA-sequencing was a useful tool that could pinpoint

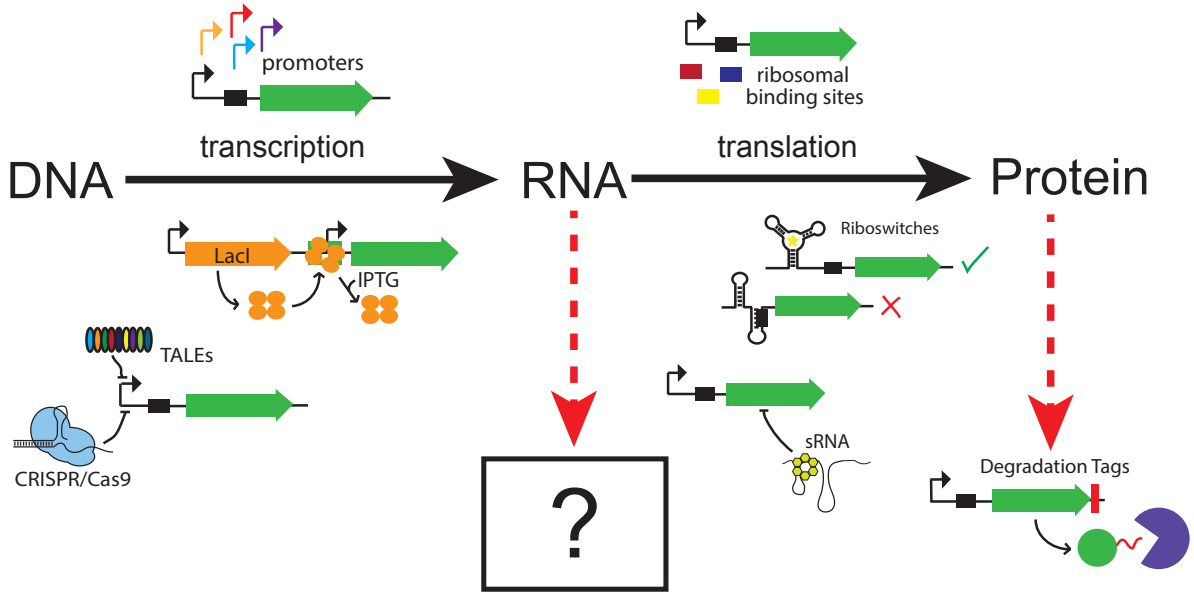


Figure 1.1: Synthetic biology tools built to control gene expression. The central dogma is shown as well as genetic tools and the steps where they exert control.

ribonuclease cleavage sites and enabled the discovery of novel RNase III cleavage sites in well-studied *E. coli*. We observed the degradation of the transcripts following cleavage events, and found some transcripts were stabilized as a result of cleavage. Our findings suggest that gene expression regulation via RNase III is complex, and RNase III plays a prevalent role in regulating metabolic enzymes.

Chapter 3 analyzes the RNases that facilitate RNA turnover in cyanobacteria and details their essentiality in PCC 7002. We explored the diversity and distribution of the multiple homologs of RNase III in many cyanobacterial strains. Analysis of RNases from 66 representative cyanobacterial genomes showed the conserved presence of both a homolog of RNase E and RNase J.

Chapter 4 explores the roles of three homologs of RNase III in PCC 7002. RNA-sequencing from the single, double, and triple RNase III mutants revealed complex gene expression changes. We describe the activity of two of these RNase IIIs in rRNA processing, and a role for the third in copy number regulation of a native plasmid. Complex changes between the single and triple

mutants highlight potential metabolic deficiencies that can be compensated for in single mutant strains but not when two or more RNase IIIs are missing.

Chapter 5 probes global RNA turnover in PCC 7002 using RNA-sequencing. Half-lives were determined for over 90% of open reading frames and degradation was very rapid, but photosynthesis genes were both highly expressed and their transcripts relatively stable. Transcripts that had a terminator with a poly-U-tract were more stable than those without a poly-U-tract, indicating the potential important role of 3' to 5' exoribonuclease activity on transcript degradation in cyanobacteria.

Chapter 6 describes the gene expression tools developed in cyanobacteria. We built a CRISPR interference system in PCC 7002 that allowed for tunable repression of both native and heterologous targets. We demonstrated the utility of this genetic tool by repressing both unessential (phycobilisome) and essential (carboxysome) native targets and used it to improve lactate production. Additionally, we characterized cyanobacterial promoters, tested ribosome binding site libraries, and demonstrated that 5' hairpins can stabilize transcripts in PCC 7002.

Chapter 7 summarizes the findings in this document and highlights future areas of study that will enable reliable prediction and design of synthetic transcripts in cyanobacteria.

The combination of both probing native gene expression control and testing genetic tools will significantly advance both the understanding of how genes are expressed in cyanobacteria and the ability to manipulate expression levels of key enzymes involved in pathways of interest.

1.3 Review of regulatory tools for controlling gene expression in cyanobacteria

To meet growing societal needs and reduce the use of fossil fuels as chemical feedstocks, researchers are engineering photoautotrophs to create chemical products from low-cost, renewable inputs: carbon dioxide, water, and sunlight. Harnessing the power of photoautotrophs has substantial benefits compared to using heterotrophs because carbon dioxide is a cheap, overabundant form of carbon that would not compete with edible sugars derived from crops (Desai et al. 2013). When selecting a photoautotrophic host, synthetic biologists are faced with two trade-offs – production cost certainty and ease of genetic engineering. Modern agricultural practices provide solid estimates of plant cultivation and harvesting costs, while the uncertainty of large-scale cyanobacteria cultivation remains a barrier to commercialization. Conversely, cyanobacteria can be genetically modified in the lab with mutation cycles measured in weeks rather than months for most crops. For this reason, metabolic engineering and synthetic biology in plants has progressed more slowly than in microbes and other heterotrophic hosts. That said, the growing number of fully sequenced plant genomes and the emerging ability to make precise insertions, deletions, and targeted mutations with Cas9 and similar nucleases will likely revolutionize the field of metabolic engineering in plants. Similar breakthroughs in cyanobacteria cultivation techniques will be needed to leverage the other physiological and processing advantages of cyanobacteria. These advantages include faster growth rates, greater salt tolerance, ability to grow on marginal lands, and ability to use waste streams as nutrients.

There is great diversity in cyanobacterial morphology, physiology, and environmental preferences. There are marine species, *Prochlorococcus* and *Synechococcus*, that are abundant in the ocean and could be responsible for as much as 25% of marine carbon fixation (Flombaum et al. 2013). There are freshwater cyanobacteria, as well as cyanobacteria found in glaciers, soil, deserts, and hot springs (Whitton et al. 2000). Different cyanobacteria species are cultivated as single cells

or as multicellular aggregates; in salt or fresh water; in the presence or absence of reduced nitrogen (i.e. nitrogen fixing or non-nitrogen fixing). Given this wide diversity, researchers have selected a few model cyanobacteria to study and engineer in detail. *Synechocystis* sp. strain PCC 6803 is a unicellular, freshwater cyanobacteria that has been studied for decades as a model of photosynthesis (Ikeuchi et al. 2001). Another freshwater cyanobacterium, *Synechococcus* sp. strain PCC 7942 has been extensively studied especially as a model of the circadian clock. There have been recent efforts to study and engineer *Synechococcus* sp. strain PCC 7002, a halotolerant, unicellular cyanobacteria, because of its faster doubling time compared to PCC 6803 and PCC 7942 (Bernstein, McClure, et al. 2016). *Anabaena (Nostoc)* sp. strain PCC 7120 is a cyanobacteria capable of fixing nitrogen through the formation of heterocysts (Kumar, Mella-Herrera, et al. 2010), a trait that would reduce (if not eliminate) the need for providing reduced nitrogen to cultures. In the remainder of this document, we will largely focus on gene regulatory tools used in these model organisms (**Table 1.1**) because these species are where regulatory tools have been developed and deployed. Most regulatory tools were built to interrogate natural phenomena found in cyanobacteria or to engineer cells to produce valuable chemicals. The future will tell us whether existing tools will transfer to other novel cyanobacteria or if corresponding tools will need to be created to leverage advantageous traits in non-model species.

Cyanobacteria are attractive hosts for converting carbon dioxide and sunlight into desirable chemical products. To engineer these organisms and manipulate their metabolic pathways, the biotechnology community has had to develop genetic tools to control gene expression. Many native cyanobacterial promoters and related sequence elements have been used to regulate genes of interest and heterologous tools that use non-native small molecules to induce gene expression have been demonstrated. Overall, IPTG-based induction systems seem to be very leaky and have low fold induction ratios in cyanobacteria. A variety of other induction systems have been optimized to enable tighter control of gene expression. There are substantial differences in performance of many of these promoters and induction systems in cyanobacteria compared to *E. coli*, and we hypothesize

Table 1.1: Cyanobacteria strains and abbreviations

Cyanobacterial Strain Name	**Abbreviation**
Anabaena (*Nostoc*) sp. PCC 7120	PCC 7120
Gleobacter violaceus	G. violaceus
Leptolyngbya sp. strain BL0902	BL0902
Nostoc punctiforme	*N. punctiforme*
Prochlorococcus marinus subsp. *pastoris* str. CCMP1986	CCMP1986
Prochlorococcus sp. MED4	MED4
Synechococcus PCC 6301	PCC 6301
Synechococcus WH7803	WH7803
Synechococcus elongatus PCC 7942	PCC 7942
Synechococcus elongatus UTEX 2973	UTEX 2973
Synechococcus sp. strain PCC 7002	PCC 7002
Synechocystis sp. strain PCC 6714	PCC 6714
Synechocystis sp. strain PCC 6803	PCC 6803
Synechocystis sp. strain WHSyn	WHSyn
Thermosynechococcus elongatus BP-1	BP-1

that these differences are due to fundamental differences in physiology between organisms.

There are many successful examples of making native and foreign metabolites in cyanobacteria, and comprehensive lists of performance metrics (titer, rate, yield) can be found elsewhere (Nozzi, Oliver, et al. 2013; Wang, Wang, et al. 2012; Case et al. 2016; Lai et al. 2015). Unfortunately, titers and productivities lag significantly behind model heterotrophs engineered to produce the same compounds. It remains to be seen whether this discrepancy is due to differences in metabolism between photoautotrophs and heterotrophs, or arises from gaps in knowledge and engineering tools. While yeast, *E. coli*, and other heterotrophs have been the subject of metabolic engineering for several decades, a significantly smaller community of researchers has only recently pursued work with cyanobacteria. This difference can be seen in the standardized media and cultivation equipment used with heterotrophs and the diverse set-ups used to grow cyanobacteria. Furthermore, culturing cyanobacteria requires optimization of many additional environmental variables including light source and intensity, timing of light/dark cycling, CO_2 concentration, and vessel geometries. Despite these obstacles, there have been a number of significant steps taken to increase production

of target chemicals including enhancing flux through desired pathways, eliminating competing pathways, and developing tolerance to high product concentrations (Nozzi, Oliver, et al. 2013). All three of these strategies are heavily reliant on the ability to manipulate gene expression.

The well-developed tool sets for gene expression combined with the fast generation times of bacteria has enabled significant progress in engineering them to create non-native chemicals and protein products. There has been significant work to build and expand tool sets to control gene expression in cyanobacteria, because often tools developed in other organisms (*E. coli*) do not behave similarly. We will cover many examples where directly using genetic tools from other organisms into cyanobacteria has yielded unpredictable and undesirable results. There are a great many factors and steps that facilitate gene expression (**Figure 1.2**), some of which may be unique and cyanobacteria-specific. Many of the successful gene expression tools in cyanobacteria have been created by systematically optimizing parameters (e.g. DNA and protein sequences) from underperforming embodiments. In this chapter, we will discuss several unexplored areas of cyanobacterial physiology (the transcriptional landscape, a highly over-represented repeat sequence, presence of multiple copies of the chromosome, as well as prevalence of small proteins), that could be playing significant roles in controlling gene expression and confounding transfer of existing regulatory tools. With future research in these areas we believe improved gene regulatory tools will be built and greatly accelerate the use of cyanobacteria as a chemical production platform

Promoters

Endogenous Promoters

Many native promoters have been used to express heterologous genes in cyanobacteria and comprehensive lists can be found elsewhere (Wang, Wang, et al. 2012). There has been extensive study of many of the photosystem promoters (Nair et al. 2001; Mulo et al. 2012; Takahashi et al. 2010), but these promoters have limited utility for many metabolic engineering applications

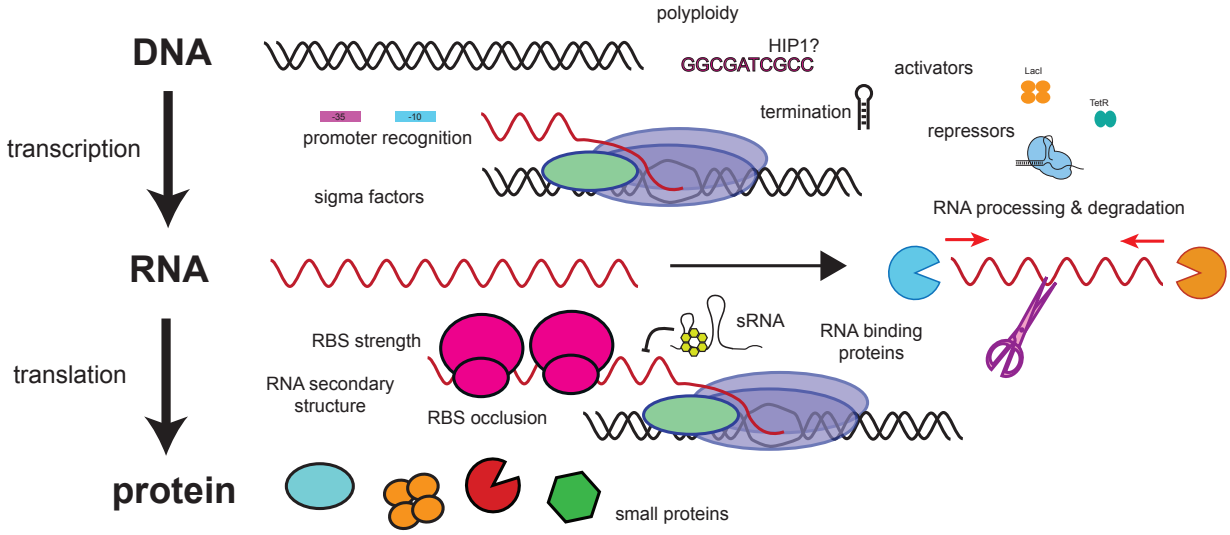


Figure 1.2: Overview of key factors governing gene expression

because their activity is light-responsive. Although many native promoters have been tested on reporters or genes of interest there has been little systematic study of native promoters to determine relative strengths or sequence elements that could be useful in producing predictable expression of a heterologous gene.

A few native promoters from PCC 7002 have been used to compare expression levels from a common reporter. Potential promoters with a range of expression levels as measured by RNA-seq were chosen and expression of heterologous yellow fluorescence protein (YFP) was tested. There was no correlation between expression level of native sequence and experimentally measured YFP levels, and most native promoters produced no more YFP than the background wild type fluorescence (Ruffing et al. 2016). This suggests that native promoters cannot be simply fused to heterologous genes without additional design.

Similarly, native promoters from PCC 7942 that produced high levels of protein (measured by proteomics) were used to attempt to increase expression of limonene synthase for the production of limonene. Varying levels of limonene were seen with different promoters, but only one native promoter produced as much limonene as the heterologous P_{trc} promoter with a synthetic RBS and

none outperformed this strain. A heterologous pea *psbA* promoter produced about 13-fold more protein and produced the most limonene (Wang, Chen, et al. 2016). Many native cyanobacterial promoters have been used, but sometimes they contain unknown regulatory elements that have unintended effects on gene expression.

Genetic Instability

Using endogenous promoters and regulators can cause genetic instability. It's been suggested that genetic instability, where strains suddenly drop in productivity or performance, is much more widespread than is reported and that researchers may not be incentivized to investigate failed results (Jones et al. 2014). One published example in PCC 7942 was an ethylene producing strain containing the ethylene forming enzyme (*efe*) that produced ethylene but looked yellow-green. Healthy, green/blue colonies soon appeared that no longer made ethylene, and these colonies were shown to have insertions at sites remarkably similar to the HIP1 sequence (see below) that caused a truncated protein (Takahama et al. 2003). This genomic instability of the ethylene-producing strain was not seen in PCC 6803 although this could be due to the use of a codon-optimized version of the gene (Guerrero et al. 2012).

Another example of genetic instability was seen in a strain of PCC 7002 engineered to produce mannitol. Cultures would sometimes suddenly stop producing mannitol and start to grow at a faster growth rate closer to WT. One mutant was isolated that had an inactivating point mutation in one of the two heterologous enzymes (Jacobsen et al. 2014). Spontaneous inactivating mutations were also seen in a PCC 6803 strain that contained a gene to make lactic acid and also expressed a transhydrogenase (Angermayr et al. 2012). Mutations causing single amino acid changes that lowered enzyme activity was seen in a PCC 7942 strain engineered to produce isopropanol (Kusakabe et al. 2013).

The causes of these mutations and genetic instability is unknown and could be related to DNA repair and mutations caused by expressing a heterologous pathway. Another factor causing genetic instability in cyanobacteria is the maintenance of many copies of the chromosome (see Polyploidy

section). There are countless examples of the difficulty and inability to obtain fully segregated strains, and variable productivities and phenotypes are seen when strains are not fully segregated. These observations motivate the development and use of more tightly-controlled gene expression regulation and highlight the advantages of using heterologous sequence elements that don't contain unknown native cyanobacterial control mechanisms.

Heterologous Promoters

Heterologous promoters have been tested to control gene expression to avoid issues with native regulation and potential strain instability due to repeated sequences. Widely used promoters in *E. coli* such as λP_L (BBa_R0040), λP_R (BBa_R0051), P_{Lac} (BBa_R0010) have been shown to have little to no detectable expression in PCC 6803 (Huang, Camsund, et al. 2010). Very little fluorescence could be detected from reporters expressed from P_{lac}, P_R, P_{trc10}, and P_{tet} in PCC 6803 (Huang, Camsund, et al. 2010). Changing elements such as the -10 element to the consensus 5'-TATAAT-3' has improved expression from heterologous promoters that originally worked well in *E. coli* but not in cyanobacteria (Huang and Lindblad 2013). However, it's also been suggested that a fully consensus promoter is undesirable and may lead to low activity because of the large number of contacts between RNA polymerase and the promoter (Hook-Barnard et al. 2007). Heterologous promoters may also perform differently in different cyanobacteria strains. Heterologous promoters P_{trc}, P_{LlacO1}, P_{conII}, P_{J23101}, and P_{J23119} were all tested in PCC 7942 and P_{conII} and P_{J23119} produced the highest expression levels (Huang, Shen, et al. 2016).

Heterologous artificial promoters have been tested in cyanobacteria and show variable results. Members of the Anderson promoter library (designed in *E. coli*, part of the BioBrick collection from the iGEM Registry) has been tested in both PCC 7002 (Markley et al. 2015) and in PCC 6803 (Camsund et al. 2014). In PCC 7002 there was relatively poor correlation when constructs were tested in both cyanobacteria and *E. coli* (R^2= 0.43426). Promoter libraries are a useful tool to get a large range of expression levels, but it is necessary to test them in a new host because they do not

seem to perform similarly across organisms.

Differences in cyanobacterial transcription from *E. coli*

Heterologous promoters may not function as expected due to the fundamental differences in transcription. Cyanobacteria have a strong consensus -10 element but seem to lack consensus or motifs in the -35 region (Vijayan et al. 2011). It has been suggested that cyanobacteria could be reliant on sigma factors that do not require the consensus -35 element. A study of native promoters in MED4 also observed the conserved -10 element but lacked a consensus -35 element. The authors suggested that the -35 regions may be divided into conserved subclasses (Vogel et al. 2003).

Cyanobacterial promoters have been divided into three groups based on the DNA sequence elements they contain and which sigma factors recognize them. Type I promoters (σ70) contain both a -10 element (consensus 5'-TATAAT-3') and a -35 element (consensus 5'-TTGACA-3'). Type II promoters contain the -10 element but lack a -35 element, instead containing target sites for other transcriptional activators. An example of a type II promoter is *glnB* P2 in PCC 6803 that contains the -10 element and the GTAN$_8$TAC binding motif recognized by NtcA which mediates enhanced transcription under nitrogen-limiting conditions. Type III promoters do not have consensus -10 or -35 elements and most likely are involved in responses to various stresses and stimuli through binding of type III sigma factors (Camsund et al. 2014).

Sigma factors are important proteins that facilitate binding of the core RNA polymerase with promoters to initiate transcription, and they are divided into three groups and recognize the corresponding type promoters. σ70 (Group 1) is the primary sigma factor in *E. coli* responsible for initiating transcription at housekeeping genes and cyanobacteria have a conserved homolog, *sigA* (Caslake et al. 1996; Brahamsha and Haselkorn 1991). Group 2 sigma factors are nonessential, and respond to environmental stimuli. Examples are the heat-shock responsive SigD, dark-responsive and nitrogen starvation-responsive SigB, and less well-studied SigC and SigE. Group 3 sigma factors have alternative structures and respond to stress and include SigF, SigG, SigH, and SigI.

SigF has been implicated in control of pili genes but little else is known about others (Imamura et al. 2009). The lack of knowledge of these important transcription factors as well as the coordinated networks between them make predicting transcriptional levels based off of promoter sequence extremely difficult.

Differences in core transcription machinery may also be responsible for the observed differences in expression of promoters between *E. coli* and cyanobacteria. The cyanobacterial RNA polymerase (RNAP) is composed of different subunits than that of *E. coli*; instead of one β' subunit (RpoC) like in *E. coli* cyanobacteria have both γ (RpoC1) and β' (RpoC2) (Imamura et al. 2009). Additionally, RpoC2 has large insertions between domains G and H. The reasons for this difference in architecture has not been explained or described in terms of gene expression control.

Other differences between cyanobacterial and *E. coli* RNAPs have been observed by testing purified RNAP complexes in vitro. RNAP from PCC 7942 and BP-1 both showed lower levels of abortive transcription and less miscincorpoation than *E. coli* RNAP (Imashimizu et al. 2011). Additionally, transcription of the same transcript was significantly slower with cyanobacterial enzymes (taking 164 seconds to complete the transcript compared to 26 seconds). These cyanobacterial RNAPs also paused more frequently and cleaved its transcript faster when NTPs were absent. Mn^{2+} is typically thought to be toxic because RNAP tends to misincorporate noncognate bases, but cyanobacteria maintain approximately 2 orders of magnitude higher intracellular levels of Mn^{2+} than *E. coli* because of its key role in photosynthesis (Keren et al. 2002). It was proposed that cyanobacterial RNAP maintains high fidelity even in the presence of high Mn^{2+} levels with expense of slower elongation (Imashimizu et al. 2011). If similar trends are also seen in vivo, these fundamental differences in transcription could have profound effects of gene expression and native mechanisms used to control gene expression.

Differences in RNAP-interacting factors may also be responsible for observed differences in transcription between cyanobacteria and other model organisms. The *E. coli* protein DksA has been shown to bind RNAP and help stimulate transcription from λP_R (Łyzeń et al. 2009). No

homologs for DksA have been found in cyanobacteria. Additionally, ppGpp affects gene expression at many promoters including λP_R (Łyzeń et al. 2009). ppGpp has been shown to be a regulator in cyanobacteria just as it is in model heterotrophs like *E. coli*, but alternate signals can affect its synthesis. In *E. coli* ppGpp levels rise as part of the stringent response during stresses. Contrastingly, ppGpp has been shown to be involved to adaption to darkness in PCC 7942 (Hood et al. 2016) and heterocyst formation in PCC 7120 (Zhang, Lin, et al. 2013).

Many promoters, including P_{lac}, require cAMP receptor protein (CRP) for transcription. These promoters contain a CRP binding site and positioning of this binding site relative to the -10 element is important to ensure that CRP and RNAP are on the same side of the DNA and can interact to cause activation. The *E. coli* P_{lac} promoter has a CRP-binding site -61.5 bases from the transcription start site (Busby et al. 1999), which is almost exactly 5 α-helical turns from the -10 element. A cAMP receptor protein has been identified in PCC 6803 (SYCRP1) and was shown to bind the consensus *E. coli* CRP sequence in the presence of cAMP (Yoshimura et al. 2000). Several SYCRP1-dependent promoters have been identified as well as their putative CRP-binding sites, but not all have proper spacing to ensure CRP and RNAP localize to the same side of the DNA (Hedger et al. 2009). Study of these important transcription factors in cyanobacteria may reveal global differences in specificity and function and help explain differential gene expression observed between organisms.

Global analyses have revealed intriguing aspects of the transcription landscape in cyanobacteria. Most of the genome seems to be transcribed with 88% having reads from RNA-seq align to either strand and 82% of the non-coding regions being transcribed in PCC 7942 (Vijayan et al. 2011). Transcriptome reads mapping to almost the entire genome ($\geq$94%) was seen for PCC 6803, PCC 7002, and CCMP1986 (Shi, Wang, et al. 2016). This shows remarkable similarity to plant chloroplasts where over 99% of the plant chloroplast genomes from rice, maize, *Arabidopsis*, and green algae *Chlamydomonas reinhardtii* are transcribed (Shi, Wang, et al. 2016). The model for transcription in chloroplasts involves haphazard initiation and termination resulting in many variable precursor transcripts that are then processed to their final 5' and 3' ends by endo and

exoribonucleases. To our knowledge there has been no in-depth study to see if a similar process could be occurring in cyanobacteria.

Other differences in transcription in cyanobacteria compared to *E. coli* include a distinct pause site approximately 63 nucleotides downstream of the transcription start site as observed by RNA pol ChIP data in PCC 7942 (Vijayan et al. 2011). This is slightly different than has been reported in *E. coli* where the majority of RNAP is associated with the promoter (Reppas, Wade, et al. 2006). Additionally, a preference for an adenine was seen at the transcriptional start site in PCC 7942 (Vijayan et al. 2011). These differences in both core RNAP machinery as well as factors that influence transcription initiation may help explain differences in promoter strength between *E. coli* and cyanobacteria.

Known Transcriptional Regulators

Global regulators play major roles in controlling gene expression, and there has been a small subset of global transcriptional regulators studied in cyanobacteria. One global transcriptional regulator, NtcA, has been shown to regulate nitrogen metabolism. NtcA belongs to the CRP (catabolite gene activator or cyclic AMP receptor protein) family and has been identified in nearly all cyanobacteria strains mentioned in this chapter (Herrero et al. 2001). DNase footprinting revealed that NtcA bound to $GTAN_8TAC$ which takes the place of the -35 element in NtcA-responsive promoters (Luque et al. 1994). Activation of these promoters is facilitated through an increase in the binding constant of RNAP and an increase in the rate constant for isomerization from the closed to open complex (Herrero et al. 2001). The consensus sequence for PCC 7120 NtcA was reported to be slightly different ($TGTN_{9\text{ or }10}ACA$) and proposed to be both a positive and negative transcriptional regulator, like many other members of the CRP family. This PCC 7120 consensus sequence is also located upstream of the large subunit of RuBisCO, *rbcL*, suggesting it may be playing other regulatory roles.

Another transcriptional regulator involved in sensing nitrogen compounds, NtcB, is often located

near NtcA and nitrate assimilation genes (*nir* operon) (Flores et al. 2005). In PCC 7120 NtcB has been shown to bind upstream of the *nir* operon and is a booster to enhance expression of the *nir* operon in combination with NtcA (Frias et al. 2000). There have been conflicting reports of conditions where NtcB activation is seen. In PCC 7942 NtcB activation of the *nir* operon is only seen in the presence of nitrite while in PCC 6803 and PCC 7120, activation is seen independently of the presence of nitrite (Aichi et al. 2001).

Other transcriptional regulators have been shown to involved in regulating genes involved in the CO_2 concentrating mechanism. CcmR and CmpR are both LysR family transcriptional regulators that sense and respond to CO_2 concentrations. Transcriptional activator, CmpR, has been identified in PCC 7942 and PCC 6803 and has been shown to bind to the upstream region of the *cmp* operon that encodes a high-affinity bicarbonate transporter (Omata et al. 2001). Ribulose-1,5-bisphosphate and 2-phosphoglycolate were able to enhance binding of CmpR to its targeting, suggesting that these metabolites are how this regulator senses CO_2 levels (Nishimura et al. 2008; Daley et al. 2012). The transcriptional repressor, CcmR, represses a suite of targets including *sbtA*, *bicA*, *porB*, as well as itself (Woodger et al. 2007). Its binding motif as well as novel regulon members have been identified through network analysis of PCC 7002 grown under many different conditions (McClure et al. 2016). α-ketoglutarate and $NADP^+$ have been identified as co-repressors of CcmR (Daley et al. 2012).

Other transcriptional regulators sense both nitrogen and carbon and mediate transcriptional responses. PCC 6803 has two cyanobacterial AbrB-like proteins (cyAbrBs); cyabrB1, which appears to have an overlapping or complementary regulon to NtcA (Ishii et al. 2008), and cyabrB2, which regulates parts of the high affinity bicarbonate transport system and overlaps with CcmR and CmpR regulons (Orf et al. 2016).

Many transcriptional repressors respond to iron and the redox status of the cell. The SufR transcriptional repressor has been identified in PCC 6803, PCC 7002, PCC 7120, *N. punctiforme*, BP-1 as well as others. SufR represses transcription of the suf operon that encodes proteins involved

in the assembly of iron-sulfur clusters. This repression is removed under conditions of oxidative stress or iron limitation (Wang, Shen, et al. 2004). FurA is a global cyanobacterial regulator that binds a 19 to 23 bp consensus sequence in the presence of iron. It can act as either an activator or repress in PCC 7120 (González et al. 2014) and similar results have been found for PCC 7942 (Ghassemian et al. 2017). PedR is a LuxR type transcriptional regulator that senses the redox state of the photosynthetic electron transport chain and facilitates gene expression change (Horiuchi et al. 2010).

This is just a small subset of the major transcriptional regulators in cyanobacteria. There remain many uncharacterized transcriptional regulators, but elucidating both their sensing mechanism, targets, and functions is complicated by their many complementary and overlapping targets. Recently, 32 putative transcriptional regulators were studied in PCC 6803 by screening knockouts grown under a variety of conditions and using liquid chromatography-mass spectrometry based metabolomics (Shi, Zhang, et al. 2017). There are clearly many unknown factors that could be influencing expression of a gene of interest.

Induction/Repression Systems

Whether studying native functions or implementing synthetic biology strategies it is desirable to be able to turn on and off gene expression with desired timing and strength. Induction and repression systems are well-established in model heterotrophs and can respond to an environmental shift or presence/absence of a small molecule. Unfortunately, directly moving these systems into cyanobacteria has not been universally successful (discussed below). Therefore, significant efforts have been made to adapt and optimize them for use in cyanobacteria.

It is extremely difficult to compare induction systems both within and across cyanobacteria strains and species. Different reporters have been used, integrated at different chromosomal locations, with different selectable markers. Comparisons between different strains is difficult due to the different culturing conditions with varying media compositions, light intensities, temperatures,

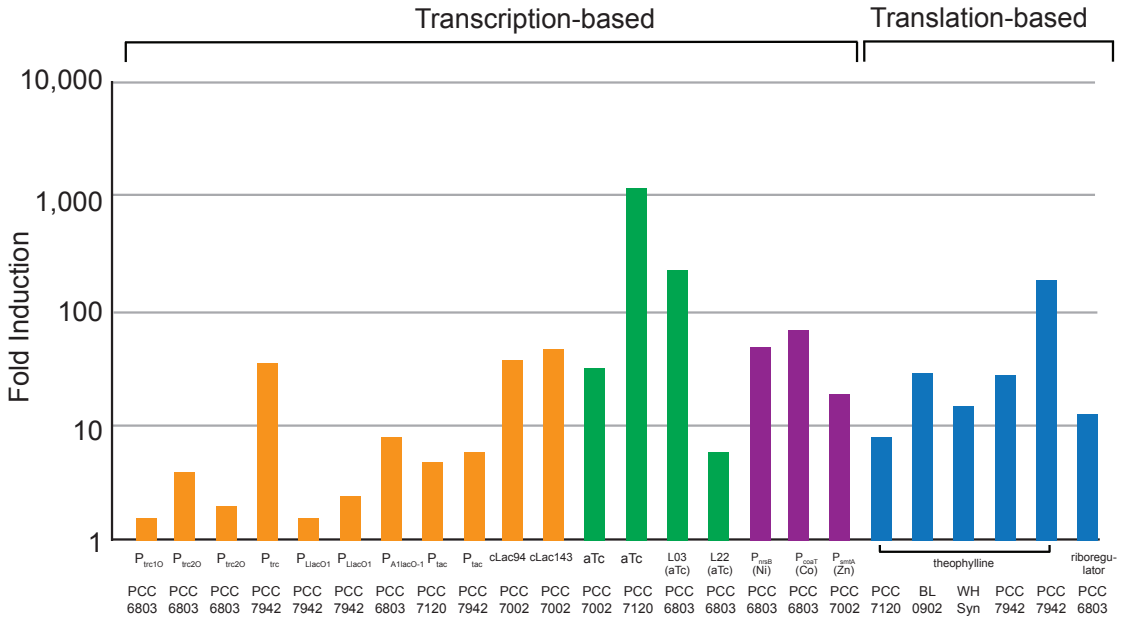


Figure 1.3: Fold induction ratios from inducible systems in cyanobacteria. We've color-coded bars by inducer (IPTG, aTc, metals, and theophylline), and separated them by induction mechanism (transcription vs translation).

culture vessels and CO_2 concentrations. Most of the engineered genetic tools have been designed and optimized with fluorescent reporters as the output due to their ease of quantification. Most papers report fold-induction ranges between uninduced and induced strains, which gives little indication of overall ranges of expression. Additionally, these fold inductions take different forms: fluorescence of reporter proteins, protein quantification with western blots, transcript levels via quantitative PCR, measurements of product titers, or changes in cellular physiology.

A compilation of the fold induction data from all of the induction systems outlined in the following two sections can be seen in **Figure 1.3**. We've divided them into both transcription-based and translation-based induction systems. There is a wide range of induction ratios seen with various inducers and in various strains. Unfortunately, this comparison does not show overall expression levels due to the indirect measurements of induction in various publications. Even systems with low fold induction ratios can be extremely useful and may have tighter repression than systems with a larger dynamic range of induction.

Induction via transcription

Expression of many cyanobacterial genes has been shown to respond to environmental stimuli: high-light responsive *psbA* promoter, copper-regulated *petE* promoter, nitrate/nitrite-inducible *nirA* promoter, and the nickel-responsive *nrsA* promoter (Ma et al. 2014). Some of these promoters have been used to induce expression of a gene of interest. A native PCC 6803 *cpcG2* promoter that is up-regulated in green-orange light and down-regulated in red light was used to create a green-light inducible lytic system in PCC 6803 (Miyake et al. 2014). This light-responsive promoter was placed in front of a T4-derived lysis gene and was shown to induce cell lysis upon exposure to green light resulting in a reduced growth rate. Additionally, reporters driven from Ni^{2+}-inducible P_{nrsB} and Co^{2+}-inducible P_{coaT} showed 50 and 70-fold induction respectively, but have very limited ranges of inducer concentrations (Peca et al. 2007). A Zn^{2+} responsive system was built in PCC 7002 that utilizes the PCC 7942 SmtB repressor and SmtB-regulated promoter P_{smtA} and achieved 19-fold induction (Perez et al. 2017).

Systems that respond to these large environmental shifts or natural inducers have a limited utility for induction of proteins of interest because native targets would be affected as well. Often, it is ideal to have no other cellular or environmental changes besides the induction of the protein of interest, so cellular effects can be attributed to the protein or interest and not global shifts in regulation and metabolism. Efforts have been made to use non-native induction systems and optimized to achieve activity in cyanobacteria.

Directly moving these induction systems from other organisms has been unsuccessful and there has been significant difficulty with IPTG-inducible promoters. The widely used IPTG-inducible, P_{trc} did have high expression in PCC 6803, but only led to 1.6-fold induction when LacI was present and IPTG was added (Huang, Camsund, et al. 2010). Leaky expression from the P_{trc} promoter was also seen without ITPG when a tagged protein was purified in PCC 6803 (Ng et al. 2000). In PCC 7942 an IPTG induction system using P_{trc} was sufficiently leaky to complement a mutant strain even in the absence of the inducer (Mutsuda et al. 2003). Leaky expression of heterologous enzymes was

sufficient for isobutanol production in PCC 6803 without inducer (Varman et al. 2013). However, there are other examples in PCC 7942 where 360-fold induction or higher was seen with P_{trc} (Geerts et al. 1995).

Many versions of IPTG inducible systems were very leaky in the absence of IPTG, resulting in low fold induction ratios. The P_{tac} (IPTG-inducible combination of *E. coli*'s P_{trp} and P_{lacUV5}) was inducible in PCC 7120, but induction was only 5-fold (Elhai 1993). Induction of heterologous genes in the 2,3-butanediol pathway in PCC 7942 using P_{LlacO1} was only 1.2 to 1.6-fold and showed very significant leaky expression in the absence of IPTG (Oliver, Machado, et al. 2013). Use of IPTG-inducible trp-lac promoter to control a sucrose permease gene in PCC 7942 showed 6-fold induction as measured by quantitative reverse transcriptase PCR (Ducat et al. 2012). A variant of P_{lac}, $P_{A1lacO-1}$ that contains a second lac operator sequence, showed stronger repression than P_{trc}, but became leakier at higher cell densities (Guerrero et al. 2012). However, there are examples where similar fold inductions ($\sim$2.5-fold) are seen in both *E. coli* and cyanobacteria (Nozzi and Atsumi 2015).

A systematic study varying the spacing and number of the lac operator sequences in PCC 6803 showed that there was weak repression when only a single lac operator sequence was used (P_{trc1O} and similar variants). However, when the second operator was added (P_{trc20} and similar variants) there was very significant repression, as high as 400 times, when LacI was not present, but these strains could not be induced. They observed the same reliance on the 11 bp periodic placement of the operator sequences as seen in *E. coli*, ensuring LacI and DNA looping can occur on the same side of the DNA. However, very significant differences in repression and induction were observed, that remain unexplained (Camsund et al. 2014).

There seems to be a trade-off between using heterologous IPTG-inducible systems from *E. coli* and native cyanobacterial induction systems. Different induction systems were used to drive expression of enzymes in ethylene metabolism in PCC 6803 and it was shown that P_{trc} produced high expression and ethylene levels, but was not tightly regulated like the same construct expressed

in *E. coli*. Contrastingly native metal-inducible promoters (P_{petE}, P_{coa}, and P_{smt}) were tightly repressed, but had low expression even when induced (Guerrero et al. 2012). Similarly, a native cyanobacterial promoter controlled by a metal-sensing repressor, P_{smt}, had the highest induction ratio in PCC 7942 when compared with P_{LtetO1}, $P_{ConII\text{-}ribo}$, P_{trc}, P_{LlacO1}, and P_{BAD} (Huang, Shen, et al. 2016).

Directly moving induction systems from *E. coli* has not always led to favorable expression levels or tight control of expression levels, so several induction systems have been optimized and modified for use in cyanobacteria. Two IPTG-inducible systems have been optimized for PCC 7002 (Markley et al. 2015). Initial systems were optimized by changing the -10 elements, the spacing between the -10 element and the LacO operator sequence, as well as the spacing between the -35 and -10 element. Both constructs were titratable between 1 μM – 10,000 μM IPTG but had different ranges of expression. cLac143 had higher expression and was somewhat leaky in the absence of IPTG ($\sim$5% of the constitutive version), but had a 48-fold range of expression. cLac94 had a 38-fold range expression, but does not reach the same expression level when fully induced. These two constructs used the cyanobacterial phycocyanin truncated promoter (P_{cpt}) flanked by LacO operator sequences to drive expression of the gene of interest. LacI expression was controlled by P_{MB2}, a synthetic promoter derived from BBa_J23119. Tighter repression was seen by using *lacI* W220F that has been shown to be less leaky possibly due to tighter DNA binding (Gatti-Lafranconi et al. 2013). Induction was also seen when P_{cpt} was replaced with P_{trc}, but a smaller dynamic range was achieved. These optimized IPTG-inducible systems have some of the largest fold induction ranges of those tested so far.

Anhydrotetracycline (aTc) induction systems have been developed for PCC 7002 (Zess et al. 2016), PCC 7120 (Higo et al. 2016), and PCC 6803 (Huang and Lindblad 2013). The aTc induction system in PCC 7002 had a 32-fold induction and was titratable with concentrations of aTc between 10 ng/mL and 1000 ng/mL aTc. Although the 32-fold induction is not as high as some other inducible systems, the tightness of this system has advantages compared to other leakier but higher

expression systems. The aTc induction system was critical to achieve inducible repression using CRISPR interference in PCC 7002 (**Chapter 6**).

The aTc induction system in PCC 7120 achieve >1,200-fold induction and had a dynamic range with between 0 ng/mL and 2,000 ng/mL (Higo et al. 2016). There was a decrease in growth of PCC 7120 with 2000 ng/mL aTc, but the growth rate appeared to be unaffected with 200 ng/mL aTc. This system was adapted to respond to nitrate by placing the *nirA* nitrate-specific promoter in front of *tetR* and over 200-fold induction was seen in the absence of the inducer when cells were grown in nitrogen-free conditions in comparison to media supplemented with nitrate.

The aTc induction system for PCC 6803 was extensively optimized and achieved 230-fold induction under light-activated heterotrophic growth (Huang and Lindblad 2013). Expression was titratable with aTc concentrations between 0 ng/mL - 4 ng/mL aTc. This induction system was created by systematically altering the -10 element sequence and the bases between the -10 element and the transcription start site. A guanine in the 2nd position downstream of the -10 element was shown to be important for high promoter strength, that might be caused by altered lifetime of the promoter-RNAP closed complex.

These systems had some of the largest fold induction ratios (**Figure 1.3**), but their use for long term induction is limited due to aTc's sensitivity to light. The dynamic range of induction in PCC 7002 dropped significantly after both 24 and 48 hours exposure to light (Zess et al. 2016). Induction persisted for at least 72 hours in PCC 7120, possibly because the cyanobacterial cells were absorbing light and protecting from aTc degradation (Higo et al. 2016). aTc has also been shown to be sensitive to temperature, with a half-life of only 15 hours at typical cyanobacterial cultivation temperature (37°C) compared to 30 hours 30°C (Politi et al. 2014). The lability of the inducer in the culture conditions is often forgotten and could be responsible for some of the different dynamics seen above.

Induction via translation

Induction can also occur at the level of translation, where the translation of a transcript can respond through a molecule that alters the occlusion or accessibility of the ribosome binding site. Theophylline-responsive riboswitches have been tested and used in a number of diverse cyanobacteria. A set of six riboswitches originally designed for and shown to be effective in a diverse group of Gram-negative and Gram-positive bacteria (Topp et al. 2010) have been tested in cyanobacteria. In PCC 7942, riboswitch E* (a slight change in the RBS from E) showed 190-fold induction of a luciferase reporter while maintaining minimally leaky expression (Nakahira et al. 2013). Expression had a wide range of induction with concentrations between 0 and 2 mM theophylline. This riboswitch was successfully used to regulate circadian clock *kaiC* expression.

The same set of theophylline riboswitches was tested in a group of 4 diverse cyanobacteria strains (PCC 7942, BL0902, PCC 7120, and WHSyn). As was seen in the original diverse group of bacteria, various fold inductions, leakiness, and gene expression ranges were seen between the six riboswitches. Fold induction ratios in cyanobacteria seemed significantly lower than those reported for the diverse set of other bacteria with cyanobacteria fold induction ranges reaching maximally 30-fold (Ma et al. 2014) whereas several of the previously tested bacterial strains reached 100 or 2000-fold (Topp et al. 2010).

It is thought that cyanobacterial small RNA (sRNA) play a large role in controlling gene expression. sRNAs are very prevalent with 371 reported in PCC 6803 (Kopf et al. 2015), but only a few of these sRNA seem to be widely conserved (NsiR1, PsrR1, Yfr1, HliR1, PsiR1, and PsiR2) (Kopf et al. 2015). The lack of development of sRNA tools for cyanobacteria is surprising, with only two reports of synthetic sRNAs used to mediate gene expression. A trans-acting sRNA system based on the *E. coli* IS10 RNA-IN/OUT system was adapted for use in PCC 7002 and achieved 70% attenuation (Zess et al. 2016), slightly less than the 90% attenuation in *E. coli* (Mutalik et al. 2012). sRNA control has also been implemented in cyanobacteria through the use of riboregulators. Riboregulators consist of a cis-repressed mRNA (crRNA) that forms a stem loop that occludes the

RBS. A trans-activating RNA (taRNA) is complementary to part of the stem loop and hybridizes with the crRNA and enhances ribosome binding. By driving the crRNA from the trc promoter and the taRNA from the Ni^{2+} inducible promoter P_{nrsB}, they saw 13-fold induction of the GFP reporter upon the addition of Ni^{2+} in PCC 6803 (Abe, Sakai, et al. 2014). Induction systems based on the level of translation have proved to be a useful tool in cyanobacteria and provide an avenue for gene expression control independent of the complex and unstudied aspects of cyanobacterial transcription machinery.

Ribosome binding sites

The ribosome binding site (RBS) is a major factor controlling the amount of translation of a given gene. The complementarity of the RBS to the 16S rRNA in the ribosome is a major determinant of protein expression as well as the accessibility of this site given the secondary structure of the surrounding RNA sequence. Most cyanobacteria mentioned here (PCC 7002, PCC 6803, PCC 7942, PCC 7120, and PCC 6301) have the portion of the 16S rRNA that pairs with messenger RNA with a sequence of ACCTCCTTT which would suggest optimal pairing with mRNA with a central region of the RBS being AGGAGG (the consensus prokaryotic sequence).

Thermodynamic models have been developed to predict translation efficiency and have been quite successful in both *E. coli* and other organisms (Espah Borujeni et al. 2014). The most widely used software is the RBS Calculator (Salis et al. 2009), but other translational rate calculators include the RBS Designer (Na et al. 2010) and the UTR Designer (Seo et al. 2013). These calculators use statistical thermodynamic models that calculate free energies for interactions between the mRNA and rRNA and utilize RNA folding software to predict mRNA secondary structure. These calculators are capable of both designing sequences for a specified translation rate (forward engineering) and predicting the location and strength of a given RBS (reverse engineering). Each of these calculators tested predicted translation initiation rates and observed fluorescence or β-galactosidase activity of reporters in *E. coli* and saw good correlations using the forward engineering mode ($R^2 = 0.84$ for

the RBS calculator (Salis et al. 2009), R^2 = 0.87 for the UTR Designer (Seo et al. 2013)88, and R^2 = 0.87 for the RBS Designer (Na et al. 2010)). These calculators have been very useful and enabling, but there are limitations including the effects of antisense RNA, RNA binding proteins, translational coupling, or differences in translational elongation (Reeve et al. 2014).

There are many examples showing how manipulation of the RBS has enabled desirable changes in cyanobacteria protein expression. Using an optimized RBS in PCC 7120, L0908, PCC 6803, and PCC 7942 increased expressed of both CFP and YFP by several fold (Taton, Unglaub, et al. 2014). Although there are limited examples, RBS predictions do not seem to be as reliable in cyanobacteria as they are in cyanobacteria. Fluorescence from a 11-member library of RBSs in front of a YFP reporter was measured and compared with the predicted translation initiation rate using the RBS calculator, UTR Designer, and RBS Designer (Markley et al. 2015). Once an outlier that had a large increase in transcript level compared to all other members was removed, R^2 values were respectively 0.74, 0.78, and 0.04. Altogether, RBS design software is not perfect, but is an extremely useful tool to both study native regulation and achieve desirable expression levels of heterologous proteins.

CRISPR interference

CRISPR interference (CRISPRi) has proved to be a powerful method to repress native and heterologous gene repression in bacteria. CRISPRi utilizes the native RNA based defense system that bacteria use to target invading nucleic acids (Qi et al. 2013). The system most widely used is derived from the *Streptococcus pyogenes* type II system that utilizes a single nuclease protein (Cas9) and a small guide RNA (sgRNA) complementary to the target sequence. CRISPRi utilizes a deactivated Cas9 (dCas9) that has point mutations in its active sites so it can no longer make double-stranded breaks in DNA. Genes targeted with CRISPR interference are thought to be repressed because of tight binding of the dCas9/sgRNA complex to the DNA that causes a steric clash, dislodging RNAP (Gilbert et al. 2013). This heterologous system has been implemented in both PCC 7002, PCC 6803, and PCC 7942 and has been used to re-direct flux toward desired products.

Several of these cyanobacteria have a native CRISPR system or multiple systems, but there has been little exploration in their function or mention of cross-talk with CRISPRi.

In PCC 6803 a number of constitutive and inducible promoters were tested to drive expression of dCas9 and sgRNA and their efficiencies were measured by the output of reporter green fluorescent protein (GFP) (Yao et al. 2016). They observed significant repression of GFP whenever both dCas9 and the sgRNA was present, and the best combination was using the aTc-inducible P_{L22}. They observed 96% of the maximal GFP fluorescence without aTc, and only 6% when aTc was added, but there was still significant repression with intermediate levels of aTc. Other promoters caused maximal repression (>99%) even without the presence of the inducer, highlighting that low levels of expression are necessary to achieve inducible repression. They targeted native *phaE* for repression and saw no PHB accumulation under nitrogen-starvation conditions which normally accumulates under these conditions and functions as a carbon and redundant storage polymer. They also targeted *glgC*, an enzyme involved in glycogen synthesis and showed that under nitrogen starvation the *glgC* repression strain only accumulated 25% of the glycogen compared to WT. They showed efficient repression of four targets at the same time. Repression of one target in PCC 6803 (*slr0091*) was only 2-fold whereas other targets were about 10-fold. This weak repression could have been because this gene is very weakly transcribed than the others, or because it is part of an operon.

An inducible and titratable system for CRISPRi was made in PCC 7002 (**Chapter 6**). Repression was seen immediately whenever dCas9 and the sgRNA was present, and the promoters and ribosome binding sites were changed to attempt to relieve this repression in the absence of the inducer. The system with the best titratable range had both dCas9 and the sgRNA under the control of the aTc-inducible promoter, P_{EZtet} (Zess et al. 2016). While this system only showed ∼70% of the maximal protein expression without the inducer, it was titratable with intermediate levels of aTc and only ∼10% of the maximal protein expression with maximum aTc concentrations. Significant repression of targets was seen even when the promoter of dCas9 was removed and when the sgRNA was truncated down to only 15 bases. Native phycobilisome (*cpcBACDEF*) and carboxysome genes

(*ccmK$_1$K$_2$LMN*) were also targeted and showed characteristic decreases in absorbance at 635nm and requirement for high CO_2 concentrations for growth respectively. This repression scheme was used to increase flux towards lactate by targeting glutamine synthetase (*glnA*) and over 2-fold more lactate was made in this strain than the baseline engineered strain. Interestingly, increased lactate production was also seen in a strain that contained the sgRNA targeting *glnA* but lacked dCas9.

A CRISPRi system was built in PCC 7942 to target native genes and increase succinate production. Even when dCas9 was under control of inducible P_{smt} leaky expression was sufficient to repress targets in the absence of the inducer. Repression of *glgC*, *sdhA*, and *sdhB* (separately) using CRISPRi increased succinate production by $\sim$12.5-fold (Huang, Shen, et al. 2016).

These CRISPRi systems have enabled both increased titers of desired products as well as provided platforms to manipulate and study native gene regulation. Combined, these three systems show that CRISPRi is extremely effective at repressing gene expression, stable (effective for maximum time tested in PCC 7942: 21 days (Huang, Shen, et al. 2016)), and reversible (Yao et al. 2016) (**Chapter 6**). These studies also show that sgRNAs can efficiently target both the template and non-template strand, but most effective repression is seen when the target is near the transcription start site or near the beginning of the coding sequence. One remaining challenge is the fact that many bacterial and cyanobacteria genes are located within operons, limiting the application of CRISPRi. Nonetheless, CRISPRi is a very promising tool, enabling regulation of gene expression *in trans*.

RNA stability and degradation

Of the many steps in the central dogma of biology, RNA turnover is the least considered when designing gene expression modules. In cyanobacteria, RNA turnover is rapid and may play a substantial role in determining gene expression levels. Global half-lives of transcripts have been determined in *Prochlorococcus* MED4 using microarrays, and the median half-life was 2.4 minutes (Steglich et al. 2010). The observation that RNA is very labile and that RNA is turned over rapidly in

bacteria has been long observed, but the half-life reported for MED4 is on the shorter end compared with other recent global RNA studies (*B. subtilis* $\sim$ 5 minutes (Hambraeus, von Wachenfeldt, et al. 2003), *Bacillus cereus* 2.4 minutes (Kristoffersen et al. 2012), *E. coli* 6.8 minutes (Selinger et al. 2003); 4.7 minutes (Bernstein, Khodursky, et al. 2002); 2.5 minutes (Chen, Shiroguchi, et al. 2015), *Chlamydia trachomatis trachoma* 15-17 minutes (Ferreira et al. 2017), *Mycobacterium tuberculosis* 9.5 minutes (Rustad et al. 2013)). The reason for this especially rapid turnover remains unexplained although it has been suggested that RNA turnover could be linked with cell doubling time. This seems unlikely when the doubling times and average/median global transcript turnover is compared (Steglich et al. 2010).

This rapid turnover of RNA in cyanobacteria could be a result of its somewhat unique composition of ribonucleases (RNases). Cyanobacteria have homologs of both RNase E (endoribonuclease that often initiates turnover in *E. coli*) and RNase J (5'$\rightarrow$3' exoribonuclease responsible for initiation of mRNA degradation in *B. subtilis*). Both RNase E and RNase J appear to be essential in PCC 7002 because attempts to delete either were unsuccessful (**Chapter 3**). This combination of RNase E and J exists in several other bacteria including *Sinorhizobium meliloti*, *Mycobacterium smegmatis*, *Rhodobacter sphaeroides*, and *Streptomyces coelicolor*. Studies in these organisms have shown that both RNase E and RNase J play important roles in rRNA maturation (Taverniti et al. 2011; Rische et al. 2012) as well as in the production of antibiotics (Bralley et al. 2014) and SAM metabolism (Baumgardt et al. 2017). A global model of RNA degradation in organisms containing both RNase E and J has yet to be proposed and universal tools for modulating RNA turnover are not available.

Another difference between the cyanobacteria RNA degradation machinery and other more well-characterized systems in the apparent lack of a 5' pyrophosphohydrolase, RppH (Chapter 3). RppH is part of the Nudix (nucleoside diphosphates linked to x) protein family that removes pyrophosphate from the 5' end of transcripts leaving a monophosphate (Deana et al. 2008). This processing event affects transcript degradation because both RNase E and RNase J can sense the 5' of a transcript and cleave more efficiently in the presence of the 5' monophosphate (Mackie 1998;

Callaghan et al. 2005; Pei et al. 2015). While no proteins with 5' pyrophosphohydrolase activity have been identified in cyanobacteria, RppH could still be present and be an important element controlling transcript degradation. All RppH proteins identified so far have been part of the Nudix family, but it is difficult to determine the substrates of these protein family members based on amino acid sequence. RppH homologs have been identified previously by purifying these proteins and testing their activities in vitro (Bischler et al. 2017). Cyanobacteria possess many proteins annotated as members of the Nudix family and it is yet to be seen if one may have the activity of RppH. Recent reports show that RppH may not be the only enzyme involved in the processing of the 5' end, and that a yet-identified protein may be responsible of initial cleavage from a 5' triphosphate to a 5' diphosphate and serve as the precursor for RppH (Luciano, Vasilyev, et al. 2017). Given that ribonucleases play a substantial role in regulating gene expression, it is important to identify the full complement of RNA processing enzymes in cyanobacteria.

Transcriptional Termination

Transcription termination has not been widely studied in cyanobacteria. Observations that the cyanobacterial genome is nearly fully transcribed, points to potential differences in transcriptional termination from other model organisms. In bacteria, termination is either rho-dependent, where the protein Rho recognizes a target sequence and dissociates RNAP, or rho-independent where an RNA structure in the transcript destabilizes RNA-RNAP interactions. In rho-dependent termination Rho recognizes a Rho-utilization site (RUT) in recently transcribed RNA, wraps the RNA inside the Rho hexamer, moves along the transcript, and is thought to pull the RNA out of the RNAP active site. The site of termination is not well defined and can be as far as 120 bases from the RUT site (Nudler et al. 2002). It is unlikely that cyanobacteria use rho-independent termination because PCC 7942, PCC 6803, and PCC 7002 do not have an obvious homolog of Rho (Vijayan et al. 2011). Many assume that cyanobacteria rely on rho-independent termination also known as intrinsic termination. Transcripts that are terminated via intrinsic termination have a 3' secondary

structure that is composed of a GC-rich palindromic sequence followed by stretch of T's (Carafa et al. 1990). The stretch of T's is thought to induced transcriptional pausing, allowing time for the hairpin to form which in turn dislodges key interactions between three nucleic acid binding sites in RNAP (the RNA:DNA T-A hybrid, RNA-binding site of just added ribonucleotides, and upstream RNA-binding site of RNA exiting RNAP) (Nudler et al. 2002). Intrinsic termination structures can be found for almost 50% of protein-coding genes in *E. coli* (Lesnik et al. 2001).

It was originally presumed that cyanobacteria must use rho-independent termination because of the lack of a Rho homolog, but genome wide predictions of hairpin formation near stop codons showed that termination in cyanobacteria may not involve hairpin formation. In *E. coli* there is a significant sharp drop in free energy approximately 30 bases downstream of stop codons indicating the presence of a hairpin. Genome-wide free energy calculations were made for many organisms including PCC 6803, and no sharp dip in free energy was seen (Washio et al. 1998). The lack of dips in free energy values was also seen in several other genomes including *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Helicobacteri pylori*, *Borrelia burgdorferi*, *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum*, *Aquifex aeolicus*, *Pyrococcus horkoshii*, *Mycobacterium tuberculosis*, and *Treponema pallidum* (Washio et al. 1998). Dips in the free energy prediction near the 3' end of transcripts, however, were seen in PCC 7942 (Vijayan et al. 2011).

Transcriptome studies suggest that cyanobacteria transcribe a large portion of non-coding regions as well as sequences within coding sequences, but on the opposite strand (Shi, Wang, et al. 2016). Cyanobacteria have intriguing parallels in terms of their transcriptional landscapes when compared with plants chloroplasts. It is a well-established in chloroplasts that transcription is very inefficient, and it has been proposed that transcription initiation and elongation are very haphazard, creating a wide array of varying length transcripts (Shi, Wang, et al. 2016). Many transcripts in chloroplasts have been shown to be variable in length with a mixture of 3' extensions (Stern et al. 1987).Transcripts are thought to be processed by proteins to have the proper 5' and 3' ends.

This phenomenon has been analyzed globally using RNA-sequencing and also been studied on an individual transcript basis. Similar results have been shown in plant mitochondria where variable 3' extensions are trimmed back to stem loop structures by exonucleases (Dombrowski et al. 1997).

It's been observed that many chloroplast genes have inverted repeats at the 3' end of the transcript (including *psbA*, *rbcL*, *petD*, and *atpE*) (Stern et al. 1987). These 3' inverted repeats have been shown not to cause termination like 3' secondary structure does on many bacterial transcripts, but instead serve as processing sites. Without these 3' inverted repeats these transcripts are rapidly degraded (Stern et al. 1987).

Transcriptional termination can play a role in controlling gene expression of genes that are located in operons. Terminators are not necessarily efficient and can allow various amounts of read-through. In PCC 7002 the phycocyanin operon (*cpcBCADEF*) encodes six genes that encode structural subunits of the phycobilisome, linker proteins and assembly enzymes, and has been shown to make eight to ten different length transcripts (Zou et al. 1999) (**Figure 1.4**). There are predicted stable stem loop structures at several locations within this transcript that could serve as terminators or processing sites. Support for their role as inefficient terminators includes that only 1-2% of the transcripts include *cpcE* and *cpcF* located at the 3' end of the operon (Zou et al. 1999). The structural components of the phycobilisome are encoded in the first two genes of the operon, followed by two linker proteins, and finally enzymatic lyases for attachment and their decrease in expression levels seems to match the stoichiometry needed for phycobilisome function (Vijayan et al. 2011).

Similar regulation of the phycobilisome operon is seen in PCC 7120, except with slight differences in operon organization and the number of different transcripts made. The PCC 7120 operon contains ten open reading frames (*cpcBACDEFG$_1$G$_2$G$_3$G$_4$*), and several different transcripts are seen (Belknap et al. 1987; Grossman et al. 1993). The major transcript includes *cpcBA*, but full-length transcripts as well as partial transcripts have also been observed, including some that suggest an internal *cpcC* promoter. Although the presence of processing events could not be ruled

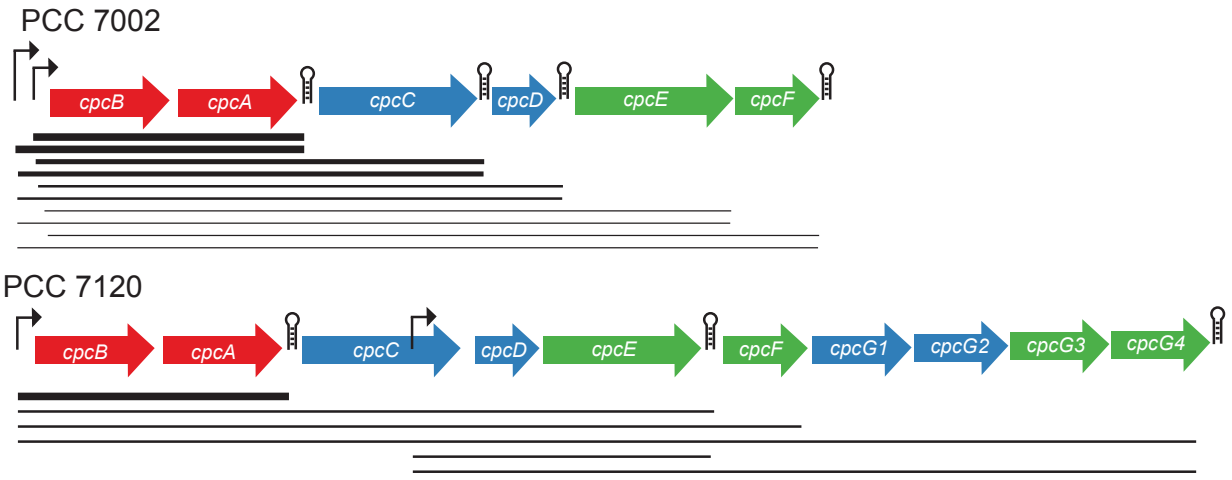


Figure 1.4: Organization and transcripts observed from *cpc* operons in PCC 7002 and PCC 7120. Genes are color-coded for their functions (red: structural components, blue: linker proteins, and green: attachment enzymes). Promoters are marked as well as regions that predicted to form significant secondary structures. Observed transcripts are marked below with the thickness indicating relative abundance.

out, it was suggested that these transcripts arose through partial termination (Belknap et al. 1987). Most of the locations where partial transcriptional termination is occurring are also predicted to form strong secondary structures. This partial read-through seems to be controlling expression levels of important parts of this light harvesting complex.

Partial termination may be playing a large role in regulating gene expression from the *cpc* operon, an operon extensively studied because of its important role in light harvesting and photosynthesis. The need to for analysis of transcriptional termination in cyanobacteria has been mentioned before (Ramey et al. 2015), and perhaps difference in transcriptional termination could play major roles in controlling gene expression.

Advances in Genome Editing

One significant advantage of working in cyanobacteria is that many of the commonly used strains are naturally competent and readily take up foreign DNA: PCC 6803 (Barten et al. 2017),

PCC 7002 (Stevens and Porter 1980), PCC 7942 (Golden et al. 1984), and others. DNA-uptake in cyanobacteria is thought to involve a type IV pilus apparatus that may be responsible for DNA translocation as a single strand (Schuergers et al. 2015). Many established protocols make use of this natural competency and involve incubating cyanobacteria cells with a plasmid or linear DNA piece containing regions homologous to the chromosome prior to plating on media with a selective pressure corresponding to a selection marker in the foreign DNA.

Traditionally, genomic integrations have been made by providing a plasmid or PCR product that contains the desired sequence surrounded by homology arms. Higher efficiency has been seen with longer homology arms, but they can be as short as 250 bp (Ruffing et al. 2016). Several studies have examined the length relationship between homology arm length and integration efficiency, and generally 500-1000 bases of homology are used.

Integrating constructs onto the genome may be more stable than maintaining them on a plasmid, but deciding the location of integration can be challenging. Ideally genes could be inserted in neutral sites that have no effect on physiology when grown in any conceivable environmental condition. Historically in PCC 7002 *desB* was used as an integration site (Ruffing et al. 2016), but it's expression at low temperatures and role in lipid desaturation (Sakamoto et al. 1997) make it an unattractive candidate. Many have used *glpK*, encoding glycerol kinase, as a neutral insertion site because of its original annotation as a pseudogene. Re-sequencing of PCC 7002 showed that *glpK* is intact and an error was made in the original sequencing (Peterson et al. 2012). Genomic integrations have also been introduced between A0935 and A0936 (termed NS1) (Ruffing et al. 2016). Given the abundance of transcription from the cyanobacteria genome, it may not be possible to identify a truly neutral site. That said, systematic studies of gene expression around the genome are not available for all cyanobacteria and many other hosts, a knowledge gap that could provide immediate help in designing engineered strains.

Several methods have been developed to obtain markerless genome edits, but many of these systems require multiple transformation and homologous recombination steps. Levansucrase,

encoded by *sacB*, confers sensitivity to sucrose, is one of the most popular counterselectable markers and has been adapted for uses in many organisms (Reyrat et al. 1998), including cyanobacteria. *sacB* was used as early as 1990 in PCC 7120 by integrating a selectable plasmid containing *sacB* onto the chromosome with a single crossover event and then selecting for a replacement event with the desired sequence (Cai et al. 1990). Unfortunately this technique produced a significant amount of spontaneous mutants as well as a mixture of five different insertion sequences at the target locus. Recently, a faster *sacB* system has been developed in PCC 6803 that utilizes a single vector containing kanamycin resistance gene and *sacB* nested within a split gene of interest with partial homology regions. This method requires two homologous recombination events: (1) insertion of the entire cassette and (2) recombination fusing the gene of interest and excising the kanamycin resistance-*sacB* region, but the time needed to produce mutant strains is greatly reduced without the need for a second conjugation or transformation step (Viola et al. 2014). Another one of the most popular markerless counterselectable markers, the *rsp12* (*strA*) system that relies on the resistance and sensitivity of different alleles to streptomycin, has been used in PCC 7942 (Matsuoka et al. 2001). This system unfortunately requires working in a mutant *rsp12* background. A novel counter selection system utilizing the sensitivity to acrylic acid in the presence of native *acsA* gene has been developed in PCC 7002 (Begemann et al. 2013).

The adaptation of the CRISPR system as a tool for genome editing has been an enabling technology in bacteria and higher organisms. CRISPR Cas9 systems are useful in cyanobacteria to allow genomic integrations or changes without the use of selectable markers and can be used to alleviate problems with segregation. One of the first examples of a CRISPR Cas9 genome editing system was in UTEX 2973 (Wendt et al. 2016). The system was based on one developed for use in *Streptomyces lividans* that uses a plasmid that is not functional at 34°C or higher. This plasmid was needed because Cas9 was shown to be toxic in UTEX 2973 when cells were transformed with a replicating RSF1010-based plasmid. This toxicity was still seen when the promoter and ribosome binding site of Cas9 was removed. Growing cells at a non-permissive temperature (38°C)

while forcing them to maintain the plasmid due to the presence of the antibiotic, appeared to allow enough transient expression for genome editing and genome edits appeared to be segregated in the first patch. The system is all contained on a single plasmid encoding Cas9, the gRNA, and repair template and can be introduced into cyanobacteria with triparental mating with a conjugal and helper plasmid. The editing plasmid could be cured from the edited strain by re-streaking without the antibiotic, but unfortunately this took 10 rounds of patching.

Another CRISPR Cas9 system was used in PCC 7942. Cas9 and repair templates were provided on two separate plasmids and they saw successful integration at higher efficiencies than traditional transformations and found that they could use shorter homology arms (400 bp vs 700 bp) (Li, Shen, et al. 2016). They found that Cas9 cleavage seemed to help with segregation, but strains were still not segregated until 3 rounds of passaging on the antibiotic. Cas9 was expressed from a plasmid that appeared to only be transiently present because it was not detected after 9 days. Notably, Cas9, and tracrRNA, crRNA were all expressed from native *Streptococcus pyogenes* promoters. This CRISPR Cas9 system was effective, but genomic edits still contained an antibiotic resistance marker, limiting its utility to make some precise edits.

Cpf1, a type V CRISPR nuclease, has also been used for genome editing in cyanobacteria. The Cpf1 system differs from type II Cas9 CRISPR system in that Cpf1 processes pre-crRNA (Fonfara et al. 2016) and introduces a staggered cut 17 bases from the PAM, and the PAM (TTN) is located 5' of the crRNA (Zetsche et al. 2015). Cpf1 appeared less toxic than Cas9 because more colonies were recovered even when compared to Cas9 that lacked a promoter. Successful editing events were shown in UTEX 2973, PCC 6803, and PCC 7120 using a RSF1010 based plasmid containing Cpf1, gRNA, and repair template (Ungerer et al. 2016). Only $\sim$20% of the initial colonies were edited and segregated, but subsequent patching on the antibiotic increased the recovery of segregated mutants. They showed markerless point mutations, knock-outs, and insertions. Continued advancements to improve the efficiency and minimize the time needed to make genome edits will help accelerate the development of gene regulatory tools.

Use of plasmids

In other model organisms, plasmids are used to express heterologous protein. Their use is much less widespread in cyanobacteria both because of the lack of characterized plasmids that can replicate in cyanobacteria as well as the difficulties introducing plasmids into the cell. The natural competence of cyanobacteria is advantageous for introducing foreign DNA with selectable markers via homologous recombination, but unfortunately is not useful for introducing plasmids. The mechanism of DNA uptake and natural competence in cyanobacteria is relatively unknown, but likely contains core DNA uptake genes as well as hundreds of regulatory genes as in other model organisms used to study natural competence (Blokesch 2016). DNA uptake is thought to involve an extendable (and potentially retractable) type IV pilus that translocates single-stranded DNA that is then decorated with protective proteins and then acted on by recombinases. This process imports only single-stranded DNA, preventing its utility for transformation of plasmids.

Both integrative and replicative plasmids have been used in cyanobacteria, but their use lags significantly behind their use and utility in other organisms. There has been a number of shuttle plasmids built that contain both *E. coli* and cyanobacteria-based origins (Wang, Wang, et al. 2012). *E. coli* origins are more standard but only a limited number of cyanobacterial origins have been used: pDU1 in PCC 7120, pDC1 in *N. punctiforme*, pUH24 and pANS in PCC 7942, pAQ1 in PCC 7002, pBA1 in PCC 6301, and pFDA in *Fremyella diplosiphon* UTEX 481 (Wang, Wang, et al. 2012; Taton, Unglaub, et al. 2014). In most of these cases native origins have been used and re-introduced on a new vector in the same host. However, pDU1 was originally isolated from *Nostoc* PCC 7524 but shuttle vectors containing pDU1 have been shown to replicate in many other strains of cyanobacteria including PCC 7120 (Wolk et al. 1984), *Anabaena* M-131 (Wolk et al. 1984), *Fischerella muscicola* PCC 7414 (Stucken et al. 2012), *Chlorogloeopsis fritschii* PCC 6912 (Stucken et al. 2012), *Chroococcidiopsis sp.* str. 029, 057, and 123143, and *Oscillatoria* MKU 277 (Ravindran et al. 2006). There is a significant lack of understanding of the maintenance, copy number control, and replication mechanisms of cyanobacterial plasmids and if these aspects are

similar or different from plasmids native to other bacterial species.

The most widely used plasmid in cyanobacteria is broad host range plasmid RSF1010 (Wang, Wang, et al. 2012)8. RSF1010 is a broad host range plasmid in the IncQ compatibility group that can replicate in over 30 Gram-negative species (Frey et al. 1989). This plasmid on its own is not transferable via conjugation, but it contains the necessary elements (*oriT* and *mob* mobilization genes) to be transferred in combination with a 'helper plasmid.' Conjugation protocols are standard in many strains of cyanobacteria and typically consist of biparental or triparental matings (Huang, Camsund, et al. 2010; Brahamsha 1996; Taton, Lis, et al. 2012). Plasmid based expression has significant advantages, but there are not many widely used or characterized plasmids that replicate in many cyanobacterial hosts.

Standardization of genetic tools

One of the reasons gene expression regulatory tools have only been built in tested in a few strains is because of the lack of standardization of cyanobacteria genetic tools. There have been significant efforts to standardize genetic tools in synthetic biology. The BioBrick 'Registry of Standard Biological Parts' contains thousands of promoters, ribosome bindings sites, terminators, coding sequences, and origins that are surrounded by standardized restriction enzyme sites for simple assembly of parts. There have been a few cyanobacterial plasmids built using the BioBrick system, but they have not gained much popularity. The plasmid, pPMQAK1, was built by adding the BioBrick multiple cloning site and antibiotic resistant cassettes to the backbone of the broad-host range plasmid RSF1010. The plasmid was shown to replicate in *E. coli*, PCC 6803, PCC 7120, and *N. punctiforme* (Huang, Camsund, et al. 2010).

A broad-host-range vector system based off of RSF1010 was built with and used to create cyanobacterial shuttle vectors (Taton, Unglaub, et al. 2014). This system utilizes GC-adaptors flanked by restriction enzyme sites, so that elements can be digested out of donor plasmids, purified, and assembled using Gibson assembly. Four different sites with overlapping regions were used that

could be used for an *E. coli* origin, cyanobacterial origin, antibiotic resistance marker, and expression cassette for the gene of interest. These elements can also be replaced with homology regions to direct homologous recombination onto the chromosome. This design scheme and sequence elements have been integrated into a web service, CYANO-VECTOR (http://golden.ucsd.edu/CyanoVECTOR/) that can be used to design and export GenBank files.

The lack of standardization of parts for gene expression in cyanobacteria likely stems from the widespread practice of integrating cassettes onto the chromosome of cyanobacteria. With the ease of Gibson assembly (Gibson et al. 2009), most labs design and assemble custom plasmids. Genomic integration is a popular method of expressing genes of interest because transformation of plasmids to date has been achieved through conjugation with helper plasmids which takes approximately the same amount of time and reagents as using natural competence and homologous recombination. An effort to make gene regulatory tools accessible across strains and labs would greatly enhance current tool building and engineering efforts.

Unexplored Physiology

The development of predictable gene regulatory tools in cyanobacteria may be limited by unexplored physiology. There are several cyanobacteria-specific elements that remain relatively unexplained that could be influencing gene expression and could explain why directly moving elements from other organisms has been somewhat unsuccessful. We highlight three aspects of cellular physiology that may be playing significant roles in influencing gene expression.

Highly iterative palindrome 1 (HIP1)

The extreme prevalence of one sequence in cyanobacterial genomes remains unexplained. The palindrome 5'-GCGATCGC-3' or the extended version, 5'-GGCGATCGCC-3', is extremely overrepresented in many cyanobacterial genomes and has been named highly iterative palindrome

1 (HIP1). This sequence was first described in PCC 6301 in 1993 as the two flanking sequences surrounding a deletion event that conferred a tolerance to cadmium (Gupta et al. 1993). Since then the overrepresentation of this sequence has been discovered in many cyanobacterial strains and has been shown to have been gained once during evolution and then lost in marine cyanobacteria (Delaye et al. 2011).

HIP1 has been shown to propagate through nucleotide substitutions. Comparisons of several conserved regions between *E. coli* and PCC 7942 show that *E. coli* contains partial HIP1 sequences (on average 5.2 of the 8 nt) that require minimal changes to the HIP1 sequence that result in synonymous or conservative amino acid changes (Robinson, Cranenburgh, et al. 1997). Deletions flanked by HIP1 sequences have been observed many times. Direct tests comparing deletion frequencies in *E. coli* and PCC 7942 show that excision frequencies surrounded by HIP1 sequences are 2-fold higher than those surrounded by control palindromic sequences. Additionally, the frequency of excision was similar in both *E. coli* and PCC 7942, suggesting that this phenomenon is not due to a cyanobacterial-specific recombinase (Robinson, Cranenburgh, et al. 1997). It has been suggested that these deletion events could be due to HIP1 sequences forming hairpins or other structures that result in replication slippage and illegitimate replication. A similar hypothesis has been suggested for repeated sequences found in radiation resistant *Deinococcus radiodurans*, that is thought to survive double-stranded breaks through extensive recombination of its many chromosomes (Minton 1994).

It's possible that HIP1 could be involved in gene expression control, but potential mechanisms and regulatory partners remain elusive. Comparative genomics of many cyanobacterial genomes show that it is not preferentially located in coding or noncoding regions and has no preference for reading frame. Additionally, attempts to see trends between HIP1 prevalence and location within a transcript, GC content, codon usage, and transcript abundance have been unsuccessful (Xu 2015). So far, no HIP1-specific proteins or complexes have been identified via electrophoretic mobility-shift assays (Robinson, Cranenburgh, et al. 1997). We urge the cyanobacterial community

to be cautious of DNA motif finders such as MEME (Bailey et al. 1994) because they will most likely find HIP1 enrichment no matter what native cyanobacterial sequences are entered. The overabundance of the HIP1 sequence remains unexplained and could be influencing gene expression or other processes that have large implications for growth and study of cyanobacteria.

Polyploidy

It has long been reported that many strains of cyanobacteria contain multiple copies of its chromosome, but it's implications on gene expression control is often ignored. PCC 7942 and WH7803 have been shown to be oligoploid containing 3-4 copies of its chromosome per cell irrespective of growth phase (Griese et al. 2011) or light-dark/continuous light cycles (Mori et al. 1996). Other strains such as PCC 6803 have been shown to be highly polyploid with as many as 218 copies of its chromosome per cell in exponential phase. In PCC 6803 the number of copies of its chromosome changes and only $\sim$ 60 copies per cell are present in linear and stationary phase. Previous studies report only 12 copies of the chromosome in PCC 6803 (Labarre et al. 1989). PCC 6803 has the highest ploidy level of any bacterium besides the very large *Epulopiscum* sp. that contains tens of thousands copies of its chromosome (Griese et al. 2011). Discrepancies in chromosomal as well as plasmid copy number may come from lysis efficiency issues as well as the method of detection (PCR vs total DNA quantification). There has also been mention that these chromosome copy number discrepancies could be due to cultivation of these strains in many labs over the past 20 years that could be resulting in mutations.

The ploidy of cyanobacteria could confer resistance to UV radiation; both PCC 6803 and PCC 7942 were more resistant to UV radiation than *E. coli*. Additionally, PCC 6803 (highly polyploid) is much more resistant to X-ray irradiation than PCC 7942 (oligoploid) (Domain et al. 2004). The ploidy of cyanobacteria leaves many unanswered questions regarding gene expression. It is unknown when genes are expressed from all chromosomes or selectively from different chromosome copies. The polyploid nature of cyanobacteria has unknown implications on gene expression.

The presence of multiple copies of the genome is challenging for the introduction of foreign genes and making targeted mutations. It has only been recently appreciated that it is imperative to show segregation of strains, to ensure predictable and reliable expression patterns and production results. In some early papers using cyanobacteria to create products, there is no mention of segregation while in others the authors acknowledge that the strains were indeed not segregated, making it almost impossible to re-create these strains and verify these results. PCR, using primers flanking the inserted gene of interest, is often used to demonstrate segregation of strains, but this is not always sufficient to demonstrate segregation. There have been cases where the PCR product made with flanking primers indicates the proper mutated size, but it is still possible to amplify the native WT sequence using an internal primer (**Chapter 3**). A combination of both internal and external primers is necessary to demonstrate segregation. The presence of multiple copies of the genome in cyanobacteria is challenging, and the cyanobacteria community must remain vigilant to ensure thorough and meticulous strain construction

Role of small proteins

Small proteins often escape annotation by automated pipeline systems despite their important functions, and many of these small proteins have important roles in photosynthesis. These small proteins are sometime known as sproteins, SPs, or μ-proteins with arbitrary cutoffs ranging from $<$100, $<$80, and $<$50 amino acids. These small proteins have been implicated in many processes such as being secreted as signals, serve as toxins, membrane components, chaperones of metals or nucleic acids, regulators, or stabilizing factors of large protein complexes (Hobbs et al. 2011).

Issues with small protein annotation likely arise because short DNA sequences are likely to contain both start and stop codons. Random DNA sequences have been modeled in yeast, and the frequencies that different length sequences would contain ORFs showed that a large number of short ORFs could be occurring by chance (Fickett 1995). Due to this, most gene annotation algorithms have set a cutoff for ORFs of greater than 100 amino acids to lower then number of false positive

annotations (Su et al. 2013).

Only 80.8% of the entire PCC 6803 predicted proteome has been detected even when the data was combined from over 50 independent proteomic studies (Gao, Wang, et al. 2015). Hydrophobicity as well as low or condition-specific expression (especially proteins encoded on plasmids) may cause the inability to detect expression (Gao, Wang, et al. 2015), but a major class of proteins that is not-represented is small proteins. Proteomic approaches fail to identify many small proteins for several reasons. Small proteins contain few or no trypsin cleavage sites, necessary for many mass spectrometry analyses. Additionally, many small proteins may be lost in the filtering process where two positively charged peptides are necessary to match to the database. Modifications such as phosphorylation, acetylation, glycosylation, and oxidation may prevent detection (Gao, Wang, et al. 2015) (all of which have been observed in PCC 6803).

Many short proteins have been shown to play important roles in photosynthesis including 19 proteins <50 amino acids that interact with photosystem II (*psbM*, *psbT*, *psbI*, *psbL*, *psb30*, *psbJ*, *psbM*, *psbX*, *psbY*, *psbN*, *psbF*, and *psbK* (Kashino et al. 2002)). Other small proteins are involved in photosystem I (*psaM*, *psaJ*, and *psaI*) as well as electron transport (*petL*, *petN*, *petM*, and *petG*). The smallest protein currently annotated in cyanobacteria is *petN* (29 amino acids) (Hobbs et al. 2011). The presence and expression of many of these small genes was completely unknown and only found through purification of photosystem II complexes. Purification of PSII from PCC 6803 revealed a total of 31 distinct proteins, 12 of which were less than 8 kDa, and 5 of which were novel (Kashino et al. 2002). Six of these small proteins remained unidentified, highlighting the challenges of both predicting and studying these small proteins. The important role these small proteins play is highlighted by the fact that many are conserved in cyanobacteria, algae and plants.

There have been attempts to predict the presence and expression of small proteins in cyanobacteria by examining transcriptomic data and comparing the conservation of possible coding sequences across different species and organisms. 146 small proteins were predicted to be shared in both PCC 6803 and PCC 6714 and a smaller subset of 42 proteins shared between both cyanobacteria

strains and *Arabidopsis thaliana*, and 29 proteins shared between the cyanobacteria strains and *E. coli* (Baumgartner et al. 2016). Expression of a test set of 5 of these genes were confirmed with immunoblotting. Two of these small proteins were originally thought to be non-coding RNAs (Baumgartner et al. 2016). Challenges to discovering the expression of these small proteins include the time it takes to make markerless mutations such as a C-terminal tag as well as their regulated expression under specific environmental conditions. There have been a number of short proteins that have been found to play important roles in photosynthesis, and we postulate that other unknown and uncharacterized sequences may play large roles in cellular physiology.

Conclusions

There have been significant advances in gene regulatory tools in cyanobacteria, but lack of knowledge of basic cellular processes as well as cyanobacteria-specific physiology may still play a large role in controlling heterologous gene expression. Endogenous cyanobacteria promoters have been used to drive gene expression as well as heterologous promoters. Adapting native sequences for expression of heterologous transcripts can cause problems with genetic instability. Many groups have tried using promoters widely-used in *E. coli*, but the majority of them are leaky and have low fold induction values, especially IPTG-based systems. Induction systems have been used both on the level of transcription and translation with varying results. It is difficult to compare induction data and measures of tightness of repression in different publications, and currently there is no method to compare overall expression levels between publications and strains. The CRISPR system has expanded the collection of cyanobacteria gene regulatory tools both in use as a repression system (CRISPRi) and for genome editing.

Gene expression tools often do not operate as expected in cyanobacteria, and we explore several relatively under-studied areas relating to cyanobacteria physiology that could be causing these inconsistencies. It's been observed that most of the cyanobacterial genome is transcribed, but implications for gene expression and gene expression tools remains unknown. Additionally, there

has been extremely little study of cyanobacterial transcription termination. There seem to be intriguing similarities in the transcriptional landscape between cyanobacteria and plant chloroplasts, and continued study in this area might be quite fruitful. We conclude with three topics that are cyanobacteria-specific that may be influencing gene expression: the unknown role of the highly prevalent repeat sequence (HIP1), the maintenance of many copies of chromosome, and the role and prevalence of small proteins.

2 Chapter 2: Identifying RNase III targets in *Escherichia coli*

Authors and Contributors:

Gina Gordon: manipulated and analyzed RNA-sequencing data to identify RNase III cleavage sites. Conducted in vitro RNA cleavage assays and pinpointed cleavage sites. Wrote the manuscript and made figures.

Dr. Jeffrey Cameron: conceived experiment and collected samples for RNA-sequencing. Assisted in developing criteria to identify cleavage sites. Purified RNase III, and conducted crude cell lysate pyruvate dehydrogenase activity assays. Edited the manuscript.

Dr. Brian Pfleger: Provided guidance, edited, and reviewed the manuscript.

Adapted from: Gordon, G.C., Cameron, J.C., and Pfleger, B.F. (2017). RNA Sequencing Identifies New RNase III Cleavage Sites in *Escherichia coli* and Reveals Increased Regulation of mRNA. *mBio* 8:e00128-17. doi: 10.1128/mBio.00128-17.

2.1 Introduction

The concerted action and overlapping functions of endoribonucleases, exoribonucleases, and RNA processing enzymes complicates the study of global RNA turnover and recycling of specific transcripts. More information about ribonuclease specificity and activity is needed to make predictions of transcript half-life and to design synthetic transcripts with optimal stability. RNase III does not have a conserved target sequence but instead recognizes RNA secondary structure. Prior to this study, only a few RNase III target sites in *E. coli* were known, so we used RNA-sequencing to provide a more comprehensive list of cleavage sites and to examine the impact of RNase III on transcript degradation. With this added information on how RNase III participates in transcript regulation and recycling, a more complete picture of RNA turnover can be developed for *E. coli*. Similar approaches could be used to augment our understanding of RNA turnover in other bacteria.

Ribonucleases (RNases) are best known for their role in processing stable RNA (e.g. rRNA, tRNA) to their mature, functional forms. In addition, RNases are critical for the inactivation and recycling of protein-encoding mRNA transcripts. In *Escherichia coli* and *Bacillus subtilis*, the best-studied models of bacterial RNA turnover, endonucleases and exonucleases act in concert, often facilitated by formation of a multi-enzyme complex, termed a degradosome, to inactivate and recycle transcripts to single nucleotides (Py, Causton, et al. 1994; Commichau et al. 2009). A common, generic model of mRNA turnover begins with a rate-determining, inactivating cleavage by an endoribonuclease followed by exoribonuclease-mediated turnover of secondary transcript fragments (Hui et al. 2014). That said, specific mechanisms of mRNA decay are known for a limited number of transcripts and predictions of how an unstudied transcript will decay are not available. This knowledge gap is due to the experimental complexity created by the presence of redundant decay pathways, the interaction of multiple enzyme classes, and the strong influence of environmental factors on RNase activity in each case. Furthermore, RNase activity can be blocked by specific protein inhibitors (Lee, Ahn, et al. 2003; Kim et al. 2008), impaired by the presence of RNA secondary structures, or accelerated by processing of other RNases. In addition, the

substrate preference (i.e. sequence, structure) of endonucleases is poorly understood. RNases can demonstrate highly selective cleavage specificity in vivo but often significant nonspecific activity in vitro (Nicholson 1999). Our imprecise understanding of factors that influence RNase activity prevent prediction of preferred cleavage sites, transcript abundance, and RNA half-life from genomic sequence. This study begins to address this gap by cataloging the native RNA targets of RNase III in *E. coli*.

RNase III is a widely distributed endoribonuclease that plays many roles in processing bacterial and viral RNAs, yet a complete catalog of its targets and regulatory effects remains unavailable. RNase III acts as a homodimer with two Mg^{2+}-dependent active sites that cleave double-stranded RNA leaving a 5' phosphate and 3' hydroxyl with a 2 nucleotide overhang (Court 1993). RNase III is best known for its role in processing ribosomal RNA to release mature 16S and 23S rRNAs (Schlessinger et al. 1974; King et al. 1984; Young et al. 1978; Court et al. 2013). RNase III is essential in *B. subtilis* due to its role in cleaving the duplex of two RNAs encoding toxins on two prophages (Durand et al. 2012). In *E. coli*, RNase III (*rnc*) regulates its own expression by binding and cleaving a secondary structure in the 5' UTR causing rapid degradation of the *rnc* message (Matsunaga, Dyer, et al. 1996). The negative autoregulation is possible because RNase III accounts for only 0.01% of the total protein (Bardwell et al. 1989), and can be effectively titrated away from its own message by abundant targets such as rRNA. Another well-known example of RNase III regulation of mRNA is *pnp* which encodes PNPase, a 5'-3' exoribonuclease. RNase III cleaves a secondary structure in the 5'UTR leading to rapid degradation of the *pnp* message (Régnier and Portier 1986). RNase III can also affect protein expression through a translational control mechanism where the ribosome binding site (RBS) is occluded in RNA secondary structure until RNase III cleavage makes it accessible (e.g. native *E. coli* alcohol dehydrogenase, *adhE*; cIII in lambda phage (Aristarkhov et al. 1996; Altuvia et al. 1987)). There are also many examples of RNase III cleavage that alter RNA maturation and the abundance of RNA duplexes. RNase III targets several locations in bacteriophage lambda (Altuvia et al. 1987; Plunkett et al. 1989; Steege

et al. 1987), T7 phage (Dunn et al. 1973), and T3 phage (Majumder et al. 1977). In addition to maturation of viral RNAs, RNase III also participates in the bacterial immune system (type II CRISPR/Cas systems) by releasing mature guide RNAs from CRISPR encoded arrays (Deltcheva et al. 2011). Cleavage of antisense RNA by RNase III can control plasmid copy number (Blomberg et al. 1990). Eukaryotic RNase III family members, Dicer and Drosha, have been shown to be central players in RNA interference and microRNA processing (Bernstein, Caudy, et al. 2001; Lee, Ahn, et al. 2003). Beyond RNA cleavage, RNase III can alter secondary structure through binding to double-stranded RNA without cleaving (Drider et al. 2005). With such an expansive list of viral and cellular RNA targets, we wondered if other native mRNAs were subject to RNase III cleavage and how cleavage impacted transcript half-life.

In this study, we performed RNA-sequencing on an *E. coli* model system to confirm known sites, identify novel targets, and determine the impact of cleavage events on transcript degradation and metabolic phenotypes. To find cleavage sites, we compared the abundance of sequencing reads across the transcriptome of a wild-type *E. coli* and an *rnc*- deletion mutant. The RNA-sequencing approach provided wider coverage and unprecedented resolution of mRNA abundance at each position in the transcriptome compared to prior studies that used qPCR, Northern blots, or microarrays (Stead et al. 2011). In addition to data collected from exponentially growing cells, we observed the effects of RNase III cleavage on transcript degradation by collecting samples in a time course after stopping nascent transcription with rifampicin. Combined, our datasets allowed us to pinpoint novel RNase III cleavage sites located within transcripts encoding important metabolic enzymes, near alternative promoters, and within transcription attenuators.

2.2 Identifying RNase III Sites using RNA-sequencing

Determination of Global RNase III Targets

To study RNase III processing of the *E. coli* transcriptome, we collected total RNA from two strains of exponentially growing *E. coli*, MG1693 (WT) and SK4455 (*rnc*-). Samples were harvested over a 20-minute timecourse following addition of rifampicin to each culture. Ribosomal RNA was reduced from these samples and the resulting pool was processed for RNA sequencing. The resulting sequencing reads were mapped to the *E. coli* genome and quantified per base using IGV tools (Robinson, Thorvaldsdottir, et al. 2011).

We examined two well-known RNase III processing sites, 5' of *pnp* and *sucA* (**Figure 2.1**) to identify hallmarks of cleavage in our RNA-seq data. Prior to rifampicin addition, WT samples (blue) exhibited drastically reduced read coverage compared to the read coverage of *rnc*- (red) at processing sites. The difference led to dramatic spikes in the ratio of *rnc*-:WT reads near the site of RNase III cleavage (green). Beyond the cleavage site, the timecourse post rifampicin addition demonstrated a degradation pattern of the *pnp* transcript consistent with previous descriptions (Matsunaga, Simons, et al. 1997); RNase III cleavage results in a faster decay of the downstream secondary transcript (**Figure 2.2**). In *rnc*- samples, the downstream transcript is more stable with a half-life of $\sim$6 minutes compared to $\sim$4 minutes in WT. Previous studies using S1 nuclease mapping reported the half-life of the *pnp* transcript was 8 minutes in *rnc*- and 1.5 minutes in WT (Takata et al. 1989); others have reported a half-life of $>$40 minutes in *rnc*- (Portier et al. 1987). The RNA profiles of several other known RNase III targets were similarly changed (**Figure 2.3**).

Based on these observations, we devised a search algorithm for identifying putative RNase III processing sites. Specifically, we plotted the ratio of reads from *rnc*- to reads from WT and searched for sequences where the *rnc*- strain was dramatically more abundant than the WT in short windows – i.e. sites where cleavage would reduce RNA abundance in WT. From this dataset, we created a list of potential cleavage sites that had a ratio of reads between *rnc*- and WT of 30 in both

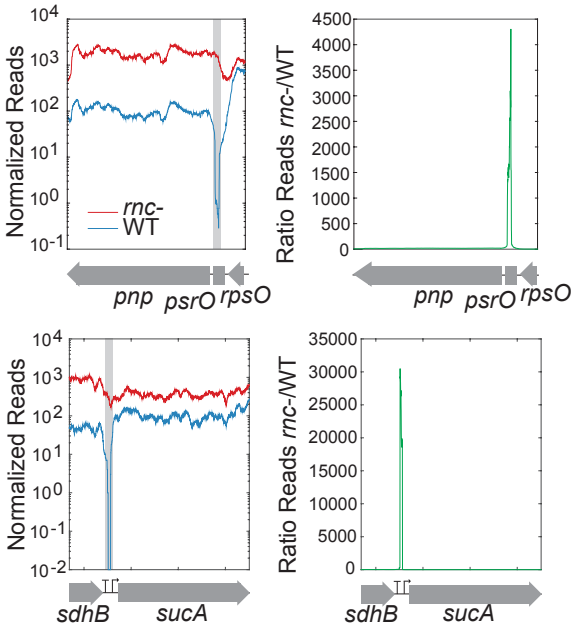


Figure 2.1: Identification of RNase III cleavage sites by comparison of RNA-sequencing coverage between wild type (WT) and RNase III mutant (*rnc*-). Normalized coverage across genomic locations of known cleavage sites near *sucA* and *pnp* are shown as well as the location of the known cleavage site (vertical dashed line). Novel cleavage sites were identified by looking for peaks in the ratio of reads between *rnc*- and WT.

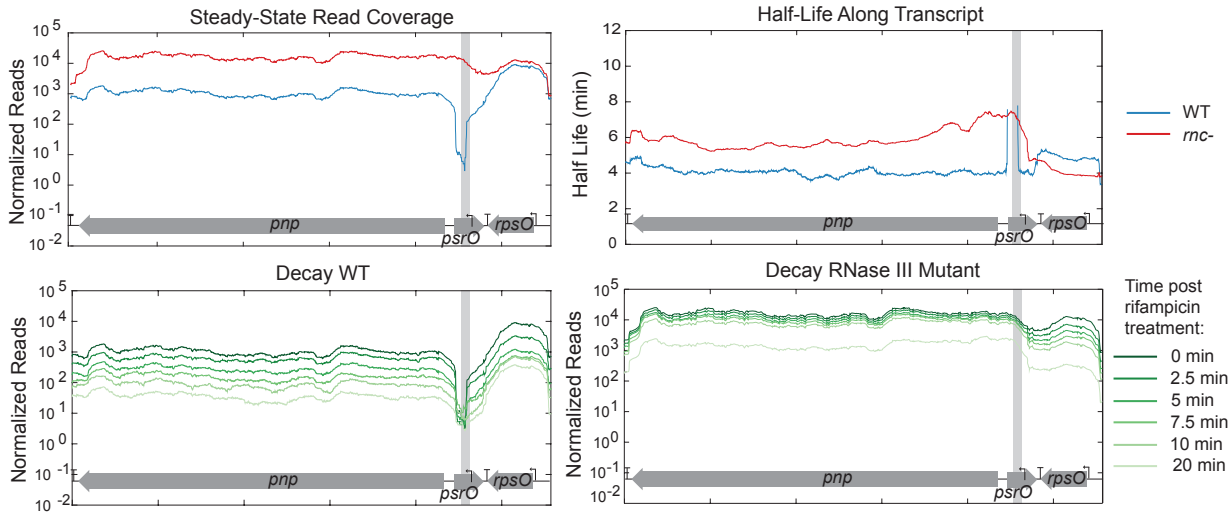


Figure 2.2: RNA-seq coverage around known RNase III cleavage site 5' of *pnp*. Degradation throughout time is shown both for WT and *rnc*- with time 0 minutes shown in dark green and lighter green at 20 minutes. Half-lives in both the WT and *rnc*- were calculated and displayed only when the R^2 value was ≥ 0.7. The cleavage site is identified with a vertical dashed line.

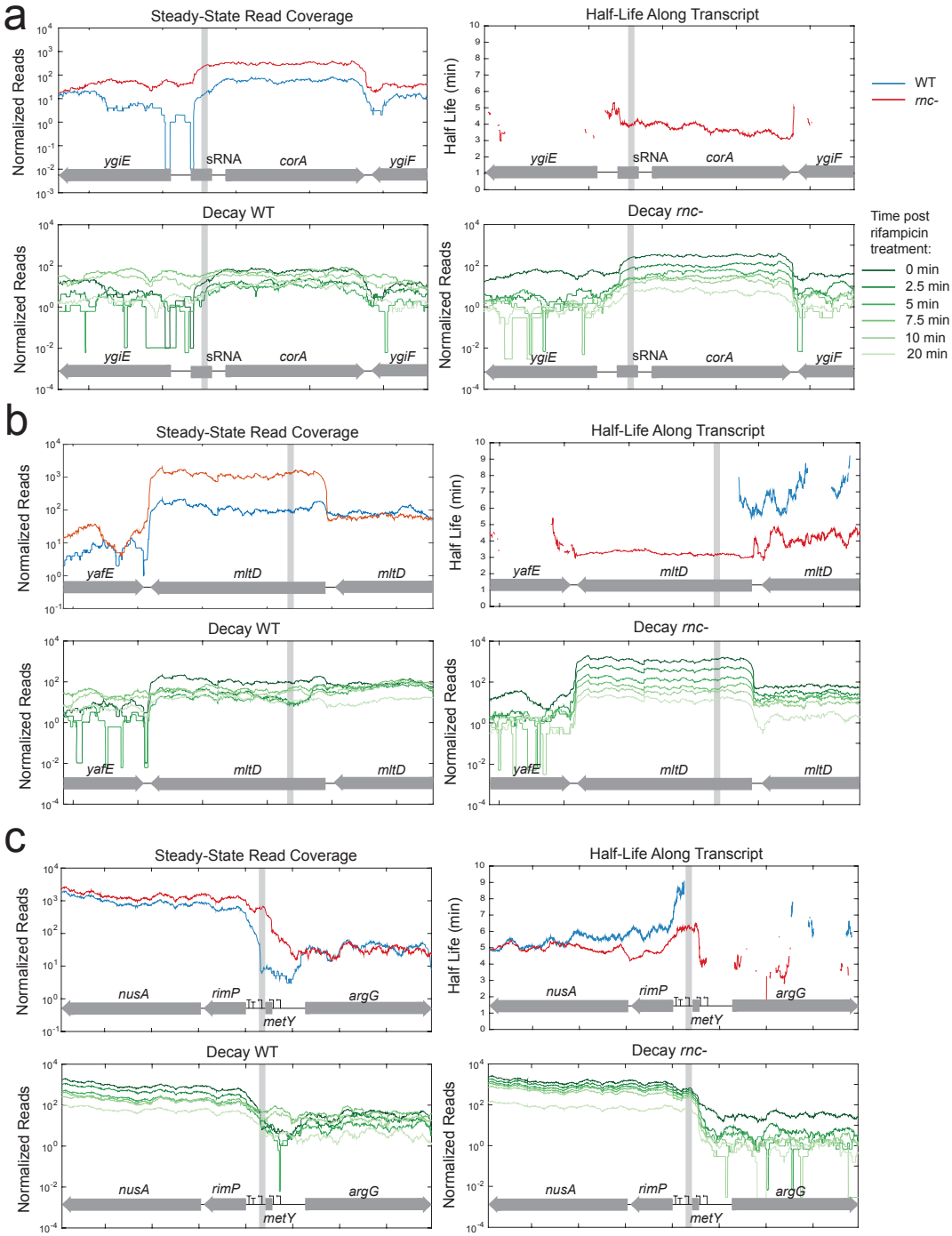
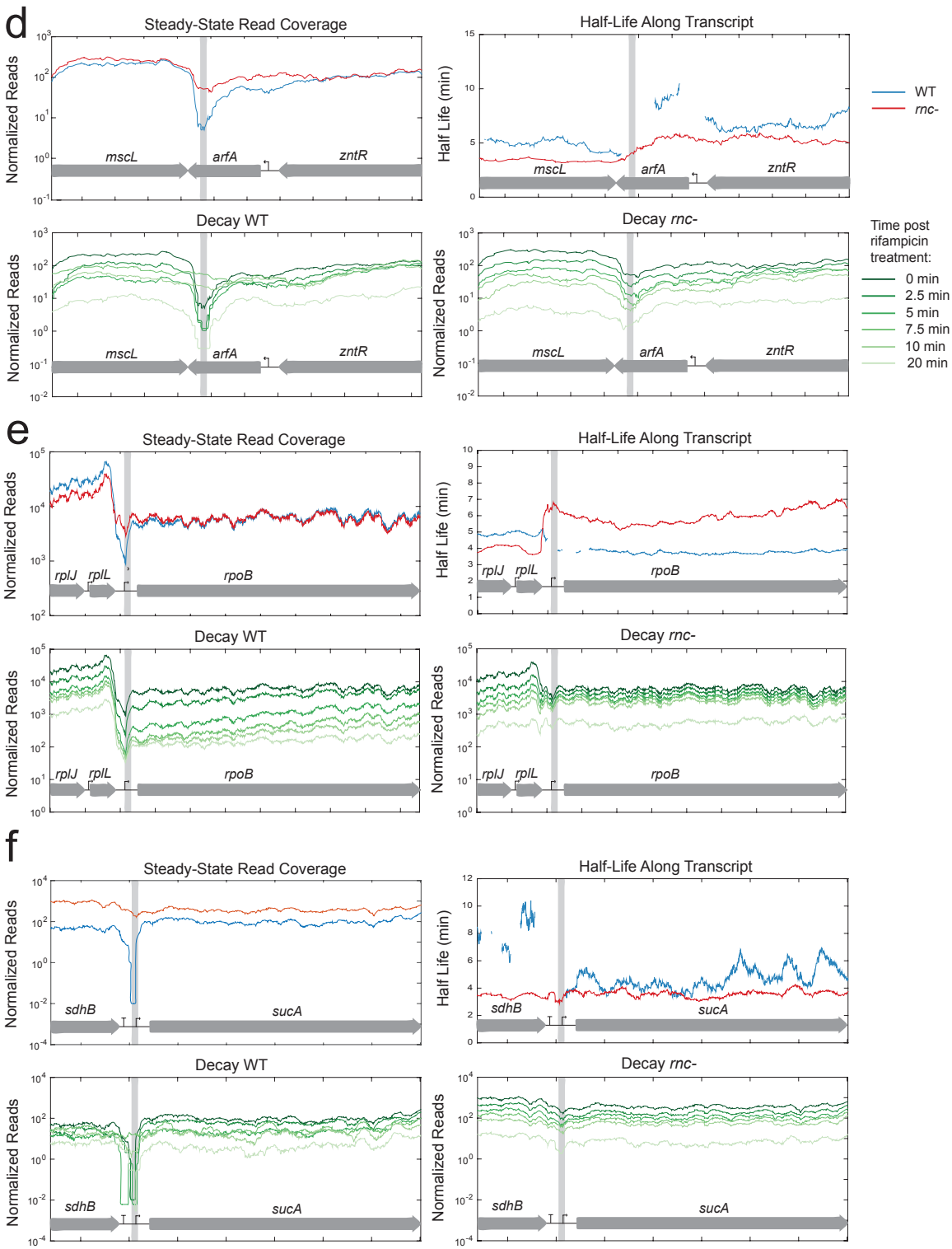
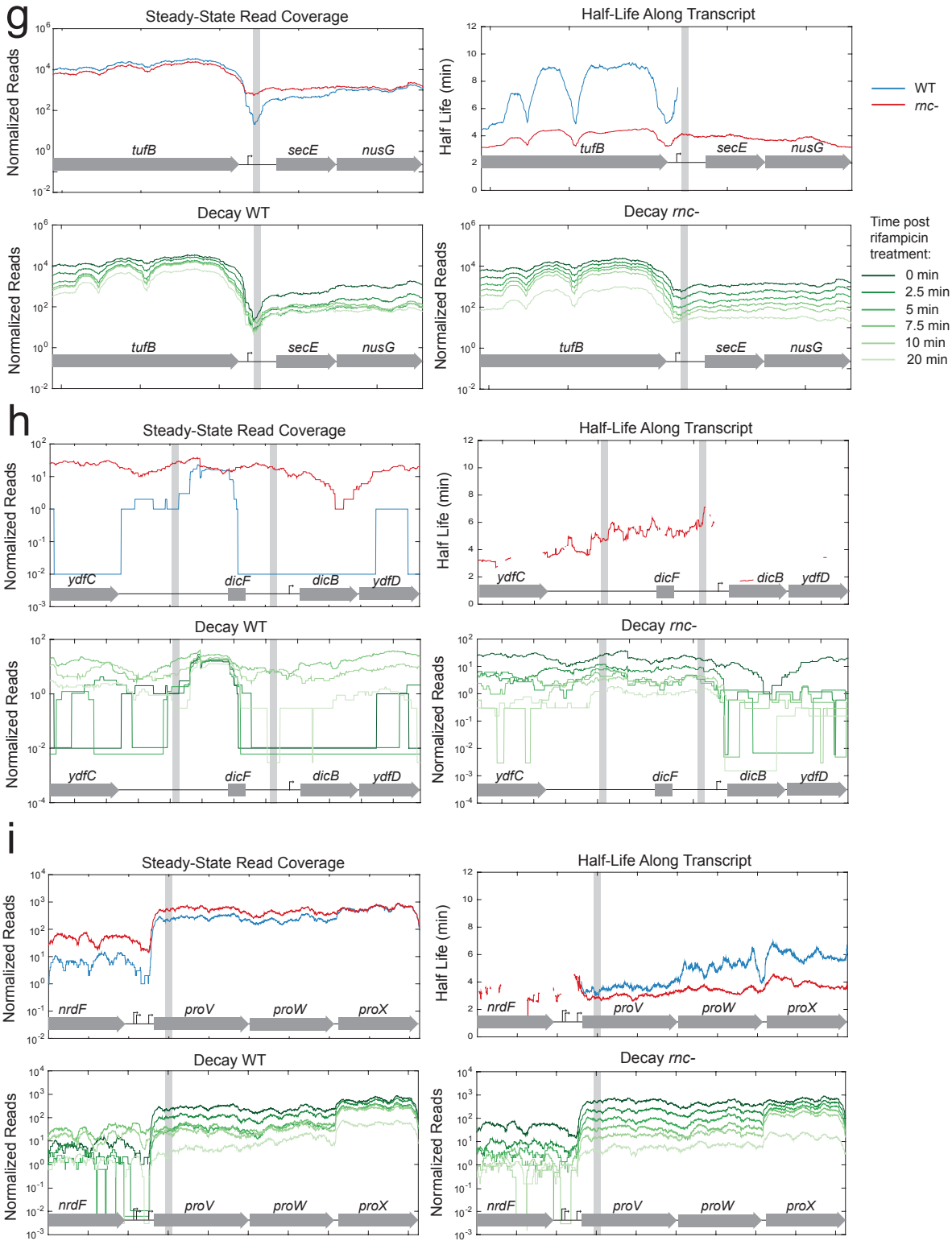
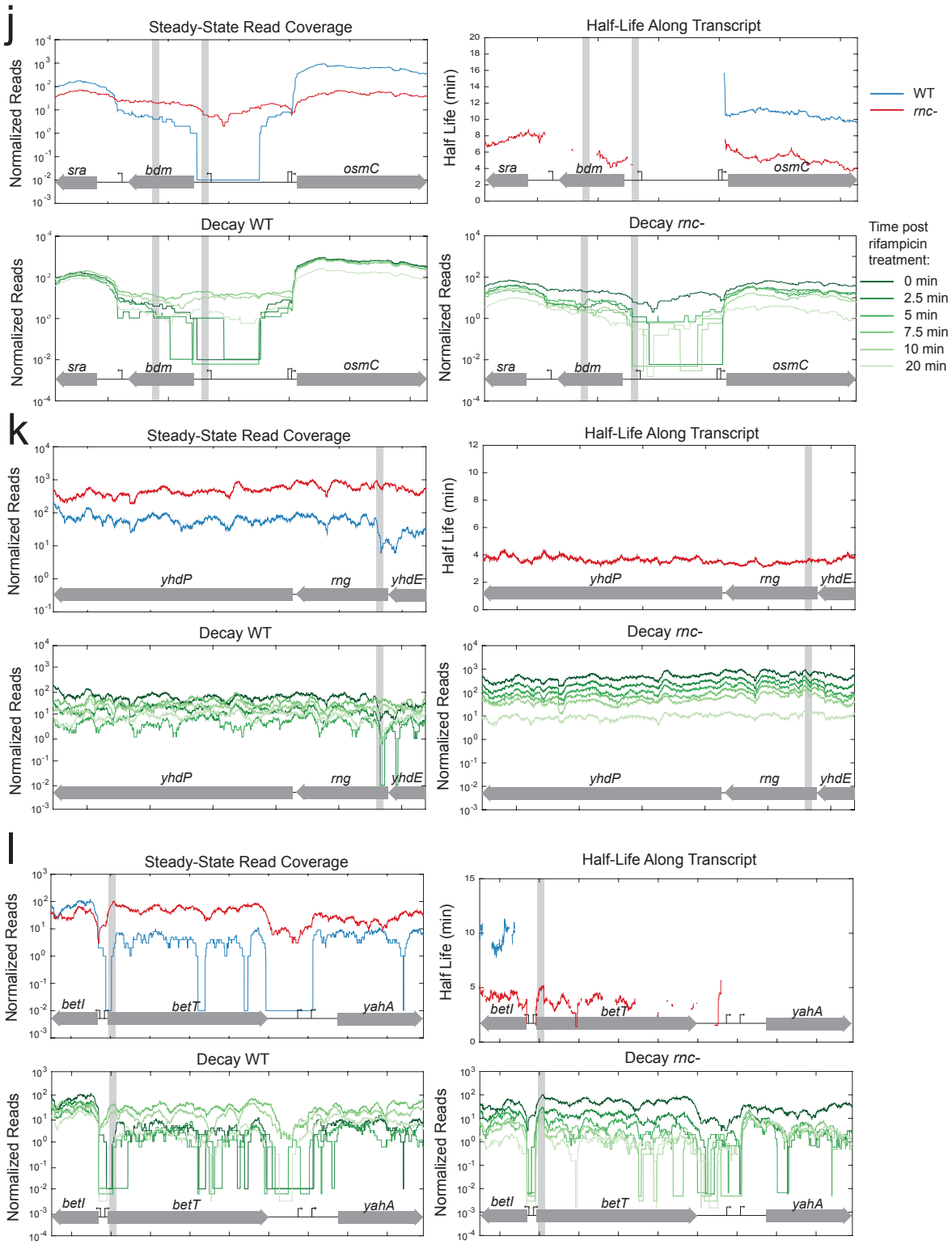
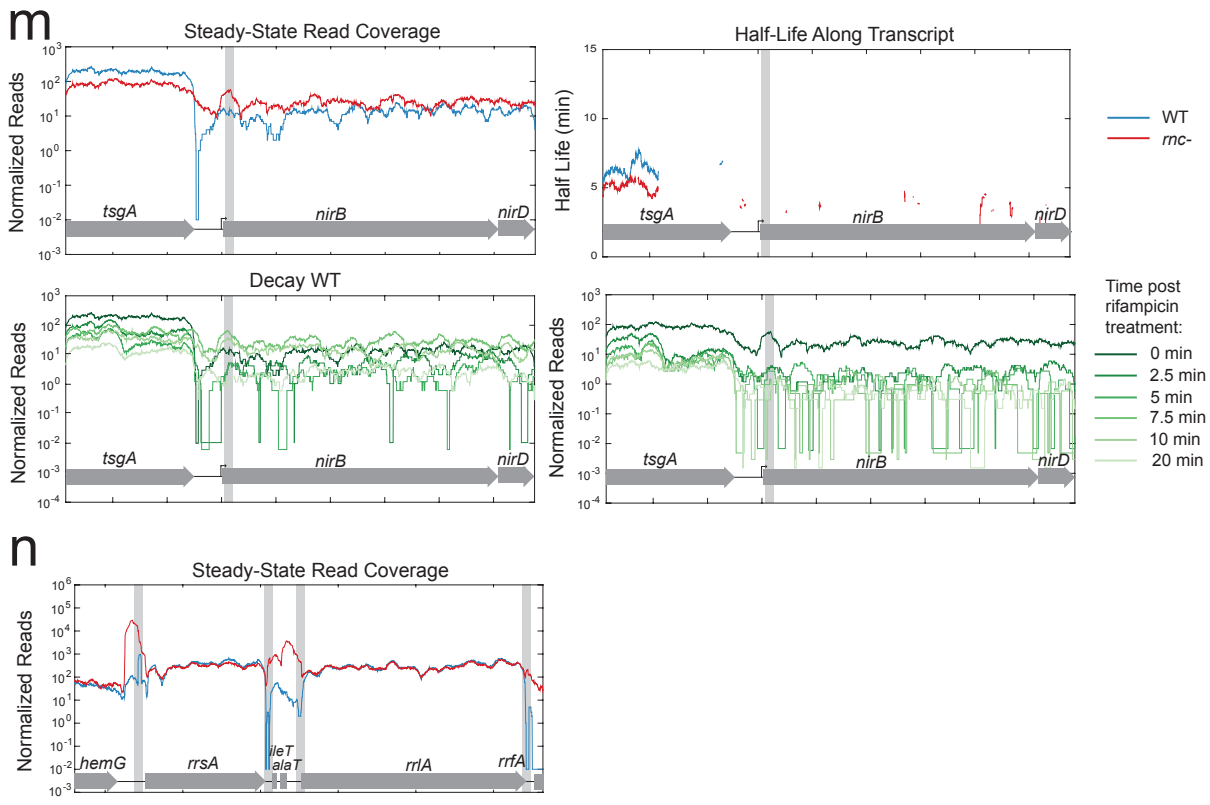


Figure 2.3: Steady-state read coverage and decay throughout time in WT and *rnc-* for known RNase III processing sites (a) *corA* (Lim, Sim, Sim, et al. 2012), (b) *mltD* (Lim, Ahn, et al. 2014), (c) *metY* (Régnier and Grunberg-Manago 1989), (d) *arfA* (Garza-Sánchez et al. 2011), (e) *rplL* (Barry et al. 1980), (f) *sucA* (Cunningham et al. 1998), (g) *tufB* (Downing et al. 1990), (h) *dicF* (Faubladier et al. 1990), (i) *proU* (Kavalchuk et al. 2012), (j) *bdm* (Sim, Yeom, et al. 2010), (k) *rng* (Song et al. 2014), (l) *betT* (Sim, Lim, et al. 2014), (m) *nirB* (Stead et al. 2011) and (n) a rRNA operon (Schlessinger et al. 1974; King et al. 1984; Young et al. 1978). A dashed vertical line marks the known cleavage site.


d
Steady-State Read Coverage
Half-Life Along Transcript
Normalized Reads
Half Life (min)
mscL
arfA
zntR
WT
rnc-
Decay WT
Decay rnc-
Time post rifampicin treatment:
0 min
2.5 min
5 min
7.5 min
10 min
20 min
e
rplJ
rplL
rpoB
f
sdhB
sucA


g
Steady-State Read Coverage
Half-Life Along Transcript
Normalized Reads
Half Life (min)
tufB
secE
nusG
WT
rnc-
Decay WT
Decay rnc-
Time post rifampicin treatment:
0 min
2.5 min
5 min
7.5 min
10 min
20 min
h
ydfC
dicF
dicB
ydfD
i
nrdF
proV
proW
proX


j
Steady-State Read Coverage
Half-Life Along Transcript
Normalized Reads
Half Life (min)
sra
bdm
osmC
WT
rnc-
Decay WT
Decay rnc-
Time post rifampicin treatment:
0 min
2.5 min
5 min
7.5 min
10 min
20 min
k
yhdP
rng
yhdE
l
betI
betT
yahA


m
Steady-State Read Coverage
Normalized Reads
tsgA
nirB
nirD
Half-Life Along Transcript
Half Life (min)
WT
rnc-
Decay WT
Time post rifampicin treatment:
0 min
2.5 min
5 min
7.5 min
10 min
20 min
n
Steady-State Read Coverage
hemG
rrsA
ileT
alaT
rrlA
rrfA

replicates. This ratio was chosen because this was the lowest ratio we observed for a known site, *tufB*. Only one site was listed within a 400 bp window because without this specification there were numerous locations selected that corresponded to shoulders of potential sites. The list comprised the majority of known processing sites as well as novel putative cleavage sites (**Appendix II**). The top 20 novel cleavage sites as well as the genes that are affected by the cleavage event are listed in **Table 2.1**. Sites were widely distributed across the genome falling both within coding regions and in untranslated regions.

Verification of novel RNase III sites with in vitro cleavage assays

We performed in vitro cleavage assays to confirm processing of 19 putative RNase III cleavage sites selected using the following criteria: 1) top ranking based on search algorithm, 2) RNA-seq pattern resembling known cleavage sites, and 3) predicted secondary structure containing double stranded region greater than the minimum RNase III target size of 22nt. For in vitro tests, synthetic RNA templates were designed such that putative cleavage sites were positioned $\sim$150 nucleotides from the 5' end to ensure the cleavage product was clearly visible when separated from uncut RNA on a denaturing 5% polyacrylamide gel. A T7 promoter was included in the 5' end of the primer used to create PCR products for use as templates in in vitro transcription reactions. RNA made from in vitro transcription was quantified and incubated with purified *E. coli* MG1655 RNase III. Reactions were activated by the addition of $MgCl_2$ and incubated at 37°C for 5 minutes before quenching with excess EDTA. In vitro reaction products were separated on a 5% polyacrylamide urea gel. Using this approach, we demonstrated RNase III cleavage of 11 putative sites near *aceEF*, *pflBA*, *pheM*, *ahpFC*, *ndh*, *hisL*, *sdhC*, *dctA*, *tnaA*, *proP*, *yajQ-panE* (**Figure 2.4**). For each putative cleavage site, the band corresponding to the full-length transcript decreased in intensity or disappeared after addition of both enzyme and $MgCl_2$. Analogously, new bands, shorter in length, appeared in each of these samples. Two known cleavage sites (*pnp* and *mltD*) were tested as well as two negative controls not known to be processed by RNase III (gene H from *Enterobacteria* phage

Table 2.1: Known RNase III cleavage sites and top putative RNase III cleavage sites as identified by RNA-seq coverage differences between WT and *rnc*-. Positions with zero reads were given a pseudocount of 0.01 reads. * indicates a multiple cleavage sites were identified within a region and combined into one entry in the table. 'Y' indicates there was observed cleavage of an in vitro transcribed transcript when incubated with purified enzyme while 'N' indicates a lack of observed cleavage.

Known Cleavage Sites Identified in this study								
Location	Annotation	Ratio 1	Ratio 2	WT reads Rep. A	WT reads Rep. B	ΔRnc reads	Genes Affected	Reference
4,166,537	intergenic (*murI-rrsB*)	415	587	41	29	17025	rRNA (*rrsB*)	(Young et al. 1978)
4,035,410	intergenic (*hemG-rrsA*)	314	925	53	18	16647	rRNA (*rrsA*)	(Young et al. 1978)
3,311,286	*psrO*	3782	3782	3	3	11347	*pnp*	(Régnier and Grunberg-Manago 1989)
2,731,278	intergenic (*rrsG-clpB*)	487	534	23	21	11206	rRNA (*rrsG*)	(Young et al. 1978)
223,653	intergenic (*gmhB-rrsH*)	271	491	29	16	7853	rRNA (*rrsH*)	(Young et al. 1978)
3,428,875	intergenic (*rrsD-yrdA*)	340	234	20	29	6791	rRNA newline (*rrsD*)	(Young et al. 1978)
3,943,398	intergenic (*rrsC-gltU*)	1309	2619	4	2	5237	rRNA (*rrsC*)	(Young et al. 1978)
4,209,733	intergenic (*rrsE-gltV*)	4959	2480	1	2	4959	rRNA (*rrsE*)	(Young et al. 1978)
4,168,221	intergenic (*rrsB-gltT*)	123700	123700	0.01	0.01	1237	rRNA (*rrsB*)	(Young et al. 1978)
4,168,637	intergenic (*gltT-rrlB*)	120	419	7	2	837	rRNA (*rrlB*)	(King et al. 1984)
3,427,176	intergenic (*ileU-rrsD*)	77500	775	0.01	1	775	rRNA (*rrsD*)	(Young et al. 1978)
3,397,879	*yhdE*	114	98	6	7	684	*rng*	(Sim, Yeom, et al. 2010)
225,355	intergenic (*rrsH-ileV*)	64300	643	0.01	1	643	rRNA (*rrsH*)	(Young et al. 1978)
4,037,115	intergenic (*rrsA-ileT*)	62600	62600	0.01	0.01	626	rRNA (*rrsA*)	(Young et al. 1978)
3,318,176	intergenic (*rimP-metY*)	102	609	6	1	609	*metY*	(Bardwell et al. 1989)

4,037,523	*rrlA*	149	223	3	2	446	rRNA (*rrlA*)	(King et al. 1984)
758,515	*sucA*	30500	305	0.01	1	305	sucA	60
4,207,874	intergenic (*purH-rrsE*)	29000	290	0.01	1	290	rRNA (*rrsE*)	(Young et al. 1978)
3,946,623	intergenic (*rrlC-rrfC*)	22700	227	0.01	1	227	rRNA (*rrlC*)	(King et al. 1984)
2,729,186	intergenic (*rrlG-gltW*)	55	221	4	1	221	rRNA (*rrlG*)	(King et al. 1984)
4,040,443	intergenic (*rrlA-rrfA*)	20300	203	0.01	1	203	rRNA (*rrlA*)	(King et al. 1984)
4,212,954	intergenic (*rrlE-rrfE*)	19500	65	0.01	3	195	rRNA (*rrlE*)	(King et al. 1984)
3,941,526	intergenic (*yieP-rrsC*)	16200	41	0.01	4	162	rRNA (*rrsC*)	(Young et al. 1978)
4,171,559	intergenic (*rrlB-rrfB*)	12100	12100	0.01	0.01	121	rRNA (*rrlB*)	(King et al. 1984)
3,423,873	intergenic (*rrfD-rrlD*)	11700	117	0.01	1	117	rRNA (*rrlD*)	(King et al. 1984)
2,726,273	intergenic (*rrfG-rrlG*)	11500	58	0.01	2	115	rRNA (*rrlG*)	(King et al. 1984)
Putative Cleavage Sites Identified in this study								
Location	Annotation	Ratio 1	Ratio 2	WT reads Rep. A	WT reads Rep. B	ΔRnc reads	Genes Affected	Results of cleavage assay
122,961	intergenic (*pdhR-aceE*)	6977	1550	2	9	13954	*aceE*, *aceF*, *lpdA*	Y
639,695	intergenic (*ahpC-ahpF*)	900	900	4	4	3598	*ahpC*	Y
1,799,278	intergenic (*pheM-rplT*)	639	799	5	4	3196	*pheM*, *pheS*, *pheT*	Y
4,330,944	*proP*	66	83	15	12	992	*proP*	Y
3,889,207	*tnaA**	301	129	3	7	903	*tnaC*, *tnaA*, *tnaB*	Y
755,579	*sdhD**	545	42	1	13	545	*sdhC*, *sdhD*	Y
3,683,389	*dctA*	74	74	7	7	518	*dctA*	Y
2,171,807	intergenic (*gatR-gatD*)	213	47	2	9	426	*gatR*, *gatD*	N
3,609,652	*yhhQ*	58	77	4	3	231	*yhhQ*	N
951,147	intergenic (*pflA-pflB*)	15800	15800	0.01	0.01	158	*pflB*	Y
2,351,931	*glpT*	14300	143	0.01	1	143	*glpT*	N
2,237,178	*mglC**	133	13300	1	0.01	133	*mglC*, *mglA*	N

4,307,817	*alsE**	126	63	1	2	126	*alsB*, *alsA*, *alsE*, *alsC*, *alsK*	N
2,090,031	*hisL*	11900	60	0.01	2	119	*hisL*	Y
1,166,043	intergenic (*ndh*)	10800	10800	0.01	0.01	108	*ndh*	Y
159,578	*gluQ**	10400	52	0.01	2	104	*gluQ*	N
444,573	intergenic (*panE-yajQ*)	48	32	2	3	96	*yajQ*, *panE*	Y
1,622,816	intergenic (*yneM-mgrR*)	9100	91	0.01	1	91	*yneM*, *mgrR*	N

ϕX174 and *ompA* a highly structured RNA in *E. coli*). Bands corresponding to full length RNA for both negative controls remained intact after 5 min of incubation with purified RNase III. In 8 out of the 19 putative sites, bands corresponding to full-length transcript remained after five minutes (**Figure 2.5**) suggesting that these sequences are not processed by RNase III in vitro. It is possible that these sites may be cleaved in vivo through interactions with other RNAs or altered secondary structure. Using Mfold (Zuker 2003), we were able to identify predicted long double-stranded regions for all eleven confirmed cleavage sites. We confirmed these locations by performing 5' RACE on RNA extracted from samples at the end of in vitro cleavage reactions (**Figure 2.6**). Three sequences, near *pheM*, *ndh*, and *panE*, contained multiple cleavage sites that were individually identified by sequencing 5' RACE fragments subcloned into pGEM. Several cleavage sites were found adjacent to or within other regulatory elements such as promoters (*aceEF* and *pflBA*) and attenuators (*hisL* and *pheM*). We noticed that three of the verified sites were close to leader peptides (*hisL*, *pheM*, and *tnaC*) and wondered if RNase III was involved in the turnover of other leader peptides. We performed RNase III cleavage assays on ten other leader peptides, but saw that none were processed by RNase III in vitro (**Figure 2.5**).

RNase III cleavage impacts mRNA transcript degradation rates

The expansion of known RNase III cleavage sites allowed us to examine the effect of RNase III cleavage on transcript decay. We found that in 10 out of the 18 putative sites highlighted in **Table 2.1**, RNase III cleavage events led to destabilization of WT transcripts compared to *rnc*-.

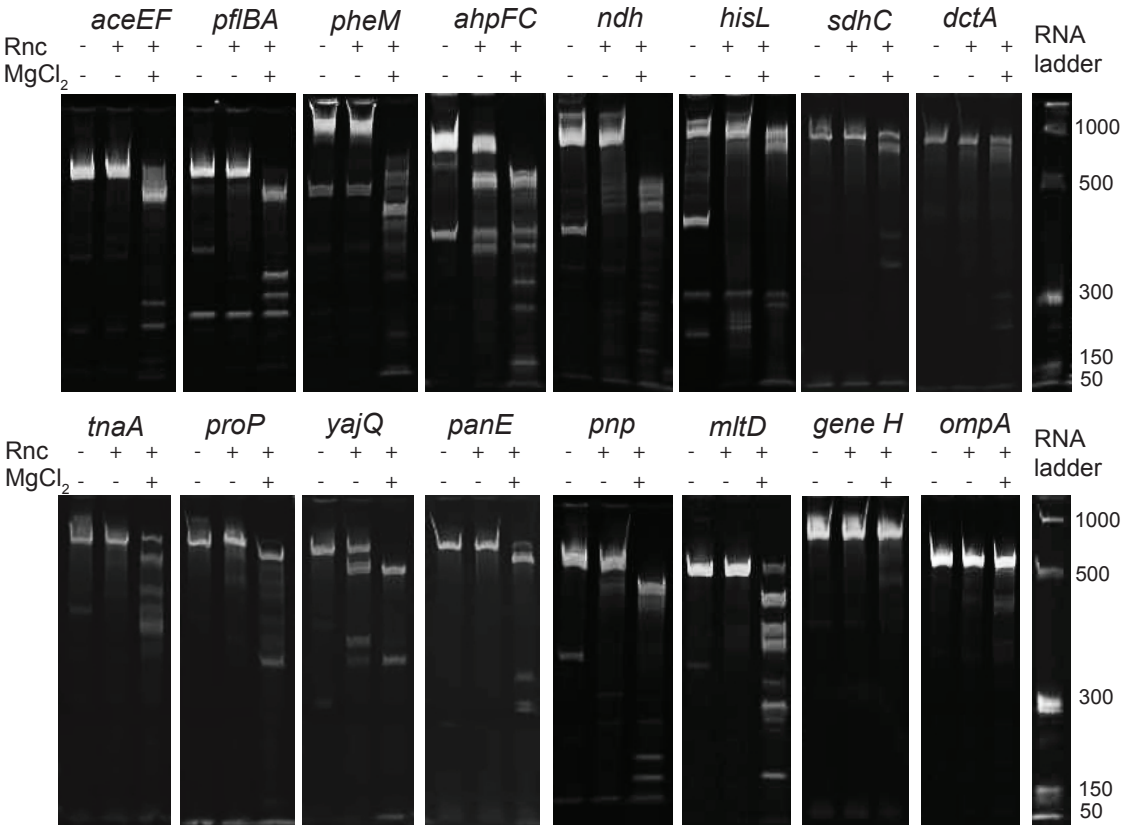


Figure 2.4: RNase III cleavage of in vitro transcripts. Five micrograms of in vitro transcribed RNA was incubated with purified RNase III and activated with $MgCl_2$ at 37°C for 15 min. Five hundred nanograms of RNA was removed before the addition of enzyme and $MgCl_2$. Samples were run on a denaturing polyacrylamide gel and imaged with ethidium bromide.

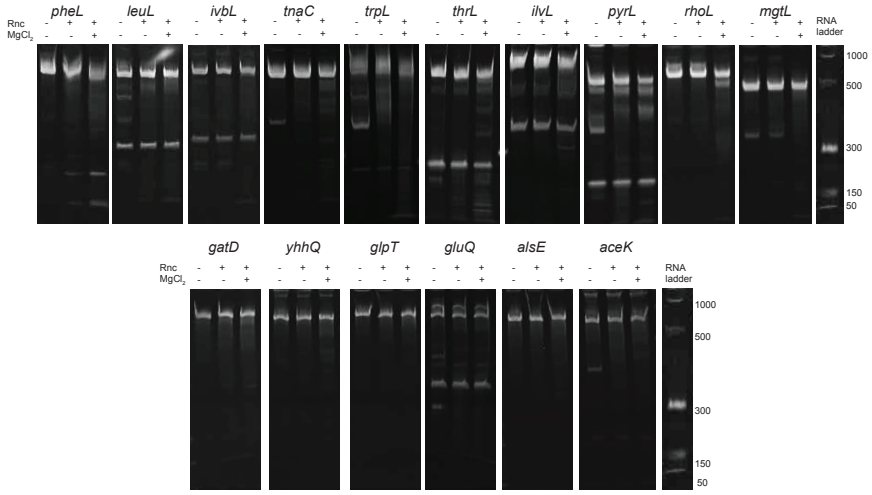


Figure 2.5: Lack of RNase III cleavage of in vitro transcripts. The top panel shows other leader sequences that were examined because we found RNase III processing of other leader transcripts. The bottom panel shows putative targets identified using the RNA-seq data. 5µg of in vitro transcribed RNA was incubated with purified RNase III and activated with $MgCl_2$ at 37°C for 15 minutes. 500 ng of RNA was removed before the addition of enzyme and $MgCl_2$. Samples were run on a denaturing polyacrylamide gel and stained with ethidium bromide and imaged.

For example, the half-life of the WT *aceEF* transcript was ∼3 minutes and the half-life of this transcript in *rnc-* was ∼5 minutes (**Figure 2.7**). This is the same pattern seen with known RNase III cleavage sites such as *pnp* (Matsunaga, Dyer, et al. 1996), where half-lives were shorter in WT than in *rnc-*. Other RNase III cleavage sites following this pattern included those near *tnaC*, *dctA*, *pheM*, *sdhC*, *yhhQ*, *glpT*, *aceK*, *gluQ*, and *proP* (**Figure 2.8A-I**). One cleavage site we identified, *proP*, has been previously described, but only under conditions where *proP* was transcribed from the osmoregulated P1 promoter (Lim, Sim, Lee, et al. 2015). Here, we identified this cleavage site in exponentially growing cells, highlighting that this cleavage event also occurs without the presence of osmotic stress. The steady-state *proP* transcript levels in this study agreed with what was previously described, showing an increase in abundance in *rnc-* compared to WT. Our data also supports their conclusion that cleavage of the *proP* transcript by RNase III leads to degradation and a faster half-life compared to the uncut transcript.

In contrast, we also identified examples where RNase III processing had no impact on transcript

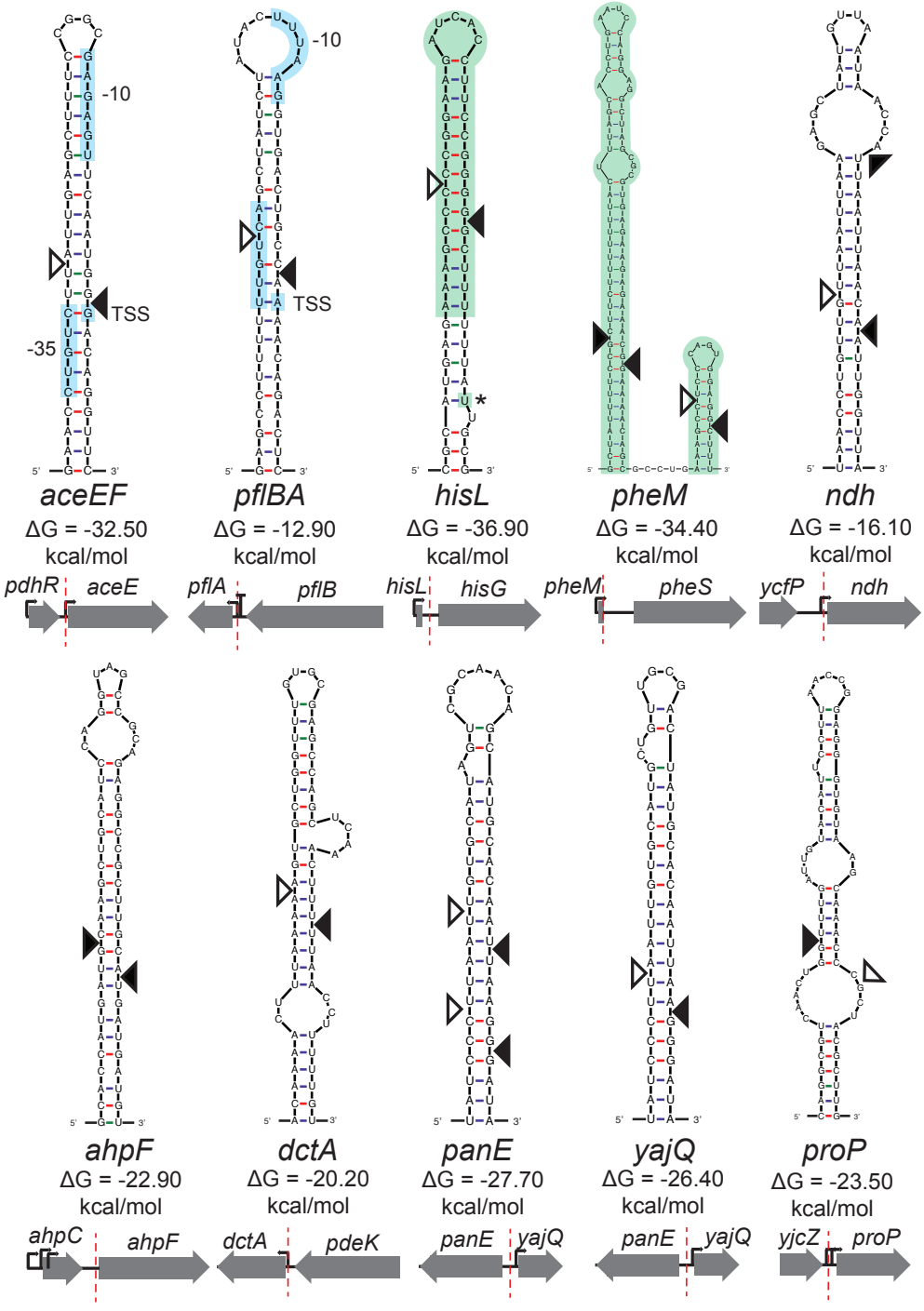


Figure 2.6: Predicted secondary structure of novel RNase III sites. Dark triangles indicate the cleavage site identified by 5' RACE. Open triangles indicate the putative location of the corresponding cut required to leave 2-nt overhangs, characteristic of RNase III cleavage. Colored sections highlight other known regulatory elements such as promoters and structures involved in transcriptional attenuation. The orientation of surrounding genes and the location of an RNase III cleavage site are depicted below the structure.

stability (e.g. upstream of *ndh*, **Figure 2.9**) or led to an increased half-life of the resulting secondary transcript. For example, we observed that RNase III cleavage at a site between *pflB* and *pflA* led to stabilization of the *pflB* transcript in WT compared to *rnc*- (∼5.5 minutes for WT and ∼3.5 minutes for *rnc*-). The cleavage site is located very close to the promoter region of *pflA* (**Figure 2.10A**). Transcription likely originated upstream, presumably from the promoter in front of *pflB* and proceeded through a strong rho-independent terminator (Sawers et al. 1989). The cleavage event at the 3' end of the *pflB* transcript may serve to ensure the separation of the RNAs encoding *pflB* and *pflA* and mimic the transcript formed from the promoter preceding *pflA*. Similarly, cleavage at a site between *ahpC* and *ahpF* stabilizes the *ahpC* transcript, ∼14 minute half-life in WT and ∼6 minutes in *rnc*- (**Figure 2.10B**), while not changing the stability of *ahpF*. Additionally, cleavage at a site in the 5' UTR of *yajQ* stabilizes the transcript in WT relative to *rnc*- (∼6 minutes for WT and ∼ 3 minutes for *rnc*-) (**Figure 2.11**). Only the stability of *yajQ* was affected, but it is possible that *panE* might be influenced by this cleavage event. Due to the complementary nature of these hairpins, very similar structures could form from transcript of the plus strand that encodes *yajQ* and the minus strand that encodes *panE* (**Figure 2.6**).

RNase III mediated processing of transcripts encoding attenuation sensing peptides

The placement of several RNase III cleavage site may indicate additional levels of regulation. Two genes containing verified novel RNase III cleavage sites, *hisL* and *pheM*, are regulated by transcription attenuation. Both *hisL* and *pheM* encode short peptides with an abundance of histidine and phenylalanine respectively. There are a series of stem loops that partially overlap these peptides that can form two alternate structures; one leading to termination and the other leading to transcription of the downstream genes encoding histidine biosynthesis genes or phenylalanyl-tRNA synthetase genes. In the case of *hisL*, we observed an increased amount of this leader transcript in the *rnc*- strain compared to WT (**Figure 2.12**). Using 5' RACE on in vitro cleavage reactions we found that this site overlapped with the E:F termination structure (Blasi et al. 1981). An RNase III cleavage

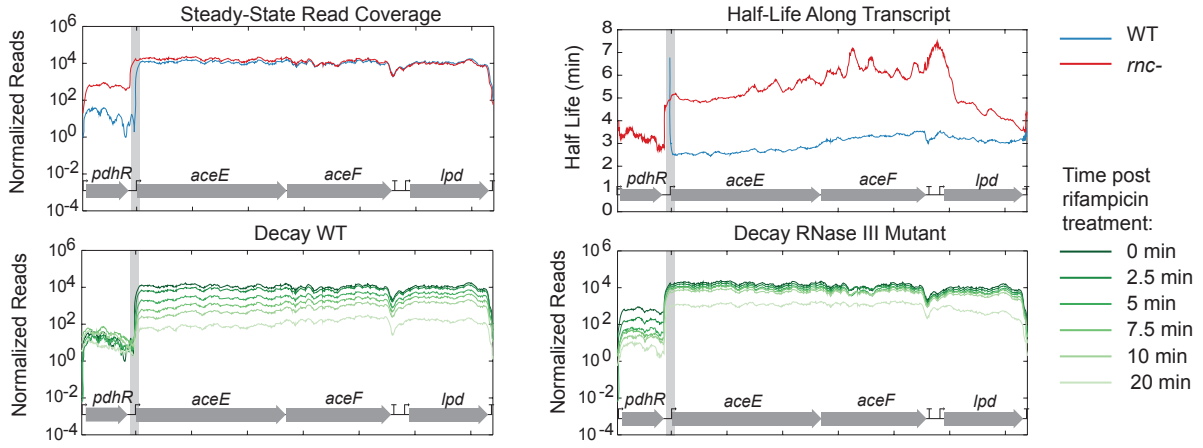


Figure 2.7: Read coverage and decay near novel RNase III cleavage site *aceEF* that shows increased degradation after RNase III cleavage and stabilization in the *rnc* deletion mutant. Normalized read coverage is displayed along the coordinates of the *E. coli* genome. Degradation throughout time is shown for both the WT and the *rnc* deletion mutant, with time zero minutes shown in dark green and lighter green at 20 min. Half-lives in both the WT strain and the *rnc* deletion mutant were calculated and displayed only when the R^2 value was ≥ 0.7.The cleavage site is identified with a gray box.

site was also identified in the *pheM* transcriptional attenuation structure (**Figure 2.13**). Using 5' RACE, we identified several different cleavage products with ends located within the termination structure. Unlike *hisL*, the leader structure did not accumulate in *rnc*-. Instead, degradation around this site resembled that of the well-known cleavage site 5' of *pnp* (**Figure 2.2**) where RNase III cleavage leads to degradation of the downstream message.

RNase III cleavage site affects key metabolic node: pyruvate dehydrogenase

Many of the novel target sites identified in this study are linked to genes encoding metabolic enzymes. For instance, we observed a significant difference in the half-life of the *aceE-aceF-lpd* transcript, encoding the multi-subunit pyruvate dehydrogenase complex (PDH), between WT and *rnc*- strains. Given the critical role of pyruvate dehydrogenase, we examined the differences in PDH activity in WT and *rnc*-. Samples were taken from WT and *rnc*- grown in LB (with 50 µg/ml thymine) and M9 (with 50 µg/ml thymine, 0.4% glucose, and 0.2% casamino acids) during

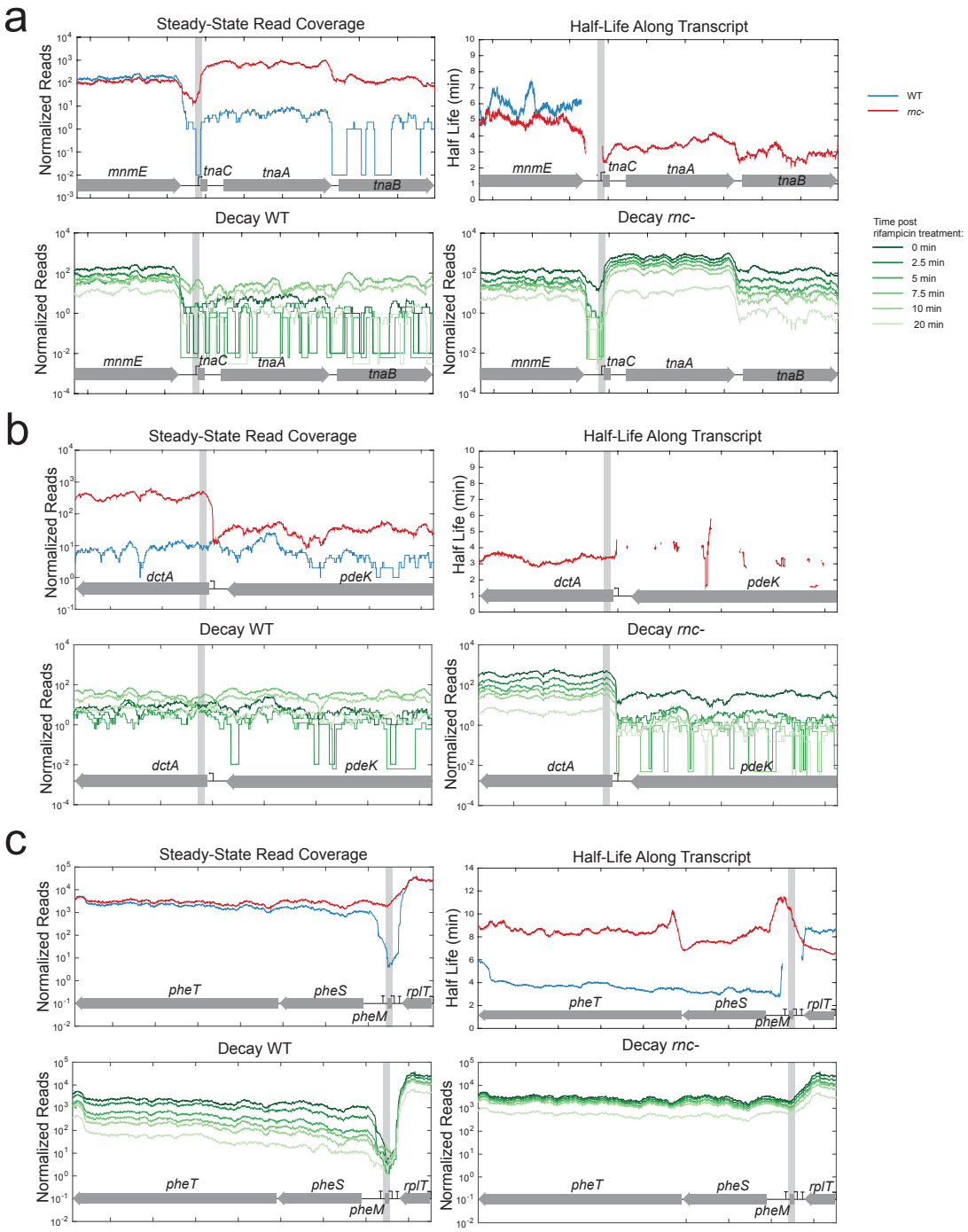
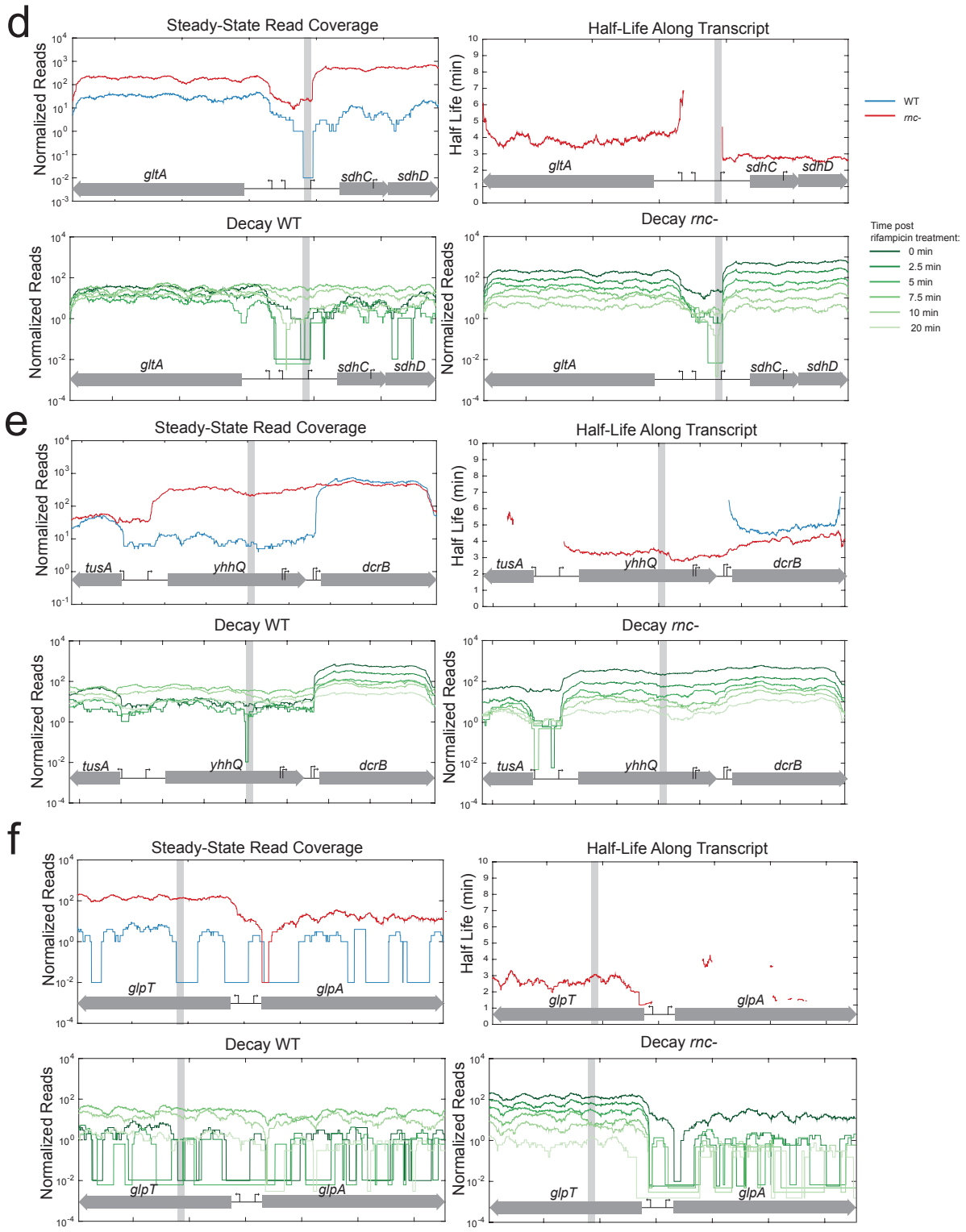
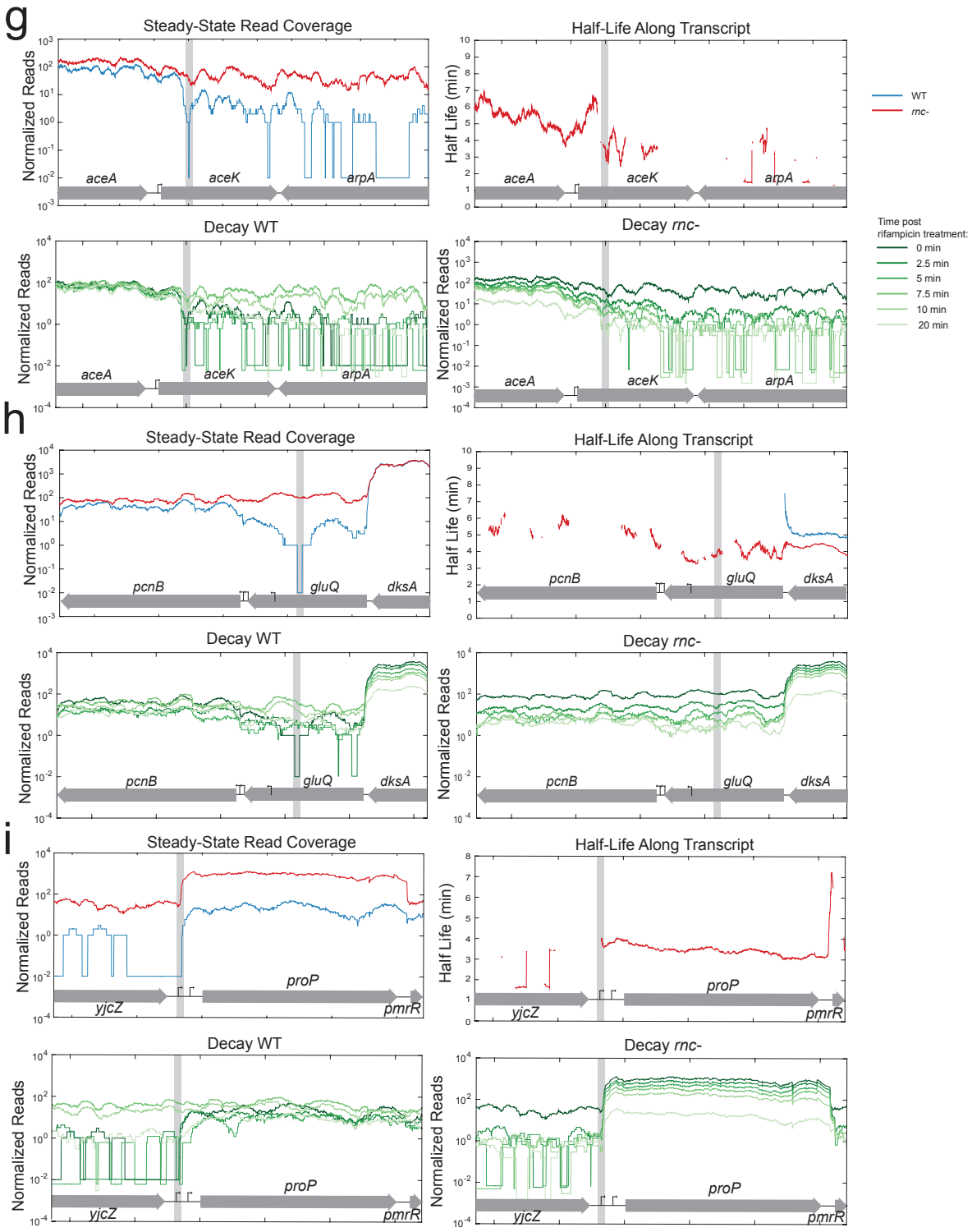


Figure 2.8: Steady-state read coverage and decay throughout time in WT and *rnc*- for *tnaC*, *dctA*, *pheM*, *sdhC*, *yhhQ*, *glpT*, *aceK*, *gluQ*, and *proP* transcripts (a-i). All transcripts in WT seemed to be degraded more quickly than in *rnc*-. We hypothesize that for several transcripts in WT the half-lives could not be calculated because degradation happened very rapidly.


d
Steady-State Read Coverage
Half-Life Along Transcript
Normalized Reads
Half Life (min)
gltA
sdhC
sdhD
WT
rnc-
Decay WT
Decay rnc-
Time post rifampicin treatment:
0 min
2.5 min
5 min
7.5 min
10 min
20 min
e
tusA
yhhQ
dcrB
f
glpT
glpA


g
Steady-State Read Coverage
Normalized Reads
aceA
aceK
arpA
Half-Life Along Transcript
Half Life (min)
WT
rnc-
Decay WT
Decay rnc-
Time post rifampicin treatment:
0 min
2.5 min
5 min
7.5 min
10 min
20 min
h
pcnB
gluQ
dksA
i
yjcZ
proP
pmrR

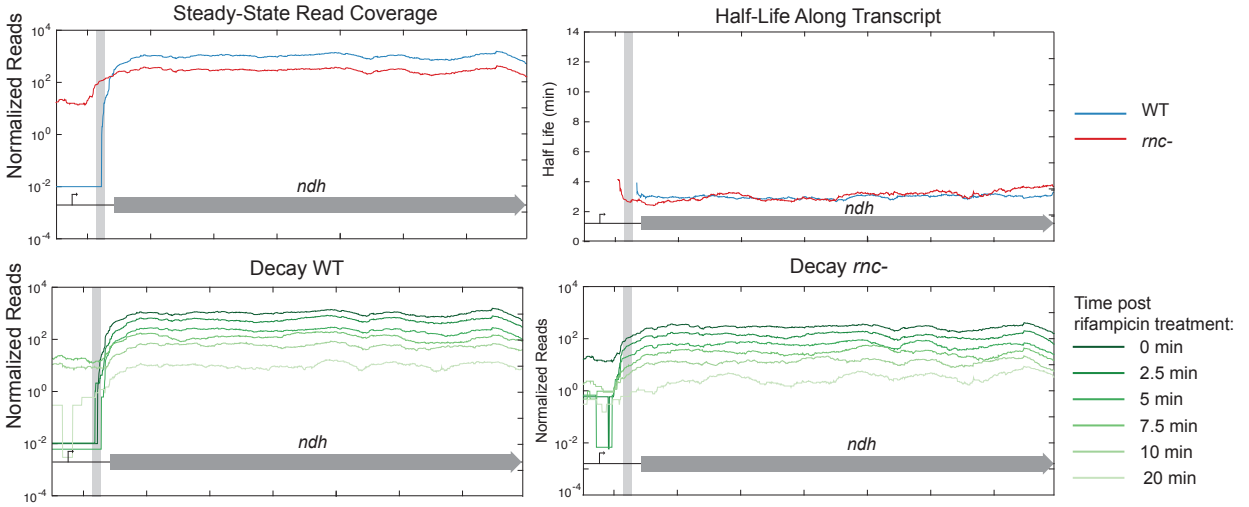


Figure 2.9: Steady-state read coverage and decay throughout time in WT and *rnc-* for *ndh* transcript. There was no difference in half-life between WT and *rnc-*.

exponential, early stationary, and late stationary phases. PDH activity was measured in a coupled enzymatic assay by monitoring pyruvate-dependent NADH formation in cell lysates. Increases in PDH activity were seen for *rnc-* during early stationary phase, but not during exponential or late stationary phase (**Figure 2.14**). RNase III processing of the *aceE-aceF-lpd* transcript could be responsible for altering PDH activity, presumably by altering the stability and abundance of the transcript (as seen in **Figure 2.7**).

Discussion

Mechanisms of RNase III-mediated regulation

In addition to confirming known sites, our data identified eleven novel RNase III sites verified by in vitro cleavage assays. Together the RNase III sites can be grouped into four mechanisms of regulation. First, cleavage by RNase III in the 5' UTR can decrease mRNA abundance of protein-coding transcripts via a decrease in half-life of the downstream message. Previously known examples of this phenomenon include *rnc*, *pnp*, and *metY*. Here, we found similar degradation

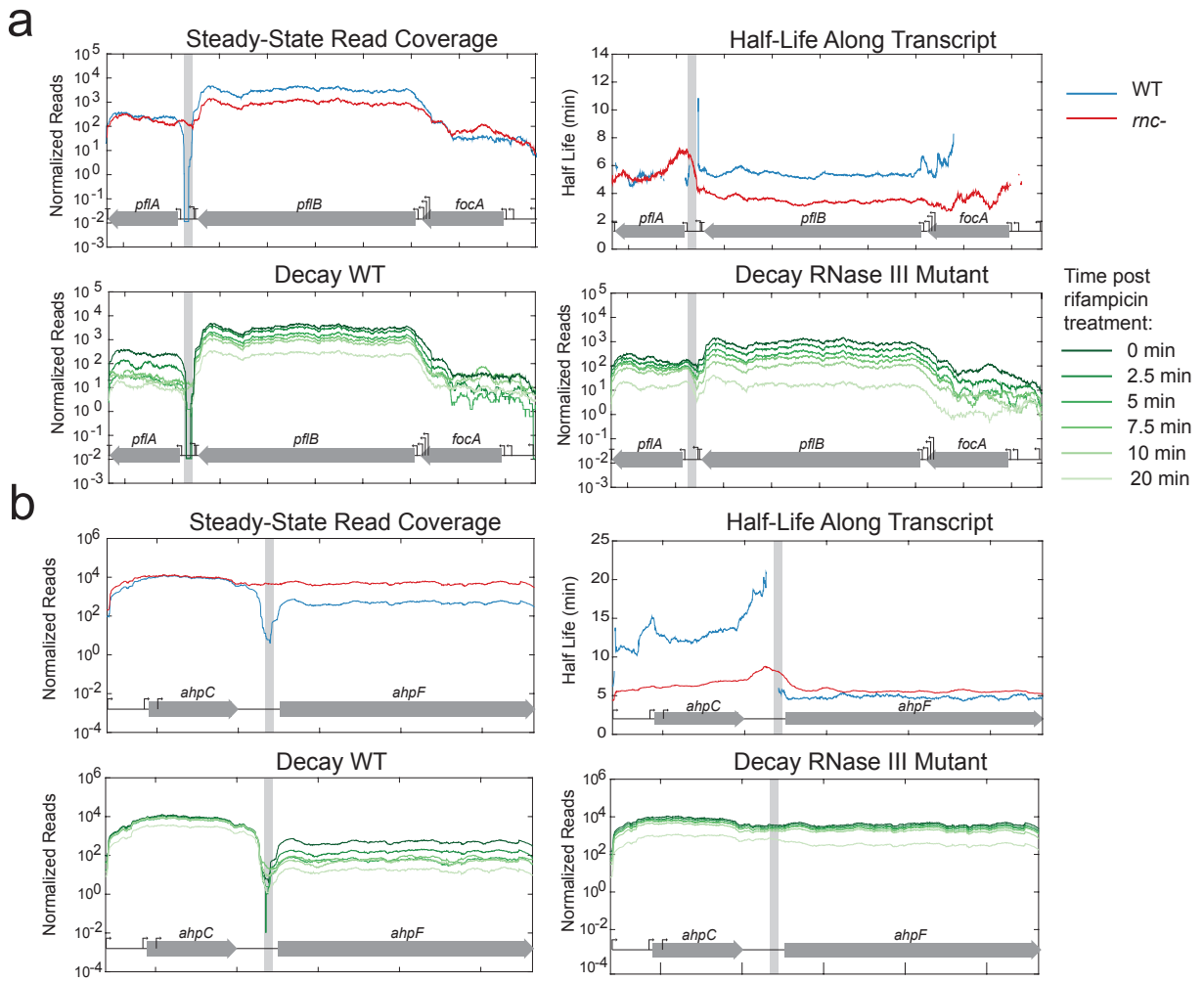


Figure 2.10: Steady-state read coverage and decay throughout time in the WT strain and *rnc* deletion mutant for *pflBA* (a) and *ahpFC* (b) transcripts that show increased stability in the WT.

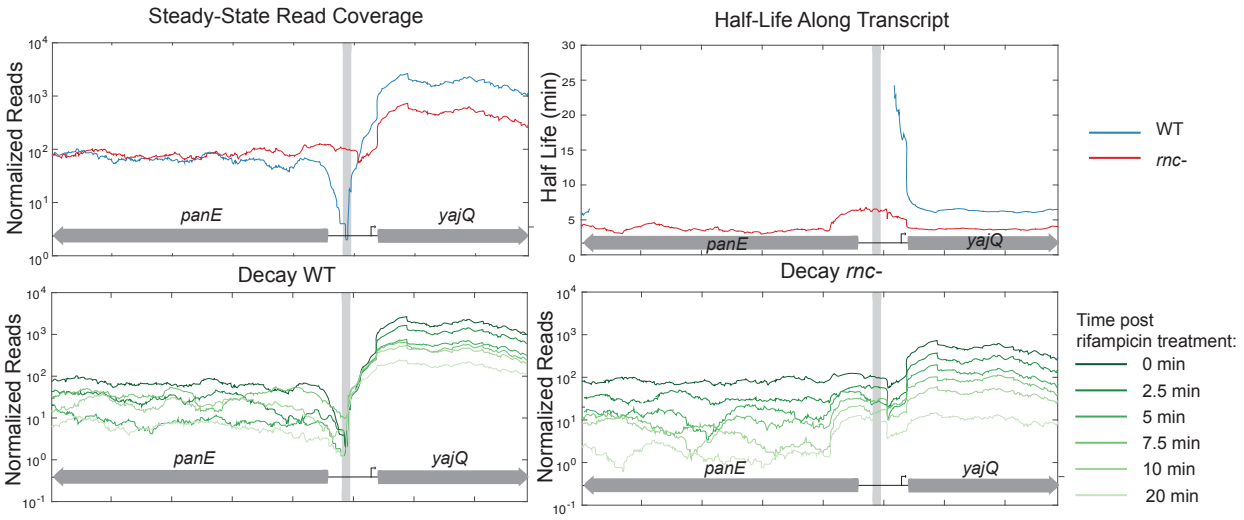


Figure 2.11: Steady-state read coverage and decay throughout time in WT and *rnc-* for *yajQ* and *panE* transcripts. Due to the lack of stranded RNA-seq data we were unable to determine if the dip in coverage was resulting from processing of the *yajQ* transcript on the positive strand or the *panE* transcript on the negative strand. There was only a different in half-life of the *yajQ* transcript.

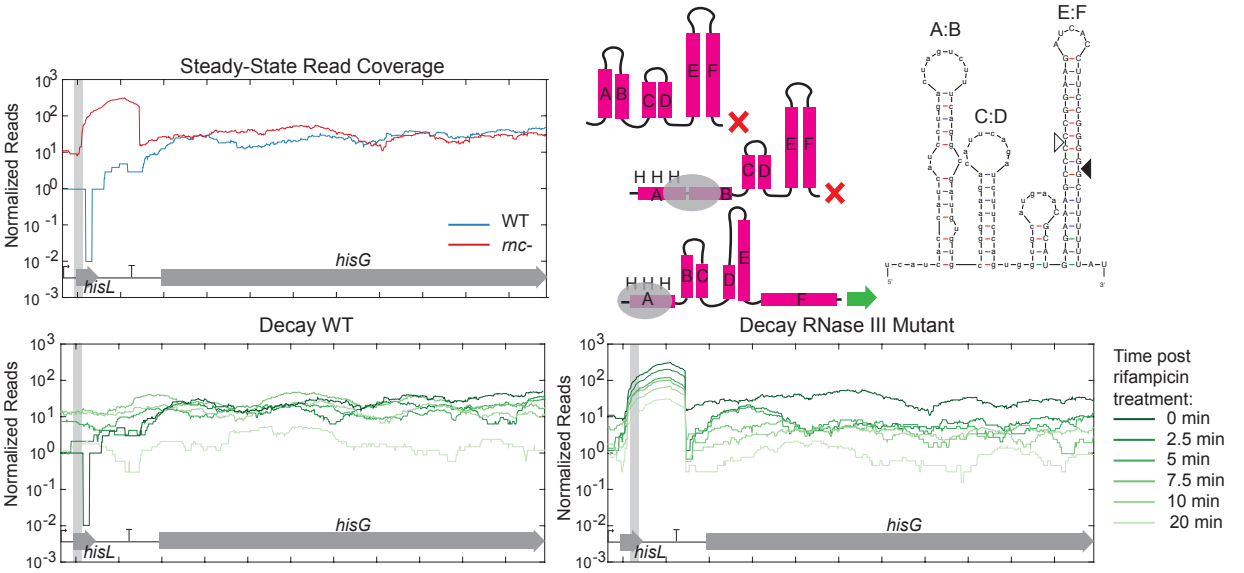


Figure 2.12: RNase III initiates decay of terminated *hisL* leader transcript. Steady-state read coverage and decay throughout time in the WT strain and *rnc* deletion mutant for *hisL* are shown in addition to the predicted secondary structures of the termination and attenuation structures. The verified RNase III cleavage site is indicated with a filled triangle.

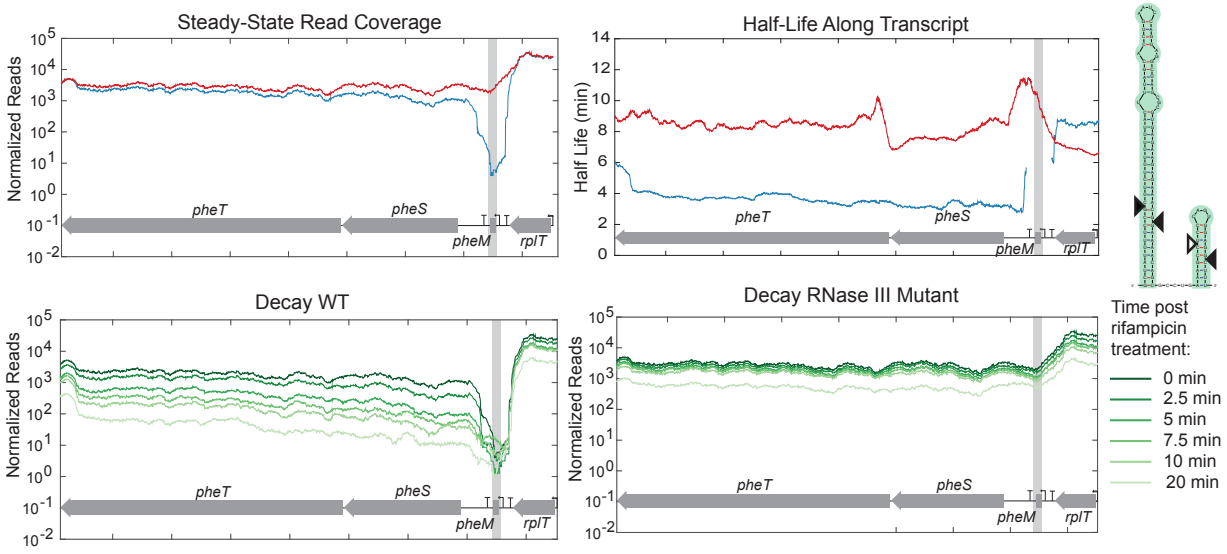


Figure 2.13: RNase III initiates decay of terminated *pheM* leader transcript. Steady-state read coverage and decay throughout time in the WT strain and *rnc* deletion mutant for *pheM* are shown in addition to the predicted secondary structure of the termination structure. The verified RNase III cleavage site is indicated with a filled triangle.

patterns near *aceEF*, *proP*, *tnaC*, *dctA*, *pheM*, *hisL*, *sdhC*, *yhhQ*, *glpT*, *aceK*, and *gluQ*. For several of these transcripts (*proP*, *dctA*, *tnaC*, *sdhC*, *yhhQ*, *glpT*, *aceK*, *gluQ*) we hypothesize that the degradation was rapid because there is no decrease in read coverage throughout the time course and the transcript abundance is low. We suspect that rapid decay occurred during the 3.5 minutes between rifampicin addition and the initial sampling. Rapid transcript degradation following RNase III cleavage has been explained by the loss of cis-acting stabilizing elements – e.g. hairpins at the 5' end of transcripts blocking RNase E activity (Emory et al. 1992) or hairpins at the 3' end of transcripts blocking exonucleases polynucleotide phosphorylase (*pnp*) and RNase II (Coburn et al. 1996). Alternatively, RNase III can inactivate transcripts by cleaving within coding regions (e.g. *nirB* (Stead et al. 2011), *arfA*, (Garza-Sánchez et al. 2011), and *mltD* (Lim, Ahn, et al. 2014)). In this study, we found no verified novel sites within coding regions.

Second, RNase III can lead to increased mRNA abundance and/or expression of target genes. For instance, RNase III cleavage removes occlusion of the RBS controlling expression of alcohol

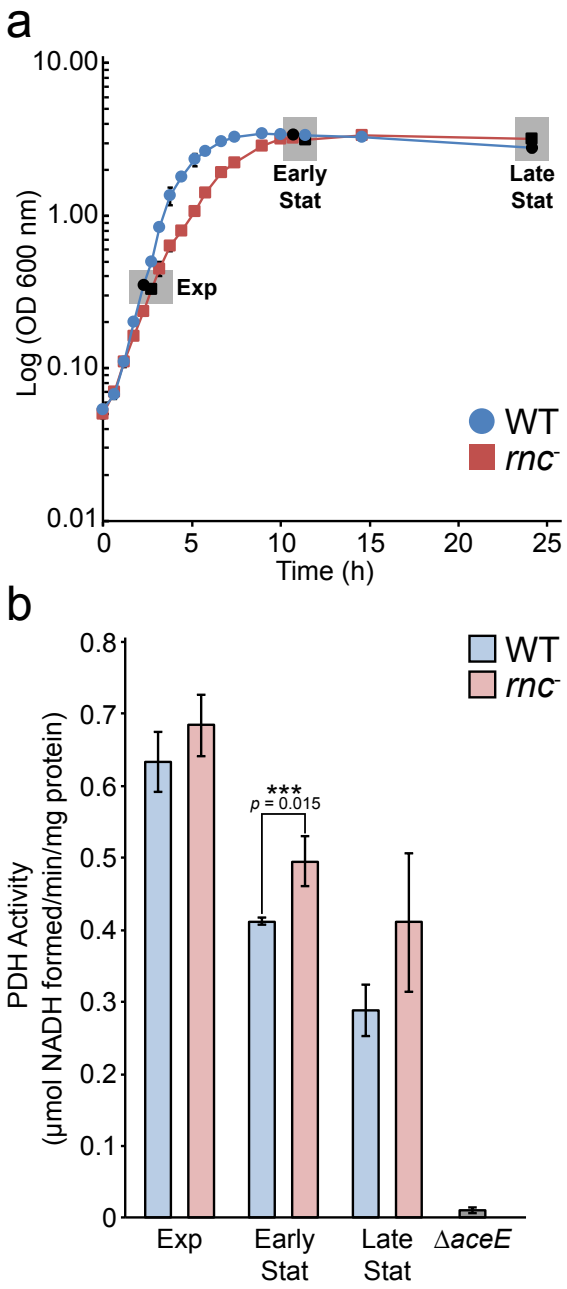


Figure 2.14: Pyruvate dehydrogenase activity in WT strain and *rnc* deletion mutant crude cell lysates. (a) Growth of MG1693 and rnc deletion mutant in LB medium supplemented with thymine (50 *micro*g/ml). Cells were harvested during exponential (Exp), early stationary (Early Stat), and late stationary (Late Stat) growth for enzyme assays. (b) Pyruvate dehydrogenase (PDH) activity was measured spectroscopically as pyruvate-dependent NADH formation using lysates extracted from cells harvested at different growth phases. Lysates from a Δ*aceE* strain (JW0110-2; Keio Collection), grown to stationary phase in M9 minimal medium supplemented with glucose (0.2%) and acetate (0.1%), were used as a negative control because they lack PDH activity. Values represent means of results from three biological replicates, and error bars indicate ± standard deviations (***, $P < 0.05$, Students *t* test, two-tailed).

dehydrogenase (*adhE*) (Aristarkhov et al. 1996). In the case of *ahpF*, *pflB*, and *yajQ*, novel sites identified in this study, there is an increase in stability of the processed transcript compared to the unprocessed transcript. We propose that these cleavage events leave stabilizing hairpins that protect secondary transcripts from degradation. Increased stability of RNase III-cleaved transcripts could be due to creation or placement of new stabilizing structures that impede further RNA turnover – e.g. placement of hairpins 3' of *pflB* and *ahpC* to block exonucleases and 5' of *yajO* that could slow RNase E processing (**Figure 2.15**). The importance of this novel mode of regulation is highlighted by the functions of the encoded proteins. AhpC and AhpF are located in the same operon and encode the subunits of alkyl hydroperoxide reductase, where AhpC forms a decameric ring and interacts with an AhpF dimer (Dip et al. 2014). Proper regulation and stoichiometry of these two genes transcribed in a single operon may be critical because they are among the most highly abundant proteins in the cell (Ishihama et al. 2008; Link et al. 1997). Another transcript that shows increased stability upon RNase III cleavage was pyruvate formate lyase (PflB), which is central in anaerobic metabolism. All together, these examples show that RNase III cleavage events can lead to both positive and negative regulation of gene expression.

Third, we observed two novel cleavage sites within transcripts known to be involved in transcriptional attenuation (*hisL* and *pheM*). Here, formation of a RNA secondary structure leads to termination of the leader transcript that would otherwise read-through the terminator and transcribe downstream genes. We pinpointed the *hisL* and *pheM* cleavage sites to these known secondary structures. Therefore, we hypothesize that RNase III is involved in recycling the terminated leader transcript in order to reduce translation of the terminated leader transcript that would be in excess of that needed to sense charged tRNA. We observed a buildup of the *hisL* leader transcript in the RNase III mutant (**Figure 2.12**), but not the *pheM* leader transcripts (**Figure 2.13**). Instead, the *pheM* transcript resembled that of other cleavage sites where the downstream transcript decayed more rapidly. We hypothesize that RNase III does help turnover *pheM* leader transcripts, but in the growth conditions in which cells were harvested for RNA-seq the major transcript was the

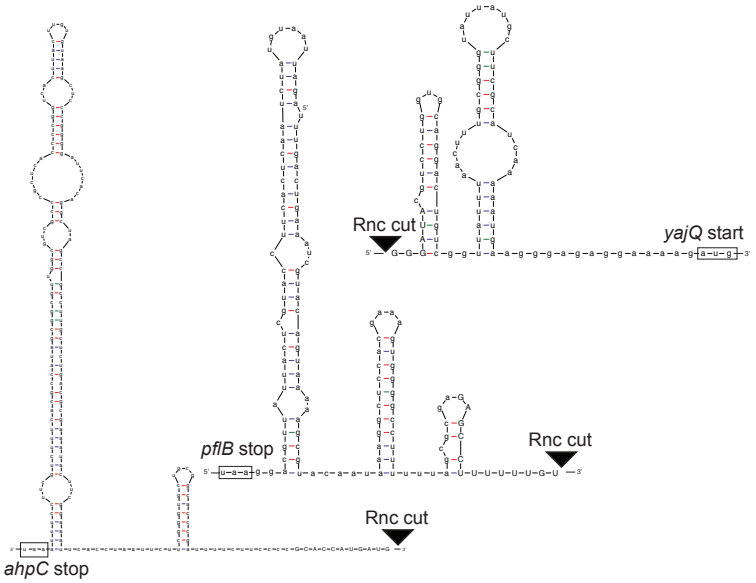


Figure 2.15: Predicted secondary structure of transcripts following RNase III cleavage. Cleavage events at the 3' end of the *pflB* and *ahpC* transcripts favor the formation of long hairpins at the 3' end (Zuker 2003). Cleavage at the 5' end of the *yajQ* transcript results in a transcript with a predicted hairpin at the 5' end.

read-through and not the attenuated transcript.

Fourth, we observed cleavage sites in several untranslated regions that produce products that are nearly identical to those produced by an alternate promoter. There is one published example of RNase III cleaving the *metY-nusA-infB* resulting in the downstream transcript resembling a transcript that originated from the internal P2 promoter (Régnier and Grunberg-Manago 1989). Two of our verified novel RNase III cleavage sites, near *aceEF* and *pflBA*, are also located very close to promoters. Both of the predicted hairpins contain the -35, -10, and transcription start site (TSS). These hairpin structures would only form from a transcript that began before the previous gene. In the case of *aceEF*, this cleavage event results in a downstream transcript that is identical to the mapped TSS of this promoter (Olvera et al. 2009). RNase III cleavage of the *pflBA* transcript leads to the downstream transcript having a single additional 'A' at the 5' end compared to a transcript that would have originated from the promoter preceding *pflA* (Sauter et al. 1990).

RNase III alters metabolic activity

RNase III-mediated regulation can have a significant impact on metabolism. Here, we observed and verified a cleavage site upstream of *aceE*, *aceF*, and *lpd*, genes encoding pyruvate dehydrogenase (PDH) and a site upstream of *pflB*, encoding pyruvate formate lyase. We observed an increase in the *aceEF-lpd* transcript levels in the RNase III mutant (**Figure 2.7**) and observed that RNase III cleavage led to a reduction in PDH activity only during early stationary phase and not exponential or late stationary phase. These new findings complement prior work that demonstrated RNase III cleavage affected expression of central metabolic enzymes, including alcohol dehydrogenase (*adhE*), oxoglutarate dehydrogenase and succinate dehydrogenase (*sdhCDAB-sucABCD*). This connection between ribonucleases and metabolic activity is not unprecedented. RNase E mutants (*rne-*) cannot grow on M9 with glucose without yeast extract or a mixture of amino acids, which indicated a defect in the connection between the glycolytic and TCA pathways in *rne-* (Tamura et al. 2013). This was explained by a drastic decrease in phosphoenolpyruvate carboxylase production. We observed a similar growth defect when *rnc-* was grown in M9 media with glucose, but this growth defect could be rescued by adding 0.2% casamino acids (data not shown). Both RNase E and RNase III are involved in the regulation of enzymes that connect glycolysis and the TCA cycle (**Figure 2.16**). In addition to these two examples, there are a significant number of examples of how ribonucleases regulate metabolic enzymes via RNA stability and other mechanisms. RNase R has been shown to hydrolyze the *prpBCDE* operon encoding enzymes in the 2-methylcitrate pathway enabling propionate degradation during oxic but not anoxic conditions (Simonte et al. 2016). Additional support for the link between RNases and central metabolism includes the presence of the degradosome where key glycolytic enzyme enolase complex with RNase E (Miczak et al. 1996). RNase G has also been associated with the degradation of the enolase transcript (*eno*) as well as the alcohol dehydrogenase transcript (*adhE*) (Kaga et al. 2002). Other metabolic enzymes that were profoundly affected by deletion of RNase G include phosphoglucose isomerase (pgi), glucokinase (*glk*), and triosephosphate isomerase (*tpiA*) (Lee, Bernstein, et al. 2002). Other links

between RNA metabolism and central metabolism include the ability of citrate to bind PNPase and act as either an allosteric regulator stimulating PNPase activity or act as an inhibitor when binding the catalytic site, causing significant effects on central metabolism (Nurmohamed et al. 2011). RNA-sequencing has proved to be a useful tool for identifying ribonuclease cleavage sites and can hopefully be used to better understand ribonuclease activity and specificity so that a more thorough picture of RNA turnover can be developed. We have identified novel RNase III cleavage sites in *E. coli* and describe four modes of regulation. Two of these novel cleavage sites may affect the abundance of key metabolic enzymes suggesting that ribonucleases may play an underappreciated role in ensuring proper regulation of central metabolic enzymes.

2.3 Materials and Methods

Sample Collection

We identified cleavage sites of RNase III by collecting and sequencing RNA from both an RNase III mutant (SK4455; *rnc*-14::ΔTn10 *thyA715 rph-1*) (Stead et al. 2011) and the parental strain (MG1693; *thyA715 rph-1*) from which the RNase III mutant was derived (Bachmann et al. 1980). Samples from exponentially growing cells (Lysogeny broth, 37°C with shaking) were harvested after the addition of rifampicin to stop nascent transcription. The first timepoint was taken 3.5 minutes after addition of the antibiotic to give time for the diffusion of rifampicin into the cells and binding of the antibiotic to RNA polymerase. This sample was called 0 minutes, and further samples were taken at 2.5, 5, 7.5, 10, and 20 minutes from wild-type *E. coli* and the RNase III mutant. A partial biological replicate of wild-type was taken (0, 2.5, and 7.5 minutes), to examine the reproducibility between samples.

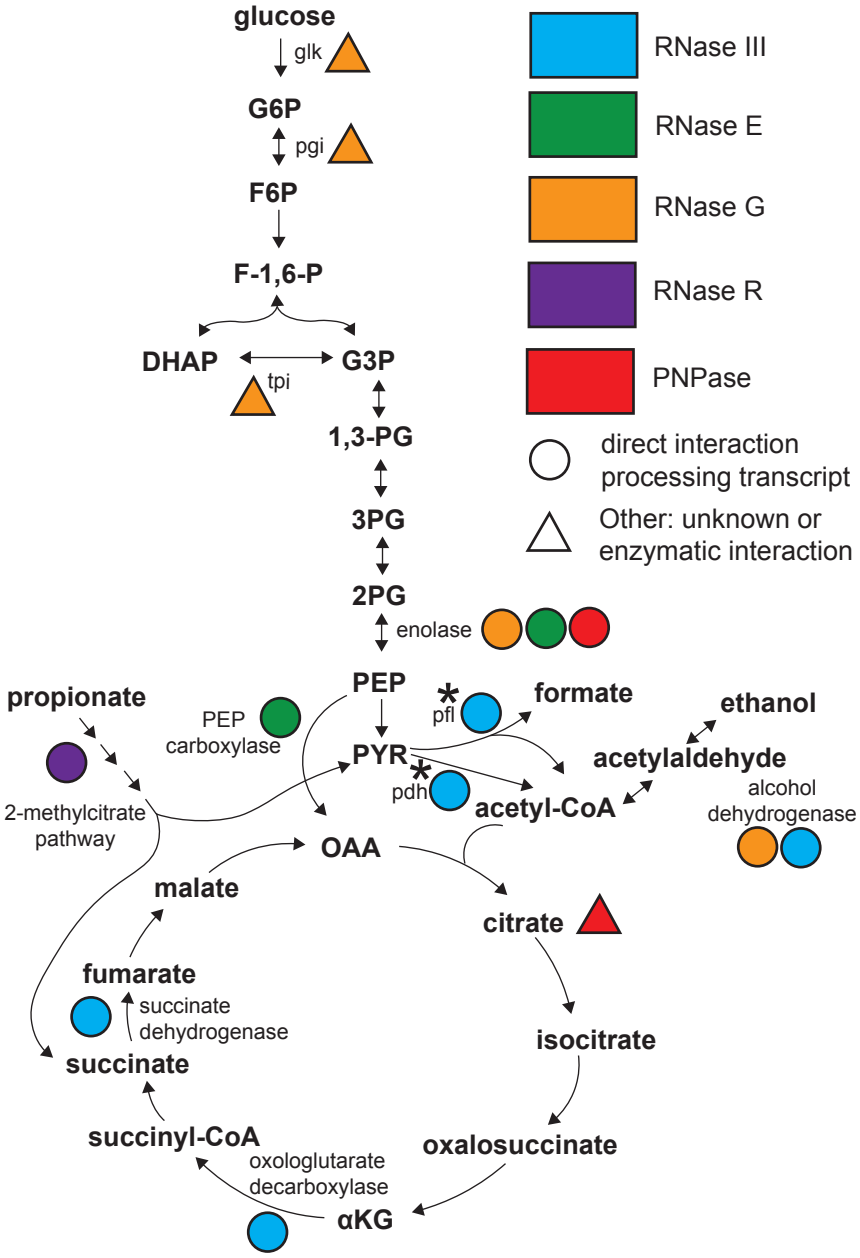


Figure 2.16: Influence of RNase activity on central metabolism. Circles indicate that the correspondingly colored RNase acts directly on the transcript encoding that enzyme (blue, RNase III; green, RNase E; orange, RNase G; purple, RNase R; red, PNPase). Colored triangles indicate that a particular RNase influences the abundance and activity of the enzyme via an unknown or indirect method. Triangles near metabolic intermediates indicate interactions of metabolites with RNases that influence RNase activity. A star indicates regulation identified in this study. Abbreviations: pgi, phosphoglucose isomerase; tpi, triosephosphate isomerase; pfl, pyruvate formate lyase; pdh, pyruvate dehydrogenase; G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F-1,6-P, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; G3P, glyceraldehyde 3-phosphate; 1,3-PG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; α KG, alpha-ketoglutarate.

RNA isolation and RNA-sequencing

Total RNA was extracted and DNase-treated according to published protocols (Khodurksy et al. 2003) and submitted to the UW Biotech Center where EpiCentre Ribo-Zero Gram Negative Ribosomal RNA reduction kits were used to remove rRNA from 2.5 µg of total RNA. 15 libraries were prepared (TruSeq RNA Library Prep) 2x100bp and sequenced at the UW DNA sequencing facility on the HiSeq2500 (Illumina). For all 15 samples there were between 14-23 million reads with the average Phred score above 35. Adapters were trimmed using Cutadapt and reads were aligned to the genome (NC_000913.3) using bwa 0.7.7. Samtools was used to index sort aligned reads. The Integrated Genomics Viewer (IGV) was used to visualize data (Robinson, Thorvaldsdottir, et al. 2011). The number of reads that aligned to each position of the genome was calculated using the count function of IGV tools with the window size set to 1. Normalization factors were calculated as the average of three stable genes (*ssrA*, *ssrS*, *rnpB*) and used to normalize successive timepoints to the time zero sample. Bases without any aligned reads were given a pseudocount of 0.01. RPKM was calculated using counts from HTSeq. Raw sequence files can be found in the Sequence Read Archive (SRP100654) and read counts for each base in the genome for each sample can be found in the Gene Expression Omnibus (GSE95318).

Expression and Purification of *E. coli* RNase III

Hexahistidine-tagged RNase III derived from *E. coli* MG1655 used was expressed and purified as previously described (Amarasinghe et al. 2001). Briefly, specific primers (RNase III.F and RNase III.R) were used to amplify the coding sequence of RNase III from *Escherichia coli* MG1655 genomic DNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and the resulting fragment was cloned into a BamHI/NdeI fragment of pET28b+ using Gibson Assembly (Gibson et al. 2009) to generate the pET28b-His-RNase III vector for expression in BL21 cells.

For expression, colonies from a fresh plate of BL21 cells transformed with pET28b-His-rnc

were pre-cultured in 5 ml of LB and transferred into 250 ml of pre-warmed LB supplemented with Kan (50 mg/L) in a 1L baffled flask and grown at 37°C with shaking (250 rpm). Protein expression was induced with IPTG (1 mM final concentration) when the cells reached an OD 600 nm = 0.3-0.4. Cells were harvested by centrifugation (6000 rpm, 10 min, 4°C, JA.10 rotor) after ∼4 hr induction, aspirated, and stored at -20°C until affinity purification.

Cells were suspended in 30 ml ice-cold buffer A (500 mM NaCl, 20 mM Tris-HCl pH 7.9) + 5 mM imidazole and sonicated on a Fisher Scientific Sonic Dismembrator Model 500 for 4 minutes (30s on, 40s off; small tip, 50% power) on ice. The clarified lysate and insoluble pellet fractions were harvested following centrifugation at 6000 x g (JA14.5 rotor) for 20 min. at 4°C. The pellet was washed in buffer A + 5 mM imidazole, solubilized in buffer A + 6 M urea for 3 hours at 4°C, and centrifuged at 10,000 g (JA14.5 rotor) for 15 min. at 4°C. The 6 M urea supernatant was applied to a 5 ml HisTrap (GE Healthcare Life Sciences) column pre-equilibrated in buffer A + 5 mM imidazole, and then washed with 10 column volumes buffer A + 5 mM imidazole and 6 column volumes buffer A + 60 mM imidazole. The protein was eluted with buffer containing 1 M NaCl, 400 mM imidazole, and 20 mM Tris-HCl (pH 7.9) (Elution 1, 5 ml; Elution 2, 5 mL; Elution 3, 5 mL; Elution 4, 15 mL). Elution fractions 2 and 3 were pooled, placed in a 12 mL Slide-A-Lyzer cassette (10 kDa MWCO, Life Technologies), and dialyzed against 2 L buffer (1 M NaCl, 400 mM imidazole, 60 mM Tris-HCl, pH 7.9) at 4°C for 2 h, and then dialyzed against 2L of the same buffer, but without imidazole. The sample was then dialyzed against buffer (1 M NaCl, 60 mM Tris-HCl (pH 7.9), 1 mM EDTA, 1 mM dithiothreitol (DTT)) for 16 h at 4°C. The protein was concentrated (Amicon 10 kDa MWCO, Millipore) to ∼1.5 mg/ml as estimated by absorbance at 280 nm (ε = 14,440 M^{-1} cm^{-1}). Proteins were resolved by SDS-PAGE (Any kD Mini-PROTEIN protein gel, Bio-Rad) and stained with GelCode Blue (Thermo Scientific) (**Figure 2.17**). The purity of the protein was estimated to be ∼90% based on densitometry. The purified protein (0.77 mg/ml) was stored in aliquots at -20°C in 50% glycerol (v/v), 0.5 M NaCl, 30 mM Tris-HCl (pH 7.9), 0.5 mM EDTA, 0.5 mM DTT) until use.

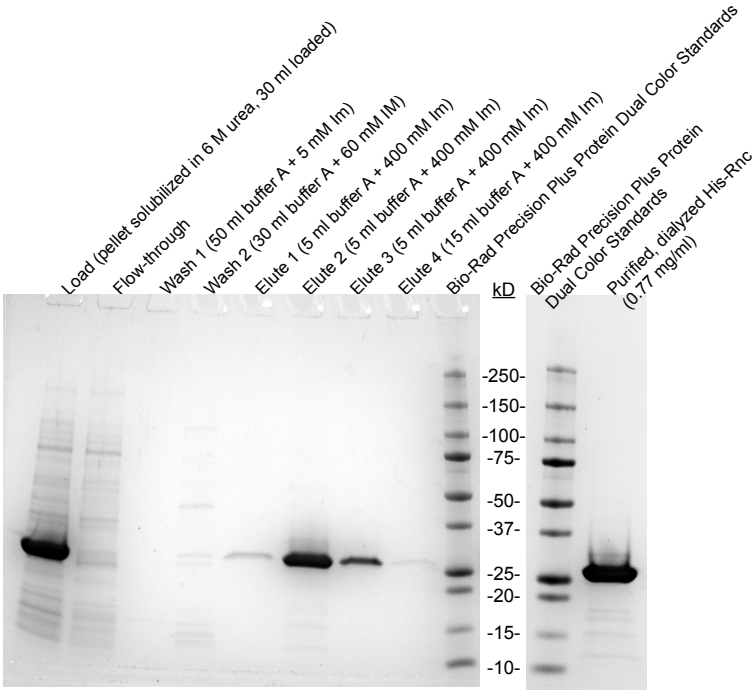


Figure 2.17: Purification of *E. coli* RNase III by immobilized metal affinity chromatography.

In vitro RNA cleavage assays

DNA primers were designed to amplify regions surrounding potential cleavage sites with the forward primer containing the T7 promoter at the 5' end of the primer. All oligonucelotides are listed in **Table 2.2**. DNA was amplified with Phusion High-Fidelity DNA Polymerase (New England Biolabs) and product size was verified on a 1% DNA agarose gel. PCR products were purified by ethanol precipitation and RNA was transcribed with the T7 RiboMAX^{TM} Express Large Scale RNA Production System (Promega). RNA was quantified with a Qubit 2.0 using the RNA HS Assay Kit (ThermoFisher Scientific). 5 µg of RNA was briefly heated at 100°C for 30 seconds before being placed on ice. RNA was diluted in cleavage reaction buffer (30 mM Tris pH 8, 160 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT) for a total reaction volume of 100 µl. 10 µl was removed before the addition of 0.70 µg His-RNase III and before the addition of $MgCl_2$. Cleavage was initiated with addition of 10 mM (final concentration) $MgCl_2$ and placed at 37°C for 15 minutes. Reactions

were quenched by addition of 40 mM EDTA (final concentration) in 2X urea sample loading buffer (BioRad 1610768). RNA was run on a 5% polyacrylamide 8 M urea gel and visualized with ethidium bromide (Fisher Scientific BP1302-10).

Table 2.2: Oligonucleotides used to verify RNase III cleavage sites

Name	**Purpose**	**Sequence**
Rnc.F	amplify Rnc	gccgcgcggcagccatATGAACCCCATCGTAATTAATC
Rnc.R	amplify Rnc	cggagctcgaattcggatcctaTCATTCCAGCTCCAGTTTTTTC CGAGTTTCAGAA
JCC259	*pnp* in vitro template	TAATACGACTCACTATAGGGttaagaaggagatataccTTCTTG
JCC107	*pnp* in vitro template	GCCATTGAACGGAATACCAG
GG274	*pflAB* in vitro template	acgacgaacaaaaccgttgg
GG275	*pflAB* in vitro template	CGAAATTAATACGACTCACTATAGGcgctgactaaagaacagcagca
JCC402	*pdhR* in vitro template	cgaaatTAATACGACTCACTATAGGGCGTCGTCTGGA GCAACGAAAGA
GG273	*pdhR* in vitro template	TGCGTGCACGGAAGAAGTGG
GG276	*pheM* in vitro template	agtgcagcgctttccagttc
GG277	*pheM* in vitro template	CGAAATTAATACGACTCACTATAGGcgtgaccgtcgtcaacgtaa
GG278	*hisL* in vitro template	CGAAATTAATACGACTCACTATAGGatcagttgaataaacattca cagagac
GG279	*hisL* in vitro template	ccctgaatacgggtcagcag
GG280	*ahpFC* in vitro template	CGAAATTAATACGACTCACTATAGGaagtttgcccggctaaatgg
GG281	*ahpFC* in vitro template	gcgggttcagtacgctcatc
GG282	*ndh* in vitro template	CGAAATTAATACGACTCACTATAGGccctcgtcgcatcaggtaac
GG283	*ndh* in vitro template	cgccattaccagggtgtcat
JCC423	gene D in vitro template	cgaaatTAATACGACTCACTATAGGGTAAGAAATCATGAG TCAAGTTACTGAAC
JCC424	gene D in vitro template	TTACATCACTCCTTCCGCACGTAATTTT
JCC429	gene H in vitro template	cgaaatTAATACGACTCACTATAGGGTGAGGTGATT TATGTTTGGTGCTATTG
JCC430	gene H in vitro template	TTATTTCCTAGACAAATTAGAGCCAATACC

GG310	*mltD* in vitro template	tcagacgctgcatcatattc
GG311	*mltD* in vitro template	CGAAATTAATACGACTCACTATAGGcttttaagcaactattgac acacac
GG312	*ompA* in vitro template	tagtgtctgcacgccatacc
GG313	*ompA* in vitro template	CGAAATTAATACGACTCACTATAGGgattctcttctgtaaattg tcgctg
GG396	*tnaA* in vitro template	CGAAATTAATACGACTCACTATAGGtctacaacagggcaaagcg
GG397	*tnaA* in vitro template	gctgtcggtcagtaaatcg
GG398	*sdhC* in vitro template	CGAAATTAATACGACTCACTATAGGgcttttcctggtaatgttt gtaac
GG399	*sdhC* in vitro template	gaacgaggatgaaatcatgtacg
GG400	*dctA* in vitro template	CGAAATTAATACGACTCACTATAGGccagggcttccttttgc
GG401	*dctA* in vitro template	gaagatgacctgcgagaaac
GG402	*gatD* in vitro template	CGAAATTAATACGACTCACTATAGGcgtgacatcgctcgtaatg
GG403	*gatD* in vitro template	gttatcgttcacacggacc
GG404	*proP* in vitro template	CGAAATTAATACGACTCACTATAGGgatgccacgactctgttcag
GG405	*proP* in vitro template	GCAATAGAACCGAAGTCCAG
GG408	*yhhQ* in vitro template	CGAAATTAATACGACTCACTATAGGcactggatggaaatcgcg
GG409	*yhhQ* in vitro template	GTCTGCGCTTTTTGCTGAT
GG410	*glpT* in vitro template	CGAAATTAATACGACTCACTATAGGcattgttatccctctgaa gttcg
GG411	*glpT* in vitro template	aataatgccaccagaatggc
GG412	*mglC* in vitro template	CGAAATTAATACGACTCACTATAGGACA GTGGCTTCTCTACCTTT
GG413	*mglC* in vitro template	CGGTGGGTTGTATTGAGC
GG414	*alsE* in vitro template	CGAAATTAATACGACTCACTATAGGGGACAACCGT TCATTCCTG
GG415	*alsE* in vitro template	GCAATTGGTTTCCAGGCA
GG416	*aceK* in vitro template	CGAAATTAATACGACTCACTATAGGCGAGTGGCTA CGGGAAAT
GG417	*aceK* in vitro template	TCTCTTCAAACAGCGGACC

GG418	*gluQ* in vitro template	CGAAATTAATACGACTCACTATAGGcggtcattgccgggtgttg
GG419	*gluQ* in vitro template	tgccaagtaacaggtcgcgc
GG420	*panE* in vitro template	CGAAATTAATACGACTCACTATAGGGTGCTTCCTG AAGATCAACTT
GG421	*panE* in vitro template	ATTAATCACGCAGTTGACTGC
GG422	*yajQ* in vitro template	CGAAATTAATACGACTCACTATAGGAACCATCTGTCT CAACCAGAT
GG423	*yajQ* in vitro template	CATGACAGCCTGCAAATCATC
GG424	*yneM* in vitro template	CGAAATTAATACGACTCACTATAGGGAATGTTTTTAT GGCCGTACTG
GG425	*yneM* in vitro template	TCCTTTGATCATCAGTTACGC
GG426	*mgrR* in vitro template	CGAAATTAATACGACTCACTATAGGGGGCAATGTAT GAGGGTAAG
GG427	*mgrR* in vitro template	GAAATCATTTATGCGCCGG
GG320	*trpL* in vitro template	CGAAATTAATACGACTCACTATAGGaagttcacgtaaaaagggtatcg
GG321	*trpL* in vitro template	ccgccacaaggtcataagag
GG322	*thrL* in vitro template	CGAAATTAATACGACTCACTATAGGatagcgcacagacagataaa aattac
GG323	*thrL* in vitro template	gaccgcgcgcttctaatac
GG338	*ilvL* in vitro template	CGAAATTAATACGACTCACTATAGGattc cttcgaacaagatgcaag
GG340	*ilvL* in vitro template	gacaatcccaggacatccac
GG324	*ivbL* in vitro template	CGAAATTAATACGACTCACTATAGGaaatattgttaaacacaaaac caacaag
GG325	*ivbL* in vitro template	tgcgggagttcttcgatatg
GG326	*leuL* in vitro template	CGAAATTAATACGACTCACTATAGGatatcgcattcatctggagc
GG327	*leuL* in vitro template	ggcgcgtttcaccatataga
GG328	*tnaC* in vitro template	CGAAATTAATACGACTCACTATAGGgtgtcttgcgaggataagtg
GG329	*tnaC* in vitro template	ctgactcggctaacgcatag
GG330	*pheL* in vitro template	CGAAATTAATACGACTCACTATAGGagtcacttaaggaaacaaa catg
GG331	*pheL* in vitro template	atcggtacgacggcatagtc

GG332	*mgtL* in vitro template	CGAAATTAATACGACTCACTATAGGtagatgctacgaatattatt ggattctcc
GG333	*mgtL* in vitro template	tccgtggcgtaagaaatagc
GG334	*pyrL* in vitro template	CGAAATTAATACGACTCACTATAGGacaatttgccgggaggatg
GG335	*pyrL* in vitro template	acttcgctaacgcctgagtc
GG336	*rhoL* in vitro template	CGAAATTAATACGACTCACTATAGGgacttcgtattaaacatacctt attaagtttg
GG337	*rhoL* in vitro template	tcgatcagcagaaccatcag

5' Race was performed on 70 µl of the in vitro cleavage reactions. Reactions were stopped with addition of 40 mM EDTA (final concentration) and RNA was extracted with phenol/chloroform and ethanol precipitation. cDNA was transcribed with Promega's GoScript^{TM} Reverse Transcription Kit (A5001) using gene specific primers according to manufacturer's instructions. cDNA was purified using a QIAquick PCR purification kit (Qiagen 28106). 1 pmol of PCR product was tailed with either dCTP or dGTP using terminal deoxynucleotidyl transferase (ThermoFisher Scientific EP0161). Products were PCR purified and 1 µl was used as template in a 40 µl PCR reaction using GoTaq Green Master Mix (Promega M712) containing the gene specific primer and primer designed to bind the tailed 5' end. PCR reactions were purified and sequenced (Functional Biosciences, Madison, WI). Some cleavage reactions yielded a mixture of products, so these were mixtures were subcloned using a pGEM-T vector system (Promega A3600).

Pyruvate dehydrogenase assay

Plates (LB + thymine 50 µg/ml) were inoculated with frozen glycerol stocks from SK4455 (*rnc*-) and the parental line (WT; MG1693) were grown overnight at 37°C. Single colonies from plates were used inoculate pre-culture tubes containing 5 ml of growth medium (LB + 50 µg/ml thymine or M9 minimal medium containing 0.4 % glucose, 0.2% casamino acids, and 50 µg/ml) and the resulting cultures were grown overnight at 37°C with shaking at ∼200 rpm. Overnight cultures were diluted to OD 600 nm = 0.05 in 50 ml pre-warmed media (LB + 50 µg/ml thymine or M9

minimal medium containing 0.4% glucose, 0.2% casamino acids, and 50 µg/ml) in 250 ml baffled shake flasks and grown at 37°C with shaking at 250 rpm). Samples were collected (∼3.3 ml/OD 600 nm) at specific timepoints (log, early stationary, late stationary) by centrifugation (5000 x g, 5', 4°C, FX6100 rotor, Beckman). Samples were aspirated, washed in 1 ml 25 mM phosphate buffer (pH 8.0), centrifuged again at 5000 x g, and the pellets were stored at 23°C.

Pyruvate dehydrogenase activity was measured in crude cellular lysates using an NADH-linked spectrophotometric assay as previously described(56). Briefly, frozen cell pellets were resuspended in 1 ml buffer (50 mM potassium phosphate buffer pH 8.0, 2 mM PMSF) on ice and then sonicated (Fisher Scientific Sonic Dismembrator Model 500) using the small tip (40% power, 1 s ON, 10 s OFF, 1.5 min total). Clarified lysates were harvested following centrifugation at 16,100 x g for 15 min at 4°C (Eppendorf microfuge) and the protein concentration determined using Bradford Reagent (Sigma) by comparison to calibrated bovine serum albumin standards (Fisher). For PDH activity, crude extract was added to a reaction mixture (1 ml total) containing thymine pyrophosphate (0.2 mM), CoA (0.1 mM), $MgCl_2 \cdot 6H_2O$ (1 mM), dithiothreitol (0.3 mM), NAD+ (2.5 mM), and bovine serum albumin (100 µg/ml). The reaction was performed at room temperature (23°C) and was initiated by the addition of pyruvate (5 mM). The pyruvate-dependent reduction of NAD+ was monitored as a function of time at 340 nm (ε = 6,220 M^{-1} cm^{-1}) and the enzyme activity is reported as µmol NADH produced $\cdot$ min^{-1} $\cdot$ mg $protein^{-1}$.

3 Chapter 3: RNases present in cyanobacteria

Authors and Contributors:

Jeffrey Cameron: conceptualized the focus of manuscript, designed the experiments, conducted bioinformatic research, and wrote the manuscript.

Gina Gordon: assisted in some bioinformatic analysis, verifying segregation of RNase deletion strains, and edited of the manuscript.

Dr. Brian Pfleger: supervised the research and edited the manuscript.

This chapter was adapted from: Cameron, J.C., Gordon, G.C., and Pfleger, B.F. (2015). Genetic and genomic analysis of RNases in model cyanobacteria. *Photosynthesis Research* 126:171-183. doi: 10.1007/s11120-015-0076-2.

3.1 Introduction

The flow of information from DNA to mRNA to protein is highly regulated to generate the appropriate quantities of enzymes and structural proteins required for growth and metabolism. Transcriptional control of gene expression has been well studied in many organisms. However, there is little concordance between transcript abundance and protein abundance (de Sousa Abreu et al. 2009). This disconnect can arise due to post-transcriptional regulation of mRNA and protein molecules (e.g. degradation or dilution). Among the many factors involved in post-transcriptional regulation in prokaryotes, the process of mRNA turnover is among the least understood. Much of our knowledge on mRNA turnover has been gained from studies of *Escherichia coli* and *Bacillus subtilis*, model Gram-negative and Gram-positive bacteria, respectively (Anderson et al. 2009). These studies have shed light on many of the ribonucleases and RNA modifying enzymes involved in RNA processing. However, even less is known about mRNA turnover in non-model bacteria including cyanobacteria. These photosynthetic Gram-negative bacteria play a major role in the global carbon cycle and are becoming an attractive platform for renewable chemical production. A detailed understanding of mRNA turnover dynamics is crucial for implementing effective metabolic engineering strategies to turn cyanobacteria into green chemical factories. Further, a better understanding of mRNA turnover will facilitate studies of basic physiology including photosynthesis and carbon fixation. A better understanding of mRNA turnover begins by identifying the relevant enzymes involved.

Bacterial mRNA decay is facilitated by exoribonucleases, endonucleases, and RNA modifying enzymes (**Figure 3.1**). In *E. coli*, mRNA degradation has been shown to proceed in the 5'-3' direction, despite the absence of a specific 5'-3' exonuclease in its genome. The directionality of degradation is initiated through a series of internal cleavages catalyzed by the single-stranded endonucleases, RNase E (Misra et al. 1979; Mackie 2013; Apirion 1973). Cleavage by RNase E is inhibited by secondary structure and by the presence of a 5'-triphosphate. Thus, removal of the inhibitory triphosphate by RppH, a NUDIX hydrolase, enhances the subsequent rate of mRNA degradation (Celesnik et al. 2007; Deana et al. 2008). Similarly, the double-stranded endonuclease,

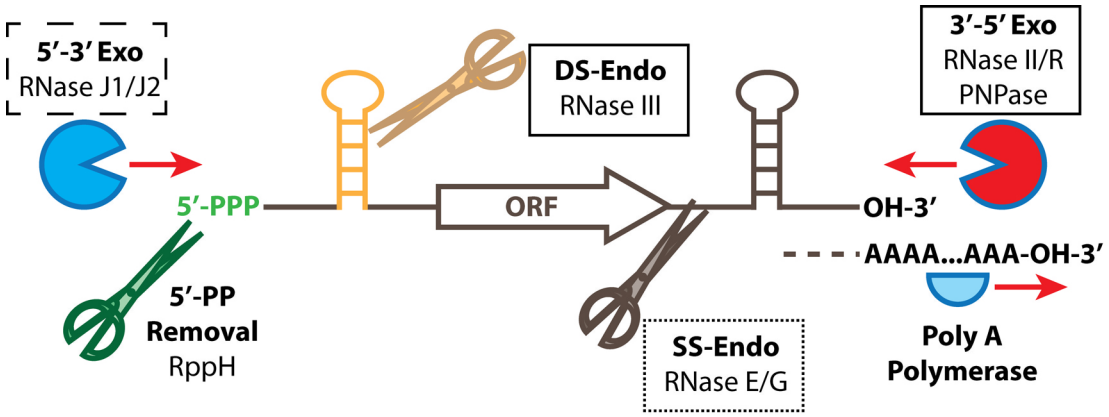


Figure 3.1: Schematic diagram of ribonucleases and RNA processing enyzmes in bacteria. Exonucleases process RNA to oligonucleotides from the 3' and 5' directions. Endonucleases process single stranded (SS) or double stranded (DS) RNA molecules, releasing additional targets for processing by exonucleases. RNA modifying enzymes can alter the rates of RNA processing by additional ribonucleases. Enzymes found in cyanobacteria, *B. subtilis* and *E. coli* are boxed with solid line. Enzymes found in *B. subtilis* and cyanobacteria are boxed with a dashed line. Enzymes found in cyanobacteria and *E. coli* are boxed in a dotted line. RNA modifying enzymes found in *E. coli* but not cyanobacteria or *B. subtilis* are not boxed.

RNase III can remove secondary structure by recognizing and cleaving a subset of structured RNA hairpins (Court et al. 2013). RNase E is a component of the degradosome, a multi-enzyme complex that is responsible for the majority of mRNA turnover in *E. coli* and other bacteria (Py, Higgins, et al. 1996; Carpousis, Van Houwe, et al. 1994; Miczak et al. 1996). The C-terminal scaffold domain of RNase E recruits PNPase, enolase, and an ATP-dependent RNA helicase, RhlB to the degradosome (Vanzo et al. 1998). RNase G, a RNase E paralog that lacks the scaffold domain and is not recruited to the degradosome, appears to play a minor role in mRNA degradation compared to RNase E (Ow et al. 2003). Following primary cleavage by RNase E, three 3'-5' exonucleases, RNase II, RNase R, and PNPase process the majority of mRNA in *E. coli* (Carpousis, Luisi, et al. 2009). Poly(A) polymerase can also increase the rate of mRNA decay by providing a template for PNPase (Mohanty et al. 2008). In contrast to *E. coli*, the majority of mRNAs in *B. subtilis* are processed by RNase J1/J2, a 5'-3' exonuclease (Condon 2010). Recently, a fourth class of RNase III (Mini-III) that lacks the double-stranded RNA binding motif (DSRM) and is involved in 23S

ribosomal RNA maturation was found in *B. subtilis* (Redko et al. 2008).

Relatively little is known about the components and mechanisms involved in mRNA turnover in cyanobacteria. Preliminary bioinformatic analysis revealed that the cyanobacterium *Synechocystis* sp. PCC 6803 (PCC 6803) contains homologs of *E. coli* and *B. subtilis* RNA degrading machinery including RNase E, RNase E/G, RNase J1/J2, RNase II/R (Condon and Putzer 2002) (**Figure 3.1**). In addition, more recent work identified the presence of the RNase III-like Mini-III in some cyanobacterial genomes (Redko et al. 2008). It has been long thought that that cyanobacteria do not contain a multicomponent mRNA degradosome complex based on the low conservation between the RNase E scaffold domains of *E. coli* and the inability to co-purify cyanobacterial degradosome components (Rott et al. 2003; Kaberdin et al. 1998). However, recent experiments indicate that the cyanobacterial RNase E homolog in PCC 6803 and *Anabaena* PCC 7120 (PCC 7210) interacts with PNPase through specific interactions with a conserved arginine rich peptide at the C-terminus of RNase E (Zhang, Deng, et al. 2014). Since cyanobacteria appear to have components similar to both *E. coli* and *B. subtilis*, and the homologs of the common RNases are often found as small gene families, the essential components of the cyanobacterial mRNA turnover machinery are not readily apparent. Therefore, in this study, we aimed to examine the RNases in commonly studied cyanobacteria and identify the major players in mRNA turnover in *Synechococcus* sp. PCC 7002.

3.2 Genetic and genomic analysis of RNases in model cyanobacteria

Genomic analysis of RNA processing enzymes in model cyanobacteria

Cyanobacteria are highly diverse organisms morphologically, ecologically, and genetically. Therefore, we first took inventory of annotated RNA modifying enzymes in the most commonly used cyanobacteria model organisms (**Table 1.1**). From primary sequence alignments and putative annotations, we were able to identify multiple homologs of *E. coli* RNase E/G, RNase III, RNase II/R, and PNPase as well as homologs of *B. subtilis* RNase J1/J2 (Mathy et al. 2007; Britton et al.

2007) and a recently described fourth class of RNase III (Mini-III) (Redko et al. 2008). A single protein with homology to the N-terminal catalytic domains of RNase E and RNase G was identified in each of the model cyanobacteria strains. RNase G resembles the N-terminal catalytic region of RNase E and does not contain the C-terminal scaffolding domain. The C-terminal domain of the cyanobacterial homologs was variable in length and exhibited low sequence homology to the RNase E scaffold domain. Thus, we classified them as RNase E/G homologs. In each model species, we identified only a single protein with homology to RNase J1 and RNase J2 (RNase J1/J2). This contrasts the model case in which *B. subtilis* contains multiple RNase J1/J2 enzymes (Even et al. 2005). The primary sequences of RNase II and RNase R are highly similar. Therefore the catalytic activity of putative sequences cannot be predicted from primary sequence alone. For this reason, we grouped the multiple cyanobacterial homologs of the RNase II/R family identified in the model organisms. Each cyanobacterial species contained at least one of each of these five classes of RNase, making it difficult to determine a priori which mechanisms would be dominant. Further, it is not known how these enzymes function together during mRNA degradation and which components are essential for the process. Therefore, we aimed to characterize each of the identified enzyme classes in *Synechococcus* sp. PCC 7002, a fast-growing strain that is being used as a platform for synthetic biology (Begemann et al. 2013) and renewable chemical production (Devroe et al. 2009).

Genetic analysis of ribonucleases in PCC 7002

In order to determine which of the identified RNase homologs are essential in PCC 7002, we attempted to replace each of the coding regions of the RNase E/G, RNase III, RNase J1/J2, RNase II/R, and PNPase with an antibiotic resistant cassette via homologous recombination and targeted gene deletion strategies as described in "Materials and Methods". Since many cyanobacteria including PCC 7002 (Murphy et al. 1990) contain multiple copies of the chromosome, deletion of essential genes often results in a heterozygous strain containing both mutant and WT alleles. However, it is possible that conditions permissible for full segregation of the mutant allele can be

Table 3.1: Ribonuclease homologs in commonly studied cyanobacteria

Genome	RNase E/G	RNase III	RNase J1/J2	RNase II/R	PNPase
Acaryochloris marina MBIC11017	AM1_1705	AM1_1717 AM1_5197*	AM1_3444	AM1_4216 AM1_E0095 AM1_D0003 AM1_E0127 AM1_B0417 AM1_4810 AM1_A0095	AM1_5920 AM1_4121
Arthrospira platensis NIES-39	NIES39_J05380	NIES39_Q02990 NIES39_N00330	NIES39_L00100	NIES39_D02610 NIES39_E01890	NIES39_R00610 NIES39_L00330 NIES39_O05010
Cyanobacterium sp. UCYN-A	UCYN_00380	UCYN_03680 UCYN_00750*	UCYN_00320	UCYN_07880 UCYN_00190 UCYN_09960	UCYN_04860 UCYN_03170
Cyanothece sp. BH68, ATCC 51142	cce_0277	cce_3135 cce_1800*	cce_0295	cce_5063 cce_5061 cce_0113 cce_3346 cce_3157	cce_1653 cce_0623 cce_3461
Gloeobacter violaceus PCC 7421	gvip208	gvip371 glr4180*	glr1019	gll1616 gll1448	gvip129 glr4144 gll3149
Microcystis aeruginosa NIES-843	MAE_15200	MAE_31570 MAE_60800 MAE_28410*	MAE_62670	MAE_23880 MAE_46880	MAE_53390
Nostoc sp. PCC 7120	alr4331	all4107 alr0280 alr1158*	all3678	all4450 alr1240	all0136 all4396
Prochlorococcus marinus pastoris CCMP 1986	PMM1501	PMM1603 PMM0949*	PMM1652	PMM0868 PMM0847	PMM1191
Synechococcus elongatus PCC 7942	Synpcc 7942_0878	Synpcc7942_1645 Synpcc7942_2120*	Synpcc 7942_1846	Synpcc7942_1911 Synpcc7942_1124 Synpcc7942_1546	Synpcc 7942_2440
Synechococcus sp. PCC 7002	PCC7002_A0788	PCC7002_A0061 PCC7002_A2542 PCC7002_A0384*	PCC7002_A1273	PCC7002_A1543 PCC7002_A0574	PCC7002_A1066
Synechocystis sp. PCC 6803	slr1129	slr1646 slr0346 slr0954*	slr0551	sll1910 sll1290	slr1984 sll1043
Thermo-synechococcus elongatus BP-1	tlr0597	tlr0428*	tlr1220	tlr1873 tlr0652	tll2193
Trichodesmium erythraeum IMS101	Tery_0483	Tery_2839 Tery_2144 Tery_1105 Tery_2312*	Tery_0589	Tery_3730 Tery_1511	Tery_1541 Tery_2502 Tery_4663
*Mini-III					

Table 3.2: Summary of PCC 7002 RNase deletion strains

Strain	Enzyme Class	Segregated	Phenotype	Reference
WT PCC 7002	n/a	n/a	n/a	PCC
Single Mutants				
ΔA0788	RNase E/G	No	n/a[a]	This work
ΔA0061	RNase III	Yes	-	This work
ΔA2542	RNase III	Yes	-	This work
ΔA0384	RNase III	Yes	-	This work
ΔA1273	RNase J1/J2	No	n/a[a]	This work
ΔA0574	RNase II/R	No	n/a[a]	This work
ΔA1543	RNase II/R	Yes	+	This work
ΔA1066	PNPase	No	n/a[a]	This work
Double Mutants				
ΔA0061/A2542	RNase III	Yes	-	This work
ΔA0061/A0384	RNase III	Yes	+	This work
ΔA2542/A0384	RNase III	Yes	-	This work
Triple Mutants				
ΔA0061/A2542/A0384	RNase III	Yes	+	This work

n/a, not applicable
Growth defect observed (+) or not (-).
[a] The phenotypes of partially segregated mutants were not analyzed.

identified. In this study, we are only reporting whether we are able to fully segregate the mutant allele in standard laboratory conditions (**Table 3.2**). Full segregation indicates that the gene is not essential for survival, whereas the inability to segregate the mutant indicates that it is required for growth in the conditions tested. We were not able to isolate fully segregated mutants of RNase E/G (ΔA0788), RNase J1/J2 (ΔA1273), RNase II/R (ΔA0574), or PNPase (ΔA1066) (**Figure 3.2**), indicating that these genes are required under laboratory conditions. Due to the inherent instability of partially segregated mutants and difficulty in accurately assessing the physiology of these strains, we focused our efforts in characterizing the physiology of mutants that could be fully segregated (**Table 3.2**), RNase III and RNase II/R.

Diversity of RNase III homologs in cyanobacteria

The well-studied *E. coli* RNase III (Rnc) contains an N-terminal RNase III catalytic domain containing the RNase III signature motif (ERLEFLGDS) and a C-terminal dsRNA binding motif (DSRM). *E. coli* contains a single RNase III homolog that is not essential for growth under standard

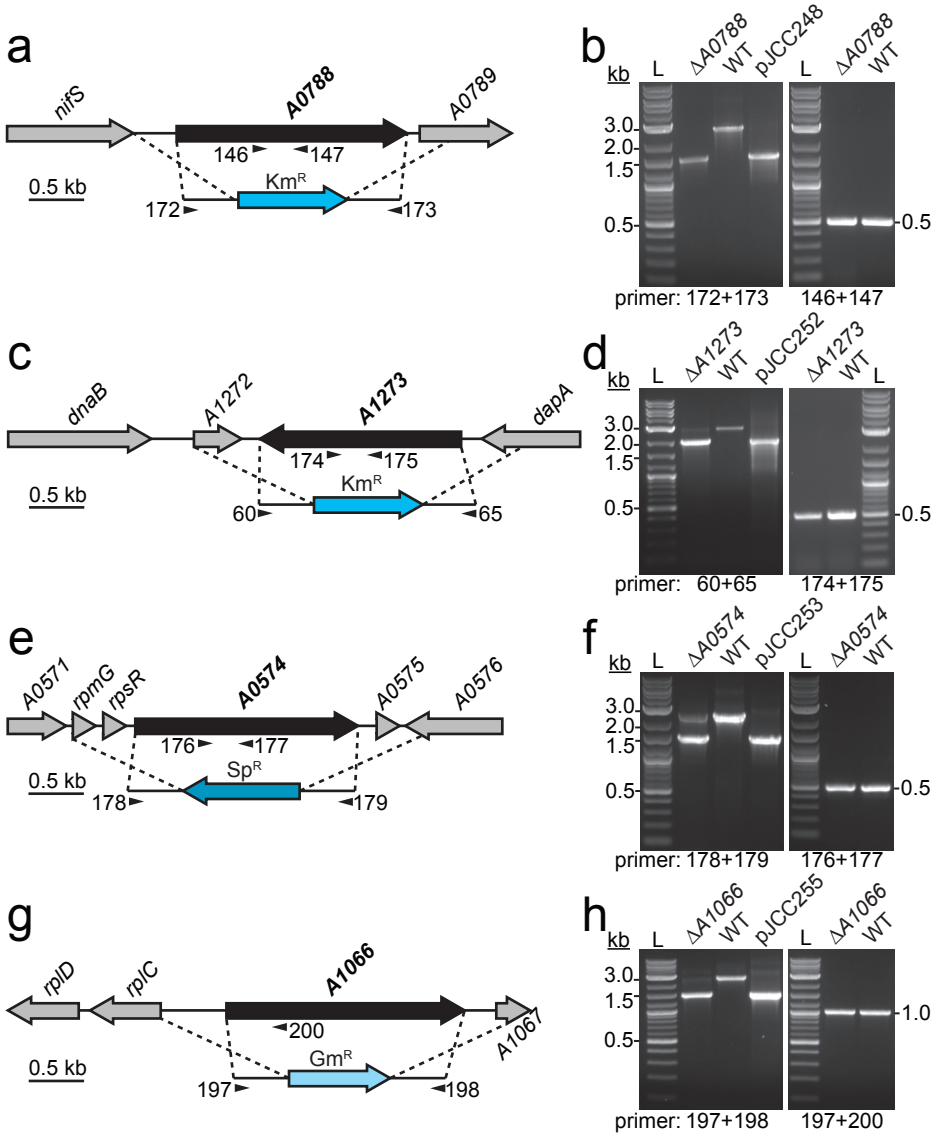


Figure 3.2: Inability to segregate deletion of A0788, A1273, A0574, and A1066 in PCC 7002. Targeted gene deletion via homologous recombination was used to replace the entire ORF of A0788, A1273, A0574, and A1066 with an antibiotic resistance cassette (a, c, e, and g). Insertion of the cassette into each of these genes was confirmed by PCR using primer pairs flanking the insert (Left gel panel in b, d, f, and h). WT genomic DNA was used as the negative control and the plasmid (pJCC#) used to generate the strain was used as a positive size control. PCR using primers that amplify the WT gene (Right gel panel in b, d, f, and h) indicates that these strains are heterozygous at the locus of interest. Arrows and numbers below antibiotic resistance cassette and gene in a, c, e, and g indicate approximate regions of flanking and gene specific primers respectively. For the gene specific reaction shown in g and h, a flanking primer and gene specific primer combination was used to amplify a 1 kb band that should not be present in a fully segregated mutant. Plasmids used as the positive control are listed in **Table 3.5**. All primers used in segregation analysis are listed in **Table 3.8**.

conditions; mutants are non-motile and have a temperature sensitivity at 45°C (Apirion and Lassar 1978; Apirion and Watson 1975). In contrast, *B. subtilis* contains an RNase III that is essential for growth (Durand et al. 2012) and a short, RNase III-like homolog lacking the C-terminal DSRM domain (YazC, Mini-III) (Redko et al. 2008). We were able to identify homologs of RNase III and Mini-III in the genomes of cyanobacteria representing all of the diverse morphological subgroups (Shih et al. 2013) (**Table 3.3**). Typically, a single Mini-III homolog was accompanied by one more RNase III homologs, however we identified several genomes that completely lacked either the full-length RNase III or the Mini-III (**Table 3.3**). PCC 7002 contains two RNase III homologs (A0061 and A2542) and a single Mini-III homolog (A0384). The functions of these enzymes in PCC 7002 are not known. Multiple sequence alignments of RNase III and Mini-III sequences from *E. coli*, *B. subtilis*, PCC 7002, *Anabaena* sp. PCC 7120 (PCC 7120) and a single RNase III homolog lacking the DSRM domain from *Leptolyngbya* sp. PCC 6406 (PCC 6406) were used to build a phylogenic tree to compare the relationships between the proteins (**Figure 3.3a**). The RNase III and Mini-III homologs could be resolved into distinct clades. Within the RNase III clade, A2542 from PCC 7002 branches more closely to Rnc than the other homolog, A0061. Alr0280 from PCC 7120, has been shown to be active in cleaving dsRNA substrates (Gao, Gong, et al. 2013) and clustered with Rnc and A2542. A0061 clusters with Alr2542, which was shown not to have ribonuclease activity when incubated with the same substrate as Alr0280 (Gao, Gong, et al. 2013). Interestingly, the RNase III-like protein from PCC 6406 (Lyn2207) that resembles a Mini-III since it lacks the DSRM domain branched within the full-length RNase III clade. We therefore attempted to determine whether there were other features that distinguished the canonical RNase III domain from the Mini-III RNase III domain. Sequence alignments revealed differences between the RNase III signature motif. All of the proteins that branched more closely with Rnc, including the PCC 6406 truncated protein exhibited an *E. coli*-like signature motif (ERLEFLGDA), whereas all of the Mini-III proteins contained a unique motif (AALAYLGDA). Based on a sequence conservation logo generated from multiple sequence alignments from 65 and 64 RNase III and Mini-III sequences,

respectively, our results indicate that classification based on the signature motif might be more robust than the presence/absence of the DSRM (**Figure 3.3a**).

Analysis of 66 cyanobacterial genomes representing the five morphological subclades revealed RNase III homologs with diverse domain structures and unknown function (**Figure 3.3b**). Most of the 66 genomes analyzed contained a single Mini-III homolog, whereas the number of RNase III homologs varied from zero to four per genome (**Figure 3.3c**). The atypical domain structures (**Figure 3.3b**) lacking the DSRM domain, containing tandem RNase III domains, or with additional domain extensions were found in organisms containing at least three RNase III homologs.

Table 3.3: Analysis of RNase III-like proteins in diverse cyanobacterial genomes. Model organisms are highlighted in bold.

Genome	#RNaseIII, #MiniIII	RNase III Locus Tag	Mini-III Locus Tag	RNase III Signature Motif	Mini-III Signature Motif
***Acaryochloris marina* MBIC11017**	1,1	AM1_1717	AM1_5197	EQLEFVGDS	AALAYIGDA
Anabaena cylindrica PCC 7122	3,1	Anacy_4655 Anacy_0811 Anacy_3962	Anacy_3641	ERLEFLGDA EQLEFVGDA EQLEFLGDC	SALAYLGDS
Anabaena sp. PCC 7108	2,1	Ana7108_1646 Ana7108_0695	Ana7108_2529	ERLEFLGDA EQLEFVGDA EQLEFLGDC	SALAYLGDA
***Anabaena* sp. PCC 7120**	2,1	all4107 alr0290	alr1158	EQLEFVGDA ERLEFLGDA	AALAYVGDA
Anabaena variabilis ATCC 29413	2,1	Ava_3081 Ava_0797	Ava_4412	ERLEFLGDA EQLEFVGDA	AALAYVGDA
***Arthrospira platensis* NIES-39**	2,0	NIES38_Q02990 NIES39_N00330		EQLEFVGDS ELLEFLGDS	
Calothrix sp. PCC 6303	2,1	Cal6303_1320 Cal6303_1680	Cal6303_1532	ERLEFLGDA EQLEFVGDA	KSLAYLGDA
Calothrix sp. PCC 7103	2,1	Cal7103_26520 Cal7103_79520	Cal7103_60540	EQLEFVGDA ERLEFLGDA	TALAYLGDA
Calothrix sp. PCC 7507	2,1	Cal7503_4153 Cal7507_3430	Cal7507_1114	ERLEFLGDA EQLEFVGDA	TALAYMGDA
Chroococcidiopsis sp. PCC 6712	2,1	Chr6712_2695 Chr6712_3177	Chr6712_4648	ERLEFLGDA EQLEFVGDA	VALAYIGDA
Chroococcidiopsis thermalis PCC 7203	2,1	Chro_4112 Chro_5889	Chro_4036	EQLEFVGDA EWLALLGDT	MALAYLGDA
Crocosphaera watsonii WH 8501	2,1	CWat_45690 Cwat_17180	CWat_7740	QQLEFVGDS QRLESLGEN	AFLAYIGDA
Cyanobacterium PCC 7702	2,1	Chl7702_0536 Chl7702_2265	Chl7702_2383	ERLEFLGDA EHLEFVGDA	TTLAYLGDA

Cyanobacterium sp. UCYN-A	1,1	UCYN_03680	UCYN_00750	QQLEFVGDS QRLESLGEN	AFLAYIGDA
***Cyanothece* sp. BH68, ATCC 51142**	1,1	cce_3135	cce_1800	QQLEFVGDS QRLESLGEN	AFLAYIGDA
Cyanothece sp. PCC 7424	2,1	PCC7424_3959 PCC7424_2664	PCC7424_3194	ERLEFLGDA QQLEFVGDA	ASLAYIGDA
Cyanothece sp. PCC 7425	1,1	Cyan7425_0389	Cyan7425_1940	EQLEFFGDA	AALAYLGDA
Cyanothece sp. PCC 7822	2,1	Cyan7822_0327 Cyan7822_1001	Cyan7822_0774	ERLEFLGDA QQLEFVGDA	ASLAYIGDA
Cyanothece sp. PCC 8801	3,1	PCC8801_3303 PCC8803_2820 PCC8801_0645	PCC8801_2425	ERLEFLGDA QQLEFLGDS ERLEYLGET	AFLAYLGDA
Dactylococcopsis salina PCC 8305	2,1	Dasca_1083 Dasca_3634	Dacsa_2494	ERLEFLGDA QQLEFVGDA	AALAYIGDA
Fischerella sp. PCC 9339	2,1	PCC9339_01064 PCC9339_03154	PCC9339_04090	ERLEFLGDA EHLEFVGDA	VALAYLGDA
Fischerella sp. PCC 9431	2,1	Fis9431_3759 Fis9431_1534	Fis9431_4808	EHLEFBGDA ERLEFLGDA	VALAYLGDA
Fischerella sp. PCC 9605	2,1	Fis9605_02642 Fis9605_04201	Fis9605_03180	ERLEFLGDA EHLEFVGDA	TALAYLGDA
Geitlerinema sp. PCC 7105	2,0	Gei7105_1223 Gei7105_5151		EQLEFIGDA ERLEFLGDS	
Geitlerinema sp. PCC 7407	2,1	GEI7407_1007 GEI7407_0526	GEI7407_3792	ERLEFLGDA EQLEFMGDA	VALAYLGDA
Gloeobacter violaceus PCC 7421	1,1	gvip371	glr4180	DRLEFLGDE	
Leptolyngbya boryana PCC 6306	1,1	Lepbo_1185	Lepbo_4252	EQLEFVGDA	AAFAYLGDA
Leptolyngbya sp. PCC 6406	3,1	LEP6406_2207 LEP6406_5336 LEP6406_1169	LEP6406_3537	ERLEFIGDG ERLEFLGDA EQLEFLGDA	QALAYIGDA
Leptolyngbya sp. PCC 7375	3,1	Lepto7375_7811 Lepto7375_5336 Lepto7375_3393	Lepto7375_2718	EIFRTLGDA ERLEFLGDA EVLEFLGDA	IALAYIGDA
Leptolyngbya sp. PCC 7376	2,1	Lepto7376_1012 Lepto7376_0760	Lepto7376_3431	EVLEFLGDA ERLEFLGDA	TALAYVGDA
Mastigocladopsis repens PCC 10914	2,1	Mas10914_2231 Mas10914_2728	Mas10914_3118	EHLEFVGDA ERLEFLGDA	TALAYLGDA
Microcoleus vaginatus FGP-2	3,1	Micva_0740 Micava_1301 Micava_0189	Micva_1151	EQLEFVGDA RLEFLGDA KRLSLLGGA	AAWAYLGDA
***Microcystis aeruginosa* NIES-843**	2,1	MAE_31570 MAE_60800	MAE_28410	ERLEFLGDA EQLEFVGDS	ASLAYLGDA
Nostoc azollae 0708	2,1	Aazo_4580 Aazo_3201	Aazo_2284	ERLEFLGDA EQLEFVGDA	SALAYLGDA
Nostoc punctiforme PCC 73102	2,1	Npun_R1331 Npun_F1233	Npun_R1191	ERLEFLGDA EQLEFVGDA	TALAYLGDA

Nostoc sp. PCC 7107	3,1	Nos7107_4879 Nos7107_2456 Nos7107_0948	Nos7107_3662	ERLEFLGDA EVLEFFGDS EQLEFVGDA	AALAYIGDA
Nostoc sp. PCC 7524	3,1	Nos7524_1070 Nos7524_4545 Nos7524_1546	Nos7524_3113	ERLEFLGDA EVLEFFGDS EQLEFVGDA	AALAYVGDA
Oscillatoria nigro-viridis PCC 7112	3,1	Osc7112_2758 Osc7112_2480 Osc_7112_2759	Osc7112_2944	ERLEFLGDA EQLEFVGDA QLLEFLGDA	AAWAYLGDA
Pleurocapsa sp. PCC 7319	3,1	Pleur7313_05633 Pleur7313_00733 Pleur7313_05876	Pleur7313_05600	ERLEFLGDA EQLEFVGDS QRLEFLGKA	IALAYIGDA
Pleurocapsa sp. PCC 7327	2,1	Ple7327_4215 Ple7327_0296	Ple7327_2958	ERLEFLGDS QQLEFVGDA	IALAYIGDA
Prochlorococcus marinus AS9601	1,1	A9601_18131	A9601_09111	EKLEFFGDA	IQLAWLGDS
Prochlorococcus marinus marinus CCMP 1375	1,1	Pro1762	Pro0747	ERLEFQGDA	LQLAWLGDA
Prochlorococcus marinus MIT 9211	1,1	P9211_17271	P9211_06951	ERLEFLGDA	LQLAWLGDA
Prochlorococcus marinus MIT 9215	1,1	P9215_18771	P9215_09411	EKLEFFGDA	IQLAWLGDS
Prochlorococcus marinus MIT 9301	1,1	P9301_17961	P9301_09091	EKLEFFGDA	IQLAWLGDS
Prochlorococcus marinus MIT 9303	1,1	P9303_26001	P9303_15791	ERLEFLGDA	LQLAWLGDA
Prochlorococcus marinus MIT 9312	1,1	PMT9312_1696	PMT9312_0850	EKLEFFGDA	IQLAWLGDS
Prochlorococcus marinus MIT 9313	1,1	PMT1949	PMT0651	ERLEFLGDA	LQLAWLGDA
Prochlorococcus marinus MIT 9515	1,1	P9515_1791	P9515_10321	EKLEFFGDA	IQLAWLGDS
Prochlorococcus marinus NATL1A	1,1	NATL1_20541	NATL1_09301	ENLEFLGDA	LQLAWLGDA
Prochlorococcus marinus NATL2A	1,1	PMN2A_1179	PMN2A_0261	ENLEFLGDA	LQLAWLGDA
Prochlorococcus marinus pastoris CCMP 1986	1,1	PMM1603	PMM0949	EKLEFFGDA	IQLAWLGDS
Prochlorococcus sp. CC931	1,1	sync_0200	sync_1634	EQLEFLGDA	LQLAWLGDA
Prochlorococcus sp. CC9605	1,1	Syncc9605_0149	Syncc9605_1154	ERLEFLGDA	LQLAWIGDA
Prochlorococcus sp. CC9902	1,1	Syncc9902_0179	Syncc9902_1305	ERLEFLGDA	LQLAWIGDA
Prochlorococcus sp. WH 7803	1,1	WH7803_0204	WH7803_1063	EQLEFLGDA	LQLAWLGDA
Prochlorothrix hollandica PCC 9006	2,1	Pro9006_0098 Pro9006_0103	Pro9006_2926	EQLEFVGDA ERLEFGDA	VLLAYVGDA

Pseudanabaena sp. PCC 6802	2,1	Pse6802_5090 Pse6802_2378	Pse6802_4789	QRLEFLGDC EKLEFFGDS	ASLAYLGDA
Pseudanabaena sp. PCC 7367	2,1	Pse7367_2482 Pse7367_1487	Pse7367_2852	QRLEFLGDA EKLEFFGDS	AALAYIGDA
Spirulina major PCC 6313	2,1	Spi6313_3729 Spi6313_2493	Spi6313_1300	ERLEFLGDA EQLEFVGDA	TALAYIGDG
Spirulina subsalsa PCC 9445	4,1	Spi9445_2508 Spi9445_1629 Spi9445_4283 Spi9445_1201	Spi9445_3328	ERLEFLGDA EQLEFVGDA ERLVVLGGM ERLEFLGDT	LALAYLGDA
***Synechococcus elongatus* PCC 7942**	1,1	Synpcc7942_1645	Synpcc7942_2120	DRLEFLGDA	QALAYLGDA
***Synechococcus* sp. PCC 7002**	2,1	PCC7002_A0061 PCC7002_A2542	PCC7002_A0384	EQLEFIGDA ERLEFLGDA	ASLAYVGDA
***Synechocystis* sp. PCC 6803**	2,1	slr1646 slr0346	slr0954	DRLEFLGDA QQLEFVGDA	VALAYLGDA
***Thermosynechococcus elongatus* BP-1**	0,1		tlr0428		AALAYFGDA
***Trichodesmium erythraeum* IMS101**	3,1	Tery_2839 Tery_2144 Tery_1105	Tery_2312	QGLANLGYV ESLEFIGDA EKLEFFGDA	SAWAYLGDA

Genetic deletion of RNase III and Mini-III homologs in PCC 7002

RNase III is essential in *B. subtilis*, but not in *E. coli*. Since the function of RNase III and Mini-III homologs in PCC 7002 are not known, we aimed to determine whether they could be genetically deleted individually and in combination. Using targeted gene deletion, we were able to replace the open reading frame of each of the RNase III (A0061 and A2542) homologs and the Mini-III homolog (A0384) (**Figure 3.4**). Insertion of the antibiotic resistances cassette into the locus of interest was checked with PCR using flanking primers (**Figure 3.3b, d, f**). Full segregation of the single mutants (ΔA0061, ΔA2542, and ΔA0384), double mutants (ΔA0061/A2542, ΔA0061/A0384, and ΔA2542/A0384), and triple mutant (ΔA0061/A2542/0A384) was confirmed by PCR using gene specific primers (**Figure 3.4g and 3.2**). These results indicate that RNase III activity is dispensable in PCC 7002.

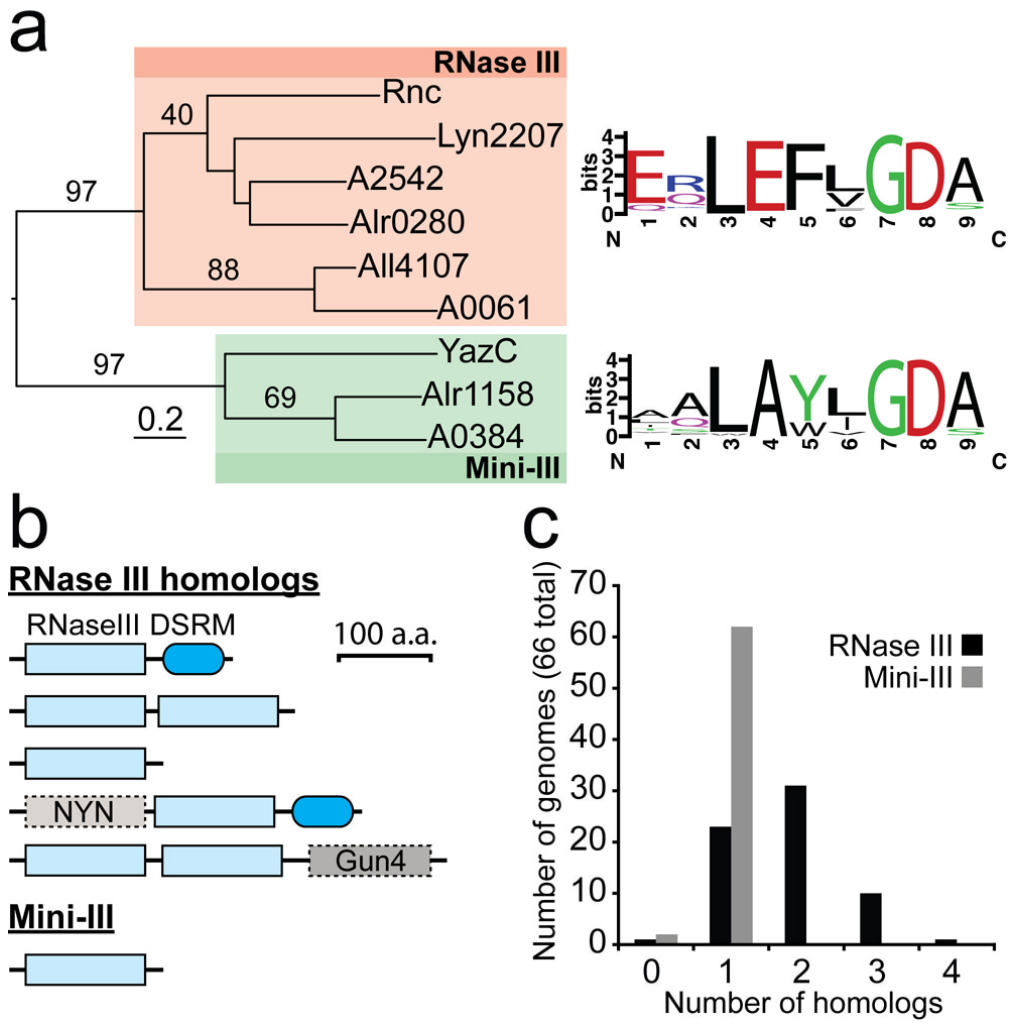


Figure 3.3: Diversity and distribution of RNase III homologs in cyanobacteria. (a) Phylogenic relationship between RNase III and Mini-III homologs in *Anabaena* sp. PCC 7120 (Alr0280, Alr4107, and Alr1158), *Leptolyngbya* sp. PCC 6406 (Lyn2207), PCC 7002 (A2542, A0061, and A0384), *E. coli* (Rnc), and *B. subtilis* (YazC). The RNase III homologs all contained the typical domain structure with a single, N-terminal RNase III domain and a C-terminal DSRM domain with the exception of Lyn2207, containing only a single RNase III domain. Conservation logos (weblogo.berkeley.edu) corresponding to the RNase III signature motif generated from multiple sequence alignments of RNase III and Mini-III homologs in 65 and 64 cyanobacterial genomes, respectively. (b) Domain architecture of RNase III and Mini-III homologs in cyanobacteria. (c) Number of RNase III and Mini-III homologs in 66 diverse cyanobacterial genomes. RNase III homologs in each analyzed genome are listed in **Table 3.3**.

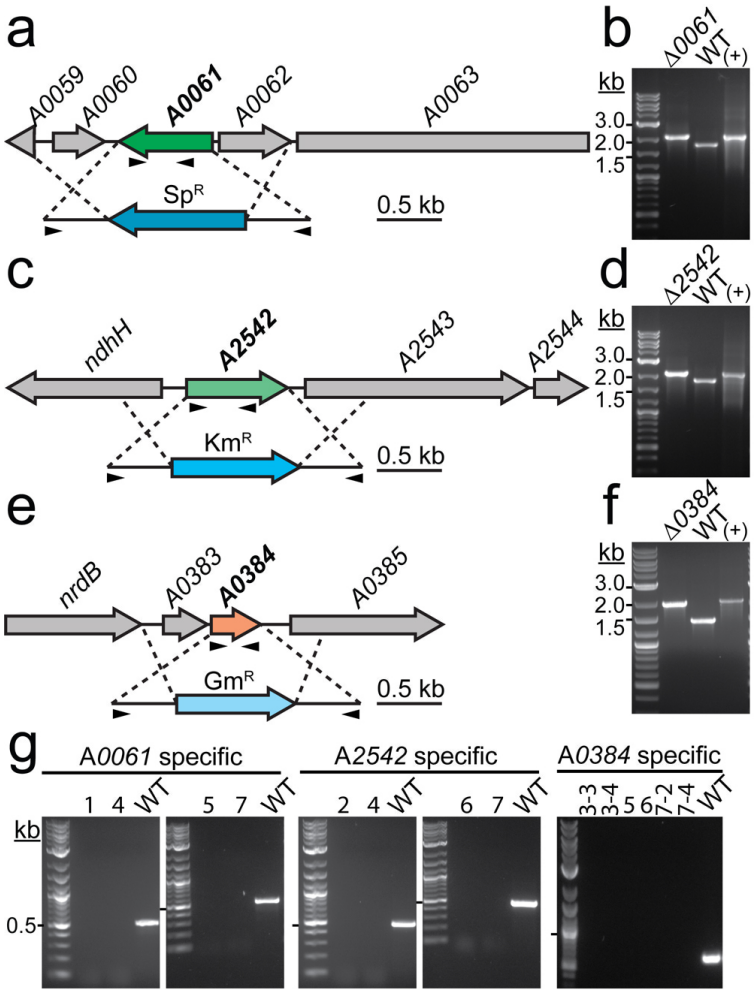


Figure 3.4: Genetic deletion and segregation analysis of RNase III homologs in PCC 7002. Targeted gene replacement via homologous recombination was used to replace the open reading frames of (a) A0061, (c) A2542, and (e) A0384 with three different antibiotic resistance cassettes as described in Materials and Methods. Insertion of the antibiotic resistance cassette into the correct genetic locus was confirmed by PCR using primers flanking the WT loci (b, d, and f). Arrows in (a), (c), and (d) below the gene and insertion cassette indicate approximate location of gene specific and flanking primers, respectively. The plasmid used to transform the mutant was used as a positive size control (+) and WT genomic DNA was used as a negative size control. WT sized bands were not observed in the mutant strains, indicating that the regions of interest are homozygous for the mutation. (g) Gene specific primers were used to confirm segregation of the insert by PCR. In this case, WT genomic DNA served as the positive control. Lanes: 1, ΔA0061; 2, ΔA2542; 3, ΔA0384; 4, ΔA0061/A2542; 5, ΔA0061/A0384; 6, ΔA2542/A0384; 7, ΔA0061/A2542/A0384. The 0.5 kb size band is indicated with a hash mark. Oligonucleotides used in cloning and segregation analysis are listed in **Table 3.8**.

Physiology of RNase III deletion mutants in PCC 7002

To determine the effect of each RNase III and Mini-III deletion on the physiology of PCC 7002, we compared the growth of each strain on solid medium and in liquid culture (**Figure 3.5**). Growth of the single mutants was not different than the WT on solid (**Figure 3.5a**) or liquid medium (**Figure 3.5b**). In contrast, the ΔA0061/A0384 (**Figure 3.5c**) and ΔA0061/A2542/A0384 (**Figure 3.5d**) strains exhibited slightly reduced growth compared to the WT. In liquid medium (**Figure 3.5c-d**), the reduced growth was more apparent during the early growth phase, whereas on solid medium (**Figure 3.5A**) growth was not observed in the most dilute patches. We also compared growth of ΔA0384 strains in an atmosphere containing ambient (∼0.04%) or enriched (0.4%) CO_2 concentrations (**Figure 3.6**). Growth of all strains was significantly improved in the enriched CO_2 atmosphere. However, the colony size of the ΔA0061/A0384 and ΔA0061/A2542/A0384 strains appeared smaller than the WT, indicating that the growth is still impacted under these conditions.

Genetic complementation of the RNase III triple mutant

Since the single mutants do not exhibit a growth phenotype, it is unlikely that the reduced fitness of the ΔA0061/A0384 and ΔA0061/A2542/A0384 strains is due to polar effects resulting from the gene deletions. To confirm that the reduced growth is directly related to the absence of the RNase III genes, we performed genetic complementation of the most severe mutant lacking all RNase III homologs (**Figure 3.7**). Analysis of double and triple RNase III mutants indicates that growth defects are only observed in strains lacking A0061 and A0384 (**Figure 3.5 and 3.6**). Thus, we hypothesized that insertion of either of these genes could restore wild-type growth ability to the triple-mutant. The A0384 gene is in an operon, so we chose to insert the native A0061 gene into the genome of the RNase III triple mutant by targeting it to the *glpK* pseudogene, which has previously been used as a neutral insertion site (Begemann et al. 2013). Full segregation of the A0061 gene in the *glpK* locus was confirmed by PCR (**Figure 3.7a**). We then compared the growth

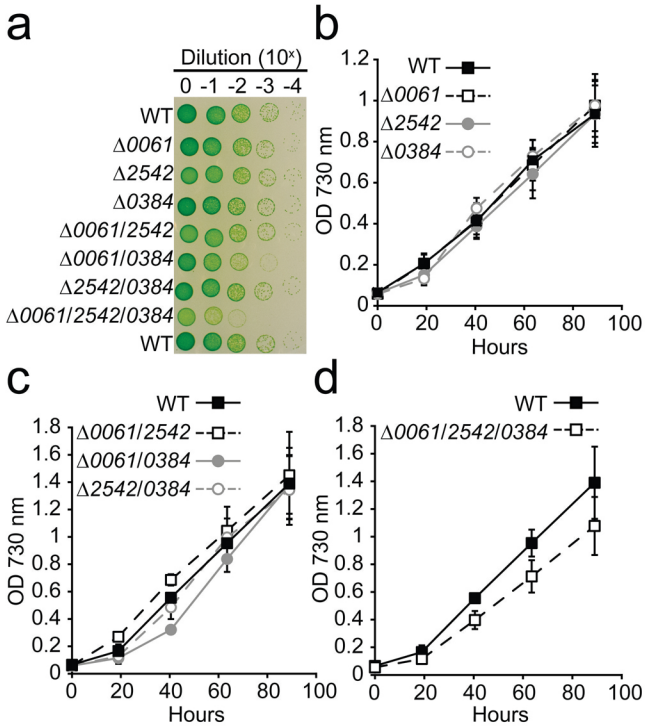


Figure 3.5: Growth of RNase III deletion mutants in liquid and on solid medium. (a) Cells suspended at an OD 730 nm = 0.05 were tenfold serially diluted four times in Media A+, spotted on solid medium, and grown for 65 h prior to being photographed. A representative of triplicate plates is shown. Liquid cultures of (b) single, (c) double, and (d) triple RNase III mutants were inoculated at an OD_{730nm} = 0.05 and grown in 12 well culture plates on a shaking platform (150 rpm). Error bars represent s.d. of three separate cultures. All growth experiments were carried out in ambient CO_2 at 38°C with a light intensity of 150 µmol photons m^{-2} s^{-1}.

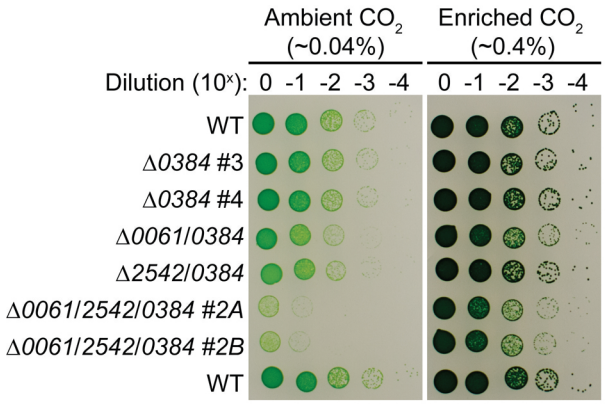


Figure 3.6: Growth of ΔA0384 strains in varying CO_2 concentrations. Strains harboring the ΔA0384 mutation were spotted on to solid medium in triplicate following serial dilutions of cells suspended at an initial OD 730 nm = 0.05. For the single and triple mutants, two biological replicates were spotted on each plate. Plates were grown in environmental chambers (38°C, 120 µmol photons m^{-2} s^{-1}) containing different CO_2 concentrations (ambient, ∼0.04%; enriched, ∼0.4%) for 86 h and then photographed.

of the resulting strain, ΔA0061+/A2542/A0384, with the WT and triple-mutant on solid (**Figure 3.7b**) and in liquid (**Figure 3.7c**) media. We calculated doubling times during exponential growth in liquid media for the WT (3.18 ± 0.21 h; n = 6), ΔA0061/A2542/A0384 (3.95 ± 0.28 h; n = 6), and ΔA0061+/A2542/A0384 (3.27 ± 0.11 h; n = 6) strains. The growth characteristics of the ΔA0061+/A2542/A0384 strain were equivalent to the WT strain on plates and in liquid culture, supporting the hypothesis that the growth defect of the ΔA0061/A2542/A0384 mutant strain is specifically due to the absence of RNase III activity.

Genetic deletion of A1543, an RNase R/II homolog, alters growth of PCC 7002

A blast search using *E. coli* RNase II and RNase R as the query resulted in the identification of two RNase II/R homologs (A0574 and A1543) in PCC 7002 (**Table 3.1**). While most genomes contained two homologs, some genomes contained more than three RNase II/R homologs. For instance, the genome of *Acaryochloris marina* MBIC11017 contained seven sequences homologous to RNase II/R **Table 3.1**. Moreover, several of these proteins exhibited unique domain architectures

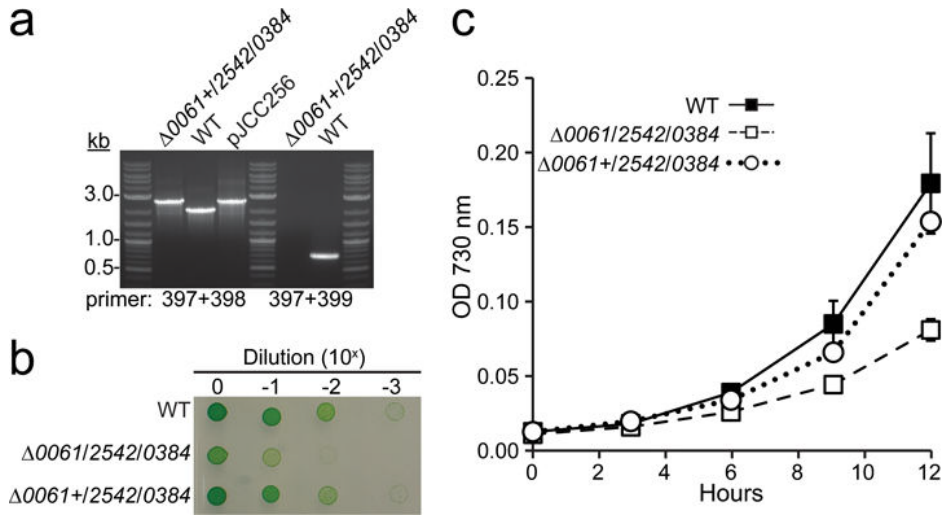


Figure 3.7: Genetic complementation of the RNase III triple mutant alleviates growth defect. The native A0061 gene was targeted to the *glpK* pseudogene as a neutral site in the ΔA0061/A2542/A0384 genetic background to generate the ΔA0061+/A2542/A0384 strain. (a) Full segregation of the A0061 gene in the *glpK* locus was confirmed by PCR using flanking (397+398) and gene specific primers (397+399). WT cells were used as the negative size control and the plasmid used in the transformation was used as the positive control (pJCC256). (b) Overnight cultures grown with bubbling in air (38°C, 200 µmol photons m^{-2} s^{-1}) were suspended to an $OD_{730\ nm}$ = 0.5 (1 cm cuvette). Tenfold serial dilutions were plated in quadruplicate on media A+ (7.5 µl/spot) and grown for 48 h (120 µmol photons m^{-2} s^{-1}) prior to being photographed. A representative image is shown. (c) For liquid growth, strains were diluted to an $OD_{730\ nm}$ = 0.01 (1 cm cuvette) and grown in test tubes bubbled for 12 h at a light intensity of 200 µmol photons m^{-2} s^{-1}. Error bars represent s.d. for three biological replicates.

such as the presence of an N-terminal DSRM domain in place of the typical S1 domain. Like PCC 7002, *Synechocystis* sp. PCC 6803 contains two RNase II/R homologs (Sll1290 and Sll1910). Recently, the function of Sll1290 was reported to exhibit RNase II, and not RNase R, activity (Matos et al. 2012). In addition, Sll1290 was shown to be essential in PCC 6803 (Rott et al. 2003). A phylogenic analysis comparing the RNase II and RNase R sequences from *E. coli* to the homologs in PCC 7002 and PCC 6803 indicates that A0574 is more closely related to Sll1290 (**Figure 3.8A**). We also found that A0574 is likely essential for growth in PCC 7002 since we were not able to obtain a fully segregated mutant (**Figure 3.2**). A1543 branched more closely to Sll1910 (**Figure 3.8a**), a protein previously reported to confer sensitivity to the carbonic anhydrase inhibitor, acetazolamide (Beuf et al. 1995; Bedu et al. 1990). In contrast to A0574, we were able to replace the A1543 gene with a Sp^R cassette (**Figure 3.8b**). Full segregation of the resulting strain was confirmed by PCR using gene specific and flanking primers (**Figure 4.10C**). The ΔA1543 strain exhibited reduced growth when cultured in liquid medium in 12 well plates (**Figure 3.9a and b**) and appeared slightly chlorotic. Comparison of absorption spectra indicates that the altered color of the ΔA1543 strain is due to reduced phycocyanin content compared to the WT (**Figure 3.9c**). Since increasing the CO_2 concentration improved the growth of the RNase III strains, we tested whether the addition of 10 mM HCO_{3^-} to the liquid medium would improve the growth of the ΔA1543 strain (**Figure 3.9b**). Although the addition of HCO_{3^-} resulted in increased clumping in the cultures, we did see an improvement in growth of the ΔA1543 strain. Genetic complementation of the ΔA1543 strain was accomplished by introducing the native gene into the *glpK* neutral site. The initial construct for complementation contained the native gene and ∼500 bp of upstream sequence. However, this construct failed to rescued the growth defect of the ΔA1543 strain (not shown). Thus, we generated an additional construct that included ∼700 bp upstream and ∼80 bp downstream intergenic regions thought to contain the native promoter and terminator sequences. Segregation of the native A1543 with additional upstream and downstream regions in the *glpK* locus was confirmed by PCR (**Figure 3.8**). Growth of the WT, ΔA1543, and ΔA1543+ strains were compared in liquid **Figure 3.8e** and

on solid media (**Figure 3.8F**). Full restoration of growth was observed in both growth regimes, supporting the hypothesis that the A1543 gene confers fitness to PCC 7002. However, the molecular mechanisms leading to the growth defect remain to be identified.

Discussion

Compared to *E. coli* and *B. subtilis*, very little is known about the components and mechanisms of the mRNA decay pathway in cyanobacteria. Genomic analysis shows that cyanobacteria share certain components of the mRNA degradation machinery with *E. coli*, and other components with *B. subtilis*. Moreover, a high level of ribonuclease diversity exists among the many cyanobacteria for which genomic information is available. However, the cellular functions and requirement of many of these ribonucleases have not yet been determined and it is clear from this study that predictions based purely on comparisons with *E. coli* and *B. subtilis* are not sufficient. This study provides a foundation to studying the function of ribonucleases and RNA modifying enzymes in the cyanobacterium *Synechococcus* sp. PCC 7002, a model platform for the photosynthetic production of renewable chemicals.

In *E. coli*, RNase E is an essential ssRNA endonuclease that initiates much of the RNA degradation in *E. coli* and is part of the mRNA degradosome. The *E. coli* RNase E contains an N-terminal catalytic domain and a C-terminal scaffolding domain that functions in recruiting degradosome components. The RNase E scaffolding domain confers enhanced cellular fitness (Leroy et al. 2002), but is not essential for viability in *E. coli* (Kido et al. 1996; Vanzo et al. 1998). A0788 and other cyanobacterial homologs typically have a C-terminal extension of the catalytic domain that does not resemble RNase E, suggesting that cyanobacteria do not contain a degradosome, though recent evidence suggests that the cyanobacterial RNase E interacts with PNPase (Zhang, Deng, et al. 2014). We were unable to generate a fully segregated mutant of the PCC 7002 RNase E homolog, A0788, which suggests that this protein is essential in cyanobacteria; whether a degradosome is essential for function in cyanobacteria remains to be determined.

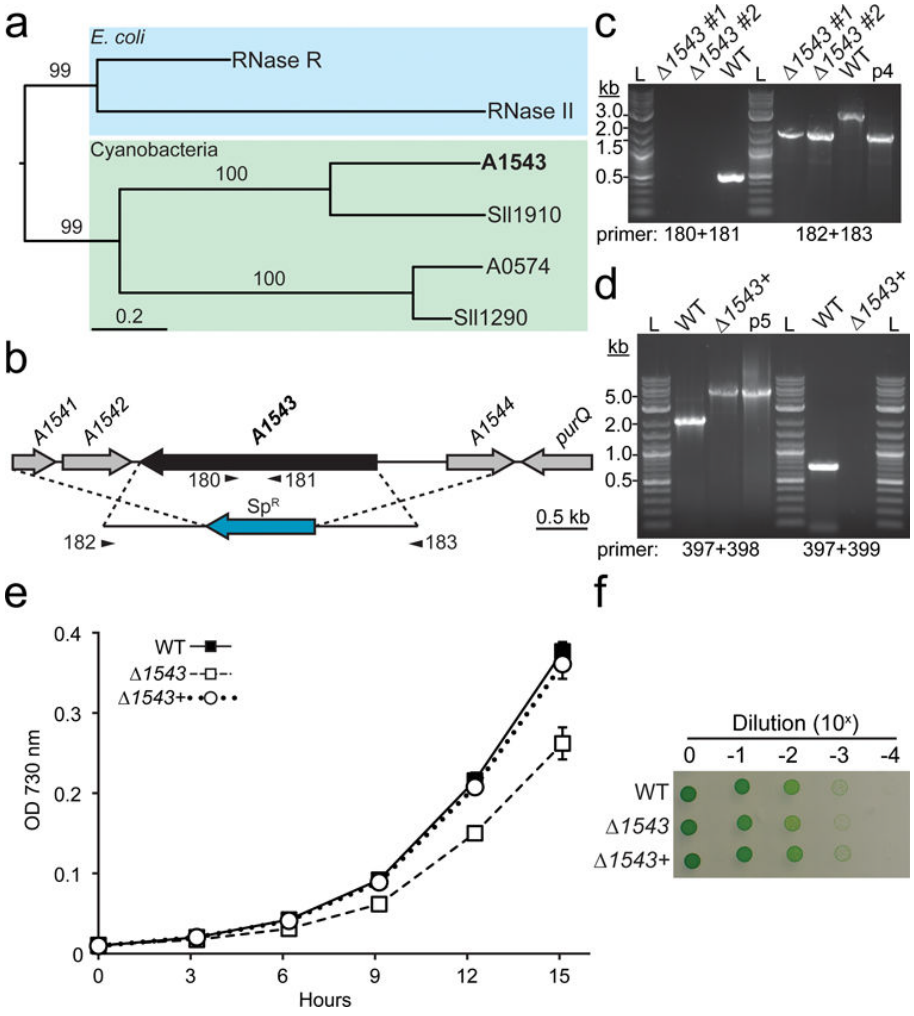


Figure 3.8: Generation and physiological characterization of a Δ1543 mutant in PCC 7002. (a) Phylogenic relationship between RNase II/R homologs in *E. coli* (RNase R and RNase II), PCC 7002 (A1543 and A0574), and *Synechocystis* sp. PCC 6803 (Sll1910 and Sll1290). Scale bar indicates the number of substitutions and values above branches indicate support from 100 bootstrap replicates. (b) The open reading from of A1543 was replaced with a Sp^R cassette via targeted homologous recombination. Gene specific and flanking primers used in segregation analysis are shown below the gene and the integration cassette, respectively. (c) Segregation of the mutation was confirmed using gene specific and flanking primers.Two independent ΔA1543 clones were analyzed (#1, and #2). WT genomic DNA was used as a control. The plasmid used to generate the Δ1543 mutant was used as a positive (+) control. (d) To genetically complement the ΔA1543 mutant strain, the native A1543 gene, including the ∼700 bp upstream and ∼80 bp downstream intergenic regions, was targeted to the *glpK* neutral site in the ΔA1543 background. Segregation of the *glpK* locus in the resulting strain (ΔA1543+) was confirmed using flanking and gene specific primers. Growth of the WT, ΔA1543, and ΔA1543+ strains in liquid bubbling culture (200 µmol photons m^{-2} s^{-1}) (e) and solid medium (120 µmol photons m^{-2} s^{-1}) in ambient CO_2 (f). Error bars represent s.d. for three biological replicates. Growth on solid medium was performed in quadruplicate; a representative image is shown.

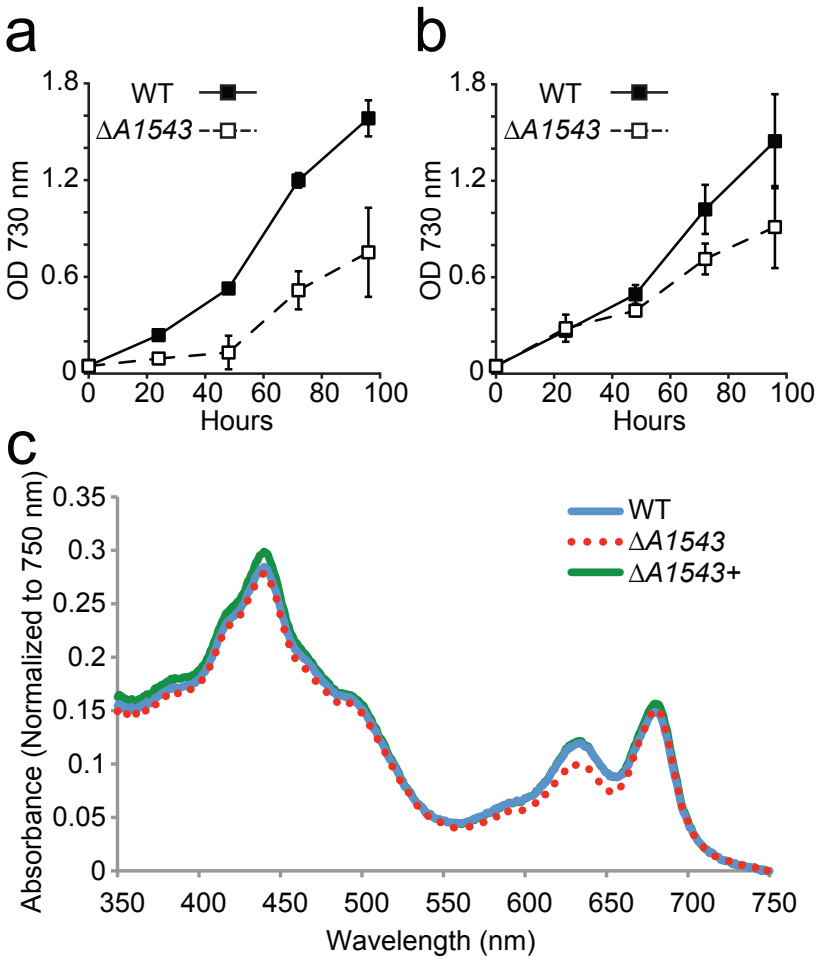


Figure 3.9: Growth of WT and ΔA1543 strains in liquid medium (12-well tissue culture plates) with shaking (150 rpm) in the absence (a) and presence (b) of 10 mM $NaHCO_3^-$. Error bars are s.d. of three cultures. (c) Absorption spectra of WT, ΔA1543, and genetically complemented ΔA1543+ strains following 24 h growth in liquid medium in culture tubes bubbled with air.

Based on previous work showing that RNase III is dispensable in *E. coli*, but required in *B. subtilis*, it could not be predicted whether RNase III activity is essential in cyanobacteria. In addition, most cyanobacterial species contain multiple homologs of RNase III, including Mini-III, a truncated version lacking the DSRM domain (**Figure 3.3**). The PCC 7002 genome encodes two RNase III homologs and a single Mini-III homolog. We were able to delete each of these genes in all single, double and triple combinations, indicating that none of these are essential for growth in laboratory conditions (**Figure 3.4**). Reduced growth of strains containing both ΔA0061 and ΔA0384 mutations could be due to redundant functions of the encoded proteins (**Figure 3.5-3.6**). An RNase III homolog in the heterocyst producing strain PCC 7120 that is closely related to A2542 in PCC 7002 was implicated in non-coding RNA mediated gene regulation (Gao, Gong, et al. 2013). Non-coding RNAs have been found to regulate many processes associated with photosynthesis in cyanobacteria (Sakurai et al. 2012; Eisenhut et al. 2012; Dühring et al. 2006). For example, (Sakurai et al. 2012) found that a cis-encoded antisense RNA can stabilize the *psbA2* mRNA, encoding the photosystem II D1 protein, by inhibiting processing by RNase E. Therefore, we were surprised that the ΔA2542 strain did not exhibit a growth phenotype, even when in combination with ΔA0061, which encodes the other full length RNase III homolog (**Figure 3.5 and 3.6**). However, we did observe a slight increase in photosystem II oxygen evolution activity in the ΔA0061/A2542 strain, but only at light intensities over 50x higher (8250 vs 150 μmol photons m^{-2} s^{-1}) than typical growth conditions (**Figure 3.10**).

Multiple RNase II/R homologs could be identified in the genomes of diverse cyanobacteria (**Table 3.1**). However, little is known about the functions of these enzymes in cyanobacteria. Between the two homologs in PCC 7002, we were able to fully segregate a ΔA1543 mutant, but the ΔA0574 strain could not be fully segregated. The ortholog of A0574, Sll1290 in PCC 6803 (67% identity) was recently shown to have RNase II activity (Matos et al. 2012); thus, it is possible that A0574 also shares this activity. A mutant of the A1543 ortholog of in PCC 6803, A1910, was isolated in a screen for resistance to the carbonic anhydrase inhibitor acetazolamide and renamed

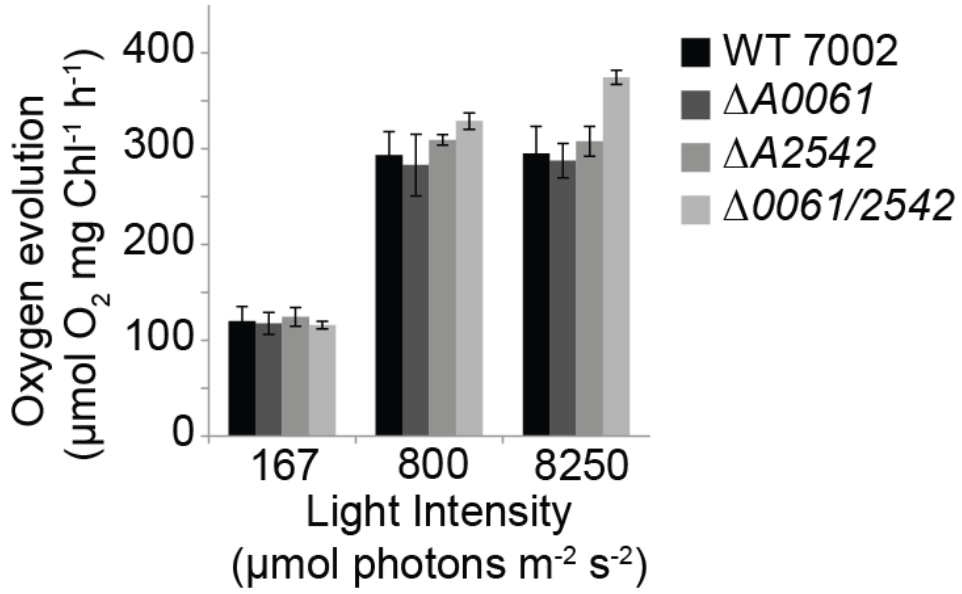


Figure 3.10: Oxygen evolution was measured as a function of light intensity on a Clark-type electrode in cell solutions suspended at a chlorophyll concentration of 5 µg/ml in A+ media containing 10 mM $NaHCO_3^-$.

Zam (Beuf et al. 1995; Bedu et al. 1990). However, the mechanism for the observed resistance to acetazolamide was not elucidated. There has been substantial work on a yeast homolog of A1543, Ssd1, a protein that is involved in translational repression through binding and sequestering of mRNAs into subcellular processing bodies (Li, Miles, et al. 2013), but it remains to be determined whether a similar function exists in cyanobacteria.

Conclusions

Ribonucleases play critical roles in post-transcriptional regulation of gene expression. However, our knowledge of mRNA turnover in photosynthetic microbes is limited. We find that genetic comparisons between *E. coli* and *B. subtilis*, models for mRNA turnover, is not sufficient to understand RNA metabolism in cyanobacteria. Using genomic analysis and genetic tools, we have defined the essential RNases in PCC 7002 and have laid the foundation for future investigations to elucidate the functional roles of these critical enzymes.

3.3 Materials and Methods

Bioinformatic analysis

Cyanobacterial ribonuclease were identified based on similarity to *E. coli* and *B. subtilis* sequences using BLASTP with an E-value threshold cutoff of 10^{-5} on the Department of Energy Joint Genome Institute Integrated Microbial Genomes (JGI-IMG) database. For RNase III, multiple sequence alignments were performed using MUSCLE on the EMBL-EBI web platform (McWilliam et al. 2013) with default parameters and ClustalW output format. The RNase III phylogenic tree was constructed with the maximum likelihood method and the default substitution method on Phylogeny.fr (Dereeper et al. 2008) using the "A la Carte" mode with the built-in curation "remove gap positions" setting. Branch support was performed using 100 bootstrap replicates and the tree was rooted to the midpoint. For the RNase R analysis, the "A la Carte" setting was used to generate multiple sequence alignments MUSCLE (v3.7), curation with Gblocks, and the phylogenic reconstruction with the maximum likelihood analysis and 100 bootstrap replicates. RNase III signature motifs were aligned with MUSCLE (v3.8) and the resulting alignments were used to generate the sequence conservation logos using Weblogo (v2.8.2) (Crooks et al. 2004). RNase III domain organization was analyzed using the sequence search function and the default E-value cutoff equal to 1.0 on the Pfam database (Punta et al. 2012). The cyanobacterial genomes used for analysis of the RNase III domain organization, signature motif, and frequency are listed in **Table 3.3**.

Growth and physiological measurements of strains

Strains were maintained on A+ media (Stevens, Patterson, et al. 1973) solidified with 1% (w/v) Bacto-Agar (Fisher). For growth of mutants, solid media was supplemented with antibiotics (Kan, 100 μg/ml; Spec, 100 μg/ml; Gent, 30 μg/ml). All growth and physiological assays were performed in media without the addition of antibiotics. For precultures used in solid and liquid growth experiments, glass culture tubes (2 x 15 cm) containing 20 ml media were inoculated with

fresh (< 1 week old) plate grown cells and bubbled with air supplemented with 0.4% CO_2 at a light intensity of 150 µmol photons m^{-2} s^{-1} in a temperature controlled chamber at 38°C. After approximately 3 days of growth, precultures were diluted to an optical density (OD) at 750 nm = 0.05 and transferred to 12 well culture plates (3 ml/well) for liquid growth assays or tenfold serially diluted in A+ media and spotted onto solid media (7 µl/spot). OD measurements were performed on a Tecan M2000 plate reader using Costar 96 well transparent plates and a volume of 200 µl. A light intensity of 150 µmol photons m^{-2} s^{-1} was used in all experiments unless indicated otherwise. Oxygen evolution was measured on a Clark-type electrode at 30°C at various light intensities achieved using neutral density filters.

Construction of mutants and segregation analysis

Mutants were generated by targeted gene deletion using homologous recombination as described (Clerico et al. 2007). Briefly, ∼500-1500 bp regions upstream and downstream of the gene of interest and an appropriate antibiotic resistance cassette were amplified by PCR using primers that contained 5' extensions with ∼15-20 bp homology with the adjacent fragment. These fragments were assembled using the Gibson Assembly method (Gibson et al. 2009) into a circular plasmid using a pUC19 backbone that was generated by digestion with XbaI. For transformations, 1 ml of three-day old liquid preculture was incubated for ∼12 hr with 1 µg plasmid in an environmental chamber with standard light condition and then plated on solid medium containing the appropriate antibiotic. After ∼4-5 days, colonies were isolated and patched onto solid medium containing antibiotic. To facilitate segregation of the inserted sequence and removal of the WT sequence, cells were passaged several times on media containing antibiotic. Segregation was analyzed by PCR with GoTaq Green Master Mix (Promega) using primers flanking the insert and with gene specific primers. Full segregation was only concluded if the WT size band was absent from the flanking and gene specific PCR reactions. A full list of primers and plasmids used to delete RNases can be found in **Table 3.5** and **Table 3.4**.

Oligo ID	Sequence (5'-3')	Description	Purpose	Reference
Δ*A0788* strain				
JC128	gctcggtacccggggatcctCGATATCGATCACGTCGC	*A0788*_upstream_F	KO/Segregation	This work
JC129	tgagcgtcagTCAAATTCCTCGATTTTTTTC	*A0788*_upstream_R	KO	This work
JC130	aggaatttgaCTGACGCTCAGTGGAACG	KmR_F	KO	This work
JC131	taattgggttTTCAGGTGGCACTTTTCG	KmR_R	KO	This work
JC132	gccacctgaaAACCCAATTACCGTTAACCTAAGG	*A0788*_downstream_F	KO	This work
JC133	tgcatgcctgcaggtcgactCGGACTTCAGCGACGGTG	*A0788*_downstream_R	KO/Segregation	This work
JC146	ACATGGATTCTCGCCATGAT	*A0788*_gene_specific_F	Segregation	This work
JC147	CGGCTTACTCGAAACCTCTG	*A0788*_gene_specific_R	Segregation	This work
JC172	TCATCTCACCCAGCCATACA	*A0788*_flanking_F	Segregation	This work
JC173	GGGCTTGTAAAACGACTTGC	*A0788*_flanking_R	Segregation	This work
Δ*A0061* strain				
JC42	gctcggtacccggggatcctTCTTTGCCCCATAAGCAG	*A0061*_upstream_F	KO/Segregation	This work
JC43	aattcactggAAAAGCGCCCGGAGATCT	*A0061*_upstream_R	KO	This work
JC44	gggcgcttttCCAGTGAATTCGAGCTCGG	SpR_F	KO	This work
JC45	tagtatttccCAAGCTTGCATGCCTGCA	SpR_R	KO	This work
JC46	tgcaagcttgGGAAATACTAAGATAGCAAAAGTAGGGAACTTTGC	*A0061*_downstream_F	KO	This work
JC47	tgcatgcctgcaggtcgactGCAGCGGCGATCCACGTC	*A0061*_downstream_R	KO/Segregation	This work
JC140	TCCCCCGACAAAAATTATGA	*A0061*_gene_specific_F	Segregation	This work
JC141	AGCAAAGTATCCCCCAGCTT	*A0061*_gene_specific_R	Segregation	This work
Δ*A2542* strain				
JC48	gctcggtacccggggatcctATGATCACGCGGATATATTG	*A2542*_upstream_F	KO/Segregation	This work
JC49	agaccccgtaGGTTAAGCGAAACTTTATAATG	*A2542*_upstream_R	KO	This work
JC50	tcgcttaaccTACGGGGTCTGACGCTCA	KmR_F	KO	This work
JC51	tctcgcccaaGATTTTGCCGATTTCGGC	KmR_R	KO	This work
JC52	cggcaaaatcTTGGGCGAGAAAATGGCT	*A2542*_downstream_F	KO	This work
JC53	tgcatgcctgcaggtcgactAAAGGCAATCTCGTTATCAAAAG	*A2542*_downstream_R	KO/Segregation	This work
JC142	GCGATCGCCTCTTCTGTTAC	*A2542*_gene_specific_F	Segregation	This work
JC143	GAATTTCACCGTGGTTTGCT	*A2542*_gene_specific_R	Segregation	This work
Δ*A0384* strain				
JC134	gctcggtacccggggatcctCAGCAAAAAATCTGTAGTTTG	*A0384*_upstream_F	KO/Segregation	This work
JC135	tatgcatgcgTGGTTTTCTCGTCCCTAT	*A0384*_upstream_R	KO	This work
JC136	gagaaaaccaCGCATGCATAAAAACTGTTG	GmR_F	KO	This work
JC137	gtcgatcataTTAGGTGGCGGTACTTGG	GmR_R	KO	This work
JC138	cgccacctaaTATGATCGACTCTCTTTTGTGTTTC	*A0384*_downstream_F	KO	This work
JC139	tgcatgcctgcaggtcgactTGGGGCGTCTTTTATTCG	*A0384*_downstream_R	KO/Segregation	This work
JC144	GTCCCTGACATTACCGCTGA	*A0384*_gene_specific_F	Segregation	This work
JC145	CAGACCGTTCTTGGGGTAAA	*A0384*_gene_specific_R	Segregation	This work
Δ*A1273* strain				
JC60	gctcggtacccggggatcctGCGCTGGCCCATTAACAG	*A1273*_upstream_F	KO/Segregation	This work
JC61	tgagcgtcagTGTAGGAAAAGGATTTGGGATC	*A1273*_upstream_R	KO	This work
JC62	ttttcctacaCTGACGCTCAGTGGAACG	KmR_F	KO	This work
JC63	aaggagttttTTCAGGTGGCACTTTTCG	KmR_R	KO	This work
JC64	gccacctgaaAAAACTCCTTAGATTTGCTTG	*A1273*_downstream_F	KO	This work
JC65	tgcatgcctgcaggtcgactGATGATAGTCTAACTCTACCTTTGATG	*A1273*_downstream_R	KO/Segregation	This work
JC174	CGCATCATCAGACGATCAAT	*A1273*_gene_specific_F	Segregation	This work
JC175	GGCGAAAAAGGTGTTCTCTG	*A1273*_gene_specific_R	Segregation	This work
Δ*A0574* strain				
JC148	gctcggtacccggggatcctCCGAATGTCGCACCAATA	*A0574*_upstream_F	KO	This work
JC149	aattcactggCAGCTTTGTTGCCTAACG	*A0574*_upstream_R	KO	This work
JC150	aacaaagctgCCAGTGAATTCGAGCTCGG	SpR_F	KO	This work
JC151	gcttttgatcCAAGCTTGCATGCCTGCA	SpR_F	KO	This work
JC152	tgcaagcttgGATCAAAAGCCGACAAAAC	*A0574*_downstream_F	KO	This work
JC153	tgcatgcctgcaggtcgactCGGAGCTAACCACCGATG	*A0574*_downstream_R	KO	This work
JC176	TCAAACGCAAAGGGGATTAC	*A0574*_gene_specific_F	Segregation	This work
JC177	ACTGAGGCCATCGTCAATTT	*A0574*_gene_specific_R	Segregation	This work
JC178	TGTTGCCCTTCATCAACAAA	*A0574*_flanking_F	Segregation	This work
JC179	TCCAAGTGGAGTTGGGATTC	*A0574*_flanking_R	Segregation	This work
Δ*A1543* strain				
JC188	gctcggtacccggggatcctCCATTGATCCAGGCGAAT	*A1543*_upstream_F	KO	This work
JC189	aattcactggTCACCCGTCTATGACTCAG	*A1543*_upstream_R	KO	This work
JC190	agacgggtgaCCAGTGAATTCGAGCTCGG	SpR_F	KO	This work
JC191	atagttttgcCAAGCTTGCATGCCTGCA	SpR_R	KO	This work
JC192	tgcaagcttgGCAAAACTATCCTTTTCGTTAGAAAGTAAAGTGAAACGACAAC	*A1543*_downstream_F	KO	This work
JC193	tgcatgcctgcaggtcgactGTGGGGCGATCGCCGGGG	*A1543*_downstream_R	KO	This work
JC180	CGCGATATTCGTACCAGTCA	*A1543*_gene_specific_F	Segregation	This work
JC181	TCAGGAAACCCTCGATGTTC	*A1543*_gene_specific_R	Segregation	This work
JC182	AAAAAGAGGGAAGCATGCAA	*A1543*_flanking_F	Segregation	This work
JC183	TATGCTGGGTTTGCTTAGCC	*A1543*_flanking_R	Segregation	This work
Δ*A1066* strain				
JC160	gctcggtacccggggatcctGAAAGAGTTGTGGTTCTTGTC	*A1066*_upstream_F	KO	This work

Table 3.4: Oligonucleotides used to make RNase mutants and verify segregation.

JC161	atgcatgcgcTATTGCGTGTTGTCCTTTTG	*A1066*_upstream_F	KO	This work
JC162	acacgcaataGCGCATGCATAAAAACTGTTGTAATTC	GmR_F	KO	This work
JC163	atgttttatatttAAACAGGCCCAGCGCCAG	GmR_R	KO	This work
JC164	gggcctgtttAAATATAAAACATCACTATTTTTCGGTG	*A1066*_downstream_F	KO	This work
JC165	tgcatgcctgcaggtcgactGTCAATGATGGCCGCTGA	*A1066*_downstream_R	KO	This work
JC200	TTCGGCTCTTCTGTGGTTTT	*A1066*_gene_specific_R	Segregation	This work
JC197	TTGGGAGTTCAAGGAACCAG	*A1066*_flanking_F	Segregation	This work
JC198	CGTTGAAAAGGGGTTTGAGA	*A1066*_flanking_R	Segregation	This work
Complementation Analysis				
JCC207	agctcggtacccgggTGAAGCGATTGGCTATGATC	*glpK*_upstream_F	for pUC19-Δ*glpK*_NS	This work
JCC208	ggggatccTTTTTTAAATGGGTTAAATTAGGTC	*glpK*_upstream_BamHI_R	for pUC19-Δ*glpK*_NS	This work
JCC209	ccatttaaaaaaggatccCCCCCTCTTGCCTACAGC	*glpK*_downstream_BamHI_F	for pUC19-Δ*glpK*_NS	This work
JCC210	caggtcgactctagagGAAACGAGATTATCTAAAACAGAAGC	*glpK*_upstream_R	for pUC19-Δ*glpK*_NS	This work
JCC263	atttaacccatttaaaaaagctcgagTACGGGGTCTGACGCTCA	KmR_F	for pUC19-Δ*glpK*:KmR_NS	This work
JCC264	taggcaagagggggggatccGATTTTGCCGATTTCGGC	KmR_F	for pUC19-Δ*glpK*:KmR_NS	This work
JCC368	atttaacccatttaaaaaaggatatcGGAAATGTGCGCGGAACC	AmpR_F	for pUC19-Δ*glpK*:AmpR_NS	This work
JCC369	taggcaagagggggggatccACGCTCAGTGGAACGAAAAC	AmpR_R	for pUC19-Δ*glpK*:AmpR_NS	This work
JCC374	aacccatttaaaaaaggatatcATGGAACCCTAAATCTGG	*A0061*_terminator_region_F	Complementation	This work
JCC375	cgcgcacatttccgatatcTTATAGAGAGTAGACGAAAATCAC	*A0061*_Promoter_region_R	Complementation	This work
JCC282	ttaacccatttaaaaaagctcgaGAACGCAATAGGAAAAGTTTAACTCAATTTATTAGTTGAG	*A1543*_promoter_region_F	Complementation	This work
JCC283	cgtcagaccccgtacCCTAGGCCACCGGGGGCC	*A1543*_terminator_region_R	Complementation	This work
JCC397	TCTCGATCCTGCAAGGGG	*glpK*_flanking_F	Segregation of Δ*glpK*	This work
JCC398	AGGCTTCATGATCAAGGGAA	*glpK*_flanking_R	Segregation of Δ*glpK*	This work
JCC399	CACCGCTGGGGCTTGTACC	*glpK*_gene_specific_R	Segregation of Δ*glpK*	This work

Uppercase nucleotides are gene-specific; lowercase indicate homologous flanking regions for Gibson assembly reaction. KO, generation of knock out mutant. NS, neutral site.

Plasmid ID	Description/Purpose	Resistance marker	Reference
pJCC248	pUC19-Δ*A0788*:Km^R	Amp/Km	This work
pJCC249	pUC19-Δ*A0061*:Sp^R	Amp/Sp	This work
pJCC250	pUC19-Δ*A2542*:Km^R	Amp/Km	This work
pJCC251	pUC19-Δ*A0384*:Gm^R	Amp/Gm	This work
pJCC252	pUC19-Δ*A1273*:Km^R	Amp/Km	This work
pJCC253	pUC19-Δ*A0574*:Sp^R	Amp/Sp	This work
pJCC254	pUC19-Δ*A1543*:Sp^R	Amp/Sp	This work
pJCC255	pUC19-Δ*A1066*:Gm^R	Amp/Gm	This work
pJCC256	pUC19-Δ*glpK*:Amp^R-*A0061*	Amp	This work
pJCC257	pUC19-Δ*glpK*:Km^R-*A1543*	Km	This work
pUC19	Cloning vector	Amp	Invitrogen
pSRA81	Source of Sp^R	Sp	(Frigaard et al. 2004)
pBBR1MCS5	Source of Gm^R	Gm	(Kovach et al. 1995)
pET28b(+)	Source of Km^R	Km	Novagen

Frigaard NU, Maresca JA, Yunker CE, Jones AD, Bryant DA (2004) Genetic manipulation of carotenoid biosynthesis in the green sulfur bacterium *Chlorobium tepidum*. *J* Bacteriol 186 (16):5210-5220

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, 2nd, Peterson KM (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166 (1):175-176

Table 3.5: Plasmids used to delete RNases in PCC 7002.

4 Chapter 4: Function of three homologs of RNase III in cyanobacteria

Authors and Contributors:

Gina Gordon: Planned the experiment, collected samples, analyzed sequencing data, designed and did in vitro experiments, made figures and wrote manuscript.

Dr. Jeffrey Cameron: Conceived the idea of the experimental setup, facilitated sample collection, extracted RNA, and provided guidance on interpreting experimental results.

Dr. Brian Pfleger: Edited the manuscript, aided in experimental design, and supervised research.

This chapter was adapted from: Gordon, G.C., Cameron, J.C., and Pfleger, B.F. (2018) Distinct and redundant functions of three homologs of RNase III in the cyanobacterium *Synechococcus* sp. PCC 7002. *Nucleic Acids Research*. doi:10.1093/nar/gky041.

4.1 Introduction

Transcript degradation plays a pivotal role in determining steady-state transcript levels. Regulation of transcript degradation is inherently critical to coordinate and control gene expression and is a central component of cellular physiology. To accurately control protein levels within a cell, detailed information on rates of transcription, transcript degradation, translation, and protein degradation must be integrated to produce models that can both precisely predict and explain protein levels. Of these four areas one is both vastly understudied and underutilized as a method to control gene expression: transcript degradation. There are many tools to control gene expression at the level of transcription (promoters, induction and repression systems, transcriptional bioswitches (Politz et al. 2013; Chappell et al. 2015; Keasling 1999), translation (ribosome binding sites, translational bioswitches (Oesterle et al. 2017; Topp et al. 2010), and protein degradation (proteolytic tags (Cameron and Collins 2014)). These tools have been applied to bacterial chemical factories and produced significant increases in titers and yields (Peralta-Yahya et al. 2012; Adrio et al. 2010; Pickens et al. 2014). Contrastingly, very few examples of controlling gene expression via engineered transcript stability exist and those that do are restricted to proof of concept in select organisms (Babiskin et al. 2011; Carrier et al. 1997; Smolke et al. 2000).

The lack of tools for controlling gene expression on the level of transcript stability most likely stems from the lack of knowledge of how these processes work in living cells. The lag in descriptive and mechanistic knowledge of transcript stability behind those of transcription, translation, and protein turnover may stem from the specific difficulties and challenges in studying RNA turnover such as the short lifetime of transcripts, the many distinct enzymes of which some form complexes, and the disparate activity of these enzymes in vitro and in vivo.

Transcript degradation is facilitated by ribonucleases (RNases) that have distinct characteristics. Models of transcript turnover involve initial cleavage facilitated by endoribonucleases that cleave RNA internally (RNase E, G, III, Y) followed by subsequent cleavage by exoribonucleases that processively remove nucleotides from the ends (RNase II, R, J, PNPase). Models of transcript

degradation have been developed primarily based on Gram-negative *Escherichia coli* and Gram-positive *Bacillus subtilis*.

Many factors and RNase activities seem to be conserved in these two models, but the specific enzymes can be different. *E. coli* transcript degradation is facilitated by RNase E and partially by RNase G while *B. subtilis* possess a somewhat similar RNase Y. *B. subtilis* has two 5' to 3' exoribonucleases (RNase J1 and J2) that facilitate degradation from the 5' end. Many bacterial groups possess a mixture of these RNases which leads to questions about how transcripts are being efficiently turned over and regulated.

We became interested in a cyanobacterial strain, *Synechococcus* sp. strain PCC 7002 (hereafter PCC 7002), for both its industrial applications in chemical production and its unique array of RNases. PCC 7002 contains both a homolog of RNase E (A0788) and RNase J (A1273) as well as three homologs of RNase III (A0061, A2542, A0384); the first two being full length RNase IIIs while the third being a Mini-RNase III that lacks the double-stranded binding domain (**Chapter 3**). These three RNase III homologs are unessential under standard growth conditions and have been deleted in all combinations (**Chapter 3**). Here, we aimed to elucidate the function of the three homologs of RNase III to determine if they played distinct or redundant roles in transcript turnover, what structures and sequences they target, and what genes and cellular activities are affected by their action.

4.2 Role of three homologs of RNase III

rRNA processing by RNase III homologs

RNase III has been shown to be involved in rRNA processing in many bacterial species (King et al. 1984; Young et al. 1978; Redko et al. 2008), yeast (Elela et al. 1996), and in plant chloroplasts (Hotto et al. 2015), so we examined the rRNA of RNase III mutants. We extracted RNA from WT and strains harboring all combinations of RNase III deletions. The distribution of rRNA fragments

was evaluated using nanoscale electrophoresis (BioAnalyzer) according to manufacturer instructions. We observed no differences in rRNA profiles in ΔA0061, ΔA2542, or ΔA0061ΔA2542 compared to WT (**Figure 4.1**). When A0384 was deleted we saw an additional peak slightly larger than the 23S peak (**Figure 4.2A**). This resembled traces of rRNA isolated from *E. coli* Δ*rnc* where the 23S rRNA has ∼25 and ∼50 bases extensions at the 5' and 3' end (King et al. 1984) as well as the case of rRNA isolated from *B. subtilis* mini-III deletion where 23S rRNA had extensions at the 5' and 3' end ranging from 2-64 nucleotides (Redko et al. 2008). We were unable to pinpoint the exact location of these proposed 23S rRNA extensions in ΔA0384 because the 5' RACE products sequenced corresponded to the mature 5' end of the 23S rRNA. There were additional peaks in ΔA0061ΔA0384 and ΔA0061ΔA2542ΔA0384, suggesting that A0061 and A0384 perform a redundant function that is altered when both are deleted. Based on the retention times, it appears that in addition to the normal rRNA species ΔA0061ΔA0384 and ΔA0061ΔA2542ΔA0384 also contain a larger rRNA fragment (A*) as well as two other fragments (B* and C*) that appear longer that corresponding fragments in the WT rRNA pool.

The 23S rRNA from PCC 7002 and other cyanobacteria is known to fragment into two pieces, but remain functional in the ribosome. The exact location of this fragmentation has not been reported, but similar fragment sizes have been observed in several species (Doolittle 1973; Shan et al. 2009). The mechanism of this 23S rRNA processing is unknown and has been described as 'slow' and more rapid in the light than in the dark (Doolittle 1973). Through 5' RACE, we located a 5' end that corresponded to base 524 of the annotated 23S rRNA suggesting that the full-length rRNA transcript is processed at this location. This site is located in a large loop between helix 24 and helix 2 in *E. coli*'s 23S rRNA secondary structure (Petrov et al. 2013).

Based on the approximate peak sizes observed in the BioAnalyzer traces, we hypothesized that the larger species when both A0061 and A0384 were absent corresponded to 23S rRNA with extra 5' and 3' extensions. We created cDNA using random hexamers and ran PCR with a series of primers to see if we could detect 23S rRNA with extended 5' or 3' regions. We were able to amplify

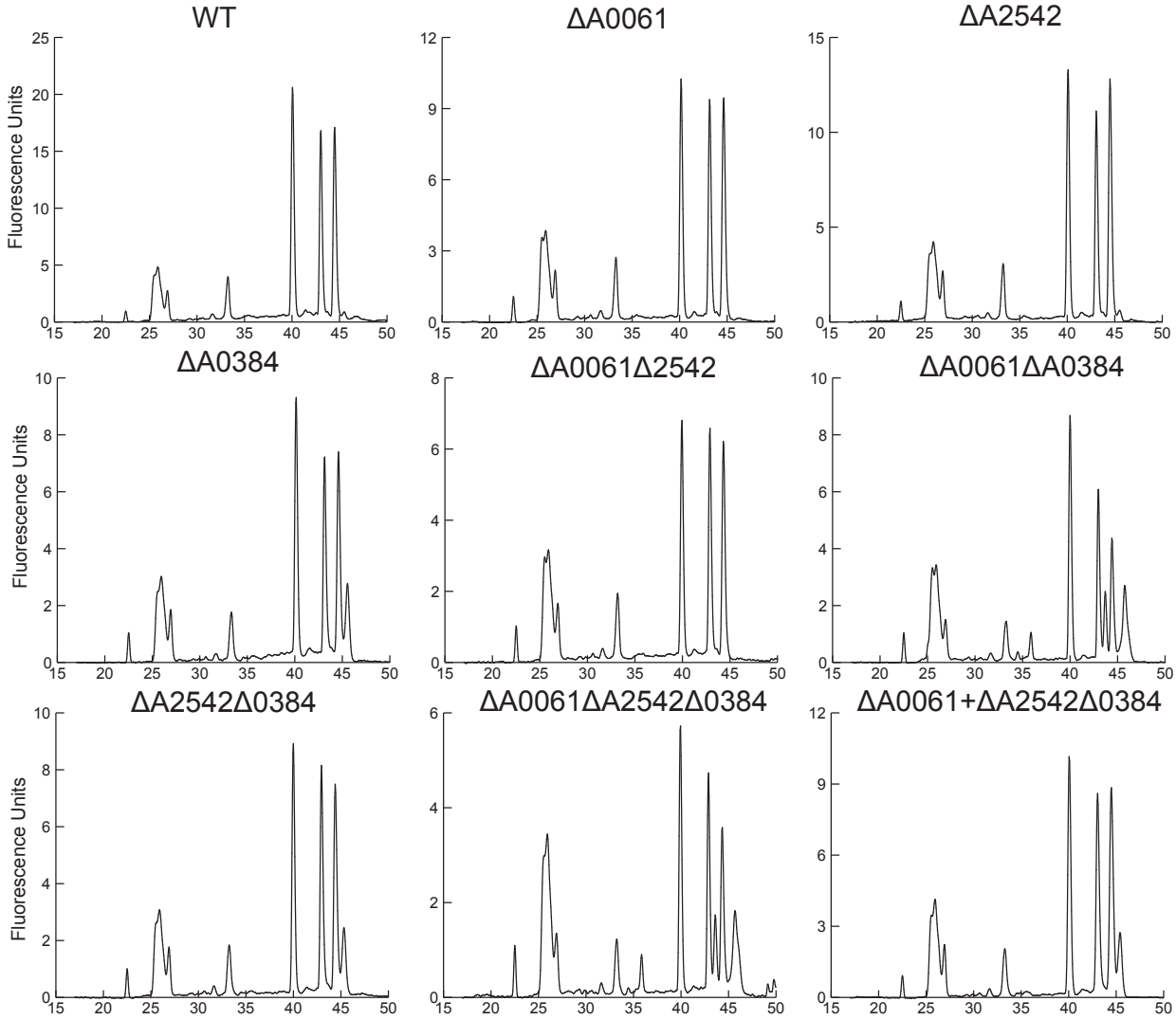


Figure 4.1: BioAnalyzer traces of rRNA isolated from WT PCC 7002 and single, double, and triple RNase III mutants. The ΔA0061+ΔA2542ΔA0384 strain is the triple RNase III mutant complemented with the A0061 gene at another locus.

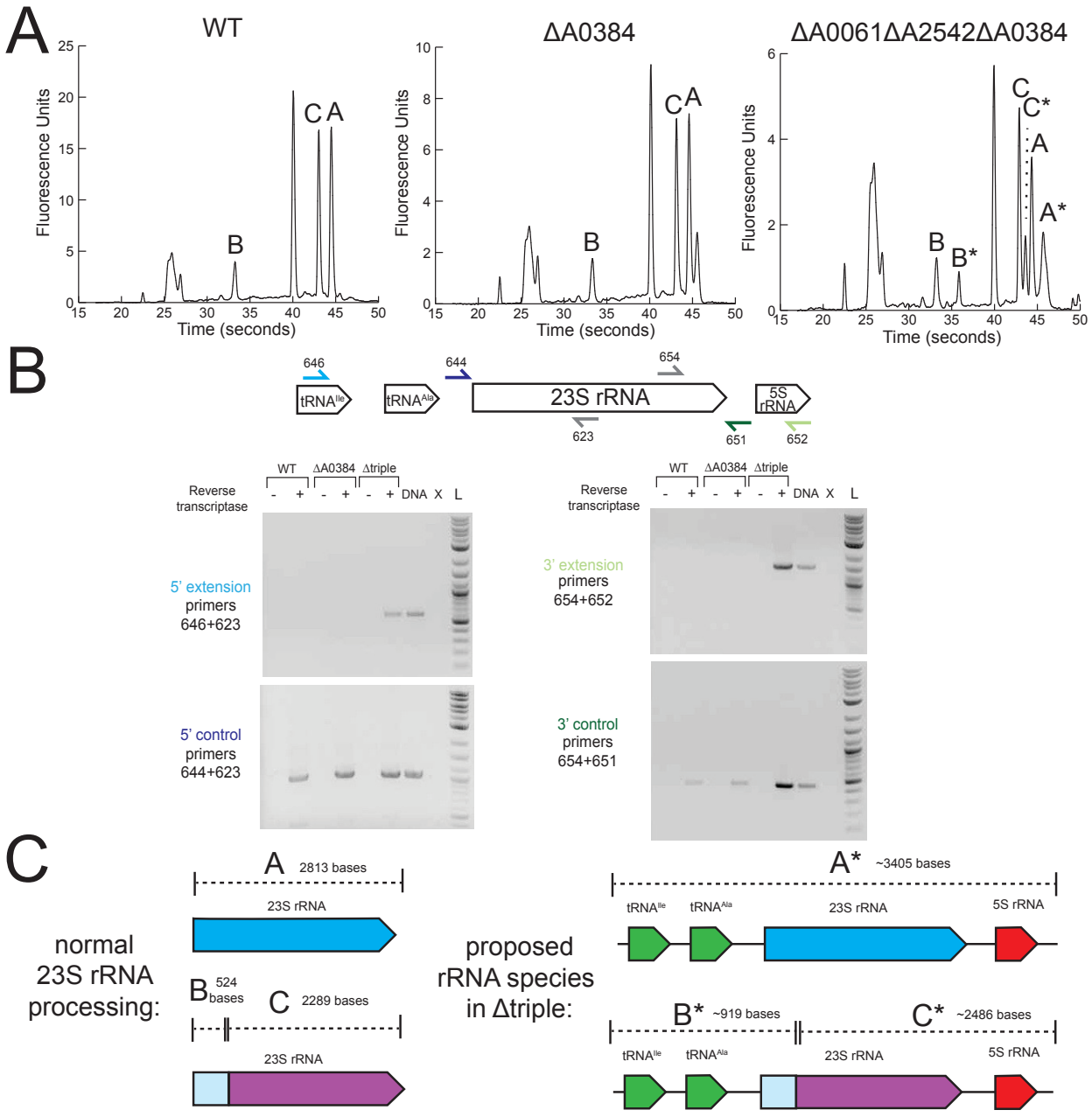


Figure 4.2: Role of RNase III in rRNA processing. (A) BioAnalyzer traces of RNA extracted from WT and select RNase III mutants. (B) Extensions of 23S rRNA at the 5' and 3' ends in ΔA0061ΔA2542ΔA0384. A diagram of the rRNA operon shows primers used to amplify cDNA made from WT and ΔA0384 and Δtriple strains. Controls using WT PCC 7002 DNA as a template (DNA), no template (X), and 2 log DNA ladder (L) are also shown. (C) Proposed rRNA species seen in the BioAnalyzer traces.The 23S peak (labeled as A) breaks into two fragments B (524 bases) and C (2289 bases). rRNA species from the triple mutant with 5' and 3' extensions are labeled with an asterisk.

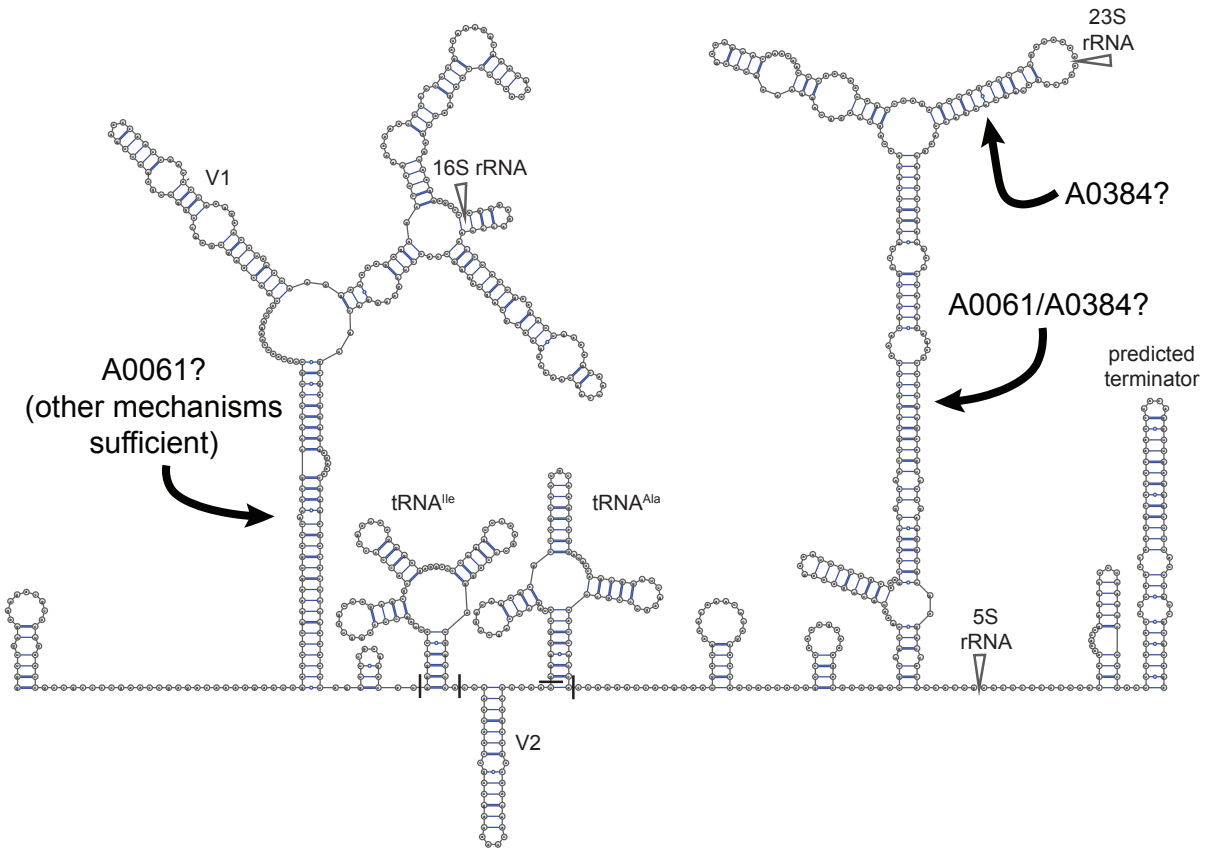


Figure 4.3: Proposed secondary structure of the rRNA operon with labels of approximate locations of RNase III cleavage sites. Sequences were folded with mfold (Zuker 2003) and visualized with VARNA (Darty et al. 2009)

products that corresponded to 5' and 3' extensions of the 23S rRNA only with cDNA created from the triple mutant when reverse transcriptase was added (**Figure 4.2B**). We were able to amplify fragments using primers as far away as located in the intergenic tRNA in the 5' direction and located in 5S rRNA in the 3' direction. These 5' and 3' extensions suggest a role of A0061 and A0384 in processing the pre-23S rRNA, which contains a long region of complementary sequences (**Figure 4.3**).

We attempted to complement *E. coli* Δ*rnc* RNase III mutant strain with the three homologs from PCC 7002, but were unable to complement this rRNA phenotype (data not shown) perhaps because

E. coli's RNase III is known to bind the ribosome and it's specificity is altered by the presence of r-proteins (Allas et al. 2003).

Differential gene expression in RNase III mutants

To explore further roles of RNase III in PCC 7002, we isolated total RNA from WT, the single, double, and triple mutants and sequenced the rRNA-subtracted transcriptome of each. To explore the impact of each RNase III, we aligned the RNA-sequencing reads to the genome and six plasmids, counted the reads that aligned to each feature, and determined differentially expressed genes compared to WT with a false discovery rate cutoff rate (FDR) of 0.005. With these parameters, approximately 6-18% of genes were up-regulated, and 8-18% of genes were down-regulated when each mutant was compared with WT. A full list of differentially expressed genes is available in **Appendix II**. We did not observe any substantial changes expression of any of the RNase III genes across any of the mutants (**Table 4.1**).

We clustered the $\log_2$-fold change data for each gene and each sample type and found that the gene expression patterns of the single RNase III mutants most resembled each other and that the double and triple mutants clustered farther away (**Figure 4.4A**). Of all the differentially expressed genes in the three single mutants, many of them are differentially regulated in all three mutants (**Figure 4.4B**). Additionally, there seems to be a significant overlap in differentially expressed genes between the ΔA0061ΔA0384 and ΔA0061ΔA2542 double mutants (**Figure 4.4C**). These results were initially surprising because we expected to see expression patterns where we could logically conclude specific functions to each homolog. Instead, these data highlight the complexities associated with gene expression regulation mediated by RNases. The following sections describe specific roles that were identified.

Table 4.1: RNase III transcript levels remain constant across mutant strains.

Strain Name	**A0061 read counts**	**A2542 read counts**	**A0384 read counts**
WT A	345	345	689
WT B	248	244	563
WT C	294	316	715
ΔA0061 A	X	301	313
ΔA0061 B	X	575	695
ΔA0061 C	X	322	348
ΔA2542 A	274	X	334
ΔA2542 B	697	X	873
ΔA2542 C	157	X	222
ΔA0384 A	283	223	X
ΔA0384 B	159	183	X
ΔA0384 C	243	190	X
ΔA0061ΔA2542 A	X	X	436
ΔA0061ΔA2542 B	X	X	931
ΔA0061ΔA2542 C	X	X	395
ΔA2542ΔA0384 A	362	X	X
ΔA2542ΔA0384 B	143	X	X
ΔA2542ΔA0384 C	273	X	X
ΔA0061ΔA0384 A	X	178	X
ΔA0061ΔA0384 B	X	155	X
ΔA0061ΔA0384 C	X	262	X
ΔA0061ΔA2542ΔA0384 A	X	X	X
ΔA0061ΔA2542ΔA0384 B	X	X	X
ΔA0061ΔA2542ΔA0384 C	X	X	X

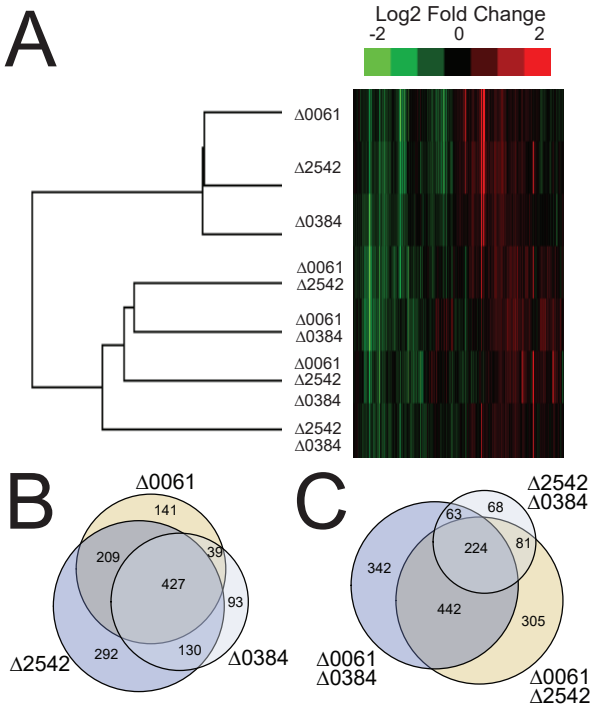


Figure 4.4: Overall expression patterns in RNase III mutants compared to WT. (A) Heatmap of clustered $\log_2$ fold changes of all genes. Venn diagram of overlapping genes that were differentially expressed in the single RNase III mutants (B) and double RNase III mutants (C) (FDR < 0.005).

Up-regulation of pAQ3 copy number in ΔA2542

When A2542 was absent (single, double, and triple mutants), we observed significant up-regulation of many genes on pAQ3, one of the six plasmids in PCC 7002. Most pAQ3 genes in strains lacking A2542 had 2-3 times higher expression level than in strains with A2542 present, except for one gene, C0011, that was approximately 16 times more highly expressed (**Figure 4.5**). Due to the up-regulation of all of the genes on this plasmid, we hypothesized that we were observing an increase in copy number of this plasmid when A2542 was absent. We used qPCR to determine the ratio of copies of pAQ3 to a control DNA spike-in and found 1.8 ± 0.4 times more copies of pAQ3 in ΔA2542 than WT (**Figure 4.5**). We saw similar values when comparing pAQ3 to a control gene on the chromosome (data not shown).

We hypothesized that C0011 and A2542 may be involved in copy number control because C0011 was the only coding sequence differentially expressed in ΔA2542 after correcting for DNA copy number. Based on the approximate start location of the C0011 mRNA as visualized in the RNA-seq data, we hypothesized that the start codon was mis-annotated, and the protein is actually 50 amino acids (coordinates 13,101 – 13,253 in NC_010477) instead of the original 63 amino acids. Protein homology searches (BLASTP) showed weak homology to CopG family proteins (*Thermus* sp. WG and *Subdoligranulum varaiable* DSM 15176). This protein has been characterized in prototype rolling circle plasmid pMV158 and its derivative pLS1, relatively small multicopy streptococcal plasmids that can replicate in both and Gram-negative hosts (del Solar, Diaz, et al. 1987). Copy number control in pMV158 and pLS1 is controlled by two components: 1) CopG, a small ribbon-helix-helix transcriptional repressor protein that binds a 50 bp operator sequence and autorepresses the CopG-replication protein transcript and 2) a small counter-transcribed RNA (ctRNA) that binds to a region of the replication protein (del Solar, Acebo, et al. 1995). The replication protein, RepB, in these plasmids is the rate-limiting factor that initiates plasmid replication by nicking the double-stranded origin (*dso*) and it's synthesis is controlled by a ctRNA that blocks an atypical ribosome binding site (López-Aguilar et al. 2013; del Solar and Espinosa 1992).

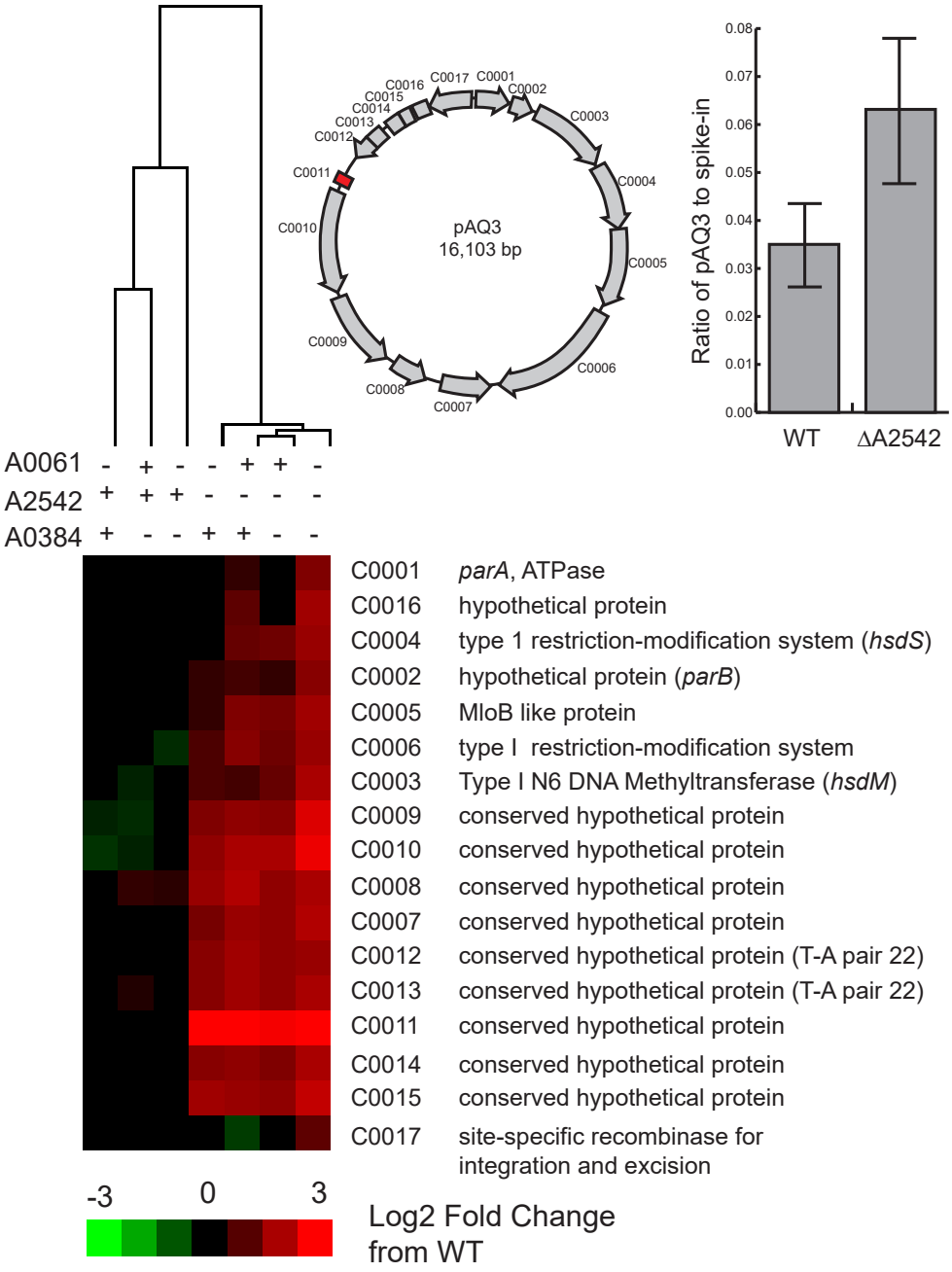


Figure 4.5: Heatmap of gene expression of coding sequences on pAQ3 in RNase III mutants compared to WT. Log_2 fold change data for each gene was clustered using Cluster 3.0 (hierarchical clustering, correlation uncentered, centroid linkage). Genes not meeting the FDR cutoff (<0.005) were given a value of zero. A schematic of the coding sequences on pAQ3 and their locations and orientations is shown. There was a significant different in pAQ3 levels compared to a DNA spike-in are shown for WT and ΔA2542 ($p < 0.05$, Student's t-test, two-sided, equal variance, n = 4). Averages are shown with the error bars showing the standard deviation.

C0011 is predicted to share a similar ribbon-helix-helix structure (as predicted by Phyre2 (Kelly et al. 2015)) and also has the highly conserved glycine in the turn between the two α-helices (Acebo et al. 1998). The gene downstream of C0011 (C0010) is in a similar position as RepB from pMV158 but has little homology. We were unable to detect the presence of a small ctRNA on the sense strand in the RNA-seq data in WT or ΔA2542 (**Figure 4.6A**), but we did see a large increase in the transcript on the antisense strand between C0011 and C0010. There is a large increase in the ratio of reads of the antisense strand when comparing ΔA2542 to WT, and this region corresponds to the location of the ctRNA in other systems. Folding this region using mfold (Zuker 2003) reveals a 86-bp structure very similar to ctRNA's from other plasmids that consists of two hairpins surrounding a 12 bp single stranded region. We hypothesize that we could not see this ctRNA in the RNA-sequencing data because the library preparation was not designed to target small RNAs ($<$100 nucleotides). We showed that it is expressed by amplifying this ctRNA from RNA samples treated with reverse transcriptase (**Figure 4.6B**). We compared ctRNA expression levels between WT and ΔA2542 but there was no statistically significant difference (data not shown).

We hypothesized that because RNase III enzymes cleave double stranded RNA, A2542 could be cutting the duplex of ctRNA and replication RNA. We used in vitro transcription to make these two species and tested purified A2542's ability to cut both the replication RNA alone as well as the annealed duplex. These reactions were initiated by the addition of magnesium chloride and samples were taken and quenched with EDTA at a series of time points. A2542 was unable to cleave replication RNA alone even after 60 minutes, but was able to cleave the duplex of the ctRNA and replication RNA (**Figure 4.6C**). The ctRNA was not cleaved, but we identified the cleavage site within the replication RNA by purifying RNA from the reaction and doing 5' RACE. A structure of the annealed duplexes is shown with the cleavage site (**Figure 4.6D**). These regions could also form a 86bp perfectly doubled-stranded region. A0384 could not cleave the duplex of ctRNA and replication RNA in vitro, but interestingly A0061 could despite no observed differences in pAQ3 gene expression (**Figure 4.7**).

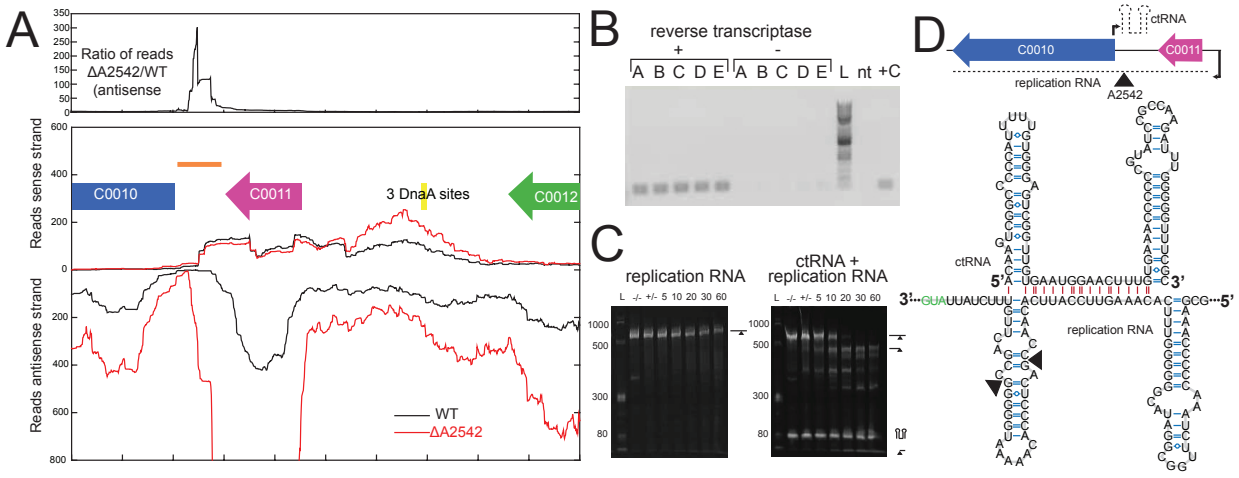


Figure 4.6: Identification of ctRNA on pAQ3. (A) RNA-seq coverage of portion of pAQ3. An example replicate is shown for WT and ΔA2542 with reads aligning to the sense strand displayed at the top and reads aligning to the antisense strand displayed in the opposite direction. The top panel shows the ratio of ΔA2542 to WT of the antisense strand. (B) Verification of expression of ctRNA by amplification off of cDNA (86 bp product). Replicates A-E made from WT RNA are shown along with a reference 2 log ladder (L), a no-template (nt), and positive control (+C). (C) *In vitro* cleavage assays of the replication RNA and the replication RNA and ctRNA with A2542. Cleavage reactions containing 5 µg of RNA were initiated by addition of $MgCl_2$ and samples were removed at 5, 10, 20, 30, and 60 minutes and run on a urea polyacrylamide gel. Initial samples were also taken before the addition of enzyme (-/-) and $MgCl_2$ (+/-). (D) A diagram of the location and orientations of these two RNAs and the predicted structure of the ctRNA as predicted by mfold (Zuker 2003). A section of the replication RNA is shown with the marked site of A2542 cleavage in the in vitro cleavage reaction. Intramolecular bonds are shown in blue and intermolecular interactions are shown in red. The structured RNA is shown, but these RNAs are perfectly complementary (being transcribed from the same DNA segment), and could exist as a perfectly double-stranded 86bp region.

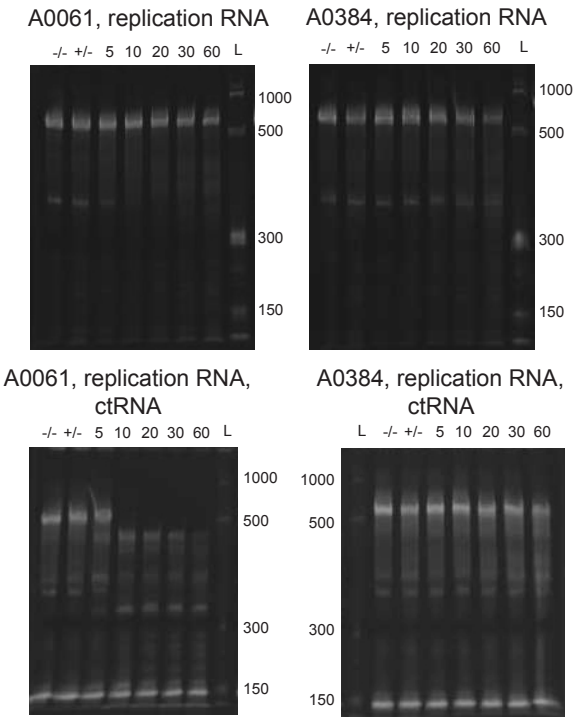


Figure 4.7: *In vitro* cleavage assays of pAQ3 replication RNA and ctRNA using purified RNase IIIs A0061 and A0384. Cleavage reactions containing 5 µg of RNA were activated by addition of $MgCl_2$ and samples were removed at 5, 10, 20, 30, and 60 minutes and run on a urea polyacrylamide gel. Initial samples were also taken before the addition of enzyme (-/-) and $MgCl_2$ (+/-).

Global gene expression changes in RNase III mutants

A set of genes, including bicarbonate transporters and NADH dehydrogenase components, were very strongly upregulated, including *sbtA*, *bicA*, *ndhF3*, *ndhD3*, *ndhD1* (**Figure 4.8**) in each of the single RNase III mutants. Many of these genes are regulated by a transcriptional repressor *ccmR* in cyanobacteria (Woodger et al. 2007). CcmR has been characterized in other model cyanobacteria and shown to repress itself as well as structural components of inorganic carbon transport (Wang, Postier, et al. 2004). The CcmR binding motif was determined for PCC 7002 (McClure et al. 2016) and we observed up-regulation in the single RNase III mutants of all genes predicted to contain this motif (**Figure 4.8** marked with *). Interestingly, the same transcripts in the double and triple mutants are not as significantly changed from WT if changed at all (**Figure 4.8** – right columns of heat map) indicating that there is more complex regulation involved.

Additional genes were also regulated in a similar manner to those containing CcmR binding motifs. This group, which has not been linked to the CcmR regulon, include carbonic anhydrase that converts bicarbonate to CO_2 inside of the carboxysome, NDH-1 components thought to be involved in CO_2 transport, NAD(P)H oxidoreductases, as well as the subunits of RuBisCO (*rbcS*, *rbcL*) and carboxysome-encoding operon (*ccmK$_2$K$_1$LMN*). Additionally, in the single mutants we observed down-regulation of several groups of genes: transhydrogenases (*pntCBA*), glycolysis genes (*pfkA*, *gapA*), protochlorophyllide reductases (*chlB*, *chlL*) and pyruvate ferredoxin oxidoreductase (*nifJ*). These patterns indicate that RNase III plays a role in regulating the key genes in carbon fixation.

Other major changes in RNase III mutants include down-regulation of prophage genes, down-regulation of r-protein genes, and down-regulation of genes involved in glycan biosynthesis and metabolism (**Figure 4.9**). We saw down-regulation of many components of the phycobilisome and saw up-regulation of phycobilisome degradation proteins. Interestingly we found differential regulation of several genes that are part of the CRISPR type III RAMP module on pAQ6 (F0044-F0047). RNase III has been shown to be involved in processing of the CRISPR precursor transcript into the mature crRNA in type II CRISPR systems (Deltcheva et al. 2011), but has not been shown

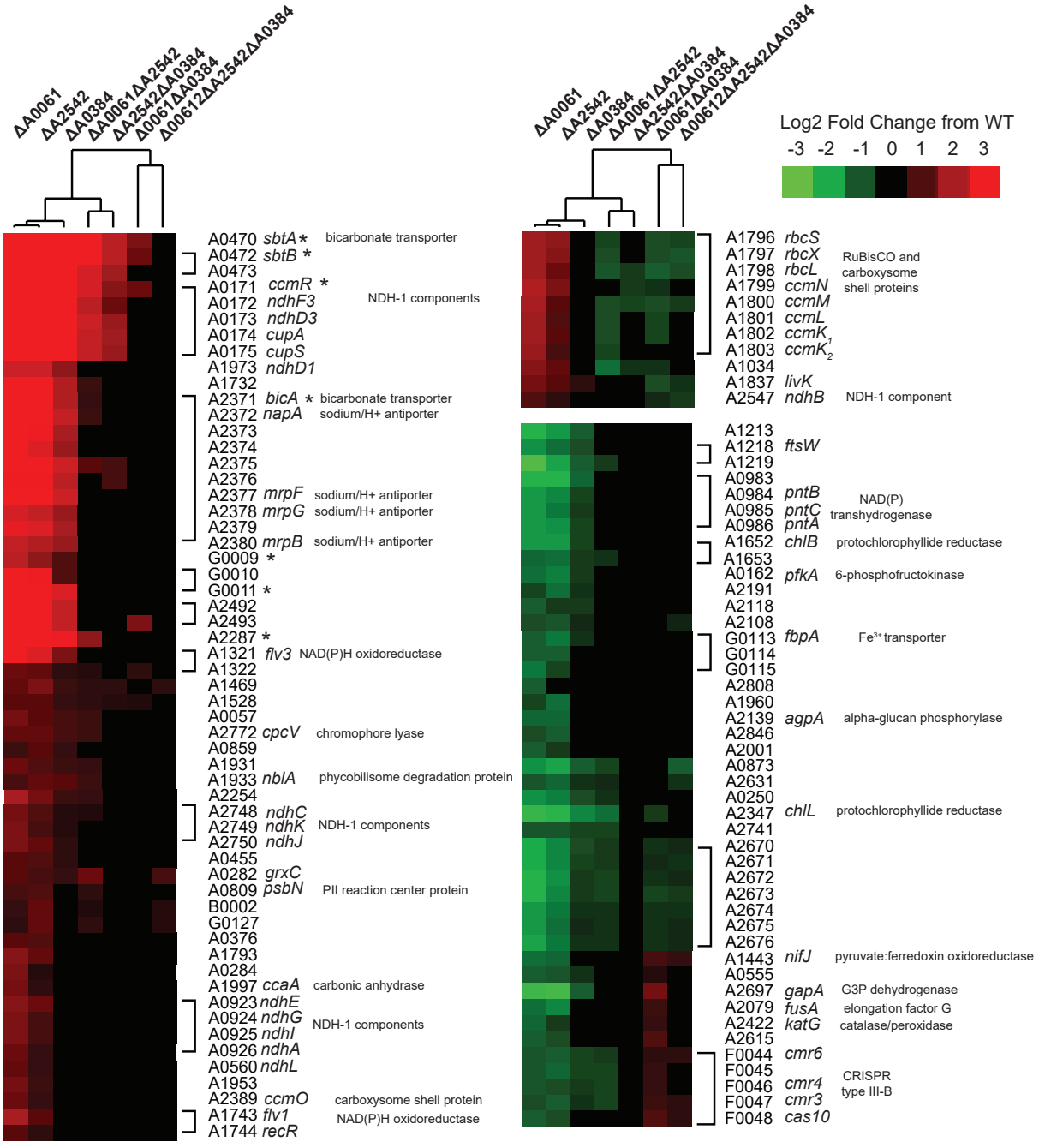


Figure 4.8: Heatmap of differentially expressed genes that showed different expression patterns in the single RNase III and double or triple RNase III mutants. Genes that were differentially expressed compared to WT with a fold-change greater than 2 or less -2 were clustered and the genes that had different expression patterns in the single mutants compared to the double or triple mutants were extracted. Genes were re-ordered to be next to those shown to or be predicted to be in the same operon. Those known or predicted to be a part of the CcmR regulon are marked (*). Genes in the same operon are grouped with brackets.

to be involved in type III RNA-targeting CRISPR systems. We were unable to detect any differences in expression of the crRNA from the type III array on pAQ6 perhaps because the RNA-sequencing did not efficiently capture RNA $<$100 nucleotides, but this potential involvement of RNase III regulating the protein components the CRISPR type III-B system has not been reported before. We saw many global shifts in related-genes, indicating a role for RNase IIIs in maintaining proper gene expression.

In vitro cleavage assays with purified RNase III homologs

We were intrigued with the somewhat contradictory observations involving the specificities and redundancies of the three RNase IIIs and decided to test their ability to cleave known RNase III targets from *E. coli*. We created RNA transcripts via in vitro transcription and tested the ability of purified cyanobacteria RNase III proteins to cut each transcript. Reactions were initiated with the addition of $MgCl_2$ and quenched with excess EDTA at a series of time points. Reaction products were run on a denaturing polyacrylamide gel and visualized with the addition of ethidium bromide. We observed cleavage of these transcripts as evidence by the disappearance of the large full-length band or the inability to cut these targets as evidence by the presence of the band after the 60-minute time course (**Figure 4.10A**). Both A0061 and A2542 could cleave many of the *E. coli* RNase III targets, but we never observed cleavage by A0384 (**Figure 4.10B**), the mini-III. One *E. coli* RNase III target, *proP*, could not be cut by any of these three enzymes. For a select few targets, we performed 5' RACE on cleavage reactions to pinpoint the exact location of these cleavage events. A0061 and A2542 cut the *pnp* transcript at the same location as the *E. coli* RNase III enzyme, but for several others A0061 and A2542 cleaved the same stem at neighboring bases (range -21 to +3) (**Figure 4.10C**). We tested the oligomerization of these three purified proteins by incubating them with the cross-linker disuccinimidyl suberate and running them on a polyacrylamide gel. Like *E. coli*'s RNase III and all other known RNase IIIs, we saw the formation of dimers of A0061 and A0384 but we did not observe dimers for A2542 (**Figure 4.11**). We hypothesize that our lack of

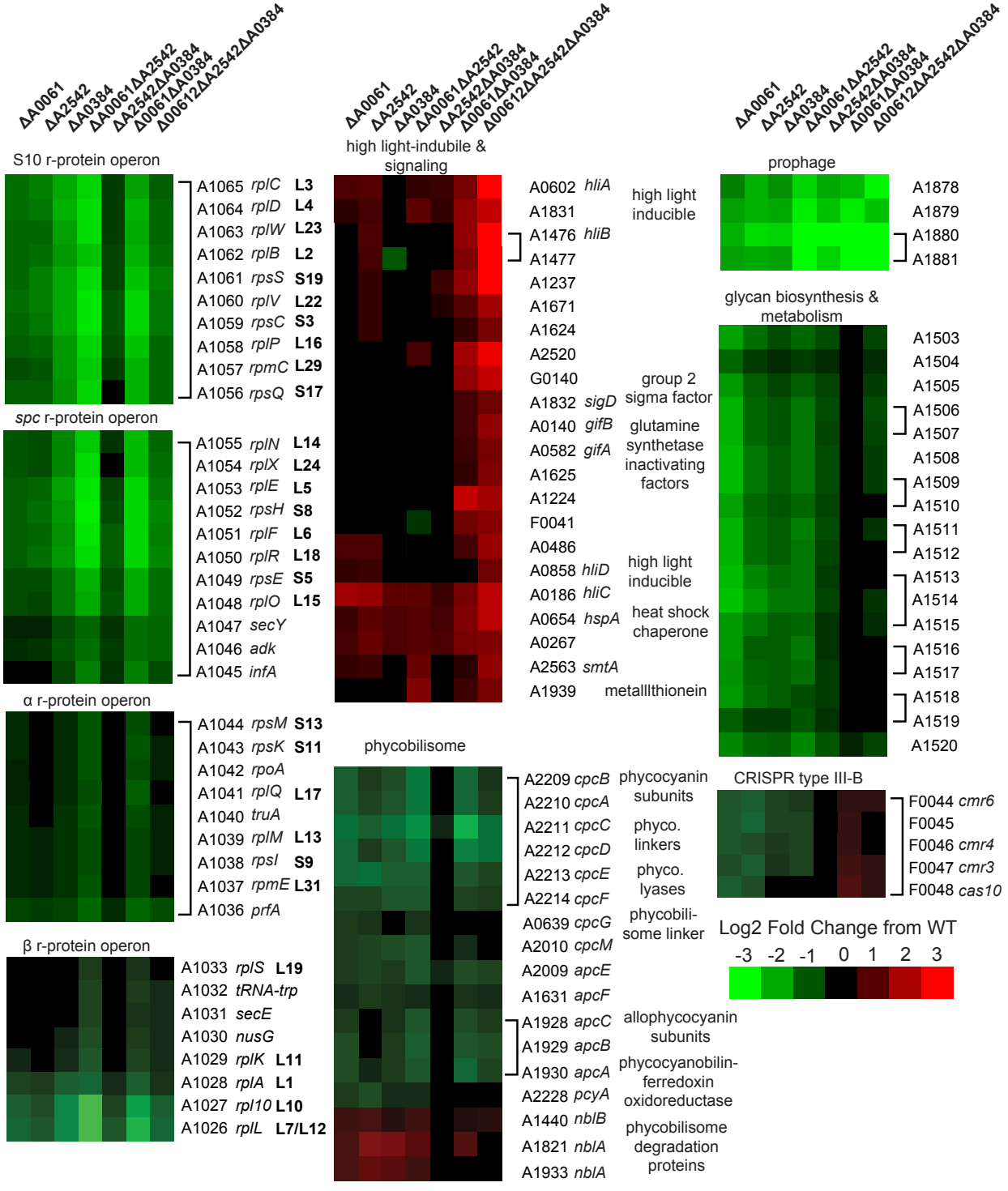


Figure 4.9: Heatmaps of clustered groups of differentially expressed genes in all samples compared to WT. Genes were re-ordered to be next to those shown to or be predicted to be in the same operon. Genes in the same operon are grouped with brackets.

observation of a dimer for A2542 may be due to structural differences or lack of closely located primary amines in the dimer. The ability of these RNase IIIs to cleave some non-native targets precisely and some non-precisely highlights the intricate specificities of these enzymes and supports the idea that these enzymes are more finely-tuned to regulate gene expression than is currently appreciated.

4.3 Discussion

rRNA processing

By analyzing the rRNA traces of these RNase III mutants we were able to determine the role and approximate location of cleavage of A0061 and A0384 on the 23S rRNA. Without A0384, there are slight extensions of the 23S rRNA at the 5' and 3' ends indicating A0384 cleavage in the paired region proximal to the mature 23S rRNA ends. A0061 seems to cleave more distal in this paired region and without A0061 or A0384, the 23S rRNA has large 5' and 3' extensions and includes both tRNAs and the 5S rRNA. From these samples, we were unable to tell if these RNase IIIs play a role in 16S rRNA processing, but in many organisms RNase III cleaves the stem surrounding the 16S rRNA but other enzymes functionally complement absence of RNase III. These results coincide with the role mini-III in 23S rRNA maturation in *B. subtilis* and the full-length RNase III in 23S rRNA processing as found in many organisms. The presence of a third homolog of RNase III in some cyanobacteria (**Chapter 3**) may indicate a specialized role beyond rRNA processing .

The analysis was initially not as straight forward due to the known rRNA fragmentation of the 23S rRNA in cyanobacteria. Using 5' RACE we were able to pinpoint the fragmentation site and show that it occurs within a predicted loop region. rRNA fragmentation is widespread in bacteria (Evguenieva-Hackenberg 2005), but still there is little known about its consequences on physiology. Some organisms have intervening sequences within their rRNA, such as *Rhodobacter capsulatus*, *Salmonella*, and *Campylobacter*, which are removed through RNase III cleavage and

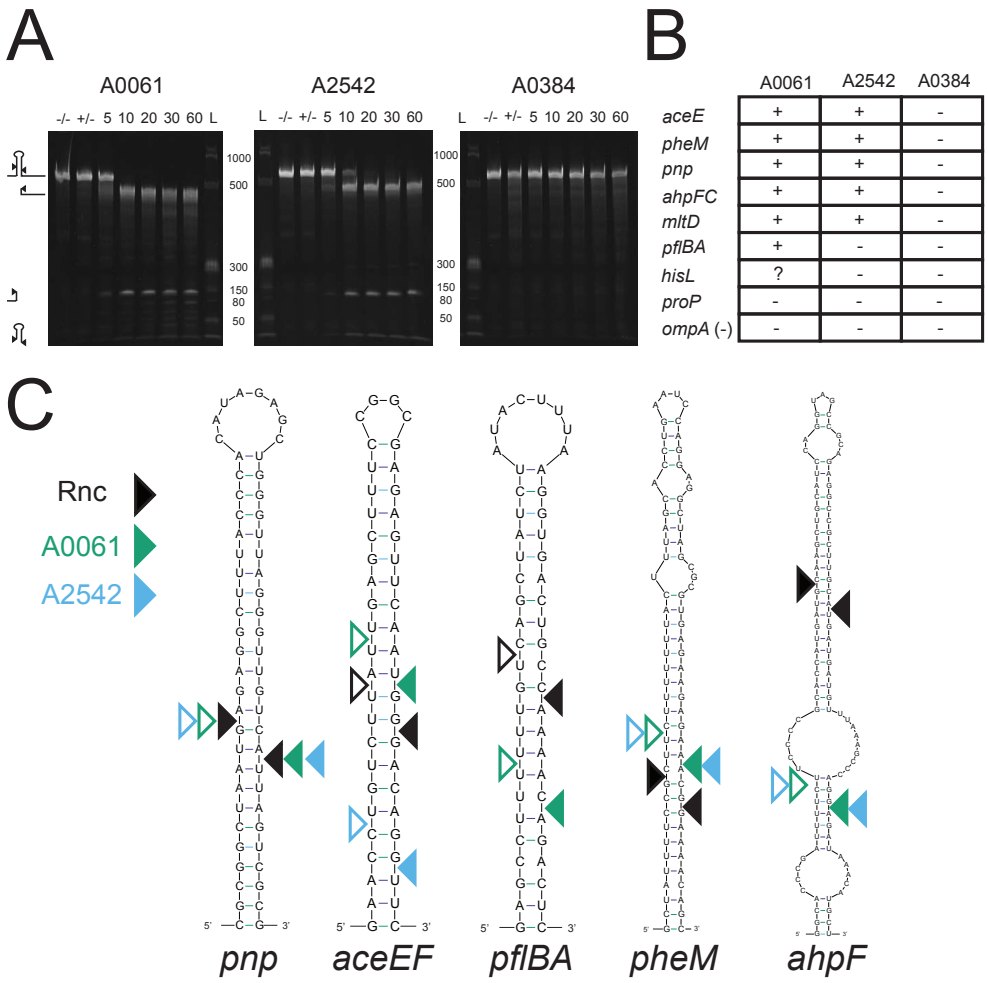


Figure 4.10: PCC 7002 RNase III's are capable of cleaving *E. coli* RNase III targets. (A) Cleavage assays of in vitro transcribed *E. coli* RNase III target (*aceEF*) incubated with purified cyanobacteria RNase IIIs Cleavage reactions containing 5 µg of RNA were initiated by addition of $MgCl_2$ and samples were removed at 5, 10, 20, 30, and 60 minutes and run on a urea polyacrylamide gel. Initial samples were also taken before the addition of enzyme (-/-) and $MgCl_2$ (+/-). (B) Summary of cleavage reactions of sequences known to be cleaved by *E. coli*'s RNase III. A highly structured RNA but not known to be cleaved by *E. coli*'s RNase III was included as a negative control (*ompA*). (C) Locations of cyanobacteria RNase III cleavage of known *E. coli* RNase III targets. Structures and locations of *E. coli* RNase III cleavage are shown as in **Chapter 2**. Filled in triangles indicate 5' ends found with 5' RACE of in vitro cleavage reactions while open triangles indicate position of the other cleavage event where the inferred other cleavage event would be to leave the characteristic 2 nucleotide 3'-overhangs.

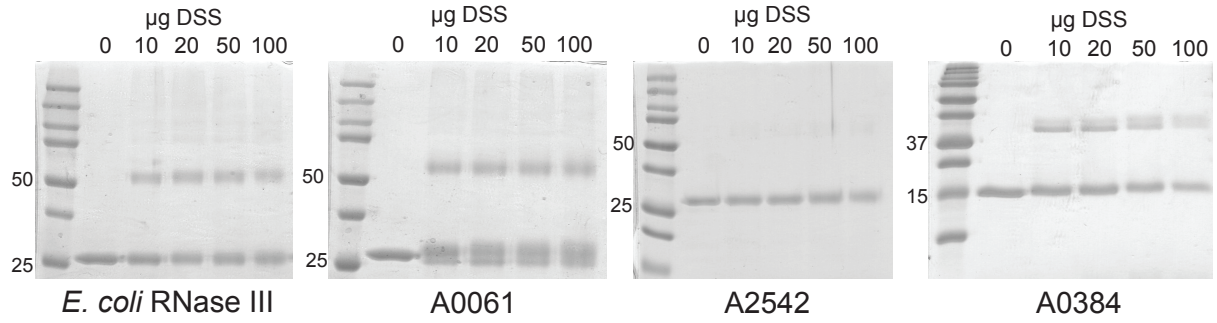


Figure 4.11: Oligomerization of purified RNase III enzymes. Dimers of *E. coli* RNase III (monomer: 27.0 kDa, dimer: 53.9 kDa), A0061 (monomer: 28.6 kDa, dimer: 57.1 kDa), and A0384 (monomer: 16.3 kDa, dimer: 32.5 kDa) were observed. Dimerization of A2542 was not observed (monomer: 29.9 kDa, dimer: 59.8 kDa).

subsequent processing by RNase E (Zahn et al. 2000). Here, we saw no evidence of RNase III involvement in 23S rRNA fragmentation in PCC 7002 and postulate that it is caused by RNase E, other endonucleases, or be naturally labile in PCC 7002.

RNase III plasmid copy number regulation

Our RNA-seq data demonstrated a significant upregulation of genes on the native plasmid pAQ3. We hypothesize that A2542 controls the copy number of pAQ3 by recognizing and cleaving the complex of a short counter-transcribed RNA (ctRNA) and a coding transcript, as we demonstrated in vitro. Initially we saw a build-up of the coding transcript only, but the similarities in gene organization with rolling-circle plasmid replication plasmids (pMV158 and pLS1) led us to detect the ctRNA that was not visible in the RNA-sequencing data. In pMV158 and pLS1 this ctRNA blocks translation of the downstream replication protein, which is the rate-limiting factor for plasmid replication initiation. No one has reported if RNase III plays a role in degrading the ctRNA and coding transcript complex in pMV158 and pLS1. We propose that in PCC 7002 A2542 is just one part of copy number regulation because in its absence copy number is only up-regulated 2-fold. Interestingly, A0061 could also cleave the complex in vitro despite seeing no differences in vivo. This could indicate differential specificities or spatial localizations in vivo (A0061 processes rRNA

while A2542 does not). A2542 could ensure proper turnover of the blocked ctRNA, coding transcript complex. This mode of regulation is unique to pAQ3 (other 5 plasmids do not contain any similar elements) and does not appear to be widely conserved in cyanobacterial plasmids.

Global mRNA expression changes in RNase III mutants

We saw many coordinated changes in gene expression within groups of genes involved in related processes. One such intriguing group included all known members of the CcmR regulon (involved in bicarbonate transport NDH-1 components thought to be involved in CO_2 transport) as well as other NDH-1 components, RuBisCO and carboxysome genes, carbonic anhydrase, NAD(P)H oxidoreductases, and other uncharacterized proteins. These genes were only up-regulated in the single RNase III mutants and not the double or triple mutants. For another group of genes we saw down-regulation in the single mutants only. This included an operon encoding NAD(P)H dehydrogenases, glycolytic enzymes, a pyruvate:ferredoxin oxidoreductase, and protochlorophyllide reductases. Taken together we hypothesize that in the mutant strains there may be issues with the redox status or sensing of the redox status of the cell. The up-regulation of genes in the CcmR regulon was the most striking and in PCC 6803 α-ketoglutarate (αKG) and NADP+ have been shown to be co-repressors of CcmR (Daley et al. 2012). Changes in αKG or NADP+ could be causing derepression of the CcmR regulon, but it seems that there is also CcmR-independent regulation of many genes with related functions. This intriguing pattern of differential gene expression between single and double and triple mutants could suggest that when a single RNase III is absent, cells can compensate with coordinated changes involving CO_2 fixation and redox balance. However, when two or all of the RNase IIIs are absent cellular objectives change resulting in different gene expression patterns.

We noticed coordinated regulation of the r-proteins in almost all of the mutants (**Figure 4.9**). Due to the role of RNase III in rRNA processing, it might make sense to down-regulate r-protein synthesis when there is mis-processing of the rRNA. Interestingly, we saw down-regulation even in samples

where we did not observe any defects in rRNA processing (ΔA0061, ΔA2542, ΔA0061ΔA2542). This is the first example of coordinated regulation of r-protein transcript levels due to the absence of RNase III. Additional regulation of r-protein synthesis is not surprising because r-protein synthesis is highly regulated both through translational feedback (Nomura et al. 1980) and at the level of transcription initiation (Lemke et al. 2011).

We observed extreme down-regulation of several genes located in a prophage. We saw down-regulation of four collinear genes (A1878-A1881) that look like they could be encoded on the same transcript. There are many examples of RNase III processing sites within phage (lambda, T7, T3), and the mini-III in *B. subtilis* was shown to be essential for the degradation of toxin genes that are part of a cryptic prophage (Durand et al. 2012). Additionally, in *E. coli* RNase III is shown to process a prophage encoded transcript into the DicF sRNA that affects cell division and other processes (Balasubramanian et al. 2016).

There were several other groups of genes where we saw coordinated changes in gene expression across RNase III mutants. We saw up-regulation of many high light inducible proteins especially in the triple RNase III mutant as well as other proteins involved in signaling including *sigD* and glutamine synthetase inactivating factors. We saw down-regulation of many genes involved in glycan biosynthesis and metabolism (A1503-A1520). Similarly, we saw down-regulation of many phycobilisome components in all samples except ΔA2542ΔA0384. In this same set of samples we saw up-regulation of several distantly-located phycobilisome degradation proteins (*nblA*'s and *nblB*).

Based on prior work in *E. coli*, we expected to identify distinct processing sites within specific RNAs when comparing samples from mutants to the WT. Instead, we observed significant changes in expression for many genes but no clear spikes in RNA levels that would indicate RNase III target sites. This observation is demonstrated in **Figure 4.12** which highlights the difference in distribution of comparing sequencing coverage between RNase III mutants and WT in both *E. coli* and cyanobacteria. In the cyanobacterial RNase III dataset we observed global, coordinated shifts

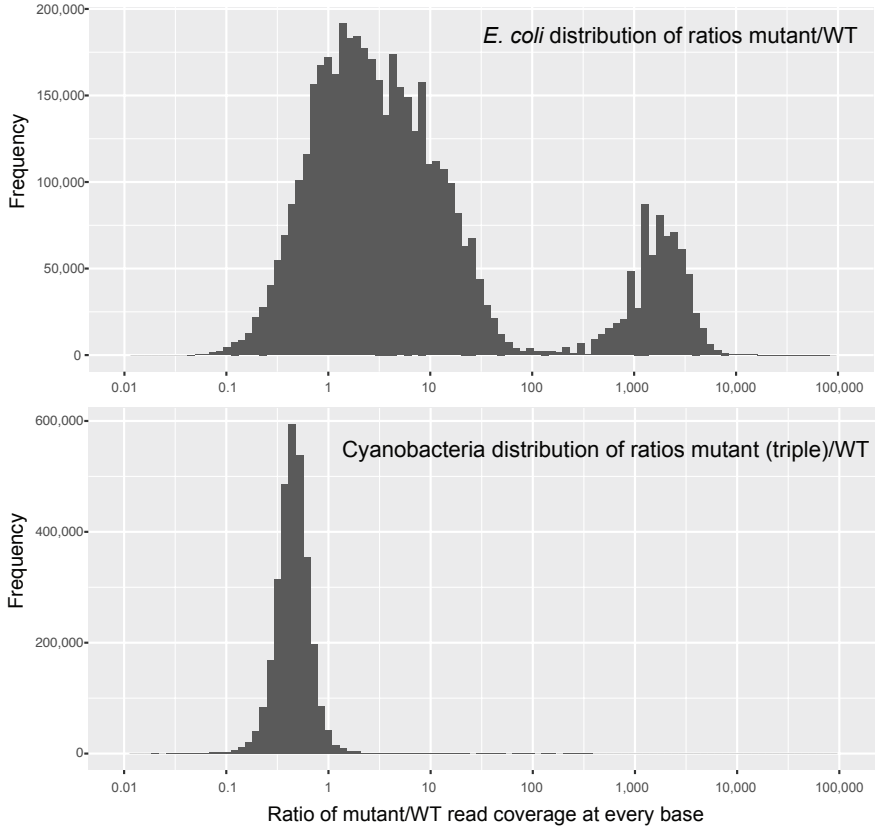


Figure 4.12: Comparison of gene expression between RNase III mutants and WT in both *E. coli* and cyanobacteria. The distribution of the ratio between the RNase III mutant and WT calculated for every base of the genome is shown from both previously published *E. coli* dataset (Chapter 2) and this cyanobacterial dataset. The *E. coli* dataset contains a large number of locations where there is a large ratio of reads in the mutant compared to WT, which included the novel RNase III cleavage sites. This population is entirely absent from the cyanobacterial dataset.

in expression of genes of related functions involve genes that are located at different locations in the chromosome, and we hypothesize that some regulator or group of regulators are responsible for this synchronized movement (transcription factor, sRNA, or other). When we see perturbations of these groups of genes in RNase III mutants, RNase III function could be directly involved in these regulators or in the sensing mechanism, or these perturbations are due to secondary effects of the lack of RNase III cleavage.

The complex changes we see in gene expression could be due to RNA binding activity without

cleavage. In bacteriophage λ RNase III is thought to bind to the upstream region of cIII and exposes it's ribosome binding site, and stimulating translation (Altuvia et al. 1987). There is also the possibility that these RNase IIIs are post-translationally modified, affecting their activity. When *E. coli* is infected with the T7 phage, the phage T7 protein kinase phosphorylates the native RNase III which stimulates it's activity up to 4-fold (Mayer et al. 1983). Another possible explanation for these complex gene expression changes is the existence of RNase III inhibitors. A protein inhibitor was found in *E. coli* (YmdB) that prevents RNase III dimerization and as a consequence, activity (Kim et al. 2008). We found no homolog of YmdB in PCC 7002, but in *E. coli* they observed other unknown YmdB independent inhibition of RNase III.

Effects of RNase III on global gene expression has been observed for a few organisms including *E. coli* and *Staphylococcus aureus* (Stead et al. 2011; Lioliou et al. 2012). RNase III has been shown to auto-regulate its own expression by cleaving a structure in its 5' UTR in many organisms including *E. coli*, *S. aureus*, and *Strepomyces coelicolor* (Lioliou et al. 2012; Bardwell et al. 1989; Xu, Huang, and Cohen 2008). We did not observe this phenomenon in PCC 7002. RNase III is also known to process the 5' UTR of PNPase in several organisms (Jarrige et al. 2001; Gatewood et al. 2011; Haddad et al. 2013), but we found no evidence of this in PCC 7002.

We recently showed the action of RNase III on transcripts encoding several key metabolic enzymes including pyruvate dehydrogenase in *E. coli* (**Chapter 2**). We did not see any RNase III-mediated regulation of any of these metabolic enzyme homologs in PCC 7002. This raises an interesting point that RNase III is involved in the regulation of key metabolic processes and enzymes in both heterotrophs and autotrophs. Transcriptional and post-transcriptional regulation of photosynthesis has been shown to be extremely important in physiological response and signaling shifts in response to changing conditions, and RNase IIIs may be a very important player influencing these interactions.

Cleavage activity and specificities of RNase III homologs

Without putative PCC 7002 RNase III targets, we tested the ability of cyanobacterial enzymes to cleave known *E. coli* targets in vitro in effort to investigate conservation of RNase III targets across species. A0061 and A2542, both full-length enzymes, could cleave many of the *E. coli* targets, but not all. Additionally, A0061 could cleave the *pflBA* target while A2542 could not. By extracting RNA after the in vitro cleavage reaction and performing 5' RACE we pinpointed the exact location of these cleavage events. For the *E. coli pnp* target, both A0061 and A2542 cleaved at the same location where the *E. coli* RNase III enzyme did (**Figure 4.10**). For all other targets we tested, A0061 and A2542 cut in nearby but not identical locations as the *E. coli* enzyme (ranging from -21 to +3 bases). For one transcript, *aceEF*, A0061 and A2542 cut at different locations within the same stem loop structure. We were not surprised to see that A0384, the mini-III, could not cleave any of these *E. coli* targets because the *E. coli* enzyme is a full-length RNase III. Expanding the list of specific targets of many different RNase IIIs may one day reveal the complex substrate enzyme recognition pairing and allow for a more predictive approach to finding RNase III target sequences.

Conclusion

We have identified the roles of two homologs of RNase III in 23S rRNA processing, one homolog's role in copy number control of one plasmid, and highlighted the complex global changes in gene expression involving CO_2 fixation, redox status, and light harvesting. We hypothesize that many of these changes are indirect effects of RNase III activity and indicate a role of RNase IIIs in either sensing environmental changes or mediating changes. This work highlights the important role of post-transcriptional regulation and its influence on gene expression. We showed both specific and redundant activities of these RNase III's on in vitro targets and show that while both A0061 and A2542 can cleave the duplex structure controlling plasmid copy number regulation in vitro, only A2542 acts on this duplex in vivo. This work highlights the continued need for in vivo

datasets to tease apart specificities and unravel the complex gene regulatory network enabled by post-transcriptional RNA processing.

4.4 Materials and Methods

Sample Collection

All strains used in this study are listed in **Table 3.2**. A detailed description of the creation of the RNase III mutant cyanobacterial strains is outlined in **Chapter 3**, but in summary, antibiotic resistance markers flanked by 500 bp homology arms were introduced via natural transformation. After passaging on the appropriate antibiotic(s), the segregation of the strains was verified. Cyanobacteria were cultured on media A^+ agar (1.5%) plates with the appropriate antibiotic or antibiotics (gentamicin 30 µg/ml, kanamycin 100 µg/ml, and streptomycin 100 µg/ml). Pre-cultures of 20 mL A^+ media were bubbled with air overnight. The OD_{730nm} was measured and 1L pyrex bottles filled with 900 mL media A^+ was inoculated with pre-culture for an OD_{730nm} = 0.01. Cultures were bubbled with air for 24 hours before harvesting 40 mL and adding 10 mL of cold stop solution (10% phenol in ethanol). Samples were centrifuged at 10,000 x g at 4°C for 10 minutes in a Beckman Coulter Avanti J-E Centrifuge. The resulting cell pellets were frozen in liquid nitrogen and stored at -80°C until further processing.

RNA isolation and RNA-sequencing

RNA was extracted as described previously (Khodurksy et al. 2003) and submitted to the University of Wisconsin Biotechnology Gene Expression Center for RNA-sequencing. rRNA was removed with a Ribo-Zero Magnetic Kit. A cDNA library was created using a TruSeq Stranded Total RNA Library Kit. DNA libraries were sequenced using an Illumina HiSeq 2500 (1x100). Samples had between 5 to 22 million reads, sufficient coverage for bacterial differential gene expression analysis (Haas et al. 2012).

Differential Gene Expression Analysis

Sequencing files were aligned to the PCC 7002 chromosome (NC_010475), pAQ1 (NC_010476), pAQ3 (NC_010477), pAQ4 (NC_010478), pAQ5 (NC_010479), pAQ6 (NC_010480), and pAQ7 (NC_010474) using Bowtie2 (v 2.2.6) and samtools (v 1.2). Counts for each feature in the general feature format files were obtained with HTSeq (v 0.6.1) (Anders et al. 2015) using the htseq-count script set on union mode and stranded set to reverse (HTseq v 0.6.1). edgeR (v 3.16.5) was used to test for differential gene expression (Robinson, McCarthy, et al. 2009). Specifically, feature counts were loaded and filtered for all genes that had at least 1 read per million in 3 samples and tested for differential expression with exactTest using tagwise dispersion. P-values were adjusted with the Benjamini and Hochberg algorithm (Benjamini et al. 1995) to estimate the false discovery rate (FDR) and genes with a FDR < 0.005 were classified as differentially expression. Raw sequence files, feature counts, and differential expression fold change data was deposited in the Gene Expression Omnibus (GSE99279). Raw data files have also been uploaded to the NCBI Sequence Read Archive (accession number SRP107964).

Expression and Purification of *E. coli* RNase III and PCC 7002 RNase IIIs

E. coli RNase III (Rnc) and three PCC 7002 RNase III homologs were amplified with primers 4.2 and inserted into pET28b using Gibson assembly (Gibson et al. 2009). Plasmids (4.3) were verified by sequencing and transformed into chemically competent *E. coli* BL21 (DE3) cells and selected on LB kanamycin plates. Cultures of *E. coli* BL21 strains containing plasmids (5 mL LB with 50 µg/mL kanamycin) grown for 16 hours were used to inoculate 600 mL of LB (50 µg/mL kanamycin) in a 2 L flask. Cultures were grown at 37°C (250 rpm) until at OD_{600nm} ∼0.3-0.4 and induced with 1 mM IPTG. 1 mL of culture was collected 0, 1, 2, 3, and 4 hours after induction to check overexpression on a SDS-PAGE gel. After 4 hours, cultures were spun down at 6,000 x g for 20 minutes at 4°C (Beckham Coulter Avanti J-E Centrifuge), aspirated, and stored at -80°C.

Purification was performed according to published methods (Amarasinghe et al. 2001), with the following deviations. Benzonase (2 µL) was added to the cells resuspended in lysis buffer and sonication was done in short bursts using a Fisher Scientific Sonic Dismembrater Model 500. Cell debris was removed by centrifugation in a Beckman Coulter Avanti J-E Centrifuge at 12,000 rpm at 4°C for 30 minutes and filtered through a 0.45 µm filter. RNase III variants were purified using an Äkta start purification system (GE) with a 1 ml HisTrap^{TM} HP column (GE Healthcare). Briefly, the column was equilibrated with 10 volumes of buffer A + 5 mM imidazole and the clarified cell lysate was added (1mL/minute). The column was washed with 15 column volumes of buffer A+ 5mM imidazole and then 10 column volumes with buffer A + 60mM imidazole. Protein was eluted with a gradient (10 column volumes) and collected in 0.5 mL fractions. A SDS-PAGE gel was run to visualize the fractions and the 5 most concentrated and pure fractions were combined. A PD-10 Desalting Column (GE Healthcare Life Sciences) was used according to manufacturer's instructions to transfer the protein into storage buffer containing 50% glycerol.

In vitro RNA cleavage assays and cleavage site mapping

RNA cleavage assays and 5' RACE were performed as previously described (**Chapter 2**). All cleavage reactions were performed as time courses with aliquots collected without enzyme, without magnesium chloride, and 5, 10, 20, 30, and 60 minutes after magnesium chloride addition. Some purified 5' RACE products were subcloned into pGEM-T Easy vectors (Promega) before sequencing.

Plasmid Copy Number Quantification

PCC 7002 (WT and ΔA2542) were cultured overnight in 20 mL media A^+ in bubble tubes and diluted to an OD_{730nm} = 0.01 in 150 mL media A^+ in 250 mL pyrex bottle with glass rod for bubbling (in quintuple). Cultures were grown for 24 hours before 40 mL was collected and centrifuged for 20 minutes at 10,000 x g in a Beckham Coulter Avanti J-E Centrifuge at room

temperature. Supernatant was removed, and samples were stored at -80°C until processing (10 days). A linear DNA spike-in was made by amplifying ϕX174 Virion DNA (NEB, N3023S) with primer JCC425 and primer JCC426 and purifying the product via ethanol precipitation. The spike-in was added immediately after addition of the TE (with lysozyme) to the frozen cell pellets at a concentration of 10 copies per cell. Cell counts in each sample were determined by measuring OD_{730nm} and converting through a standard curve of cells/mL (hemocytometer counts) vs OD_{730nm}. DNA was isolated from four replicates as in (Khodurksy et al. 2003), except that the pH of the phenol solution was adjusted to 8 using Tris alkaline buffer and purified DNA was diluted 1:1000. The relative copy number was determined by qPCR using an AriaMx Real-Time PCR System. Reactions consisted of 5 µL iQ SYBR Supermix (Bio-Rad, #1708880), 0.3 µL 10 µM primer mix, 3.7 µL of water, and 1 µL of template or standard. Cycling conditions consisted of 95°C for 3 minutes, 95°C for 15 seconds, 55°C for 30 seconds, 72°C for 1 minute, repeat to step 2 for a total of 35 times, followed by a melt curve from 55°C to 95°C. All samples and standards were run in technical duplicate. Analysis was done using Agilent Aria Software v1.3 and Cq values were determined with an absolute threshold. Sample Cq values were compared to standard curve of 5 10-fold dilutions of purified DNA templates (pAQ3: R^2 = 0.998, efficiency = 109.9%, slope = -3.11, y-intercept = 1.46; gene F: R^2= 0.999, efficiency = 101.4%, slope = -3.29, y-intercept = 4.64). Absolute quantification was calculated by dividing the ratio of pAQ3 to gene F spike-in. No template controls were well above standard curves (23 and 29 for pAQ3 and gene F respectively). Primer GG607 and primer GG608 were used to detect pAQ3 (116 bp product) and primer JCC437 and primer JCC438 were used to detect control spike-in gene F (98 bp product). Primers were designed with Primer-Blast (NCBI).

Amplifying 23S rRNA extensions and verification of ctRNA expression

1 µg of total RNA extracted from WT, ΔA0384, and ΔA0061ΔA2542ΔA0384 was used to create cDNA with the iScript^{TM} cDNA Synthesis Kit (Bio-Rad #1708891) using random primers. Control

reactions lacking reverse transcriptase were simultaneously performed. PCR reactions were set up using 5 µL GoTaq Green Master Mix (Promega #M7123), 1 µL each of forward and reverse primer (10 µM), 2 µL of water, and 1 µL of cDNA template. Reactions were run on a ProFlex PCR system with the following cycling conditions: 95°C for 3 minutes, 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute, repeat to step 2 for a total of 25 times, 72°C for 10 minutes, and storage at 4°C. PCR products were visualized on a 1% agarose gel with ethidium bromide.

For verification of ctRNA expression, aliquots were taken from WT and ΔA2542 cultures mentioned above, mixed with cold stop solution (10% phenol in ethanol), and stored for 2 weeks at -80°C until RNA was extracted using the TRIzol 95 method (Pinto et al. 2009). RNA was quantified using the Qubit RNA high sensitivity assay kit, and 1 µg of RNA and primer GG660 was used to create cDNA using the GoScript^{TM} reverse transcriptase kit (A5000, Promega) according to manufacturer instructions. The presence of the 86 bp product was detected using the same PCR reaction and conditions as mentioned above using primer GG660 and primer GG661.

Oligomerization of Purified RNase IIIs

The ability of purified RNase Ii I's to form oligomers was tested by incubating purified enzyme with a primary amine crosslinker and running on a SDS-PAGE gel. 0.5 µg of purified protein was incubated in a reaction volume of 20 µL containing conjugation buffer (100 mM Hepes pH 7.4, 250 mM NaCl, 0.1 mM EDTA, and 0.1 mM EDTA). Disuccinimidyl suberate (DSS) was added and samples were incubated at room temperature for 30 minutes before quenching with loading buffer (60mM Tris, β-mercaptoethanol, and SDS-PAGE loading buffer). Samples were run on a 12 or 18% SDS-PAGE gel.

Table 4.2: Primers used to study role of RNase III homologs in cyanobacteria.

Name	Purpose	Sequence
JCC419	Insert Rnc into pET28b	CATCACAGCAGCGGCATGAACCCCATCGTAATTAATC
JCC422	Insert Rnc into pET28b	CGGAGCTCGAATTCGGATCCTATCATTCCAGCTCCAGTTTTTTC
GG553	Insert A0061 into pET28b	ATCATCATCACAGCAGCGGCATGGTTCCTCCCCAACGT
GG537	Insert A0061 into pET28b	GGAGCTCGAATTCGGATCCTACTATGGCTGAGGCGTGAA
GG534	Insert A2542 into pET28b	ATCATCATCACAGCAGCGGCATGACAAGCTTGCCTCCC
GG538	Insert A2542 into pET28b	GGAGCTCGAATTCGGATCCTATTAAGGGGATTGGCGCG
GG558	Insert A0384 into pET28b	CATCATCATCACAGCAGCGGCATGTTTGTCCCTGACATTACCG
GG539	Insert A0384 into pET28b	GGAGCTCGAATTCGGATCCTATCAGACCCCAGACCGTTC
JCC418	Gibson backbone for pET28b	GCCGCTGCTGTGATGATGAT
GG536	Gibson backbone for pET28b	TAGGATCCGAATTCGAGCTCC
JCC425	DNA template gene F	CGAAATTAATACGACTCACTATAGGGAGGATAAATTATGTCTAA TATTCAAACTG
JCC426	DNA template gene F	TCACGAAGTCATGATTGAATCG
GG607	qPCR pAQ3	GAGTCGGAACGGTCTTGGTT
GG608	qPCR pAQ3	ATCTCCTCCGGTTCCCGTAA
JCC437	qPCR gene F	TTTTAAAGCGCCGTGGATGC
JCC438	qPCR gene F	TTTTGAGATGGCAGCAACGG
GG660	ctRNA expression F	GCGAAACCCCCAAATCTTG
GG661	ctRNA expression R	ACAAGTCGGCCCCCATTT

Table 4.3: Plasmids used to over expression RNase III enzymes

Name	Antibiotic Resistance	Purpose
pGCG90	kanamycin	overexpress His-tagged *E. coli* Rnc
pGCG81	kanamycin	overexpress His-tagged A0061
pGCG82	kanamycin	overexpress His-tagged A2542
pGCG83	kanamycin	overexpress His-tagged A0384

5 Chapter 5: Global RNA decay in cyanobacteria

Authors and Contributors:

Gina Gordon: Assisted in experimental design, collected samples, processed and analyzed RNA-sequencing data, calculated RNA half-lives, wrote manuscript.

Jeffrey Cameron: Conceptualized the experiment, conducted sampling and RNA extraction, and provided interpretations of some results.

Dr. Brian Pfleger: Supervised the research and edited the manuscript.

This chapter is being prepared for submission.

5.1 Introduction

RNA degradation is a fundamental process that controls the amount of messenger RNA present. It inherently controls gene expression because the level of any given transcript is controlled by the amount of its synthesis and degradation. A suite of enzymes called ribonucleases (RNases) facilitate RNA degradation. Although many RNases have been characterized, there is limited information on their precise in vivo targets and how they together contribute to the global turnover of mRNA.

Transcripts are rapidly turned over in prokaryotes, allowing for rapid responses to environmental conditions via differences in transcription. Half-lives of transcripts and genes are on the magnitude of minutes in prokaryotes: *Escherichia coli* 2.5 minutes, 4.7 minutes, and 6.8 minutes (Bernstein, Khodursky, et al. 2002; Chen, Shiroguchi, et al. 2015; Selinger et al. 2003); *Bacillus subtilis* 5 minutes (Hambraeus, von Wachenfeldt, et al. 2003); *Bacillus cereus* 2.4 minutes (Kristoffersen et al. 2012), *Lactococcus lactis* 5.8 minutes (Redon et al. 2005), *Mycobacterium tuberculosis* 9.5 minutes (Rustad et al. 2013), and *Chlamydia trachomatis* trachoma 15-17 minutes (Ferreira et al. 2017). Most of these reported values were calculated from microarray experiments, with more recent values calculated from RNA-sequencing data.

Global RNA turnover was recently studied in a slow-growing cyanobacteria, *Prochlorococcus* MED4. Originally it was proposed that RNA turnover rate could be linked with the growth rate of the organism, but *Prochlorococcus* MED4 had a rapid median half-life of 2.4 minutes despite only doubling once or twice per day (Steglich et al. 2010). We chose to examine RNA degradation in another cyanobacterium, *Synechococcus* sp. strain PCC 7002 that has one of the fastest doubling times of cyanobacteria ($\sim$2.5 hours) (Bernstein, McClure, et al. 2016). This strain of cyanobacteria has industrial relevance to create green chemical or protein products from CO_2 and sunlight, and there also has been a significant amount of research addressing aspects of gene expression and cellular physiology to build better genetic tools.

Another reason motivating the study of RNA degradation in PCC 7002 is its unique array of RNases that facilitate both RNA degradation and maturation. The functions of three homologs of

RNase III have been studied (**Chapter 4**), but PCC 7002 also has both a homolog of RNase E (a hallmark of RNA degradation in *E. coli*) and RNase J (a hallmark of RNA degradation in *B. subtilis*) (**Chapter 3**). The impact of the presence of both enzymes on RNA turnover remains unexplained.

In this manuscript we collected samples and did RNA-sequencing to track RNA degradation in PCC 7002 throughout time. We calculated global RNA half-lives of genes and examined how gene half-life may be related to cellular function and what sequence features may enhance transcript stability.

5.2 Global RNA degradation in PCC 7002

We examined global RNA turnover in PCC 7002 by collecting samples after the addition of the antibiotic rifampicin to stop nascent transcription. Samples were taken from 3 biological replicates before the addition of the antibiotic (0 minutes), and 0.5, 1, 2.5, 5, 7.5, and 10 minutes after addition. RNA spike-ins were added to cell pellets immediately before lysis to normalize for potential large differences in RNA pool sizes.

By aligning the number of reads that aligned to each coding sequence and normalizing between samples by the number of RNA spike-ins, we fit an exponential decay model and determined the half-life of each coding sequence. We were able to calculate half-lives for 2,949 genes (91.1%) which had a median half-life of 0.97 minutes (**Figure 5.1**). The average half-life was 1.18 minutes (standard deviation 0.73 minutes). We classified 0.5% of the genes as stable that had greater than 50% of the original read counts in the final timepoint (10 minutes). We could not calculate half-lives for 4.4% of genes because of no or low expression, and 4.1% that exhibited a poor fit of decay.

We examined the half-life of genes grouped by cellular function, and found that genes involved in energy metabolism, had a significantly longer half-life than all other groups (**Figure 5.2A**, pairwise t-test p-values between 9×10^{-6} and 2×10^{-16}, Bonferroni adjusted). Only cellular functions that contained greater than 50 genes were included. Within the energy metabolism subgroup, the photosynthesis genes had a significantly longer half-life than all other subgroups (**Figure 5.2B**,

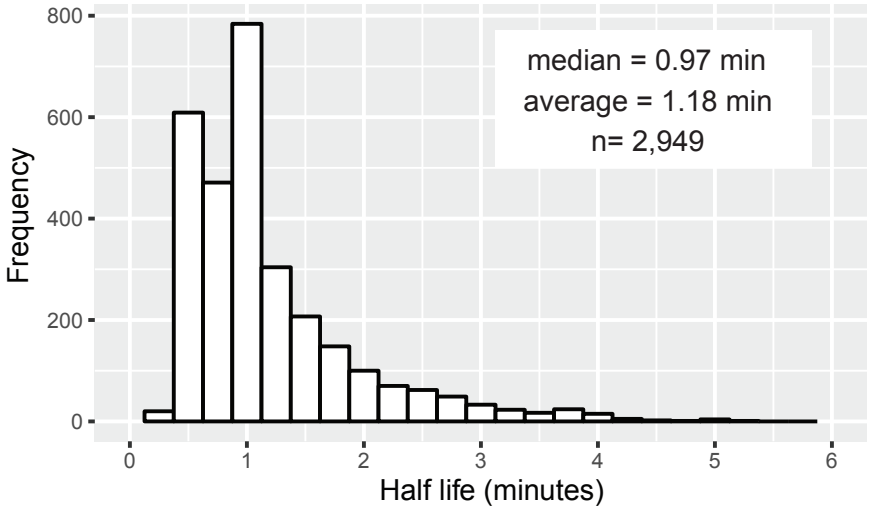


Figure 5.1: Distribution of gene half-lives in PCC 7002. The median half-life of all genes that had a calculated half-life (n = 2,949) was 0.97 minutes. The average half-life was 1.18 minutes (standard deviation 0.73 minutes).

pairwise t-test p-values all below <0.05, Bonferroni adjusted).

We looked for a relationship between gene expression level and half-life and saw no overall trend (R=0.127, Pearson correlation), but photosynthesis genes seemed to be both highly expressed and have a long half-life (**Figure 5.3A**) (R=0.400, Pearson correlation). Genes involved in energy metabolism had a significantly higher expression level than genes with other cellular functions (**Figure 5.3B**, pairwise t-test p-values between $2x10^{-16}$ and $3x10^{-11}$, Bonferroni adjusted), but photosynthesis genes did not have a higher expression level than those within the energy metabolism subgroup (**Figure 5.4**). Instead, only the antenna protein group had statistically significant higher expression than all other subgroups besides nitrogen metabolism.

While examining the read coverage of our highly stable genes, we made the observation that they had very distinct start and stop locations (**Figure 5.5A**). We looked for motifs in both the upstream (150 bp) region and downstream region (100 bp) of each gene that were enriched for genes with long half-lives compared to short half-lives. Using MEME (Bailey et al. 1994) in discriminative mode, we looked for motifs overrepresented in the sequences surrounding the 400 most stable genes

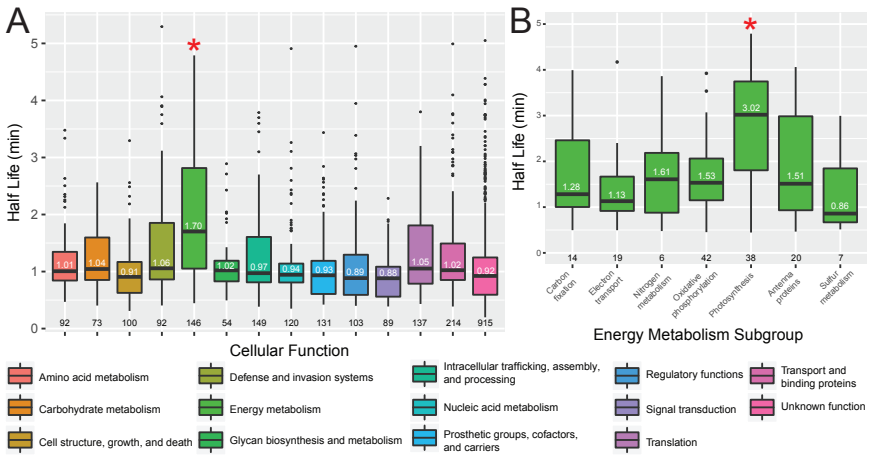


Figure 5.2: Photosynthesis genes have a longer half-lives than genes of other cellular functions. (A) Half lives of genes grouped by cellular function. Only cellular functions that contained greater than 50 genes were included. The median half-life is displayed on the boxplot and the number of genes in each group is displayed below. The energy metabolism group had a significantly greater half-life than all other groups (*, pairwise t-test p-values between $9x10^{-6}$ and $2x10^{-16}$, Bonferroni adjusted). (B) Half lives of genes in the energy metabolism subgroup. The photosynthesis group had a significantly greater half-life than all other groups (*, pairwise t-test p-values all below <0.05, Bonferroni adjusted).

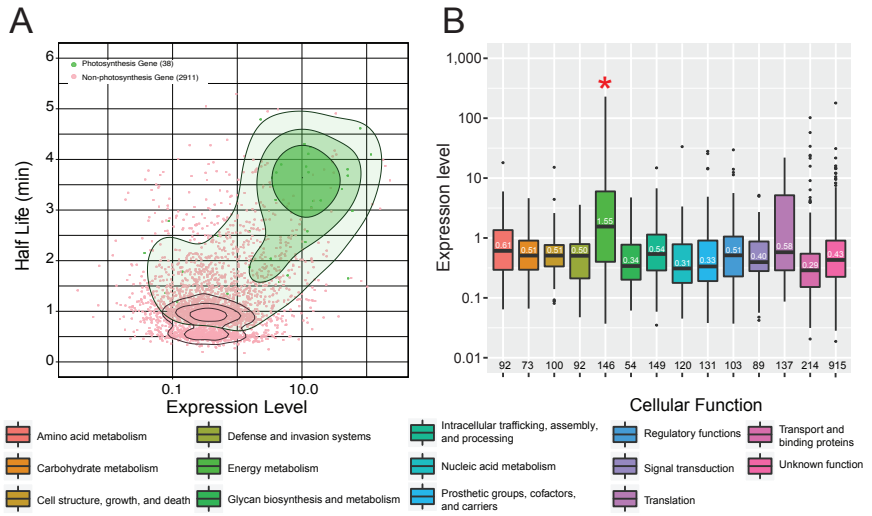


Figure 5.3: High expression of photosynthesis genes. (A) A portion of photosynthesis genes are expressed at a high level and also have a long half-life. The half-life and expression level of photosynthesis genes (green) and non-photosynthesis genes (pink) is shown along with contours of the 2D kernel density estimation. (B) have a higher expression level than genes with other cellular functions (*, pairwise t-test p-values between $2x10^{-16}$ and $3x10^{-11}$, Bonferroni adjusted). The median expression level is displayed on the box plot and the number of genes in each group is displayed below.

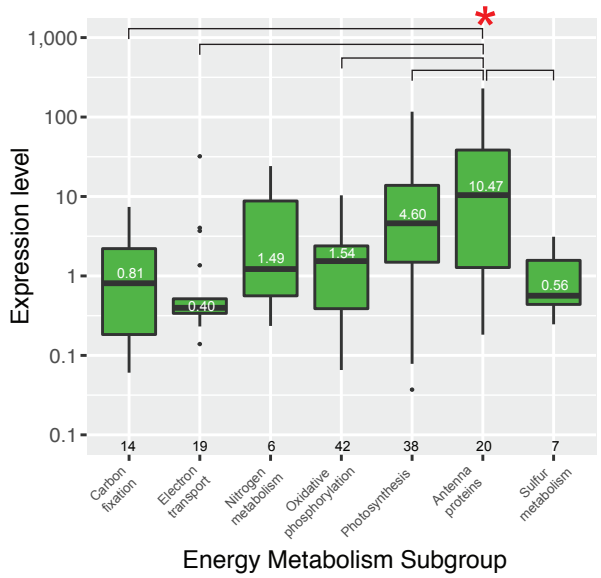


Figure 5.4: Expression level within energy metabolism subgroup. The antenna proteins group had a significantly higher expression level than all other groups besides nitrogen metabolism (*, pairwise t-test p-values < 0.05, Bonferroni adjusted).

compared to the control 400 least stable genes. No significant motifs were identified in the upstream region, but an enriched motif was found in the downstream region (**Figure 5.5B**). This 21-bases long motif had an E-value of $5.6e^{-83}$ and featured a prominent string of six consecutive 'T's.

If this motif was transcribed into RNA, it would have significant similarities to rho-independent terminators that feature a G-C rich hairpin followed by a U-rich region, so we examined if the presence of a terminator in the 3' untranslated region (UTR) influenced transcript stability. Using the Web Genome Scanner for Terminators (WebGeSTer) database (Mitra et al. 2011), we examined if genes with different types of terminators had different half-lives. We divided genes into three classes based on the absence of a terminator, presence of a L-shaped terminator with a U-tract, or the presence of a I-shaped terminator lacking a U-tract. We found that genes that had a L-shaped terminator containing a U-tract in the 3' UTR had a significantly higher half-life than both genes that contained a I-shaped terminator in the 3' UTR or lacked a terminator (**Figure 5.5D**).

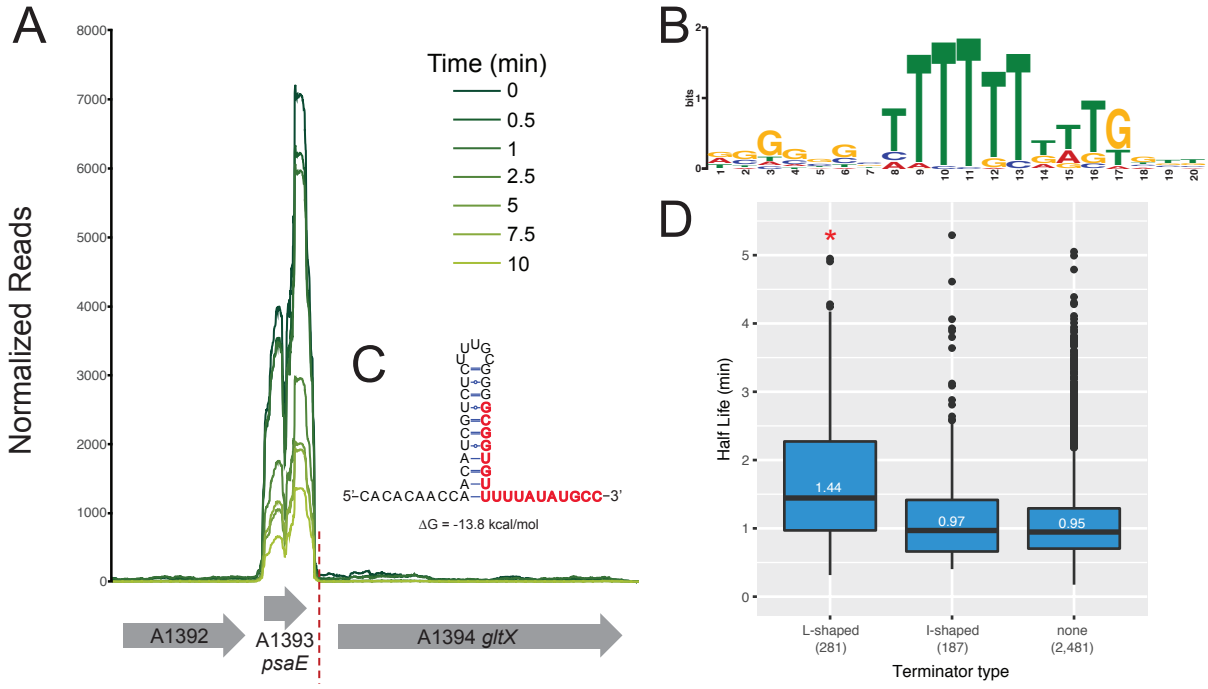


Figure 5.5: L-shaped rho-independent terminators enriched in stable genes. (A) RNA degradation of photosynthesis gene, A1393 (*psaE*) throughout time. The number of normalized reads aligning to each base is shown in the samples taken from different timepoints following addition of rifampicin (dark green to light green). (B) The motif enriched in the 100 bases downstream of the top 400 stable genes compared to the 400 least stable genes (MEME E-value of $5.6e^{-83}$). (C) The terminator sequence of A1393 predicted by RNAMotif (Macke et al. 2001) that contains the enriched motif. (D) Genes with a L-shaped terminator in the 3' UTR have an increased half-life compared to genes with a I-shaped terminator or those that lack a terminator (*, pairwise t-test p-values: L vs I: $4.8x10^{-13}$ and L vs none $<2x10^{-16}$, Bonferroni adjusted).

5.3 Discussion

RNA turnover in PCC 7002 is extremely rapid with a median half-life of coding regions of 0.97 minutes (∼58 seconds), faster than any other report of bacterial mRNA decay. Only 15 genes were classified as stable (retaining at least 50% of the read counts from the initial pre-rifampicin timepoint). Recent experiments using RNA-sequencing to study RNA degradation (Chen, Shiroguchi, et al. 2015; Kristoffersen et al. 2012) have shown that mRNA half-lives are much lower than original values determined by microarrays (Bernstein, Khodursky, et al. 2002; Hambraeus, von Wachenfeldt, et al. 2003; Selinger et al. 2003), but RNA degradation occurs more rapidly in PCC 7002 than expected.

We examined if any groups of genes that encode proteins performing different cellular functions had significantly longer half-lives. We found that genes involved in energy metabolism, specifically photosynthesis had a significantly longer half-life than all other groups. We also examined if there was a correlation between RNA expression level and mRNA half-life. Although overall there was low correlation between expression level and half-life for all genes (R=0.127, Pearson correlation) there was a slightly higher correlation between expression and half-life for photosynthesis genes (R=0.400, Pearson correlation). Genes within the energy metabolism group had a significantly higher expression level than genes with other cellular functions. The photosynthesis genes did not have a significantly higher expression level than other subgroups within energy metabolism; the antenna proteins had a statistically significant higher expression level than all other subgroups. The high expression of photosynthesis genes and all genes falling under the energy metabolism category is not by itself surprising, as energy metabolism genes encode proteins that perform this fundamental process. There are only 53 genes categorized as photosynthesis genes, of which half-life values could be determined for 38 and 2 were categorized as stable. Many of these 40 genes are both highly expressed and very stable making it hard to differentiate between if being inherently stable increases expression level and/or having high expression protects the transcript from degradation.

We wanted to identify sequence elements that confer stability, so we searched for motifs at the

beginning and end of the transcript. We used 150 bases upstream of the start codon and 150 bases downstream of stop codon because transcriptional start and stop sites are unknown. An enrichment of a sequence element in either of these regions in the most stable genes compared to the least stable genes could help explain transcript stability. We identified an enriched motif in the downstream region of stable genes that would be transcribed into a string of 'U's. This motif resembles canonical rho-independent terminators that consist of a GC-rich hairpin followed by a string of 'U's. The location of these motifs had very significant overlap with computationally predicted rho-independent terminators (Gautheret et al. 2001; Macke et al. 2001). We then explored the possibility that the presence or absence and type of terminator present may influence the stability of a transcript.

Intrinsic, or rho-independent termination, has classically been thought to consist of a GC-rich palindrome followed by a string of 'T's on the sense strand, which is transcribed into a U-tract. Intrinsic termination occurs through several steps, starting with a pause that allows the recently transcribed RNA hairpin to form within the RNA exit channel of RNA polymerase. This hairpin then disrupts the complex by unknown mechanisms (Peters et al. 2011). The U-tract has traditionally been thought to stimulate release of the transcript due to the low stability of the A-U hybrid. The presence of the GC-rich hairpin seems essential because when it is removed almost all terminator activity is eliminated, but the role and necessity of the U-tract has been controversial because its disruption only led to reduced activity (Abe and Aiba 1996).

Early algorithms designed to identify intrinsic terminators searched for the canonical structures found in *E. coli* intrinsic termination: the GC-palindrome and the U-tract. Surprisingly, in many organisms no terminators were found and it was suggested that they relied on some yet uncharacterized termination method (Washio et al. 1998). More recent algorithms designed to address this issue found that many of these organisms contain hairpin structures but lack the canonical U-tract, and were thus overlooked (Unniraman et al. 2002). Terminators have been classified into several groups with the major being L-shaped (hairpin and U-tract) and I-shaped (hairpin with no U-tract). We found that genes with L-shaped terminators had a significantly longer half-life than those with

an I-shaped terminator or lacked a recognizable terminator in the 3'UTR.

The observation that an L-shaped terminator may enhance transcript stability in PCC 7002, but not an I-shaped terminator indicates the U-tract may play an important role in determining transcript stability. This could be a direct effect through the affinity and activity of exoribonucleases on a terminal U-tract or could be the indirect, causing a difference in termination efficiency. It has been suggested that the T-tract serves as a pausing signal (possibly through backtracking) that facilitates the folding of the hairpin, irreversibly trapping the RNA polymerase complex, and disrupting all key contacts between RNA, DNA, and RNA polymerase (Gusarov 1999). It is possible that the presence or absence of the T-tract determines termination efficiency, and the presence or absence of the poly-U's at the end of a transcript influence transcript stability. The distinction between different rho-independent termination structures could be quite important in cyanobacteria because without a known homolog of rho (Vijayan et al. 2011), intrinsic termination is likely the only termination method.

The possibility that the presence of a terminator may enhance stability of a transcript is not unprecedented. The impact of both the hairpin and U-tract on transcript stability of the *E. coli crp* transcript was analyzed, and the removal of the of the G-C hairpin enhanced transcript degradation, but disruption of the U-tract did not affect transcript stability. Additionally, swapping out the terminator for other native rho-independent terminators did not change expression level or transcript stability (Abe and Aiba 1996). The presence of RNA hairpins at the 3' end of transcripts may act as barriers to exonuclease degradation, as several 5' to 3' exonucleases cannot degrade secondary structure (RNase II) or require a toehold such as polyadenylation to initiate degradation (RNase R and PNPase) (Condon 2007). It remains to see if these RNases act differently on L or I-shaped terminators, and thus influencing transcript stability.

Several aspects of this data collection and workflow were fundamental to obtaining high quality half-lives for most of the PCC 7002 genome. The short time-frame of sampling was extremely important for capturing rapidly decaying messages. Most studies of RNA degradation have taken

an initial timepoints and then not collected the first samples until 2.5 or 5 minutes after addition of rifampicin. In all of these studies, the overall median or average half-life is likely skewed as much of the early mRNA degradation may have been missed. The only study besides this one that included a 0.5 minute timepoint was that done in *E. coli*, and that also had an extremely low median half-life of 2.5 minutes (Chen, Shiroguchi, et al. 2015).

Another critical aspect of this dataset is the use of RNA spike-ins. The only other RNA-sequencing experiment to study transcript degradation that utilized RNA spike-ins was (Chen, Shiroguchi, et al. 2015). The typical RNA-sequencing pipeline normalizes to the total number of reads from each sample, which can greatly distort the results when the pool size of mRNA is shrinking throughout the time course. RNA spike-ins can be added proportional to the number of cells to normalize both for the change in RNA pool size between samples, and for incidental differences between samples introduced during library preparation. The need for RNA spike-ins to correct for differences in RNA or DNA yields from samples has been addressed previously (Chen, Hu, et al. 2016), but their use has still been quite limited. A standardized set of spike-ins has been developed for eukaryotes (the ERCC mix of 96 poly-adenylated transcripts), but there is no standardized set for use in prokaryotes. We hope that RNA spike-ins become standard for future RNA degradation studies, and that the methods described in this manuscript and (Chen, Shiroguchi, et al. 2015) will enable robust measurements of RNA half-lives in other prokaryotes.

This dataset can also be probed to ask other fundamental questions about RNA degradation such as the potential directionality of transcript degradation. Early *E. coli* measurements using microarrays indicate degradation occurs in a 5' to 3' direction by analyzing operons containing 2 or more genes (Selinger et al. 2003). Similar findings were seen in slow-growing cyanobacteria *Prochlorococcus* MED4, where the farther a gene in an operon was from the transcription start site the more stable it was (Steglich et al. 2010). This 5' to 3' directionality was also seen within the genes not located within an operon. For *B. cereus*, they observed a mix of 5' to 3' degradation, 3' to 5' degradation, and more complex patterns (Kristoffersen et al. 2012). We also observed this 5' to

3’ directionality of degradation where the half-life increased with distance along an operon (**Figure 5.6**).

A major factor influencing the observation that transcript degradation is occurring in a 5’ to 3’ direction is the use of rifampicin to stop transcription. Rifampicin binds the β' subunit of RNA polymerase and blocks the channel for the elongating RNA (Campbell et al. 2001). This blockage prevents extension of the RNA beyond 2 or 3 nucleotides, but has no effect on elongating RNA polymerases. (Chen, Shiroguchi, et al. 2015) noticed there was a delay before transcripts began to decay for locations downstream of the transcription start site. They showed that this delay disappeared when streptolydigin (an antibiotic that prevents transcription elongation) was used in place of rifampicin. The change in this delay due to residual transcription was used to calculate the elongation rate, on average 25 nucleotides per second. The effect of residual transcription after rifampicin addition was estimated to be less than 30 seconds for *E. coli* (Pato et al. 1970).

Despite these drawbacks, rifampicin has been used in all RNA-sequencing and microarray studies to examine transcript degradation due to its availability and cost. The confounding influence on half-life calculations may be minimal when initial timepoints are not taken for 2.5 or 5 minutes after rifampicin addition. It may significantly influence calculated half-lives when samples are taken early and there are measurements of rapidly decaying messages. This effect would be even more exaggerated with longer global transcription elongation rates. There is little that is known about cyanobacterial transcription rates, but in vitro data comparing RNA polymerase from *Synechococcus* sp. strain PCC 7002 and *Thermosynechococcus elongatus* BP-1 with that from *E. coli* showed that the cyanobacterial enzymes took six times longer to complete the same transcript (Imashimizu et al. 2011). Additionally, there were significant differences in levels of abortive transcription and misincorporation. The global differences in transcriptional machinery and transcription rates must also be considered when comparing global degradation rates between organisms. We hypothesize that the 5’ to 3’ directionality we see in degradation in this cyanobacterial dataset may be entirely due to the use of rifampicin and residual transcription.

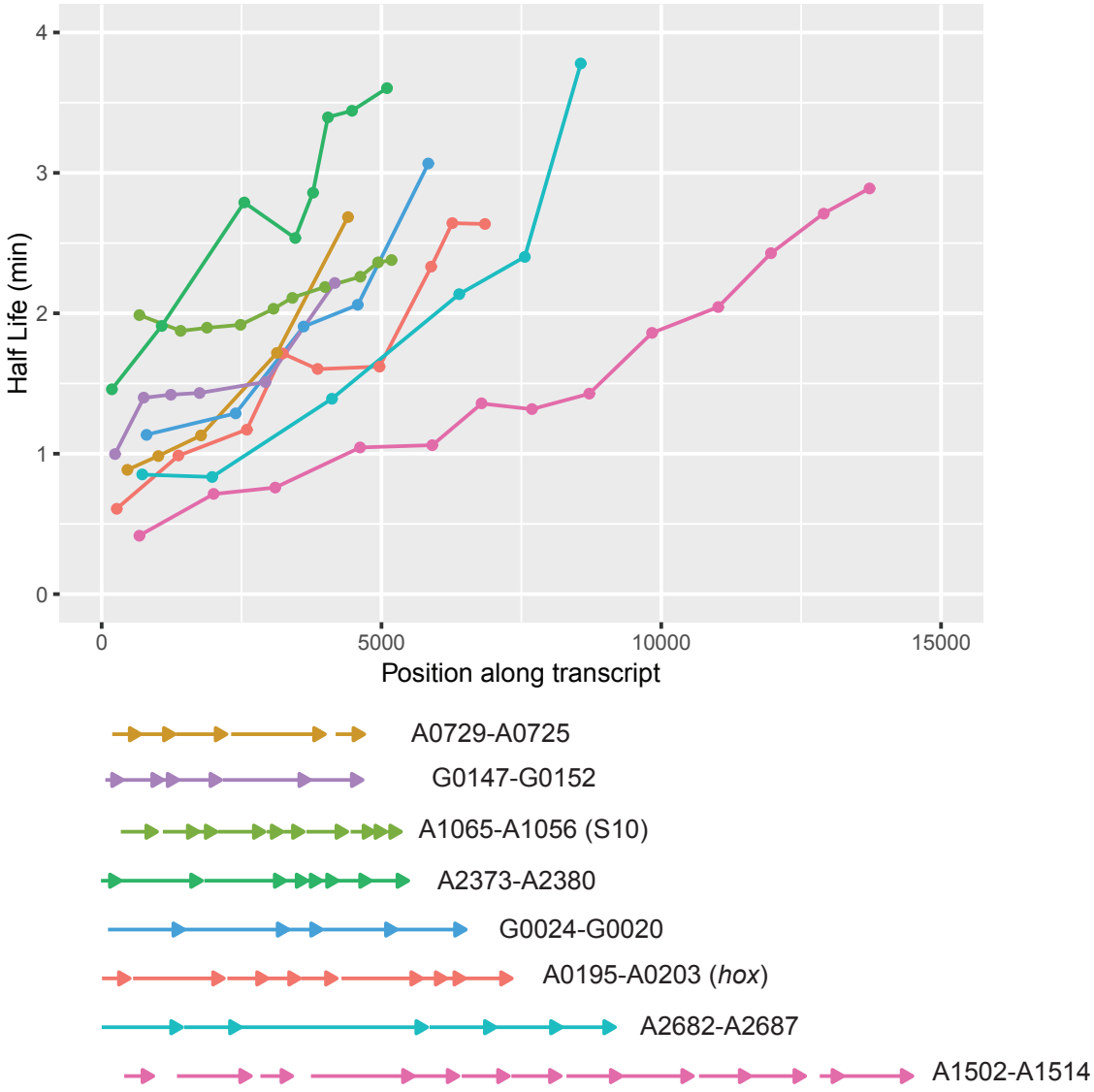


Figure 5.6: Half-life increases along transcript. The half-lives of genes from eight operons from PCC 7002 are shown with the points located in the middle of the gene. The bottom diagram shows the number of genes, their spacing, and locus tags. The approximate transcriptional start site is +1.

The conditions during sampling may have a large impact on RNA half-life. These cyanobacterial half-lives were determined at an OD $\sim$0.2 which is between late exponential and early linear phase (for most cyanobacteria very brief exponential growth followed by a linear relationship between time and optical density). We chose this growth phase because we wanted to examine actively growing and dividing cells. This could be why we saw high stability and expression of the photosynthesis genes because cells were growing rapidly and needed to produce high levels of photosynthesis machinery. Transcript degradation, however, is likely influenced by many factors including growth phase, growth rate, and environmental changes. There is an increasing body of evidence that RNase activity and expression can be influenced by inhibitors (Kim et al. 2008), temperature (Awano et al. 2008), growth phase (Chen and Deutscher 2010), and osmolarity (Sim, Yeom, et al. 2010). To gain a better understanding of global RNA degradation, it will be necessary to examine global decay rates under many different conditions and to identify which specific RNases are affected under these conditions. Ultimately a global understanding of transcript degradation would enable accurate predictions of transcript half-life given its sequence, but currently there is a lack of knowledge of how transcript stability is regulated and how different sequence elements impact RNA processing and degradation.

5.4 Materials and Methods

Sampling

WT PCC 7002 (Pasteur collection) was grown on grown on media A+ (Stevens, Patterson, et al. 1973) 1.5% agar plates. All liquid cultures were grown at 37°C and bubbled with air. A bubble tube containing 21 mL media A+ was inoculated and was grown for 24 hours. The OD_{730nm} was measured and used to inoculate 3 separate bubble tubes containing 21 mL media A+ at an OD_{730nm} = 0.1, which were grown for 24 hours. The OD_{730nm} was measured and 3 separate 1 L bottles containing 900 mL of media A+ were inoculated with pre-culture to achieve an OD_{730nm} of 0.01.

Cultures were bubbled with air at 37°C for 22 hours before sampling.

2 mL of each culture was collected for cell counts via hemocytometer and OD_{730nm} measurements. 40 mL of culture was collected and deposited into a 50 mL conical tube containing 5.0 mL stop solution (10% phenol in ethanol) and placed on ice. Rifampicin was added at a concentration of 50mg/mL (in DMSO) and samples were taken as described above at 0.5, 1, 2.5, 5, 7.5, and 10 minutes following addition. Samples were spun down in a Beckman Coulter Avanti J-E Centrifuge at 7,500xg for 10 minutes at 4°C. The supernatant was removed and samples were flash frozen in liquid nitrogen and stored at -80°C until RNA extraction (1 week).

RNA spike-in Preparation

RNA spike-ins were used for normalization to ensure read counts corresponded to RNA copy number. Six RNA spike-ins from bacteriophage ϕX174 were used as in (Chen, Shiroguchi, et al. 2015). DNA templates for in vitro transcription were created by including the T7 promoter sequence in the primer sequence (**Table A1.1**) and amplifying ϕX174 DNA (New England Biolabs, #N3023S). DNA templates were purified with ethanol precipitation and used to synthesize RNA with the T7 RiboMAX Express Large Scale RNA Production System (Promega, #P1320). RNA was purified with ethanol precipitation and quantified via Nanodrop. RNA spike-ins were combined to obtain a range of copy numbers per cell (gene H: 0.1 copies/cell; gene D: 1 copy/cell; genes F and G: 10 copies/cell; fragment 290: 100 copies/cell; and fragment 190: 1,000 copies/cell).

RNA Extraction and Sequencing

RNA spike-ins were added to samples immediately before extraction with hot phenol and DNase-treatment as in (Khodursky et al. 2003). RNA was quantified via Nanodrop, quality verified with the BioAnalyzer, and submitted to the University of Wisconsin Madison Biotechnology Gene Expression Center for library preparation and sequencing. rRNA was removed with a Ribo-Zero Magnetic Kit and cDNA was creating with a TruSeq Stranded Total RNA Library Kit. Libraries

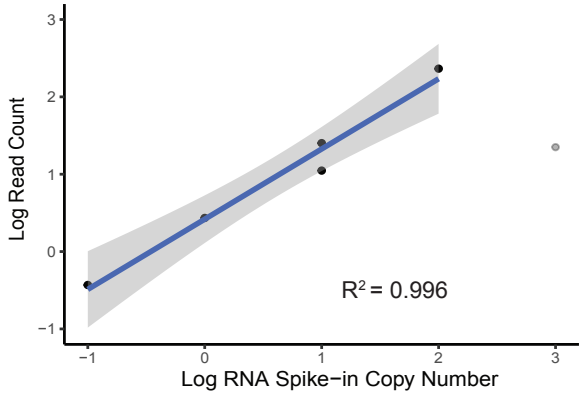


Figure 5.7: RNA-seq read counts from RNA spike-ins. The read counts from RNA spike-ins from example replicate B time 0 minutes is shown. A linear relationship between the log read count of the spike-ins (normalized for length) and the log spike-in copy number for spike-ins added between 0.1 and 100 copies/cell. The spike-in added at 1,000 copies/cell was retrospectively too high, and not used in the normalization calculation. The linear fit for the remaining 5 spike-ins is shown along with the shaded 95% confidence interval, $R^2 = 0.996$.

were sequenced on an Illumina HiSeq 2500 (1x100).

RNA spike-in Normalization and Half Life Determination

Sequencing files were aligned to the PCC 7002 chromosome and plasmids (NC_010474-NC_010480) as well as the ϕX174 genome (NC_001422) with Bowtie 2 (v. 2.2.6) and samtools (v 1.2). The HTSeq count function (v 0.6.1) was used to count the number of reads that aligned to each feature. A linear relationship between log RNA spike-in copy number and read counts was only observed for spike-ins added at a ratio of 0.01 - 10 copies/cell, so only these were used for normalization (**Figure 5.7**). All PCC 7002 genes were within this range and all samples had a good correlation between log RNA spike-in copy number and read count with all R^2 values were between 0.971 and 0.999. The sum of the reads that aligned to the spike-ins was calculated for each sample and used to create a normalization factor to keep the sum of reads that aligned to the spike-ins constant throughout the time course.

We combined data from all 3 biological replicates and used all 21 points to fit an exponential

decay model. Degradation of many genes was extremely rapid and read counts were close to zero for many of later timepoints, so we determined the best number of timepoints to include based on the fit. We used the half-life corresponding to the highest R^2 fit of using the first 3, 4, 5, 6, or all 7 timepoints. Based on our ability to calculate a half-life, we classified all genes as either 'decay', 'stable', 'low_r2', or 'no_reads'.

Motif enrichment

Using MEME (Bailey et al. 1994) we searched for enriched motifs in both the upstream 150bp region and downstream 100bp region of each gene. Using discriminative mode we looked for motifs enriched in the sequences of the 400 most stable genes compared to the control 400 least stable genes. We looked for site distributions of zero or one occurrence per sequence (zoops) and looked in the coding strand only. Top hits always included the highly iterative palindrome 1 (HIP1, 5'-GGCGATCGCC-3') sequence whose function is still unexplained yet is overrepresented in any PCC 7002 sequence.

Terminator analysis

The types and locations of terminators were extracted from the Web Genome Scanner for Terminators Database (WebGeSTer DB) (Mitra et al. 2011). The input parameters for identification of terminators was (max stem length: 12; min stem length: 4; max loop length: 8; min loop length: 3; max mismatch: 3; max distance from ORF: 270; and a species-specific ΔG cut-off of -11.618 kcal/mol.

Accession Numbers

Raw sequencing files have been deposited in NCBI's Sequence Read Archive (accession number SRP130967). Half-lives of genes and per base count data has been deposited at the Gene Expression Omnibus (accession number GSE109174).

6 Chapter 6: Tools to control gene expression in cyanobacteria

Authors and Contributors:

Gina Gordon: Designed and optimized reporter CRISPRi repression system, conducted experiments on native targets, built and tested ribosome binding site reporter libraries, built and tested RNA stability of hairpin constructs, wrote text.

Travis Korosh: Built and tested base strains and strategies to produce lactate, conceived of CRISPRi target to increase lactate production, built and tested lactate production strains.

Andrew Markley: Built promoter libraries for reporter expression and assisted in experimental design for optimization of the CRISPRi system.

Matthew Begemann: Began work to produce lactate in cyanobacteria and built promoter libraries for reporter expression.

Jeffrey Cameron: Generated ideas of native targets to show application of CRISPRi.

Dr. Brian Pfleger: Supervised research, provided scientific direction, and edited manuscripts.

Portions of this chapter were adapted from: Gordon, G.C., Korosh, T.C., Cameron, J.C., Markley, A.L.,Begemann, M.B., and Pfleger, B.F. (2016). CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metabolic Engineering* 38: 170-179. doi: 10.1016/j.ymben.2016.07.007.

6.1 Introduction

The ability to manipulate and predictably control gene expression is an essential tool for engineering metabolism and biology. Modern gene expression toolboxes include promoter libraries for initiating transcription at desired rates (Alper et al. 2005; Markley et al. 2015), transcriptional regulator/operator pairs for creating dynamic switches and circuits (Stanton et al. 2014; Zhang, Ouellet, et al. 2012), transcription terminators for insulating neighboring expression cassettes (Chen, Liu, et al. 2013) models of translation initiation for engineering ribosome binding sites (Espah Borujeni et al. 2014; Salis et al. 2009), codon optimization algorithms for optimizing gene sequences (Puigbò et al. 2007), and RNA structures that tune mRNA turnover, termination, and translation initiation (Pfleger et al. 2006). These tools have enhanced the study of natural systems and enabled the creation of engineered microbes for addressing societal challenges such as sustainable chemical production (Chubukov et al. 2014; Hara et al. 2014; Liao et al. 2016; Lynch et al. 2012; Smanski et al. 2016). Unfortunately, differences in how microbes recognize promoters, regulate gene expression, and translate protein do not allow genetic circuits and tools to be universally moved between organisms with the desired outcome. Instead synthetic biology tools must be validated in new hosts and then frequently adapted and refined for optimal functionality (Keasling 2012).

6.2 CRISPR interference in PCC 7002: physiology and chemical production

Applications of CRISPRi in PCC 7002

The majority of established tools regulate genes in *cis* and require replacement of native expression cassettes with heterologous sequences to alter the level of gene expression. Replacing sequences has the unfortunate side-effect of removing native regulation that has often evolved to optimize the protein abundance needed for a given set of natural physiological states. In many instances, the native context is ideal for one state but needs to be altered for a new unnatural state

(e.g. chemical production instead of growth). Trans acting tools that supersede native regulation in specific environments are therefore desirable additions to the synthetic biology toolbox (Copeland et al. 2014). One such tool, termed CRISPR interference (CRISPRi), takes advantage of the adaptive RNA-based defense system that in many bacteria targets and cleaves foreign nucleic acids such as viruses and plasmids (Qi et al. 2013). Diverse CRISPR-Cas systems exist in bacteria with altered locus architecture, components, maturation processes, and functions (Makarova et al. 2015). The type II system from *Streptococcus pyogenes* has been the most widely adapted to manipulate gene expression. A complex of a single guide RNA (sgRNA) containing a sequence complementary to the target and the Cas9 DNA nuclease protein can initiate double-stranded breaks. Point mutations in the two active sites can be used to create a nuclease deficient or dead Cas9 (dCas9). The complex of the sgRNA and dCas9 is still able to bind target DNA and can be used to either repress or activate gene expression (Bikard et al. 2013). There is strong evidence that repression is caused by the dCas9-sgRNA complex sterically blocking RNA polymerase elongation (Gilbert et al. 2013). The use of CRISPRi in *Escherichia coli* has led to efficient repression of targets as high as 300-fold when the sgRNA was targeted to the non-template strand near the 5' end of the gene with no detectable off-target affects (Qi et al. 2013). This repression was reversible and multiplexed to target several genes simultaneously.

CRISPRi has become an increasingly attractive alternative to other trans-acting regulators repressor proteins such as trans-activator-like effectors (TALEs) or zinc fingers, due to the simplicity of design and ease of synthesis. By altering just 20 nucleotides of the sgRNA, the system can be designed to repress any gene of interest. However, it is necessary to choose a target site with a protospacer adjacent motif (PAM) which is 5'-NGG-3' for *S. pyogenes* Cas9, but engineered Cas9 nucleases are reducing this constraint (Kleinstiver et al. 2015). Despite this limitation this technology has been used to repress reporter genes as well as enhance titers of products in various organisms: flavonoids in *E. coli* (Wu et al. 2015) and L-lysine and L-glutamate in *Corneybacterium glutamicum* (Cleto et al. 2016).

Here, we adapted this technology to an industrially relevant cyanobacterial strain, *Synechococcus* sp. strain PCC 7002 (PCC 7002) (Devroe et al. 2009; Reppas and Ridley 2010). Cyanobacteria are an attractive chassis for chemical production that enables direct conversion of carbon dioxide and sunlight into useful products (Oliver and Atsumi 2014). PCC 7002 is a promising strain of cyanobacteria because it grows rapidly, is halotolerant, is naturally transformable, can tolerate high light conditions, and has a growing synthetic biology toolbox (Markley et al. 2015; Zess et al. 2016). To add to the PCC 7002 toolbox, we developed a tunable gene repression system using CRISPRi. The system is novel because targets can be repressed to varying degrees by controlling the expression of CRISPRi components with an inducer, anhydrotetracycline (aTc). Most CRISPRi systems including one developed for another cyanobacterium *Synechocystis* sp. PCC 6803, can be turned on and off but the ability to titrate the levels of expression was not reported (Yao et al. 2016). Attempts have been made to tune repression by introducing mismatches in the sgRNA, but this involves making numerous strains to reach a variety of repression levels (Qi et al. 2013).

The ability to finely tune native gene expression will be critical for manipulating essential genes and to achieve growth regimes that decouple biomass and chemical production for maximum yield (Xu, Huang, Wu, et al. 2014). In many cases, intermediate levels of enzyme expression result in maximum product titer (Freed et al. 2015; Pitera et al. 2007). With the described CRISPRi technology we can lower expression of essential cyanobacterial genes, decrease but not abolish flux towards competing products, and manipulate cellular processes to varying degrees. Here, we demonstrate the utility of finely tuning native gene expression by downregulating the abundance of phycobilisomes. In addition, we create a conditional auxotroph by repressing synthesis of the carboxysome, an essential component of the carbon concentrating mechanism cyanobacteria use to fix atmospheric CO_2. Lastly, we demonstrate a novel strategy for increasing central carbon flux by conditionally downregulating a key node in nitrogen assimilation pathways. The resulting cells produced 2-fold more lactate than a baseline engineered cell line, representing the highest photosynthetically generated productivity to date.

Optimization of YFP repression with CRISPRi

To demonstrate a functional CRISPRi system in PCC 7002 we integrated three expression cassettes into the chromosome: 1) a fluorescent reporter, 2) an inducible dCas9, and 3) a constitutively expressed sgRNA (**Figure 6.1A**). We constructed a reporter cassette consisting of a heterologous yellow fluorescence protein (EYFP) expressed from a strong constitutive promoter and a gentamicin resistance marker. The resulting cassette was integrated into a neutral site in the genome by replacing the *glpK* pseudogene and selecting transformants on gentamicin. We integrated dCas9 under the control of the anhydrotetracycline (aTc) induction system, EZ3 (Zess et al. 2016)) into the *acsA* locus using acrylic acid as a counter selection marker (Begemann et al. 2013). The aTc-controlled promoter was critical to construct the system; no transformants were obtained when strains were transformed with a plasmid containing dCas9 expressed from the stronger cLac94 IPTG-inducible promoter (Markley et al. 2015). A single guide RNA (sgRNA) was designed to target the non-template strand of YFP at a position $\sim$175 nucleotides downstream of the start codon. An expression cassette containing the sgRNA expressed from the strong constitutive promoter, P_{J23119}, was integrated onto the chromosome at neutral site 1 (NS1) with a kanamycin resistance marker (Davies et al. 2014). We measured bulk YFP fluorescence in extracts from these strains and found that the fluorescence levels were unaffected in strains containing only dCas9 or the sgRNA (**Figure 6.1B**) even when expression of dCas9 was induced with aTc (**Figure 6.2**). In strains containing all three cassettes, YFP fluorescence was reduced to 0.05% of the unregulated control. In the presence of 1 μg/mL aTc (to induce dCas9 expression), YFP fluorescence was further reduced to 0.02% of the control. These data confirm that CRISPRi functions in PCC 7002, however the limited dynamic range of the system limits the utility in many applications.

In effort to increase the dynamic range of the CRISPRi system, we hypothesized that decreasing expression of the sgRNA and dCas9 would reduce repression in the absence of aTc. Therefore, we replaced the promoter controlling expression of dCas9 with weaker alternatives, but saw no reduction in repression in the absence of aTc (**Figure 6.3**). Similarly, we changed the promoter

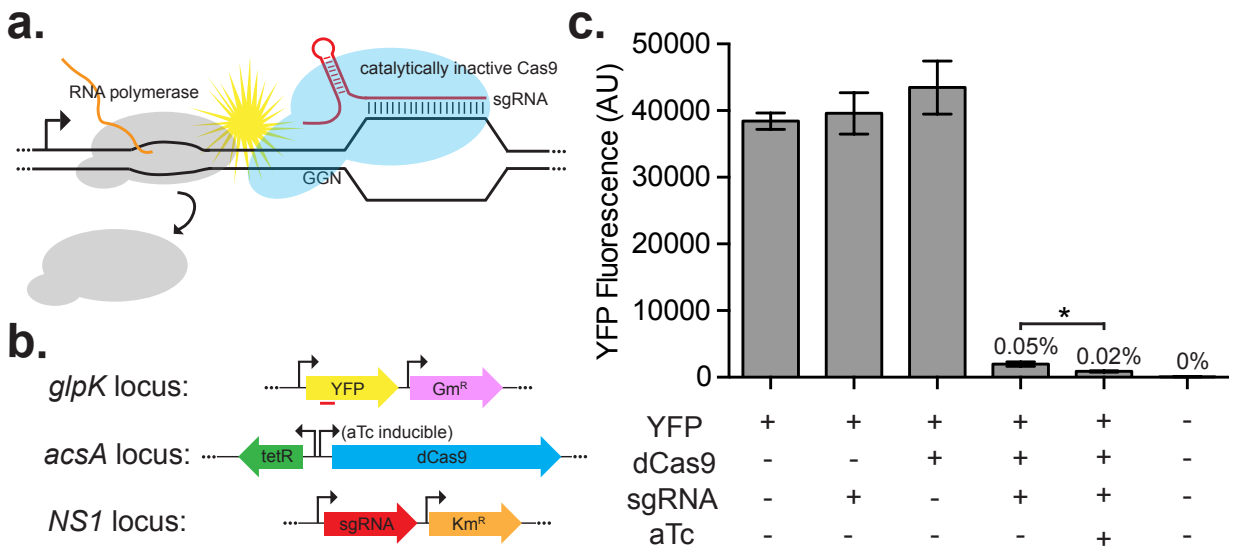


Figure 6.1: Repression of YFP reporter with CRISPRi. (A) Cartoon illustrating how the dCas9-sgRNA complex interferes with RNA polymerase, thereby repressing transcription. (B) Schematic of the three cassettes used to implement CRISPRi in PCC 7002. (C) YFP fluorescence of PCC 7002 extracts containing dCas9 expressed from the aTc-inducible promoter and sgRNA targeting YFP from constitutive P_{J23119}. Repression requires the presence of both dCas9 and sgRNA but is not dependent on the presence of the inducer, aTc. Induction of dCas9 caused a significant decrease in YFP fluorescence (*p = 0.0008, unpaired two-tailed t-test with equal standard deviation). Numbers above each bar report the percent fluorescence relative to the YFP only strain. Error bars represent standard deviation of fluorescence measured from four biological replicates.

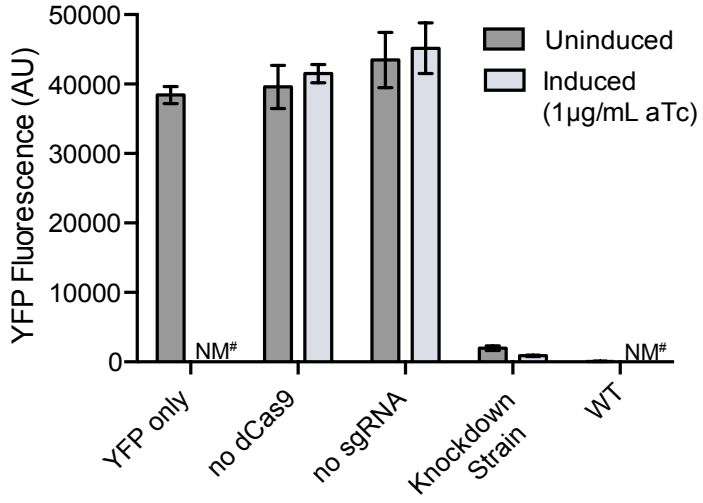


Figure 6.2: Induction with aTc does not cause repression unless both dCas9 and sgRNA are present. Bars show YFP fluorescence of strains containing dCas9 expressed from the aTc-inducible promoter and sgRNA targeting YFP from constitutive P_{J23119}. Averages of strains grown in quadruplicate are shown (error bars show standard deviation).

strength of the sgRNA from a strong promoter, P_{J23119}, to a medium constitutive promoter, P_{J23108}, a weak constitutive promoter, P_{J23117}, and the same aTc-inducible system (Zess et al. 2016) controlling dCas9 expression (**Figure 6.4**). We found that decreasing the sgRNA promoter strength reduced the repression of YFP in the uninduced culture (from 7% of the control for P_{J23119}, to 11% for P_{J23108}, and 70% for P_{J23117}), but also in the induced samples (from 3% of the control for P_{J23119}, to 6% for P_{J23108}, and 45% for P_{J23117}). In contrast, expression of the sgRNA under control of the aTc-inducible system (Zess et al. 2016) already controlling dCas9, generated the largest dynamic range. The YFP fluorescence of the uninduced sample was reduced to 46% of the control and the induced sample maintained strong repression of YFP (4% of the control).

Due to the required tet operator, the resultant sgRNA from the aTc-inducible system has an additional 7 nucleotides at the 5' end. When those same 7 nucleotides were added to the 5' of the sgRNA expressed from P_{J23119} there was no change in YFP repression (**Figure 6.5**). This finding suggests that CRISPRi in PCC 7002 may be tolerant to mis-matched extra nucleotides at the 5' end

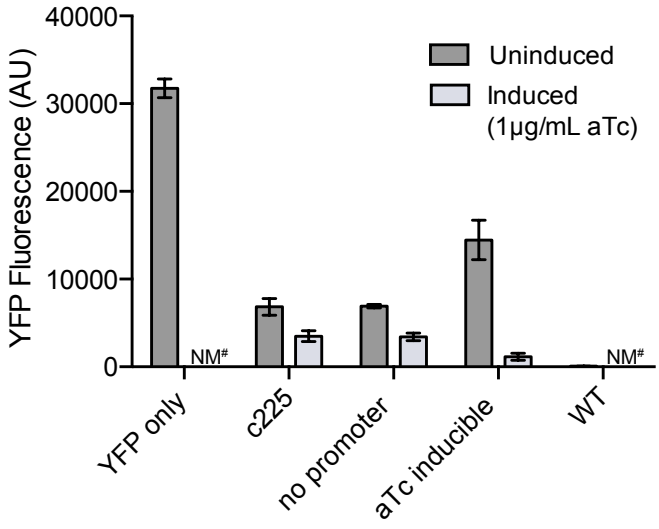


Figure 6.3: Altering the promoter of dCas9 does not relieve repression in uninduced samples. Bars show YFP fluorescence of strains containing dCas9 expressed from the c225 low constitutive promoter (Markley et al. 2015), no promoter, and aTc-inducible promoter. All sgRNA targeting YFP are expressed from constitutive P_{J23119}. Averages of strains grown in quadruplicate are shown (error bars show standard deviation).

of the sgRNA in contrast to examples in other bacteria where 5' extensions decrease the efficiency of repression (Larson et al. 2013).

We hypothesize that minimizing both dCas9 and sgRNA concentrations relieve the unintended repression when dCas9 is uninduced. In other words, very small concentrations of dCas9 may be capable of effectively silencing YFP expression when excess sgRNA is present. Attempts to reduce dCas9 expression using an alternate promoter and by deleting the promoter region were unsuccessful, suggesting that residual or promiscuous expression of dCas9 is sufficient for full repression (**Figure 6.3**).

We next sought to improve this tunable repression by decreasing repression of the target gene in the absence of aTc. We hypothesized that reducing the expression level of dCas9 by altering its ribosome binding site (RBS) would meet this goal. Using the RBS Library Calculator (Farasat et al. 2014) we designed five additional RBS with predicted decreasing translation initiation rates (labeled A-F). We achieved a range of repression in the uninduced samples although they did not correspond

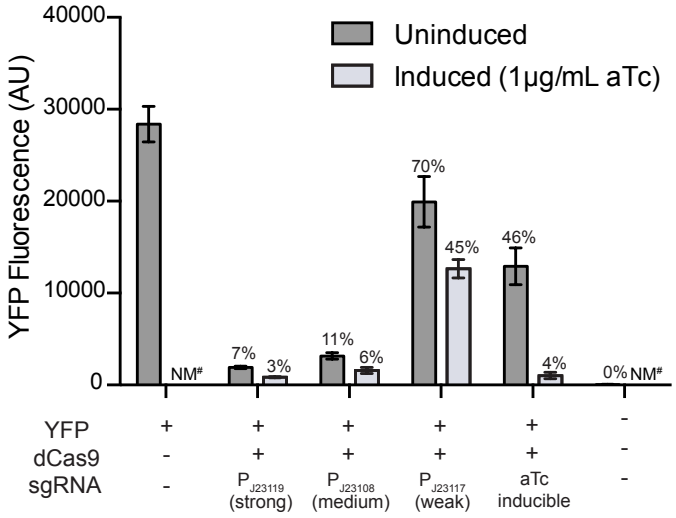


Figure 6.4: The dynamic range of the system is improved when both dCas9 and sgRNA are controlled by EZ3 tet promoters. Bars show YFP fluorescence of extracts containing dCas9 expressed from the aTc-inducible promoter and sgRNA targeting YFP from constitutive promoters of varying strength and an aTc-inducible promoter. Promoter strengths were previously characterized in PCC 7002 (Markley et al. 2015). The percent fluorescence relative to the YFP only strain is displayed above the bar. Averages of strains grown in quadruplicate are shown (error bars show standard deviation, NM$^{\#}$ indicates not measured).

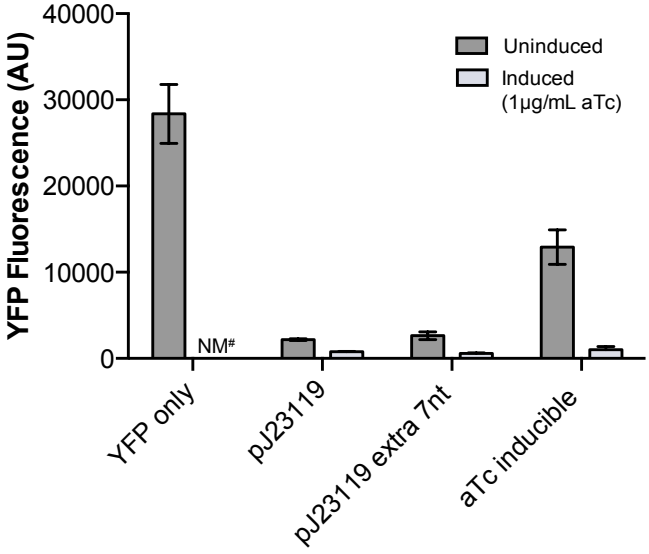


Figure 6.5: Extra nucleotides at 5' end of inducible sgRNA do not affect repression. Bars show YFP fluorescence of strains containing dCas9 expressed from the aTc-inducible promoter. The sgRNA targeting YFP are expressed from constitutive P_{J23119} and the aTc-inducible promoter. Averages of strains grown in quadruplicate are shown (error bars show standard deviation).

to the predicted translation initiation rate of dCas9 (**Figure 6.6A**). Altering the dCas9 RBS allowed us to obtain minimal repression without aTc (∼30%) while maintaining repression when maximally induced (∼90%). In an effort to further optimize the system, we created 5' truncations of the sgRNA as this has been shown to lower repression in other systems (Qi et al. 2013; Xu, Li, et al. 2015). Here, we observed a drastic decrease in repression between 15 nucleotides and 12 nucleotides, but overall dynamic range was not improved relative to the original 20 nucleotide sgRNA system (**Figure 6.7**). We proceeded to characterize the strain with the 'F' RBS controlling dCas9 expression (there was no statistical difference in YFP fluorescence between uninduced samples of 'C' and 'F' RBS). With the uninduced samples normalized to 100%, we were able to titrate YFP fluorescence to 13% when fully induced (**Figure 6.6B**). There is still some repression in the uninduced sample (∼30% of no CRISPRi control). Last, we designed sgRNA to target different locations within YFP and found similar results for sgRNA targeting both the template and non-template strands near the beginning of the transcript, with slightly reduced repression when fully induced when targeting the 3' end of the transcript (**Figure 6.8**).

Reversibility of repression

One advantage of trans-acting regulators is the potential to dynamically regulate expression of target genes between on and off states in the same culture. To examine the potential of the CRISPRi system in this regard, we performed a series of timecourse experiments wherein aTc levels were either added or removed at specific timepoints and YFP fluorescence was monitored. In the absence of aTc, YFP fluorescence increased over time (**Figure 6.9A**) as expected. When cultures were induced, (**Figure 6.9B**), YFP fluorescence decreased over time as cells grew (filled in markers). Cultures were centrifuged,washed, and resuspended in fresh media with or without aTc. In the absence of aTc, YFP fluorescence returned, albeit at a slower rate than the completely uninduced cultures (**Figure 6.9A**). In contrast, induced cultures maintained a lower level of YFP fluorescence for the remainder of the experiment. The increasing fluorescence in the samples without aTc shows

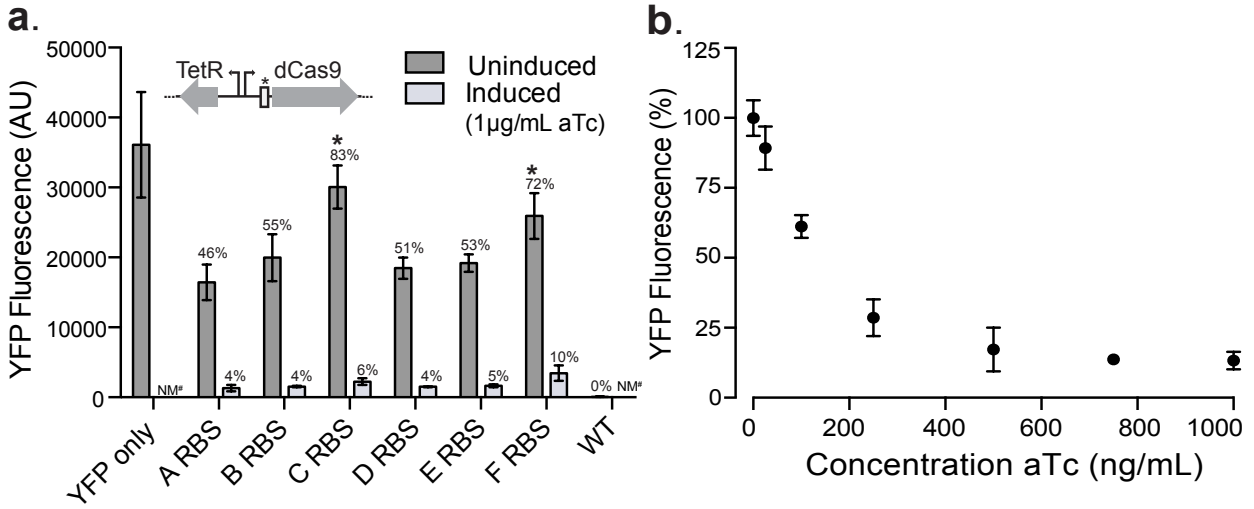


Figure 6.6: Decreasing RBS strength of dCas9 maximizes the dynamic range and enables titration of repression. (A) Bars show YFP fluorescence of cultures with different ribosome binding sites (RBS) controlling translation of dCas9. RBS' were designed to be in decreasing strength, A to F. A one-way ANOVA showed statistical significance between uninduced samples ($p < 0.0001$), but not when C and F RBS were removed. There was no statistical difference between uninduced samples of C and F RBS ($p = 0.115$ unpaired two-tailed t-test with equal standard deviation). Data are the average of biological quadruplicates and error bars show standard deviation. (B) Titration of strain with F RBS with increasing concentrations of aTc. Averages of strains grown in triplicate are shown with the uninduced sample normalized to 100% (error bars show standard deviation).

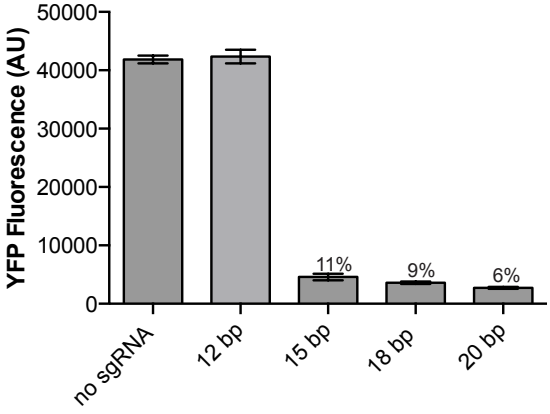


Figure 6.7: Significant repression is seen with only 15 bases in the sgRNA complementary to the target. Bars show YFP fluorescence of strains containing aTc-inducible dCas9 and sgRNA targeting YFP expressed from constitutive P_{J23119}. The region of sgRNA complementary to the YFP target site was truncated from 20 bp to 12 bp. Averages of strains grown in quadruplicate are shown (error bars show standard deviation).

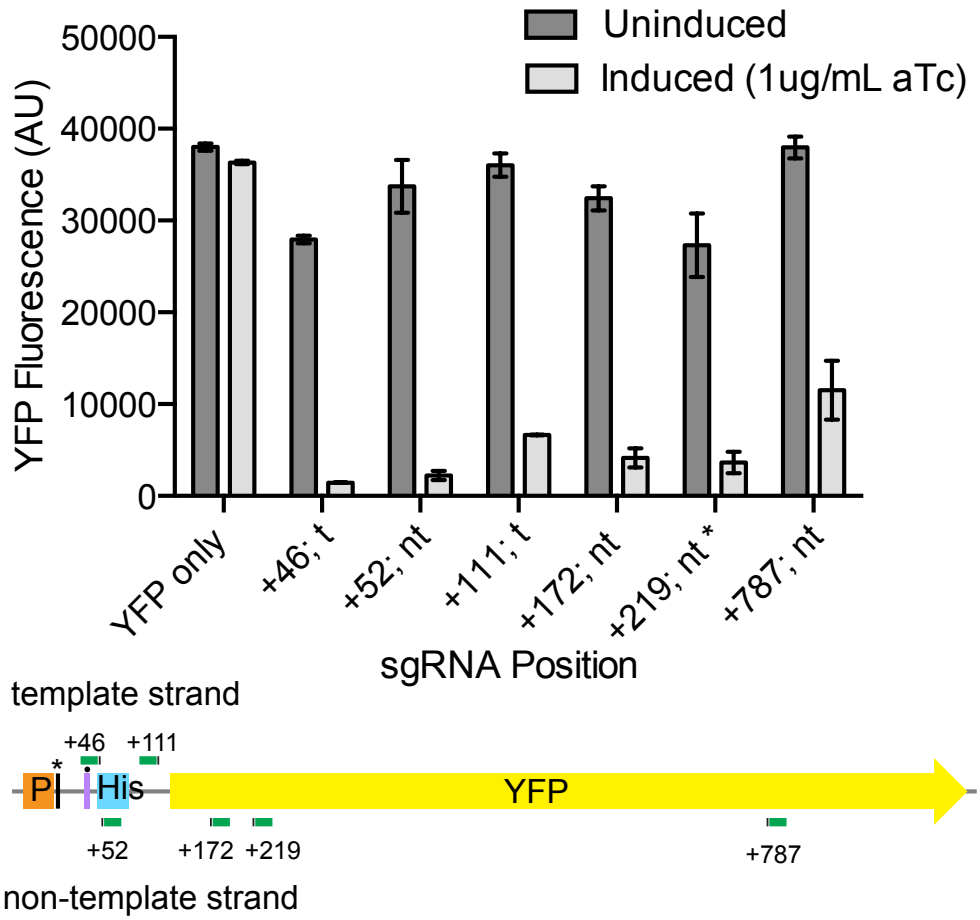


Figure 6.8: YFP repression with sgRNA targeting different locations within YFP. Bars show YFP fluorescence of strains containing aTc-inducible dCas9 and aTc-inducible sgRNA targeting YFP. sgRNA is labeled according to the location from the transcription start site to the 20 nucleotides targeting YFP ('t' indicates targeting the template strand while 'nt' indicates targeting the non-template strand). Below is a to-scale diagram of the locations of the sgRNA. P represents the promoter, * indicates the transcription start site, the dot indicating the ribosome binding site, and 'His' indicates the location of the His tag on YFP. Averages of strains grown in quadruplicate are shown (error bars show standard deviation).

that CRISPRi can be turned on and off by the addition or removal of aTc. This tool enables the study of PCC 7002 genes allowing the knockdown of genes to intermediate levels as well as the ability to track physiological changes throughout time when repressing or derepressing genes of interest.

Repression of native genes

We tested whether CRISPRi could be used to rapidly modulate two key processes in photosynthetic organisms: light harvesting and CO_2 fixation. First, we designed a system to target and repress expression of core phycobilisome genes. Phycobilisomes are multisubunit pigment-protein complexes that function as light harvesting antennae in cyanobacteria by transferring light energy to the photosynthetic reaction centers, photosystem I and photosystem II (Grossman et al. 1993; Liu et al. 2013). The phycobilisome is comprised of an allophycocyanin core attached to peripheral rods assembled from disks of phycocyanin separated by linker polypeptides. In PCC 7002, the *cpcBACDEF* operon encodes the major components of the peripheral rods including the α-(CpcA) and β-subunits (CpcB) of phycocyanin, the rod linkers (CpcC and CpcD), and two proteins involved in attachment of the phycocyanobilin pigment to the α-phycocyanin subunit (CpcE and CpcF) (Bryant et al. 1990). Since reduction of antennae size has been a promising strategy for increasing light penetration into high-density cyanobacterial cultures (Kirst et al. 2014), we designed an sgRNA to target *cpcB* and inhibit transcription of the entire *cpcBACDEF* operon. Others have shown that when the rod proteins are deleted there is a decrease in absorbance corresponding to phycocyanin at 635 nm (Lea-Smith et al. 2014). We hypothesized that repression of this operon would lead to an observable phenotype – a decrease in absorbance in the phycocyanin absorbance peak at 635 nm relative to the 680 nm chlorophyll peak (Alvey et al. 2011).

Following induction of the dCas9 in the strain containing the sgRNA targeting *cpcB*, we observed a significant decrease in the absorption peak at 635 nm, indicating a severe depletion phycobilisomes compared to the control strain (**Figure 6.10**). Dynamic modulation of the light-harvesting complex,

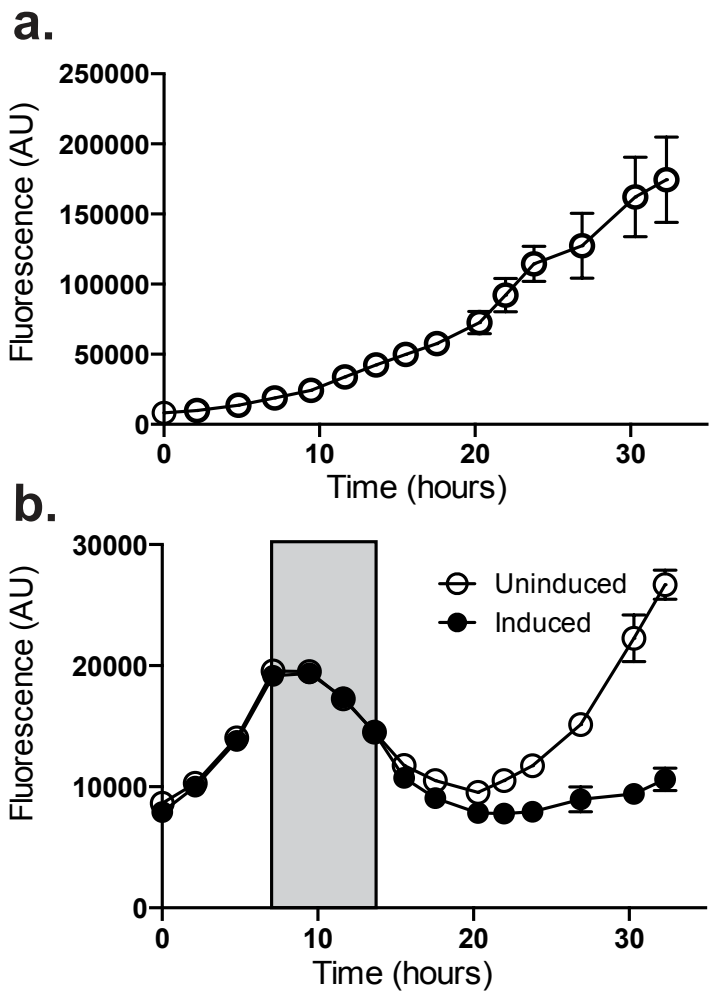


Figure 6.9: CRISPRi repression can be reversed. (A) YFP fluorescence increases over time in cultures grown without aTc. (B) YFP fluorescence increases over time before the addition of aTc (marked with arrow), where YFP fluorescence begins to decrease. Upon removal of aTc by centrifugation and washing, YFP fluorescence begins to increase again in cells without aTc (open circles), but is maintained at a lower level when aTc was re-added (dark circles). Experiments were performed in quadruplicate with the average fluorescence values marked and the standard deviation shown as the errors bars. Data for panels A and B were collected at the same time using the same detection settings.

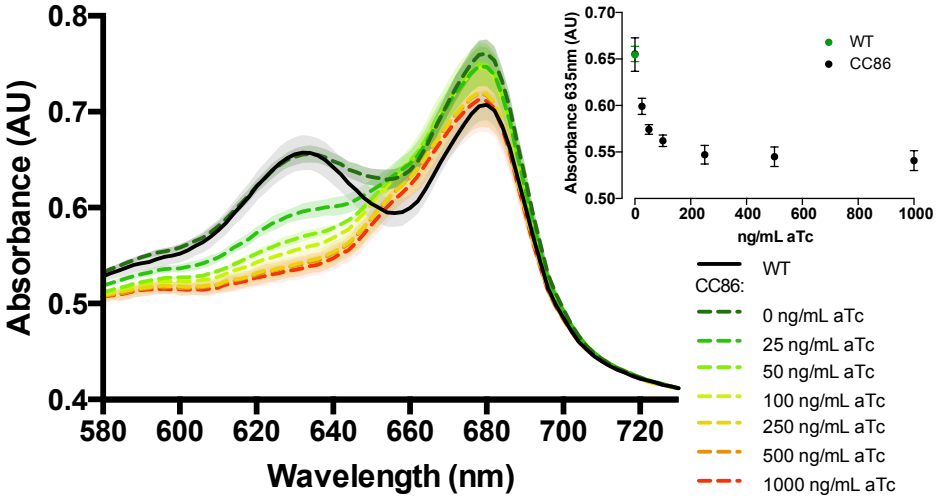


Figure 6.10: Tunable repression of the phycobilisome *cpcBACDEF* operon. A portion of the absorption spectra of PCC 7002 (CC86) cultures induced with different levels of aTc shows a decrease in the peak at 635 nm corresponding to the phycocyanin maximum absorbance peak. Average spectra are shown and were normalized to $OD_{730\ nm} = 0.4$ (standard deviations displayed with shading). The inset shows the absorbance at 635 nm with different concentrations of aTc.

as shown by our CRISPRi strategy, could be useful for improving growth at high cell density (Kirst et al. 2014) without affecting initial biomass scale-up that could be reduced in a mutant lacking phycobilisomes. Moreover, CRISPRi may be advantageous compared to making traditional deletion strains due to challenges in isolating homoplasmic strains, especially when the deletions have impacts on cellular fitness; for example, we were unable to generate a fully segregated *cpcB* strain, but could phenocopy this trait using CRISPRi (Lea-Smith et al. 2014).

Next, we tested whether essential genes could be repressed using the CRISPRi system. In all free-living cyanobacteria and some chemoautotrophic bacteria, CO_2-fixation occurs within the carboxysome, a bacterial microcompartment comprised entirely of protein that functions as a primitive organelle (Yeates et al. 2008). The carboxysome is comprised of a thin, multi-subunit, semipermeable protein shell that encapsulates the key carboxylase, Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO), and carbonic anhydrase in a unique sub-cellular environment (Cameron, Wilson, et al. 2013). The carboxysome functions as a key component of the cyanobacte-

rial CO_2 concentrating mechanism, which increases the local concentration of CO_2 at the site of CO_2-fixation within the carboxysome.

While carboxysomes are essential for growth in atmospheric CO_2 levels (0.04%), models predict that increased productivities could be achieved using un-encapsulated RuBisCO during growth in industrial conditions that provide high-CO_2 (Clark et al. 2014), including flue gas from a coal-fired power plant ($\sim$15%). In PCC 7002, the *ccm* operon (*ccmK$_1$K$_2$LMN*) encodes the hexameric shell proteins ($CcmK_1$ and $CcmK_2$), the pentameric vertex protein (CcmL), and the interior scaffold proteins (CcmM and CcmN) that make up the major structural subunits of the carboxysome (Ludwig, Sultemeyer, et al. 2000). We designed a sgRNA to target *ccmK1*, with the prediction that inhibition of the *ccm* operon would inhibit formation of the carboxysome, resulting in a strain with a high-CO_2 requiring (HCR) phenotype and RuBisCO localized to the cytoplasm instead of being encapsulated within the carboxysome shell (Cameron, Wilson, et al. 2013).

We tested this hypothesis by comparing growth of a mutant lacking the *ccmK$_1$K$_2$LMN* (Δ*ccm*) operon with a CRISPRi strain targeting *ccmK$_1$* and a control containing the sgRNA but no dCas9 on solid medium in air or elevated CO_2. Cultures were grown overnight in elevated ($\sim$10%) CO_2, diluted and re-grown in the absence or presence of aTc, and then serially diluted and spot-plated onto solid media placed in either elevated CO_2 or air. As expected, the control strain with only the sgRNA could grow in both high CO_2 and air whereas Δ*ccm* could only grow with high CO_2 (**Figure 6.11A**). While the uninduced CRISPRi strain could grow in both air and high CO_2, the presence of aTc in the pre-culture resulted in a HCR phenotype, indicating functional repression of carboxysome assembly.

To investigate whether assembly of the carboxysome was being inhibited, we expressed a C-terminal sfGFP fusion of the RuBisCO large subunit, rbcL, as a visual marker in the test strains. The uninduced CRISPRi strain resembled wild type (WT) with discrete fluorescent foci, indicating the presence of carboxysomes (**Figure 6.11B**), whereas the Δ*ccm* strain exhibited diffuse fluorescence. The induced strain resembled an intermediate between the WT and Δ*ccm* strains, with polar

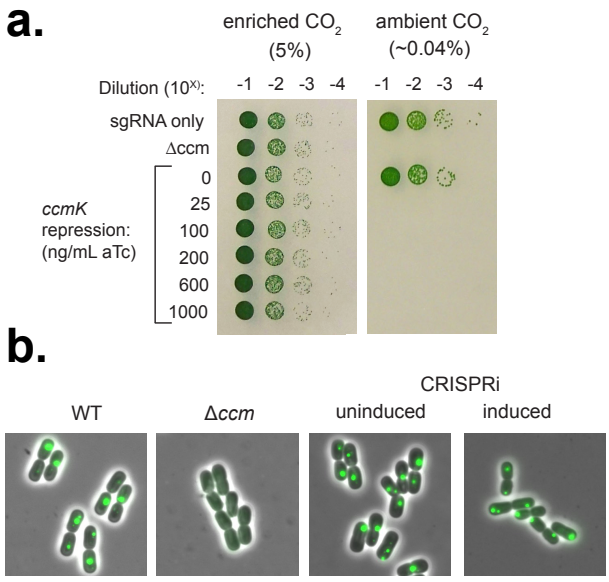


Figure 6.11: Repression of the *ccm* operon. (a) Spot plates of control and ccmK repression strains grown in enriched CO_2 and ambient CO_2. Liquid cultures were grown with or without aTc with $\sim$10% CO_2 and normalized before spot-plating on solid media (without aTc). Δ*ccm* is able to grow with enriched CO_2 but is unable to grow with ambient CO_2 The uninduced *ccmK* repression strain could survive in air but those induced could not. (b) RbcL-sfGFP was used to visualize carboxysomes in control and *ccmK* repression strains.

fluorescent foci and diffuse GFP signal. We suspect that the foci represent an intermediate in carboxysome assembly pathway initiated by molecular scaffolding of RuBisCO by CcmM; this intermediate is the result of insufficient shell proteins and exhibits an HCR phenotype (Cameron, Wilson, et al. 2013). Based on the HCR and visual phenotypes of the induced strains, we hypothesize that repression of the entire *ccm* operon is incomplete, but sufficient to inhibit assembly of functional carboxysomes.

Improving lactate production with CRISPRi

Trans-acting regulators also have the potential to implement elegant regulatory strategies in cell factories. Here, we used CRISPRi to increase the production of lactate. Lactate is an excellent model compound for studying metabolic engineering strategies in photoautotrophic bacteria because

its synthesis pathway is short (one step removed from central metabolism) and consumes cofactors within the range generated by photosynthesis during linear electron transport (Oliver and Atsumi 2014). PCC 7002 has an endogenous lactate dehydrogenase, but autotrophic cultures produce negligible levels of lactate under optimal laboratory conditions. Therefore we expressed an engineered version of the *Bacillus subtilis* lactate dehydrogenase (LDH) with a previously described single amino acid substitution (Richter et al. 2011), that switches the cofactor requirement of the enzyme from NADH to NADPH during conversion of pyruvate to lactate. An expression cassette with LDH under the control of the $P_{cLac143}$ IPTG-inducible system (Markley et al. 2015) was integrated into the chromosome at the *glpK* pseudogene locus along with a gentamicin resistance marker to generate the base strain, CC133.

In conjunction with elevated expression of optimized LDH, we predicted that limited flux from carbon fixation to pyruvate would limit lactate production. Enhanced pyruvate excretion and attenuation of phycobilisome and chlorophyll a degradation has been demonstrated under nitrogen starvation in several strains of cyanobacteria when glycogen synthesis is disrupted (Davies et al. 2014; Gründel et al. 2012; Hickman et al. 2013).We hypothesized that moderate reduction of glutamine synthetase I (*glnA*) expression would slow the rate of nitrogen assimilation through the GS-GOGAT pathway (Muro-Pastor et al. 2003). Slowing the rate of nitrogen assimilation would thereby lead to increased intracellular accumulation of α-ketoglutarate, activating the global nitrogen transcriptional activator, NtcA, a transcription factor from the cyclic AMP receptor protein class (Herrero et al. 2001). This in turn, modulates a suite of metabolic processes related to glycogen degradation and glycolysis, thereby enhancing the flux of fixed carbon to pyruvate (Osanai et al. 2006). However, severe nitrogen limitation will further increase intracellular α-ketoglutarate levels, facilitating complex formation between active NtcA and regulatory factor, PipX (Espinosa, Forchhammer, et al. 2006), negatively affecting de-novo protein synthesis and overall photosynthesis rates (Espinosa, Rodríguez-Mateos, et al. 2014; Krasikov et al. 2012). Therefore, we chose to titrate the degree of repression of *glnA* using the CRISPRi approached described above (**Figure 6.12**). As

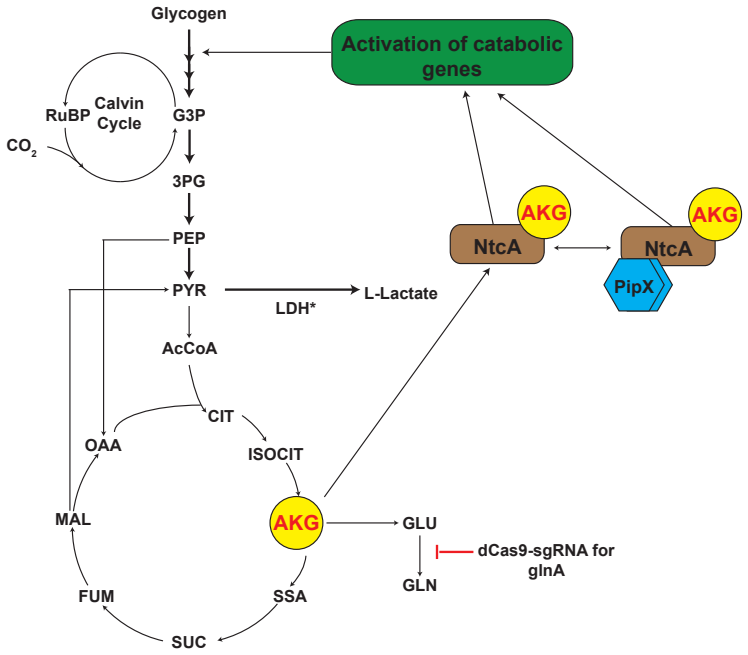


Figure 6.12: Schematic of proposed mechanism for increased lactate production. Moderate accumulation of α-ketoglutarate by *glnA* repression leads to NtcA activation and subsequent activation of catabolic genes.

additional controls, a sgRNA targeting YFP or *glnA* were integrated in the NS1 locus under the control of the aTc-inducible system, to create strains CC142 and CC130, respectively. dCas9 with the F RBS was integrated in CC130 to create strain CC131.

Interestingly, we found enhanced lactate production, as well as delayed phycobilisome and chlorophyll a degradation (**Figure 6.13**), in both strains harboring the *glnA* sgRNA when compared to the control strains in 1% CO_2 (**Figure 6.14B**) or air (**Figure 6.14D**). Integration of *glnA* sgRNA improved production rates from 0.079±0.001 mM/h in CC130 to 0.092±0.001 mM/h in CC131, both significantly improved to control strains CC133 and CC142, with rates of 0.045±0.002 mM/h and 0.029±0.002 mM/h, respectively, indicating that this effect was *glnA* sgRNA specific. This *glnA* sgRNA mediated effect may be due to a basal level of repression through cross-talk with the native CRISPR-Cas systems, unlike the other native genes targeted previously, given the difference in transcript abundance between *glnA*, *ccmK*, and *cpcB* (Ludwig and Bryant 2011; Ludwig and Bryant

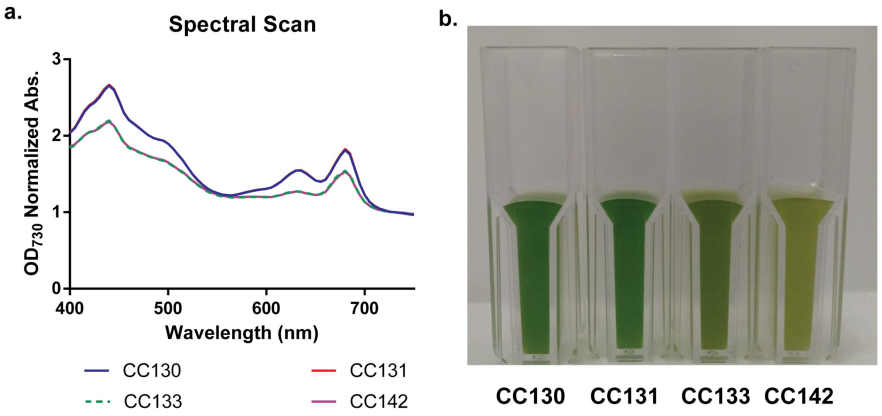


Figure 6.13: Lack of phycobilisome (635 nm) and chlorophyll a (680 nm) degradation in strains harboring *glnA* sgRNA. Representative (A) spectra and (B) image are of cultures grown in air at 96 hrs of induction, normalized to OD_{730nm}.

2012). There was no apparent difference in growth rate between any of the producing strains in 1% CO_2 (**Figure 6.14A**) or air (**Figure 6.14C**). This suggests that *glnA* repression or LDH expression did not severely impact amino acid synthesis or biomass generation. Given the significant increase in carbon products, the data also suggests that carbon storage polymers were rerouted to product, or carbon fixation rates were enhanced in CC131, analogous to cells engineered to secrete sucrose (Ducat et al. 2012).

Conclusions

CRISPRi provides a straightforward method to repress native genes of interest. We showed the functionality of this system genes in the fast-growing cyanobacterium, *Synechococcus* sp. strain PCC 7002 by repressing heterologous YFP and three native genes. We found that leaky expression of the system components was sufficient to achieve strong repression of its target and therefore the system required significant optimization to achieve the desired dynamic range of repression. We found that expression of dCas9 and sgRNA from aTc-inducible promoters allowed for a functional range of repression compared to expressing the sgRNA from constitutive promoters of various strengths. Leakiness of the system was further improved by screening RBS variants to modulate

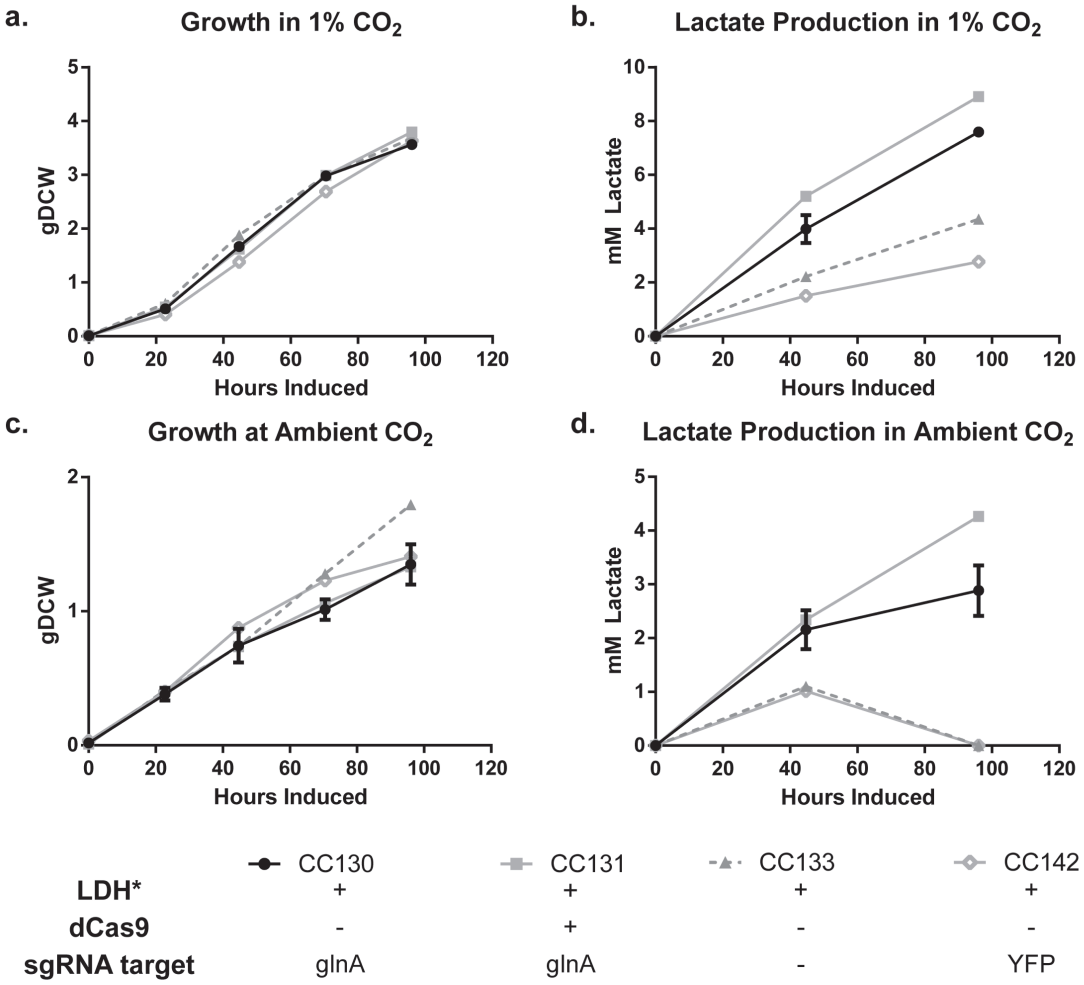


Figure 6.14: Repression of glutamine synthetase improves lactate production. Growth (A) and (B) lactate production of strains grown in 1% CO_2. Growth (C) and (D) lactate production of strains grown in air. Averages of strains grown in duplicate are shown (error bars show standard deviation).

the levels of dCas9. The resulting platform enables the rapid and titratable repression of gene expression in an industrial relevant cyanobacterium. As a proof of concept, we demonstrated repression of a YFP reporter construct, the native light-harvesting phycobilisome (*cpcBACDEF*), and the carboxysome (*ccmK$_1$K$_2$LMN*). Repression was titratable and reversible allowing genes to be turned off and on via the addition or removal of aTc.

CRISPRi is a synthetic biology tool that facilitates both the study of natural physiology as well as engineering of strains for applied purposes. Current methods for increasing titers of products include deleting genes in competing pathways, but maximum yields might be achieved by lowering expression to some intermediate level. With CRISPRi technology we can lower expression of essential genes, decrease but not abolish flux towards competing products, and manipulate cellular processes to varying degrees. Here, we improved lactate production from 0.0457±0.002mM lactate/h to 0.0927±0.001mM lactate/h by repressing glutamine synthetase, *glnA*. This was achieved without reducing autotrophic growth rates or mutating chromosomal genes. We anticipate that this approach could be similarly effective in increasing photosynthetic flux to other chemical products, especially those derived from pyruvate.

6.3 Promoter characterization

A set of characterized and predictable promoters that perform consistently on both native and synthetic sequences is essential to metabolic engineering projects. Maximum expression is not optimal in all cases, so promoter libraries containing a large range of expression are ideal. Promoter libraries such as the Anderson promoter library form *E. coli* are routinely used. Unfortunately promoter libraries cannot easily be ported between one organism to another with similar expression patterns. We tested the Anderson promoter library in both *E. coli* and PCC 7002 and saw poor correlation (R^2= 0.434, **Figure 6.15A**) (Markley et al. 2015). Similar results were seen when we compared the performance of a promoter library created by mutating a cyanobacterial promoter, *cpcB* (R^2= 0.366, Figure 6.15B).

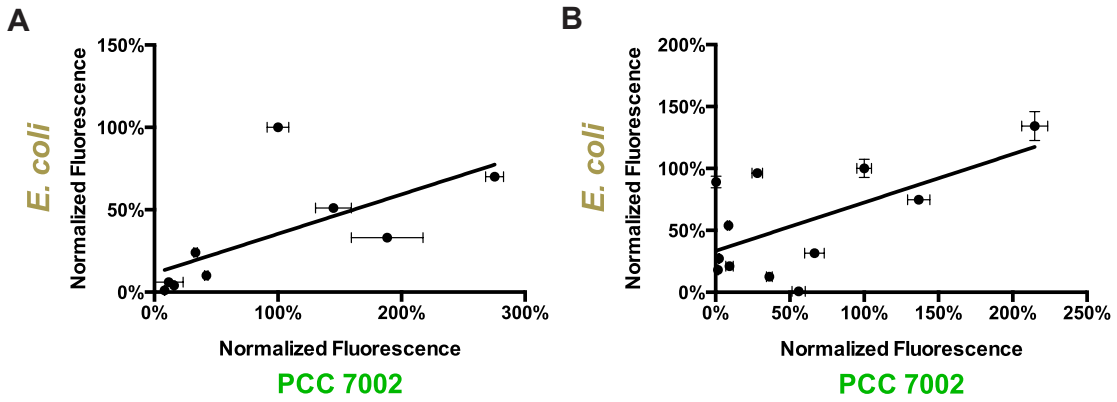


Figure 6.15: Comparison of YFP reporter libraries in *E. coli* and PCC 7002. (A) Anderson promoter library (R^2 = 0.434) (B) *cpcB* derived promoter library (R^2= 0.366)

CO_2 responsive promoter

For large scale cultivation of cyanobacteria it would not be economical to turn on and off genes in a biochemical production pathway using small, expensive molecules. Ideally gene expression could be controlled by an inexpensive environmental stimulus such as light or carbon dioxide levels. Tools that controlled gene expression using these inputs would be useful so that biomass could be grown up, before stimulating production. This type of genetic tool could be also useful to turn on genes to lyse the cyanobacterial cells to release a product.

Throughout the course of analyzing gene expression in PCC 7002 we realized the potential utility of a CO_2-responsive promoter. A LysR-type transcriptional regulator (LTTR), *ccmR*, has been identified in several model cyanobacteria and has been shown to be responsive to CO_2. CcmR controls a regulon of genes that are turned on in the presence of low concentrations of CO_2; this regulon includes bicarbonate transporters, cation/proton anti-porters. We realized that this regulator and it's sequence could be used to control the expression of genes of interest.

The CO_2-sensing mechanism for CcmR in PCC 7002 has not been described, but CcmR in related model cyanobacteria *Synechocystis* sp. PCC 6803 has been shown to be regulated by co-repressors α-ketoglutarate and $NADP^+$ (Daley et al. 2012). LysR-family transcriptional regulators

can act as repressors or activators and act with the presence of a co-inducer. They often have a recognition sequence of T-N_{11}-A that appears at a regulatory binding site (RBS) and activation binding site (ABS). These sites often appear between -35 and +20bp and between -40 and -20bp respectively. These transcriptional regulators are active as tetramers and bend the DNA in the presence of the co-inducer. There seems to be no correlation between positioning of binding sites and amount of DNA bending relating to a LTTR acting as an activator or repressor.

The binding site motif for CcmR was recently shown of PCC 7002 and used to expand the CcmR regulon (McClure et al. 2016). We took the promoter of region of bicarbonate transporter, *sbtA*, that contains CcmR binding sites and used it to drive expression of report YFP. We tested both the native ribosome binding site of *sbtA* driving YFP expression as well as a synthetic ribosome binding site that was predicted to be stronger. We grew these strains in high CO_2 (5%), air (0.4%), and transfered cultures from high CO_2 to air. We saw low YFP levels in high CO_2 as expected and high levels of YFP expression in air (**Figure 6.16**). This native system was fairly tight and showed 15-fold high expression in air than in high CO_2 conditions.

This promoter has potential use as a mechanism to control gene expression. The biomass could be grown at high CO_2 conditions and then production enzymes could be turned on by moving the culture to low CO_2 conditions. This system could also be used in reverse by using this promoter to drive expression of repressors LacI or TetR and causing genes to be repressed in air and turned on in the presence of high CO_2 levels.

6.4 RBS characterization

Alteration of the ribosomal binding site is a useful tool that can be used to manipulate gene expression. There are a number of calculators that can use both 'forward engineering' to design a RBS of an input with a desired strength or 'reverse engineering' to predict the strength of a given input. These calculators predict the translation initiation rate, using factors such as the complementarity of the RBS to the ribosome and the accessibility of the site given RNA secondary structure. These

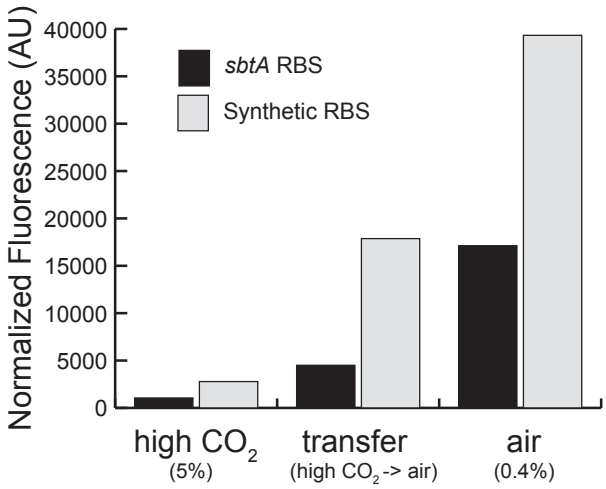


Figure 6.16: YFP fluorescence levels from strains with the *ccmR* promoter driving expression of YFP grown at different CO_2 concentrations

calculators are widely used and show promising results in the model organism, *E. coli*. We tested the utility of the RBS calculator in photosynthetic cyanobacterium PCC 7002 by creating a library of RBS sequences in front of YFP. When the experimental fluorescence values were compared to predictions of the RBS calculator (Salis et al. 2009), there was one significant outlier (**Figure 6.17A**). This presence of this RBS, AGGAGA, appears to have affected the promoter and increased the amount of transcription (**Figure 6.17B**) that could explain the increased fluorescence values.

Although the RBS calculator was a useful tool to generate a wide range of constructs with various strengths, it was less accurate than anticipated (**Figure 6.18**). These predictive tools can be used design RBS libraries for cyanobacteria, but experimental values from individual constructs may not match predictions.The reasons for this finding in cyanobacteria remains unknown, but alteration of the RBS remains an important tool to change gene expression.

6.5 5' hairpin stabilizes transcripts in PCC 7002

There have been attempts to identify sequence elements that confer stability, and one such element was identified: a hairpin at the 5' terminus of a message. This discovery was made by

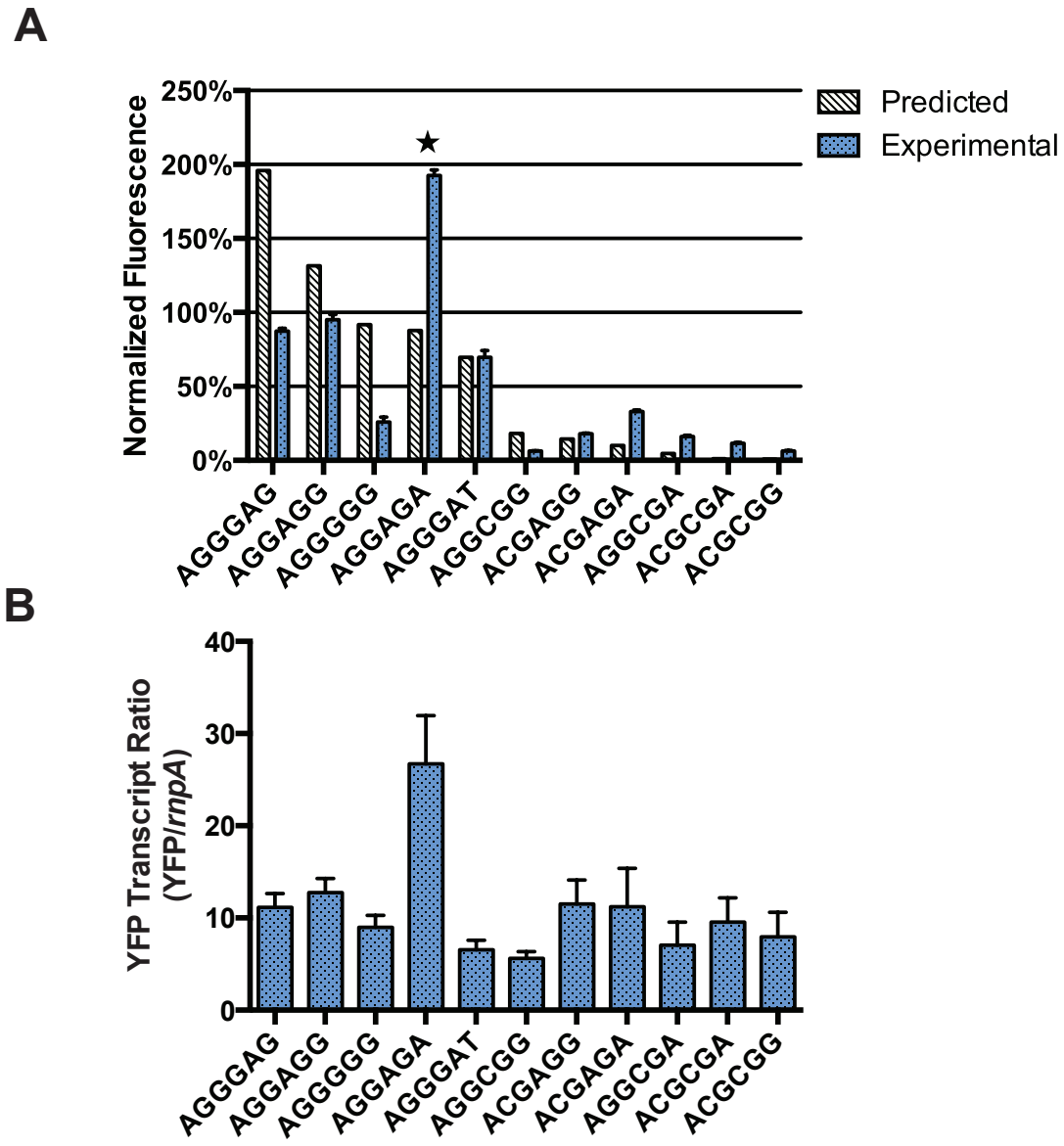


Figure 6.17: A RBS library of a synthetic YFP transcript. (A) Predicted and experimental fluorescence levels of a RBS library encoding yellow fluorescent protein. (B) Only one of the library members had increased transcript levels.

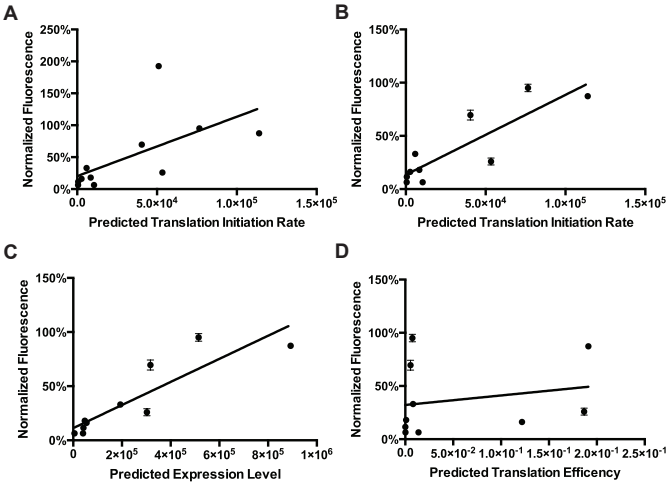


Figure 6.18: Correlation of the predicted translation initiation rates from the RBS Calculator to experimentally determined YFP fluorescence. (A) Correlation with AGGAGA included (R^2 = 0.3760) using the RBS calculator. (B) Correlation with AGGAGA removed (R^2 = 0.7443) using the RBS calculator. (C) Correlation of the experimental data set, excluding AGGAGA, using the UTR Designer (R^2 = 0.7843). (D) Correlation of the experimental data set, excluding AGGAGA, using the RBS Designer (R^2 = 0.04448). Error bars on experimental data represent standard deviation of biological replicates (n = 3).

studying the 5' UTR of the *ompA* transcript in *E. coli*. Fusing the *ompA* 5' UTR onto other messages also led to enhanced stability (Emory et al. 1992). By mutating this sequence, it was determined that the placement of the hairpin within 2-4 nucleotides from the 5' end of the transcript was essential to confer stability, but the sequence within the stem loop was relatively unimportant (Emory et al. 1992).

This finding has been used to engineer transcript stability in *E. coli*. By placing a 5' hairpin in front of a synthetic message the half-life could be increased 3-fold with corresponding increases in transcript and protein levels (Carrier et al. 1997). This concept has also been expanded by showing enhanced stability from a library of different synthetic 5' hairpins (Carrier et al. 1997).

This enhanced stability through a 5' hairpin is thought to be due to the fact that the enzyme that initiates degradation, RNase E, is 5'-end-dependent. Despite being a endoribonuclease that cleaves RNA internally, it's activity is affected by the status of the 5' end of the message. It

has reduced activity on circular transcripts that lack a 5' end as well as transcripts with annealed antisense nucleotides at the 5' end. RNase E is much more efficient at cleaving transcripts with a 5' monophosphate compared to a 5' triphosphate (Mackie 1998). Crystal structures have identified a 5' sensing region within the protein that makes contacts with the 5' end of the transcript (Callaghan et al. 2005). This 5'-end-dependent regulation was further explained when a RNA pyrophosphohydrolase was identified that could convert the 5' triphosphate to a monophosphate (RppH) (Deana et al. 2008). Recent data showing that many transcripts ends are diphosphorylated, suggests that another unidentified enzyme may be responsible for the initial removal of a single phosphate followed by subsequence action by RppH to produced the monophosphorylated end (Luciano, Vasilyev, et al. 2017). This sensing and regulation via the status of the 5' end could explain how obstruction of the 5' end of the transcript due to RNA secondary structure could lead to reduced activity of RNases.

The observation that a 5' hairpin can enhance stability of a transcript has also been observed in *B. subtilis*. The structured 5' UTR of *aprE* has been shown to be both responsible for its stability and has been used to enhance stability of a heterologous transcript (Hambraeus, Persson, et al. 2000). The conservation of the effects of a 5' hairpin on transcript stability is remarkable because of the fundamental differences in transcript degradation between *E. coli* and *B. subtilis*. Despite having different RNases that do the initial cleavage event (RNase E for *E. coli* and either RNase J or RNase Y for *B. subtilis*) transcript stability is affected by the structure of the 5' end of the message. RNase J and RNase Y have also been shown to have 5' sensing regions (Pei et al. 2015). Additionally RNase J and RNase Y have preferential activity towards the 5' monophosphorylated transcripts compared to 5' triphosphorylated transcripts (Shahbabian et al. 2009). The identification of a homolog of RppH in *B. subtilis* supports the hypothesis that regulating transcript degradation through the sensing of the 5' end of a message is conserved across diverse species.

We wondered if 5' hairpins could enhance stability of transcripts in PCC 7002. We made several reporter constructs encoding yellow fluorescent protein (YFP) and tested their stability and expression levels. A base construct without a hairpin was compared to a construct with a

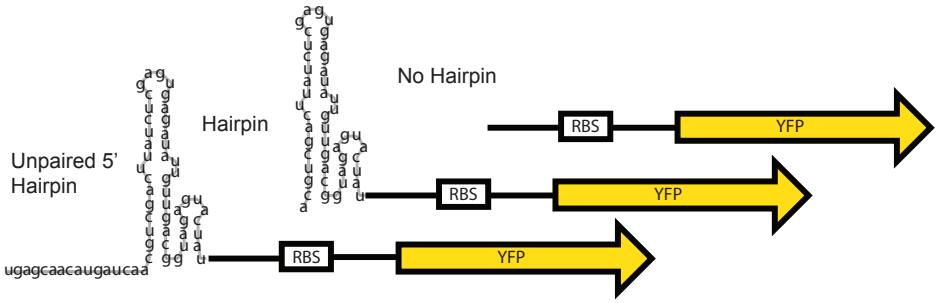


Figure 6.19: Diagram of constructs used to test effects of 5' hairpin on gene expression and transcript half-life. The hairpin used is a synthetic hairpin used in (Carrier et al. 1999).

hairpin located at the terminal 5' end (**Figure 6.19**). We also made a construct with extra bases preceding the hairpin to examine if the location of the hairpin at the terminal 5' end was important for stability. The hairpin was a synthetic hairpin (HP4) used in (Carrier et al. 1999). All constructs were expressed from a strong constitutive promoter, P_{J23119}.

We integrated all three constructs onto the genome of PCC 7002 and measured the fluorescence levels, steady-state transcript levels, and transcript half-life. We saw increased fluorescence, increased steady-state transcript levels, and an increased half-life from $\sim$1.5 minutes to $\sim$3 minutes with the 5' hairpin (**Figure 6.20**). When extra bases were added preceding the hairpin we saw similar fluorescence levels, transcript levels, and a similar half-life to the control without a hairpin. This data suggests that like in *E. coli* and *B. subtilis* the structure and accessibility of the 5' end of a transcript influences transcript stability. This finding is not surprising because PCC 7002 has homologs of both RNase E and RNase J, with matching residues in the 5' sensing pockets of the respective crystal structures. There has been no homolog of RppH identified in PCC 7002 (**Chapter 3**), but due to these findings that show 5' secondary structure can effect transcript stability it seems likely to exist. RppH belongs to the Nudix (nucleoside diphosphates linked to x) family of proteins whose substrates and products can be difficult to identify without in vitro characterization.

We wanted to see if this stabilization via a 5' hairpin could be used as a genetic tool across a range of expression levels. We added this hairpin to a library of different ribosome binding sites

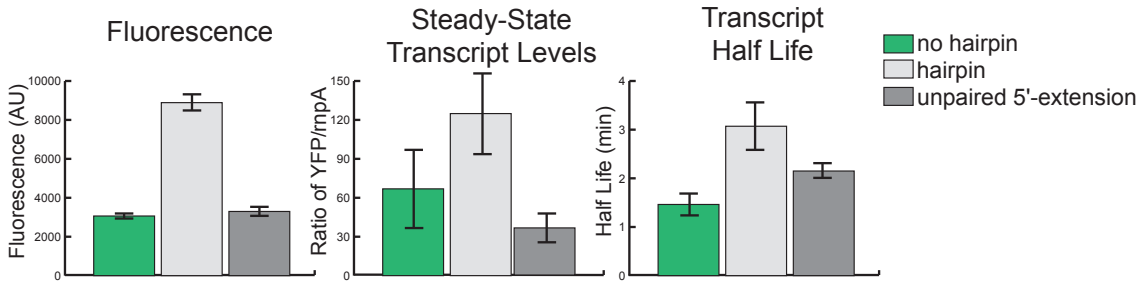


Figure 6.20: A hairpin at the 5' end of a transcript increases protein expression, steady-state RNA levels, and transcript stability. This stabilization is removed when unpaired bases are placed in front of the hairpin.

controlling the translation of YFP. We measured YFP fluorescence levels from the two libraries and found that the version with the hairpin also had higher fluorescence levels (**Figure 6.21**).

Like in other organisms, hairpins can be used to stabilize transcripts in PCC 7002. Global mRNA turnover seems to be very rapid in PCC 7002 (**Chapter 5**), and the increase in stability was fairly modest (from 1.5 to 3 minutes). Other sequence elements may need to be explored if extreme stability is desirable. This data indicates that PCC 7002 RNases likely sense the 5' end of the transcript and it's phosphorylation state affects RNase activity. Future work may reveal the other enzymes, such as RppH, that facilitate this process.

This is the first example of a genetic tool altering expression via changes in RNA stability in cyanobacteria. Regulation of heterologous or synthetic transcripts on the level of RNA stability is still fairly uncommon and overlooked even in other model organisms. For precise control of gene expression it will be necessary to control gene expression on every level; currently gene expression in often manipulated by changing promoters and levels of transcription, but RNA degradation is just as an important of process.

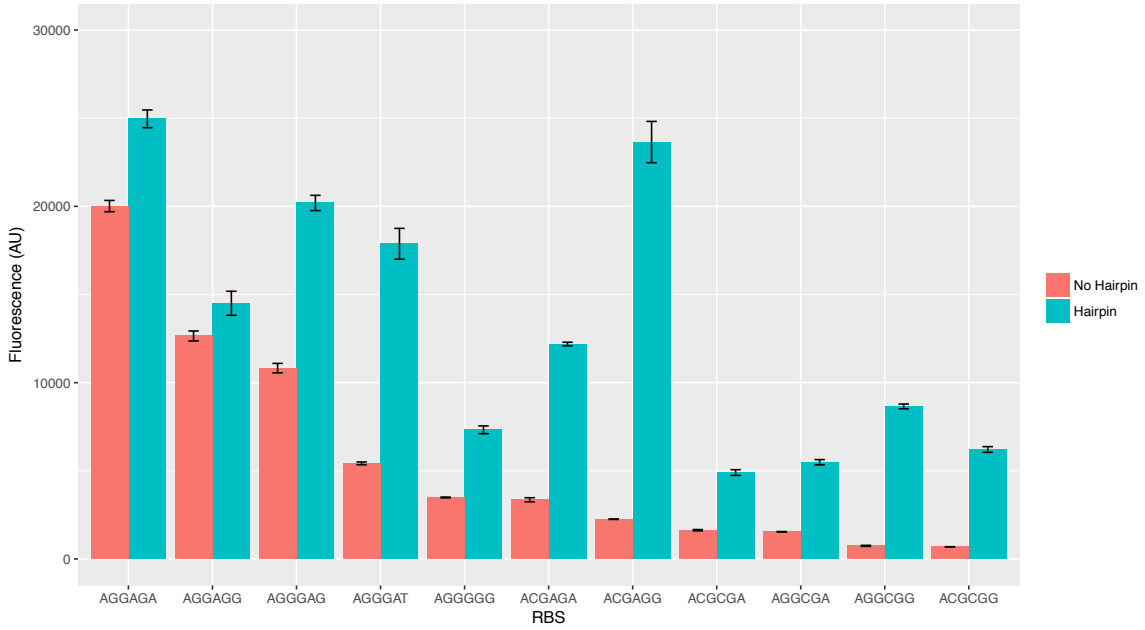


Figure 6.21: Fluorescence levels increase in RBS library with a 5' hairpin.

6.6 Materials and Methods

Chemicals, reagents, and media

Strains were grown and maintained on media A+ (Stevens, Patterson, et al. 1973) with 1.5% (w/v) Bacto-Agar (Fisher). Strains with antibiotic resistance markers were selected on media with antibiotics (kanamycin, 100 µg/mL; gentamicin, 30 µg/mL) and strains with cassettes introduced in the *acsA* locus were plated on 100 µM acrylic acid. Strains were grown in glass culture tubes (2x15 cm) with 20 mL media A+ and bubbled with either air (0.04% CO_2)or high CO_2 (10% CO_2). Temperature was maintained at 38°C and light intensity was approximately 150 µmol photons $m^{-2}s^{-1}$. Optical density was measured in a Genesys 20 spectrophotometer (Thermo Scientific) in 1-cm cuvette.

Strain construction: CRISPRi

All CRISPRi strains are listed in Table 6.2.The plasmids used to construct the CRISPRi strains are listed in **Table 6.1**.Plasmids were cloned in *Escherichia coli* DH5α. The sequences of chromosomal cassettes are available as supplementary material. Genes and sgRNA were transformed onto the chromosome of wild-type PCC 7002 using homologous recombination. dCas9 was introduced at the *acsA* locus (Begemann et al. 2013) and the sgRNA and a kanamycin resistance marker were introduced at the NS1 site (Davies et al. 2014). In some strains EYFP or the LDH with the $P_{cLac143}$ inducible promoter system were placed in the *glpK* pseudogene along with a gentamicin resistance marker (Begemann et al. 2013). Constructs were made using Gibson assembly (Gibson et al. 2009) with regions of homology added in the 5' end of the primers. Site specific mutations in LDH enzymes was based on the V39R mutation described for the L-LDH of *B. subtilis* strain 168 (Richter et al. 2011). In the case of the *B. subtilis* LDH, the point mutation was made by site-directed mutagenesis protocol (Ho et al. 1989) with overlap extension PCR and specific primers, followed by a DpnI treatment and transformation to chemically competent *E. coli* DH5α. Some sgRNA were instead made by adding the 20 nucleotide target sequence in the 5' region of an oligonucleotide primer and amplifying all of the way around a previously constructed plasmid to create one linear piece. After purification, PCR products were phosphorylated with T4 polynucleotide kinase and ligated with T4 DNA ligase (NEB) to form circular plasmids. sgRNAs targeting the non-template strand were picked with DNA 2.0's gRNA designer by entering a gene of interest. Potential target sequences with a NGG PAM sequence within the first 500 nucleotides from the start codon were used in a BLAST search against the wild-type PCC 7002 genome to ensure no substantial off target effects.

Table 6.1: Plasmids used to create CRISPRi strains

Plasmid Name	Backbone	Resistance	Reference	Contents
pEZ03	pACSA	amp	(Zess et al. 2016)	aTc-inducible system template
pdCas9-bacteria	p15A	cm	(Qi et al. 2013)	dCas9 template
pGCG26	pUC19	amp, gm	This work	YFP expressed from P_{J23119} in the *glpK* locus with gentamicin resistance
pCas2A	pACSA	amp	This work	dCas9 expressed from aTc-inducible promoter in *acsA*, A RBS: AGGAGA
pCas2B	pACSA	amp	This work	dCas9 expressed from aTc-inducible promoter in *acsA*, B RBS: TCGAGA
pCas2C	pACSA	amp	This work	dCas9 expressed from aTc-inducible promoter in *acsA*, C RBS: TGGACA
pCas2D	pACSA	amp	This work	dCas9 expressed from aTc-inducible promoter in *acsA*, D RBS: AGGACG
pCas2E	pACSA	amp	This work	dCas9 expressed from aTc-inducible promoter in *acsA*, E RBS: AGGGCG
pCas2F	pACSA	amp	This work	dCas9 expressed from aTc-inducible promoter in *acsA*, F RBS: TGGGCG
pCas5	pUC19	amp, km	This work	sgRNA targeting YFP expressed from P_{J23119} in the NS1 locus with kanamycin resistance
pCas7	pACSA	amp	This work	dCas9 expressed from low constitutive promoter, c225, in *acsA*
pCas8	pACSA	amp	This work	dCas9 without a promoter in *acsA*
pCas13	pUC19	amp, km	This work	sgRNA targeting YFP (truncated to 12 bp) expressed from P_{J23119} in the NS1 locus with kanamycin resistance
pCas14	pUC19	amp, km	This work	sgRNA targeting YFP (truncated to 15 bp) expressed from P_{J23119} in the NS1 locus with kanamycin resistance
pCas15	pUC19	amp, km	This work	sgRNA targeting YFP (truncated to 18 bp) expressed from P_{J23119} in the NS1 locus with kanamycin resistance
pCas27	pUC19	amp, km	This work	sgRNA targeting YFP expressed from P_{J23117} in the NS1 locus with kanamycin resistance
pCas28	pUC19	amp, km	This work	sgRNA targeting YFP expressed from P_{J23108} in the NS1 locus with kanamycin resistance
pCas34	pUC19	amp, km	This work	sgRNA targeting YFP expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas35	pUC19	amp, km	This work	sgRNA targeting *cpcB* expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas36	pUC19	amp, km	This work	sgRNA targeting *ccmK* expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas39	pUC19	amp, km	This work	sgRNA targeting *glnA* expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas45	pUC19	amp, km	This work	sgRNA targeting YFP with additional bases 5' end like pCas34, but from P_{J23119} in the NS1 locus with kanamycin resistance

pCas49	pUC19	amp, gm	This work	*rbcL*-sfGFP expressed from P_{ccmK_2} promoter in *glpK* with gentamicin resistance
pCas52	pUC19	amp, km	This work	sgRNA targeting YFP +52 from TSS; non-template strand; expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas55	pUC19	amp, km	This work	sgRNA targeting YFP +172 from TSS; non-template strand; expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas56	pUC19	amp, km	This work	sgRNA targeting YFP +46 from TSS; template strand; expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas57	pUC19	amp, km	This work	sgRNA targeting YFP +111 from TSS; template strand; expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas59	pUC19	amp, km	This work	sgRNA targeting YFP +787 from TSS; non-template strand; expressed from aTc inducible system in pEZ03 in the NS1 locus with kanamycin resistance
pCas61	pACSA	amp, gm	This work	IPTG inducible, cLac143 (Markley et al. 2015), LDH V39R with Gm^R in *glpK*
pCas73	pUC19	amp, km	This work	Δ*ccm*:Km, used to delete *ccm* operon

Table 6.2: *Synechococcus* sp. strain PCC 7002 CRISPRi strains

	Locus			
Strain	***acsA***	***glpK***	**NS1**	***ccm***
WT	X	X	X	X
GG47	X	YFP (pGCG26)	X	X
CC1	X	YFP (pGCG26)	sgRNA YFP from pJ23119 (pCas5)	X
CC91	dCas9 A RBS aTc (pCas2A)	YFP (pGCG26)	X	X
CC2	dCas9 A RBS aTc (pCas2A)	YFP (pGCG26)	sgRNA YFP from P_{J23119} (pCas5)	X
CC115	dCas9 A RBS aTc (pCas2A)	YFP (pGCG26)	sgRNA YFP from P_{J23108} (pCas28)	X
CC114	dCas9 A RBS aTc (pCas2A)	YFP (pGCG26)	sgRNA YFP from P_{J23117} (pCas27)	X
CC116	dCas9 A RBS aTc (pCas2A)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC107	dCas9 B RBS aTc (pCas2B)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC120	dCas9 C RBS aTc (pCas2C)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC121	dCas9 D RBS aTc (pCas2D)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC110	dCas9 E RBS aTc (pCas2E)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X

CC111	dCas9 F RBS aTc (pCas2F)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC86	dCas9 F RBS aTc (pCas2F)	X	sgRNA *cpcB* aTc (pCas35)	X
CC90	dCas9 F RBS aTc (pCas2F)	X	sgRNA *ccmK* aTc (pCas36)	X
CC104	X	X	sgRNA *ccmK* aTc (pCas36)	X
CC85	X	X	X	Δ*ccm*:Km (pCas73)
CC125	X	*rbcL*-GFP (pCas49)	X	Δ*ccm*:Km (pCas73)
CC126	X	*rbcL*-GFP (pCas49)	X	X
CC129	dCas9 F RBS aTc (pCas2F)	*rbcL*-GFP (pCas49)	sgRNA *ccmK* aTc (pCas36)	X
CC130	X	cLac94 LDH (pCas61)	sgRNA *glnA* aTc (pCas39)	X
CC131	dCas9 F RBS aTc (pCas2F)	cLac94 LDH (pCas61)	sgRNA *glnA* aTc (pCas39)	X
CC133	X	cLac94 LDH (pCas61)	X	X
CC142	X	cLac94 LDH (pCas61)	sgRNA YFP aTc (pCas34)	X
CC117	dCas9 dead from c225 (pCas7)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC109	no promoter dCas9 (pCas8)	YFP (pGCG26)	sgRNA YFP aTc (pCas34)	X
CC122	dCas9 A RBS aTc (pCas2A)	YFP (pGCG26)	sgRNA YFP add 7nt aTc (pCas45)	X
CC13	dCas9 C RBS aTc (pCas2C)	YFP (pGCG26)	X	X
CC15	dCas9 C RBS aTc (pCas2C)	YFP (pGCG26)	sgRNA YFP 12 bp P_{J23119} (pCas13)	X
CC25	dCas9 C RBS aTc (pCas2C)	YFP (pGCG26)	sgRNA YFP 15 bp P_{J23119} (pCas14)	X
CC26	dCas9 C RBS aTc (pCas2C)	YFP (pGCG26)	sgRNA YFP 18 bp P_{J23119} (pCas15)	X
CC7	dCas9 C RBS aTc (pCas2C)	YFP (pGCG26)	sgRNA YFP from P_{J23119} (pCas5)	X
CC135	dCas9 F RBS aTc (pCas2F)	YFP (pGCG26)	sgRNA YFP +52 nt (pCas52)	X
CC137	dCas9 F RBS aTc (pCas2F)	YFP (pGCG26)	sgRNA YFP +172 nt (pCas55)	X
CC138	dCas9 F RBS aTc (pCas2F)	YFP (pGCG26)	sgRNA YFP +46 t (pCas56)	X
CC139	dCas9 F RBS aTc (pCas2F)	YFP (pGCG26)	sgRNA YFP +111 t (pCas57)	X
CC141	dCas9 F RBS aTc (pCas2F)	YFP (pGCG26)	sgRNA YFP +787 nt (pCas59)	X

Strain construction: 5' hairpin

All strains used to measure the effects of a 5' hairpin and the plasmids used to create them are listed in **Table 6.3**. Plasmids were built as in (Markley et al. 2015) and integrated into the *glpK* and selected for segregation with gentamicin. All strains used in the RBS libraries are listed in Table 6.4. Plasmids were built as in (Markley et al. 2015) and integrated into the *acsA* locus using selection on acrylic acid (Begemann et al. 2013).

Measurement of fluorescence

Cultures (20 mL media A+) were bubbled with air for 24 h before being diluted in triplicate to $OD_{730\ nm}$ = 0.5 in 10 mL of media A+. Cultures were bubbled for 24 h to an $OD_{730\ nm}$ ∼1 where 1.5 $OD_{730\ nm}$-mLs were spun down at 1400xg for 10 min. Cell lysis with Bugbuster and fluorescence measurements on a Tecan M1000 plate reader were performed as described previously (Markley et al. 2015).

Reversibility test

Cultures (20 mL media A+) were inoculated from fresh plates and bubbled with air for 48 h. Cultures were diluted to ∼0.1 $OD_{730\ nm}$mL in 20 mL A+ in quadruplicate and 1 µg/mL aTc was added where indicated. A fluorescence measurement of 200 µl of cells was taken in a Tecan M1000 plate reader with a gain of 200. Cultures were spun down at 5000xg for 10 min in a Beckman Coulter Allegra X-15 R Centrifuge, washed once with A+ media, and resuspended in the same volume of A+ media to maintain a constant cell density pre- and post-wash.

Table 6.3: Strains and plasmids used to measure effects of 5' hairpin

	PCC 7002 Strain Name	**Plasmid used for strain construction**
no hairpin	AM183	pALM117
5' hairpin	AM102	pALM102
unpaired 5' hairpin	GG42	pGCG23

Table 6.4: Strains and plasmids used to build RBS libraries

	PCC 7002 Strain Name		Plasmid used for strain construction	
RBS	**No Hairpin**	**Hairpin**	**No Hairpin**	**Hairpin**
AGGAGA	AM184	AM104	pALM120	pALM105
AGGAGG	AM170	AM107	pALM115	pALM108
AGGGAG	AM169	AM105	pALM113	pALM106
AGGGAT	AM172	AM97	MB P_{J23119}	pALM91
AGGGGG	AM182	AM106	pALM114	pALM107
ACGAGA	AM183	AM102	pALM117	pALM102
ACGAGG	AM185	AM103	pALM121	pALM103
ACGCGA	AM175	AM99	pALM122	pALM97
AGGCGA	AM171	AM100	pALM116	pALM98
AGGCGG	AM174	AM101	pALM119	pALM99
ACGCGG	AM173	AM98	pALM118	pALM96

Measurement of absorbance

Cultures (20 mL media A+) were bubbled with air for 24 h before being diluted in triplicate to $OD_{730\ nm}$=0.5 in 10 mL of media A+ with aTc added where indicated. Cultures were bubbled for 24 h with air before an absorbance scan (300–750 nm) was conducted in 96 well plates with a working volume of 200 µl using a Tecan M1000 plate reader.

Spot plating

Cultures (20 mL media A+) were inoculated from plates and grown with 10% CO_2 for 24 h. Cultures were diluted to $OD_{730\ nm}$=0.5 in 20 mL A+, aTc was added where indicated, and bubbled for another 24 h with 10% CO_2. Samples were suspended at $OD_{730\ nm}$=0.01 as measured by a Genesys 20 spectrophotometer, 10-fold serially diluted in media A+, and spotted onto solid media (7 µl/spot).

Fluorescence microscopy

Cultures (20 mL media A+) were inoculated from plates and bubbled with 10% CO_2 for 24 h with aTc where indicated. Cultures were diluted to $OD_{730\ nm}$=0.05 in 20 mL media A+ and aTc was added where indicated, and bubbled for another 24 h with 10% CO_2. Cultures were diluted to $OD_{730\ nm}$=0.05 in 20 mL A+ with aTc and grown with 10% CO_2 for 5 h before visualizing. Cells (2 µl) were spotted onto 1% (w/v in media A+) agarose pads in a 16-well chamber slide (Nunc™Lab-Tek™, Scotts Valley, CA), air-dried, and covered with a 0.17mm coverslip. Images were acquired on a Zeiss Axioimager Z2 for GFP (excitation: BP 470/40 nm; beam splitter: FT 495 nm; emission: BP 525/50 nm), and brightfield using a 100x oil-immersion objective (NA=1.3). Images were analyzed with ImageJ (Abràmoff et al. 2004).

Lactate quantification

Cultures (10mL media A+) were inoculated from plates and bubbled with air for 24 h before being diluted in duplicate to $OD_{730\ nm}$=0.05 in 15 mL of media A+, induced with 1mM IPTG, and bubbled with either air (0.04% CO_2) or high CO_2 (1% CO_2). Temperature was maintained at 38°C and light intensity was approximately 250 µmol photons $m^{-2}s^{-1}$. Samples were withdrawn periodically for biomass and lactate measurements, centrifuged, and the supernatant was stored at -20°C until quantification by HPLC (Shimadzu Co., Columbia, MD, USA) equipped with a quaternary pump, autosampler, vacuum degasser, photodiode array and refractive index detector. HPLC separations were performed using an Ultra Aqueous C18 column (Restek). The HPLC operating conditions were as follows: mobile phase: 50mM KH_2PO_4 (pH 2.5 with 1% acetonitrile), flow rate 0.100 mL/min, column temperature 30°C, photodiode array detector at 210 nm, run time: 10 min, injection volume 10 mL. Organic acids were quantified by comparison with peaks generated by known amounts of lactate sodium salt (Sigma) in media A+. Dry cell weight concentrations were estimated from $OD_{730\ nm}$ values using a standard curve created from lyophilized cell pellets (1

$OD_{730\ nm}$= 1.409 g/L; R^2=0.9852). Biomass productivity was calculated over the time course using a linear fit. Lactate concentrations for each individual culture are plotted with respect to time and a linear fit was applied.

Half-life measurements

20 mL A+ starter cultures were grown overnight and diluted in triplicate to $OD_{730\ nm}$=0.05 in 20 mL and bubbled with air overnight. The OD_{730nm} was taken and an initial 3.5 mL sample was taken pre-rifampicin addition. Rifampicin (200µg/mL in DMSO) was added to stop transcription. 3.5 mL samples were taken and added to 350 µL of cold stop solution (5% phenol in ethanol) at 0.5 min, 1.5 min, 3 min, and 5 min after the addition of rifampicin. Cells were spun down at 4750 RPM for 10 min at 4°C, supernatant was aspirated, and cells were flash-frozen in liquid nitrogen and stored at -80°C until RNA was extracted.

RNA extraction and qPCR

RNA was extracted using the TRIzol 95 method (Pinto et al. 2009). cDNA was synthesized using random hexamers with iScript™ cDNA Synthesis Kit (Bio-Rad) on 500 ng total RNA per manufacturers instructions. RT-qPCR was performed on 10 µl reactions using iQ SYBR Green Supermix on a CFX Connect TM Real-Time PCR Detection System (Bio-Rad). 1 µl of the cDNA reaction was added to each reaction and purified PCR products of genes of interest were used to create standard curves for each primer pair. Primers AM33 (GGTGTTCTGCTGGTAGTGGT) and AM34 (GACGACGGCAACTACAAGAC) were used to detect YFP. Primers GG37 (AGAA-CACCGTCTAAAACACTG) and GG38 (GATTAAATTCCGAACCACTG) were used to detect the control gene *rnpA*. Denaturation at 95°C for 3 min was followed by 36 cycles of 95°C for 15 seconds, 54°C for 30 seconds, and 72°C for 1 minute. A melt curve with increments of 0.5°C every 5 seconds.

7 Chapter 7: Conclusions

Authors and Contributors:

Gina Gordon: Wrote text.

Dr. Brian Pfleger: Supervised research and edited text.

7.1 Summary of thesis research

The research described in this document probes aspects of gene expression regulation in cyanobacteria in an effort to build reliable genetic tools. Cyanobacteria are industrially relevant organisms that could be used as a green alternative in chemical production; they can produce chemicals and products from sunlight and carbon dioxide, and can be grown on non-arable land with non-potable water. They provide a good alternative so that crops can be used as a food source instead of broken down into sugar to feed heterotrophs. Although there have been many chemicals produced in cyanobacteria (Knoot et al. 2017), their titers are still low compared to many heterotrophs. One cause of these low titers may be the lack of predictable and reliable genetic tools to control gene expression.

Genetic tools are well advanced in model organisms such as *E. coli*, but unfortunately they cannot be directly ported into cyanobacteria with reliable results. In Chapter 1 we reviewed the current genetic tools available in cyanobacteria such as the use of heterologous promoters, induction systems, and CRISPR interference systems. One common finding was that genetic tools developed in other systems did not work as expected when moved directly into cyanobacteria. Many of these genetic tools had to be optimized in cyanobacteria by iterations of building and testing constructs with variations in sequence and spacing of elements. We highlight several key areas that could be contributing to differences in gene expression: major differences in the transcriptional landscape (transcriptional machinery and regulators), the high polyploidy, critical role of many small proteins, and presence of a highly over represented repeat sequence, HIP1. These fundamental aspects of cellular physiology may play important yet uncharacterized roles in gene expression.

One area that is underutilized and often overlooked when building genetic tools is the area of RNA stability. The lack of this type of genetic tool likely stems from the lack of knowledge of how specific transcripts are turned over and which enzymes facilitate this process. We've addressed this need by examining the specificity of RNases and their role in RNA processing and turnover. To advance the control of gene expression in cyanobacteria we explored native regulation of RNA

turnover and also built, tested, and optimized several genetic tools that will enable metabolic engineering and chemical production in cyanobacteria.

Progress towards understanding native regulation of RNA turnover

RNA degradation is a critical process that controls the amount of a transcript, yet we have little predictive power to explain how fast any specific transcript will degrade or the specific RNases that facilitate this process. There are advanced models of transcript degradation in both Gram-negative model organism *E. coli* and Gram-positive model organism *B. subtilis*, yet many other prokaryotes possess elements of both systems. We took two approaches to gain a better understanding of RNA turnover in cyanobacteria: the first being ribonuclease specific by determining the function of specific enzymes and the second, a global study of RNA degradation in an attempt to identify stabilizing (or destabilizing) elements.

We identified the RNases that facilitate RNA turnover in both PCC 7002 and 65 other representative cyanobacterial genomes (**Chapter 3**). All 66 cyanobacterial genomes contained one homolog of RNase E, one homolog of RNase J, and at least one homolog of RNase III, RNase II or RNase R, and PNPase. To investigate the role and essentiality of these RNases we attempted to delete each of these homologs in PCC 7002. We successfully deleted all three homologs of RNase III singly and in all combinations. We also deleted one of the two homologs of RNase II or R, but could not solidify its identity due to the high similarity of amino acid sequence. We did not obtain segregated mutants when we attempted to delete the homolog of RNase E, RNase J, or the other homolog of RNase II/R, suggesting that they may be essential. Many cyanobacterial genomes had several copies of RNase III, most with between one and three copies of the full length RNase III and an additional mini-RNase III that lacks the double-stranded RNA binding motif. We observed no growth defect in any of the PCC 7002 single RNase III mutants, but whenever two homologs (A0061 and A0384) were deleted in combination there was slightly reduced growth compared to WT. The generation of these mutant strains enabled further study of the function and roles of these RNases.

We were interested in examining the activity and specificity of an endoribonuclease, RNase III. We were initially interested in RNase III because of it's somewhat more specific target sequence than other RNases. RNase III recognizes double-stranded RNA and before these experiments, only a handful of known cleavage sites were known in well-studied *E. coli*. The RNase III enzyme and recognition sequence could be used as tool to alter the transcript stability. Initially we hypothesized that expression of the enzyme would lead to cleavage of it's target, and rapid degradation of the message. Preventing expression of RNase III (an unessential enzyme) would presumably enhance stability of the message. To explore this potential tool to regulate transcript stability, we would ideally use a heterologous RNase III that would not interfere with native processes such as rRNA maturation. We wanted to explore the function and specificities of three RNase IIIs in PCC 7002. It was unclear if they were completely redundant or had specific activities and target sites. If the target sequences of the cyanobacteria RNase IIIs were different from that of *E. coli* they could potentially be used to regulate transcript stability in *E. coli*.

We decided to leverage RNA-sequencing as a tool to probe transcript processing and RNase activity. There were no previous examples of using RNA-sequencing to pinpoint RNase cleavage sites, so we decided to do a preliminary study in *E. coli* where there were known cleavage sites that could validate our approach (**Chapter 2**). By taking steady-state samples of a WT strain and a strain that lacked RNase III as well as a series of samples after the addition of rifampicin to stop transcription, we were able to develop criteria to identify RNase III cleavage sites. We were able to validate our approach by identifying most of the known RNase III cleavage sites, but were also able to identify novel cleavage sites. We were able to show that many of our putative sites identified through analyzing the RNA-sequencing data were able to be cut by purified RNase III in vitro.

We made four major observations involving the role of RNase III in transcript degradation. The first observation was that for several cleavage sites, processing by RNase III lead to degradation of the message. This pattern of degradation had been observed before for many previously characterized RNase III cleavage sites. Surprisingly, we also observed the opposite where cleavage by RNase III

resulted in the transcript becoming more stable. Two of our novel cleavage sites were within leader sequences involved in transcriptional attenuation (*hisL* and *pheM*). The verified in vitro cleavage sites overlapped with the known termination structures, suggesting that RNase III could initiate and ensure turnover of the terminated transcripts. Finally, we observed that several RNase III cleavage events resulted in fragments that were almost identical to those produced from an alternate promoter. RNase III seems to be playing a role in separating open reading frames transcribed within the same operon and that those precise 5' ends are important for proper expression and regulation. We significantly expanded the list of known processing sites of RNase III in *E. coli* and showed that RNase III is involved in processing transcripts encoding important metabolic enzymes, including pyruvate dehydrogenase. This study showed that RNA-sequencing can be used to identify RNase cleavage sites, and that RNase III participates in complicated regulation much of which has yet to be fully described.

We planned to use the same framework and experimental setup to study the three homologs of RNase III in cyanobacteria (**Chapter 4**). We collected steady-state samples from WT PCC 7002, the three single RNase III mutants, double RNase III mutants, and the triple RNase III mutants and did RNA-sequencing. We hoped that by using the same methods and criteria developed in the *E. coli* study we could elucidate the function of each individual RNase III. Unfortunately we did not observed any locations that resembled RNase III cleavage sites in the cyanobacterial dataset. We hypothesize that these RNase IIIs may be processing small RNAs or transcription factors that are then resulting in global and synchronized movements of many regulons. We were able to identify the roles of two homologs of RNase III (A0061 and A0384) in rRNA processing and the third RNase III's (A2542) influence on plasmid copy number regulation. Additionally we described the global shifts in gene expression, specifically the up-regulation of many bicarbonate transporter and CO_2 fixation-related genes in the single mutants but not the double or triple RNase III mutants. We hypothesize that the single mutants can overcame a defect in carbon fixation and metabolism by large changes in gene expression, but the double and triple mutants cannot. To address the

similarities and differences between the cyanobacterial RNase IIIs and the *E. coli* enzyme, we tested the ability of cyanobacterial RNase IIIs to cut known *E. coli* RNase III targets in vitro. We found some differences between the cyanobacterial enzymes and that of *E. coli*, with one structure being cleaved at the exact same location while others cut at nearby locations or not at all. These data suggest that there are fundamental differences in target sequences among these enzymes, and that deciphering their specificities requires additional detailed investigations.

To complement this ribonuclease-specific study of RNA degradation in cyanobacteria, we also examined global RNA turnover (**Chapter 5**). By collecting samples from WT PCC 7002 following a timecourse after the addition of rifampicin to stop transcription, we were able to calculate half-lives for greater than 90% of the genes. Global turnover was very rapid with a median half-life of 0.97 minutes, which is faster than any other reported in prokaryotes. We examined if half-life was correlated with any specific cellular role, and found that expression of energy metabolism genes, specifically photosynthesis genes was significantly more stable compared to all other functional categories. Energy metabolism genes were highly expressed, and it remains unknown whether these transcripts are abundant because they are so stable, or if high levels of expression assist in their stability.

Using this dataset we had hoped to identify sequence elements or features that enhanced stability and could be adapted into genetic tools. We looked for motifs in the 5' UTR and 3' UTR that were enriched in stable genes compared to unstable genes. We identified a 21 base motif that was enriched in the 3' UTR that included a string of 'T's that resembled a rho-independent terminator sequence. By grouping genes by the presence, or absence and type of terminator sequence in the 3' UTR, we found that genes with a L-shaped (hairpin with U-tract) terminator had a longer half-life than those with an I-shaped (hairpin with no U-tract) or no terminator. This observation that terminators may play a role in transcript stability is not unprecedented because 3' secondary structure is thought to prevent 3' to 5' exoribonucleases from degrading the transcript. The role of the U-tract and it's presence or absence in a terminator has been debated, and it remains to be seen

if the presence of the U-tract enhances transcript stability in other organisms.

Regulatory tools for controlling gene expression

In this document we have developed and characterized several genetic tools for reliable and predictable gene expression in cyanobacteria. This includes the optimization of a CRISPRi interference system that can be used to decrease expression levels of any target in trans (**Chapter 6**). An important and novel feature of this system is that the target can be repressed to varying degrees by titrating the amount of inducer. We showed how this system can be used to repress native targets such as the unessential light-harvesting phycobilisome complex as well as the essential carbon-fixation structure, the carboxysome. We used this genetic tool to increase the production of lactate by repressing a native nitrogen assimilation gene. Lactate production increased by 2-fold over a base engineered strain and is the highest photosynthetically-produced lactate to date.

We also examined the use a 5' hairpins as stabilizing elements in PCC 7002. Hairpins at the terminal 5' end have been shown to increase stability of a transcript in both *E. coli* and *B. subtilis*, presumably due to the fact that initiating RNases sense and bind the 5' end of the transcript. It was unclear initially if this would be the case in PCC 7002 because the preferential activity based on 5' phosphorylation of RNases has not been examined in PCC 7002. Our data suggests that the RNases, both RNase E and RNase J, likely sense the 5' end of transcripts because of the stabilizing effect we see of adding a 5' terminal hairpin. The stabilization we saw was only modest (an increase in half-life from 1.5 to 3 minutes), and other tools may need to be developed to create even more stable transcripts.

7.2 Future Trends

The ability to do metabolic and synthetic biology in cyanobacteria is heavily reliant on the quality and availability of gene expression regulatory tools. We suspect that the cyanobacterial

community will continue to develop novel and more dependable gene expression regulatory systems that will enable rapid acceleration of the use of cyanobacteria for biotechnology purposes.

There are many intriguing aspects of cyanobacterial physiology that may play substantial roles in controlling gene expression. We hope researchers will draw parallels from studies of plant chloroplasts and help unravel the complex gene regulatory networks in cyanobacteria. Specifically, the presence of the overrepresented sequence HIP1 is very intriguing. Although various different roles have been suggested for its presence, we think that it most likely facilitates recombination, whether autonomously through homologous recombination or mediated through an enzyme that recognizes this sequence. Cyanobacteria readily take up foreign DNA, and nearby repetitive sequences could readily allow for excision if the foreign DNA was deleterious. While integrating constructs into cyanobacteria via homologous recombination we have obtained strains that have excised portions of the insert at flanking HIP1 sequences (data not shown). The high polyploidy of cyanobacteria and its implications for gene expression is also almost completely unexplained. Experiments that probe the distribution of gene expression from different copies of the chromosome would greatly increase our understanding of gene expression in cyanobacteria and several other organisms that are polyploid (such as *Pseudomonas putida*).

We believe that continued study of the transcriptional landscape of cyanobacteria will enable the construction of useful genetic tools. Leaky expression of genes of interest is a phenomenon seen across many cyanobacteria species (**Chapter 1**). The reasons for this leaky expression may be from fundamental differences in the transcriptional landscape, possibly a difference in transcriptional termination. There is significant need for a tightly regulated induction system that has little to no leaky expression. Due to the transcription landscape in cyanobacteria, it is likely that tools to control gene expression on the level of translation may allow stricter control.

Although it has been thought transcript degradation was fundamentally different between prokaryotes and eukaryotes, there have been surprising commonalities found in recent years (Belasco 2010). Polyadenylation was once thought to only be found in eukaryotes, but has been

shown to play an important role in bacteria (although having an opposite effect than in eukaryotes), allowing exonucleases to degrade 3' secondary structures. 5' capping in eukaryotes has been known for decades, but 5' modifications and capping in prokaryotes has only been described within the last several years. A post-transcriptionally added cap resembling NAD has been found on short transcripts in bacteria and is thought to protect the transcript from degradation until removal of the cap (Luciano and Belasco 2015). Additionally, the status of the 5' phosphorylation state of messages (whether triphosphate, diphosphate, or monophosphate) has shown to have important implications for transcript stability. Although different enzymes are utilized in eukaryotes vs prokaryotes, there seem to be remarkable conservation of principles involving transcript stabilization and degradation. We hypothesize that future studies will uncover more similarities between transcript degradation in prokaryotes and eukaryotes.

Due to our findings that a 5' terminal hairpin can increase the stability of a transcript (Chapter 6), it is likely that cyanobacterial enzymes bind and sense the 5' ends of transcripts. Like in other organisms, these enzymes likely sense the 5' phosphorylation state and suggest that there is a yet-unidentified homolog of RppH. Futures studies in cyanobacteria and other organisms may reveal that bacteria utilize other 5' capping molecules to protect specific messages from degradation.

We think that further analysis of the global RNA turnover dataset described in Chapter 5 may reveal specific sequences that can influence transcript stability. We've calculated RNA half-life for open reading frames, but this dataset can also be used to calculate half-lives corresponding to each base of the genome. Drastic differences in transcript stability within short distances may be indicative of RNase processing sites. It is likely that many factors contribute to overall transcript stability, and machine learning approaches may be useful in unraveling these complex relationships.

The findings that genes with terminators with a U-tract are more stable than genes with a terminator without a U-tract or lack a terminator lead to intriguing questions about the role of exonucleases in transcript decay. Experiments examining the stability of a transcript with different terminators could show that terminator type is directly affecting transcript stability. Cyanobacteria

may be a useful model to study intrinsic termination because it lacks a homolog of rho and may rely solely on intrinsic termination.

With continued study of gene expression and the mechanisms that facilitate this regulation, more advanced and reliable genetic tools can be built. These genetic tools will enable the creation of more robust and productive strains that can be used to create valuable products from renewable inputs.

References

Abe, H. and Aiba, H. (1996). Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabilization. *Biochimie* 78, 1035–1042.

Abe, K., Sakai, Y., Nakashima, S., Araki, M., Yoshida, W., Sode, K., and Ikebukuro, K. (2014). Design of riboregulators for control of cyanobacterial (*Synechocystis*) protein expression. *Biotechnology Letters* 36, 287–294.

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image Processing with ImageJ Second Edition. *Biophotonics International* 11, 36–42.

Acebo, P., García De Lacoba, M., Rivas, G., Andreu, J. M., Espinosa, M., and del Solar, G. (1998). Structural features of the plasmid pMV158-encoded transcriptional repressor CopG, a protein sharing similarities with both helix-turn-helix and β-sheet DNA binding proteins. *Proteins: Structure, Function and Genetics* 32, 248–261.

Adrio, J. L. and Demain, A. L. (2010). Recombinant organisms for production of industrial products. *Bioengineered Bugs* 1, 116–131.

Aichi, M., Takatani, N., and Omata, T. (2001). Role of NtcB in activation of nitrate assimilation genes in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* 183, 5840–5847.

Allas, U., Liiv, A., and Remme, J. (2003). Functional interaction between RNase III and the *Escherichia coli* ribosome. *BMC Molecular Biology* 4, 1–8.

Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., Demasi, J., Huh, K., Nakatani, Y., and Mu, K. (2005). Tuning genetic control through promoter engineering. *PNAS* 103, 3006–3007.

Altuvia, S., Locker-Giladi, H., Koby, S., Ben-Nun, O., and Oppenheim, A. B. (1987). RNase III stimulates the translation of the *cIII* gene of bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America* 84, 6511–6515.

Alvey, R. M., Biswas, A., Schluchter, W. M., and Bryant, D. A. (2011). Effects of modified phycobilin biosynthesis in the cyanobacterium *Synechoccus* sp. strain PCC 7002. *Journal of Bacteriology* 193, 1663–1671.

Amarasinghe, A. K., Calin-Jageman, I., Harmouch, A., Sun, W., and Nicholson, A. W. (2001). *Escherichia coli* ribonuclease III: Affinity purification of hexahistidine-tagged enzyme and assays for substrate binding and cleavage. 342, 143–158.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.

Anderson, K. L. and Dunman, P. M. (2009). Messenger RNA turnover processes in *Escherichia coli*, *Bacillus subtilis*, and emerging studies in *Staphylococcus aureus*. *International Journal of Microbiology* 2009, 525491.

Angermayr, S. A., Paszota, M., and Hellingwerf, K. J. (2012). Engineering a cyanobacterial cell factory for production of lactic acid. *Applied and Environmental Microbiology* 78, 7098–7106.

Apirion, D. and Lassar, A. B. (1978). A conditional lethal mutant of *Escherichia coli* which affects the processing of ribosomal RNA. *Journal of Biological Chemistry* 253, 1738–1742.

Apirion, D. and Watson, N. (1975). Mapping and characterization of a mutation in *Escherichia coli* that reduces the level of ribonuclease III specific for double stranded ribonucleic acid. *Journal of Bacteriology* 124, 317–324.

Apirion, D. (1973). Degradation of RNA in *Escherichia coli* - A hypothesis. *MGG Molecular & General Genetics* 122, 313–322.

Aristarkhov, A., Mikulskis, A., Belasco, J. G., and Lin, E. C. (1996). Translation of the *adhE* transcript to produce ethanol dehydrogenase requires RNase III cleavage in *Escherichia coli*. *Journal of Bacteriology* 178, 4327–4332.

Awano, N., Inouye, M., and Phadtare, S. (2008). RNase activity of polynucleotide phosphorylase is critical at low temperature in *Escherichia coli* and is complemented by RNase II. *Journal of Bacteriology* 190, 5924–5933.

Babiskin, A. H. and Smolke, C. D. (2011). Synthetic RNA modules for fine-tuning gene expression levels in yeast by modulating RNase III activity. *Nucleic Acids Research* 39, 8651–8664.

Bachmann, B. and Low, K. B. (1980). Linkage map of *Escherichia coli* K-12, edition 6. *Microbiological Reviews* 44, 1–56.

Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36.

Balasubramanian, D., Ragunathan, P. T., Fei, J., and Vanderpool, C. K. (2016). A prophage-encoded small RNA controls. *Molecular Biology and Physiology* 1, 1–18.

Bardwell, J. C., Régnier, P., Chen, S. M., Nakamura, Y., Grunberg-Manago, M., and Court, D. L. (1989). Autoregulation of RNase III operon by mRNA processing. *The EMBO Journal* 8, 3401–3407.

Barry, G., Squires, C., and Squires, C. L. (1980). Attenuation and processing of RNA from the *rplJL–rpoBC* transcription unit of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 77, 3331–3335.

Barten, R. and Lill, H. (2017). DNA-uptake in the naturally competent cyanobacterium, *Synechocystis* sp. PCC 6803. *FEMS Microbiology Letters* 129, 83–88.

Baumgardt, K., Melior, H., Madhugiri, R., Thalmann, S., Schikora, A., McIntosh, M., Becker, A., and Evguenieva-Hackenberg, E. (2017). RNase E and RNase J are needed for S-adenosylmethionine homeostasis in *Sinorhizobium meliloti*. *Microbiology* 163, 570–583.

Baumgartner, D., Kopf, M., Klähn, S., Steglich, C., and Hess, W. R. (2016). Small proteins in cyanobacteria provide a paradigm for the functional analysis of the bacterial micro-proteome. *BMC Microbiology* 16, 285.

Bedu, S., Peltier, G., Sarrey, F., and Joset, F. (1990). Properties of a mutant from *Synechocystis* PCC 6803 resistant to acetazolamide, an inhibitor of carbonic anhydrase. *Plant Physiology* 93, 1312–1315.

Begemann, M. B., Zess, E. K., Walters, E. M., Schmitt, E. F., Markley, A. L., and Pfleger, B. F. (2013). An organic acid based counter selection system for cyanobacteria. *PLoS ONE* 8, 1–12.

Belasco, J. G. (2010). All Things Must Pass: Contrasts and Commonalities in Eukaryotic and Bacterial mRNA Decay. *Nat Rev Mol Cell Biol* 11, 467–478.

Belknap, W. R. and Haselkorn, R. (1987). Cloning and light regulation of expression of the phycocyanin operon of the cyanobacterium *Anabaena*. *The EMBO Journal* 6, 871–884.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57, 289–300.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366.

Bernstein, H. C., McClure, R. S., Hill, E. A., Markillie, L. M., Chrisler, W. B., Romine, M. F., McDermott, J. E., Posewitz, M. C., Bryant, D. A., Konopka, A. E., Fredrickson, J. K., and Beliaev, A. S. (2016). Unlocking the Constraints of Cyanobacterial Productivity : *mBio* 7, 1–10.

Bernstein, J. A., Khodursky, A. B., Lin, P., Lin-Chao, S., and Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America* 99, 9697–9702.

Beuf, L., Bédu, S., Cami, B., and Joset, F. (1995). A protein is involved in accessibility of the inhibitor acetazolamide to the carbonic anhydrase(s) in the cyanobacterium *Synechocystis* PCC 6803. *Plant Molecular Biology* 27, 779–788.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* 41, 7429–7437.

Bischler, T., Hsieh, P. K., Resch, M., Liu, Q., Tan, H. S., Foley, P. L., Hartleib, A., Sharma, C. M., and Belasco, J. G. (2017). Identification of the RNA pyrophosphohydrolase RppH of *Helicobacter pylori* and global analysis of its RNA targets. *Journal of Biological Chemistry* 292, 1934–1950.

Blasi, F. and Bruni, C. B. (1981). Regulation of the histidine operon: translation-controlled transcription termination (a mechanism common to several biosynthetic operons). *Current Topics in Cellular Regulation* 19, 1–45.

Blokesch, M. (2016). Natural competence for transformation. *Current Biology* 26, R1126–R1130.

Blomberg, P., Wagner, E. G., and Nordström, K. (1990). Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNase III. *The EMBO Journal* 9, 2331–40.

Brahamsha, B. (1996). A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Applied and Environmental Microbiology* 62, 1747–1751.

Brahamsha, B. and Haselkorn, R. (1991). Isolation and characterization of the gene encoding the principal sigma factor of the vegetative cell RNA polymerase from the cyanobacterium *Anabaena* sp. strain PCC 7120. *Journal of Bacteriology* 173, 2442–2450.

Bralley, P., Aseem, M., and Jones, G. H. (2014). SCO5745, a Bifunctional RNase J Ortholog, affects antibiotic production in *Streptomyces coelicolor*. *Journal of Bacteriology* 196, 1197–1205.

Britton, R. A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W. C., Mathy, N., Tobin, C., Daou, R., Szyk, J., and Condon, C. (2007). Maturation of the 5' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Molecular Microbiology* 63, 127–138.

Bryant, D. A., Lorimier, R. D., Guglielmi, G., and Stevens, S. E. (1990). Structural and compositional analyses of the phycobilisomes. *Pharmacia* 7002, 550–560.

Busby, S. and Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology* 293, 199–213.

Cai, Y. and Wolk, C. P. (1990). Use of a condtionally lethal gene in *Anabaena sp.* strain PCC 7120 to select for double recombinats and to entrap insertion sequnces. *Journal of Bacteriology* 172, 3138–3145.

Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., and Luisi, B. F. (2005). Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* 437, 1187–1191.

Cameron, D. E. and Collins, J. J. (2014). Tunable protein degradation in bacteria. *Nature Biotechnology* 32, 1–8.

Cameron, J. C., Wilson, S. C., Bernstein, S. L., and Kerfeld, C. A. (2013). Biogenesis of a bacterial organelle: The carboxysome assembly pathway. *Cell* 155, 1131–1140.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104, 901–912.

Camsund, D. and Lindblad, P. (2014). Engineered transcriptional systems for cyanobacterial biotechnology. *Frontiers in Bioengineering and Biotechnology* 2, 1–9.

Carafa, Y. d'Aubenton, Brody, E., and Thermes, C. (1990). Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *Journal of Molecular Biology* 216, 835–858.

Carpousis, A. J., Luisi, B. F., and McDowall, K. J. (2009). Endonucleolytic initiation of mRNA decay in *Escherichia coli*. *Progress in Molecular Biology and Translational Science* 85, 91–135.

Carpousis, A. J., Van Houwe, G., Ehretsmann, C., and Krisch, H. M. (1994). Copurification of *E. coli* RNase E and PNPase: Evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell* 76, 889–900.

Carrier, T. A. and Keasling, J. D. (1997). Engineering mRNA stability in *E. coli* by the addition of synthetic hairpins using a 5' cassette system. *Biotechnology and Bioengineering* 55, 577–580.

Carrier, T. A. and Keasling, J. D. (1999). Library of Synthetic 5' Secondary Structures To Manipulate mRNA Stability in *Escherichia coli*. *Biotechnology Progress* 15, 58–64.

Case, A. E. and Atsumi, S. (2016). Cyanobacterial chemical production. *Journal of Biotechnology* 231, 106–114.

Caslake, L. F. and Bryant, D. A. (1996). The *sigA* gene encoding the major sigma-factor of RNA polymerase from the marine cyanobacterium *Synechoccus* sp. strain PCC 7002: Cloning and characterization. *Microbiology* 142, 347–357.

Celesnik, H., Deana, A., and Belasco, J. G. (2007). Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal. *Mol Cell* 27, 79–90.

Chappell, J., Watters, K. E., Takahashi, M. K., and Lucks, J. B. (2015). A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Current Opinion in Chemical Biology* 28, 47–56.

Chen, C. and Deutscher, M. (2010). RNase R is a highly unstable protein regulated by growth phase and stress. *RNA* 16, 667–672.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Molecular Systems Biology* 11, 781–781.

Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J. K. (2016). The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Molecular and Cellular Biology* 36, 662–667.

Chen, Y., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T., and Voigt, C. A. (2013). Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods* 10, 659–664.

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. *Nature Reviews Microbiology* 12, 327–340.

Clark, R. L., Cameron, J. C., Root, T. W., and Pfleger, B. F. (2014). Insights into the industrial growth of cyanobacteria from a model of the carbon-concentrating mechanism. *AIChE* 60, 1269–1277.

Clerico, E. M., Ditty, J., and Golden, S. (2007). "Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria". Ed. by R. E. Vol. 362. Humana Press. Chap. Circadian Rhythms. Methods in Molecular Biology.

Cleto, S., Jensen, J. V., Wendisch, V. F., and Lu, T. K. (2016). *Corynebacterium glutamicum* metabolic engineering with CRISPR interference (CRISPRi). *ACS Synthetic Biology* 5, 375–385.

Coburn, G. A. and Mackie, G. A. (1996). Overexpression, purification, and properties of *Escherichia coli* ribonuclease II. *Journal of Biological Chemistry* 271, 1048–1053.

Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Völker, U., and Stülke, J. (2009). Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Molecular & Cellular Proteomics* 8, 1350–1360.

Condon, C. (2007). Maturation and degradation of RNA in bacteria. *Current Opinion in Microbiology* 10, 271–278.

Condon, C. (2010). What is the role of RNase J in mRNA turnover? *RNA Biology* 7, 316–21.

Condon, C. and Putzer, H. (2002). The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Research* 30, 5339–5346.

Copeland, M. F., Politz, M. C., and Pfleger, B. F. (2014). Application of TALEs, CRISPR/Cas and sRNAs as trans-acting regulators in prokaryotes. *Current Opinion in Biotechnology* 29, 46–54.

Court, D. L. (1993). "RNA processing and degradation by RNase III". Ed. by J. Belasco and G. Brawerman. Academic Preork, pp. 71–116.

Court, D. L., Gan, J., Liang, Y., Shaw, G. X., Tropea, J. E., Costantino, N., Waugh, D. S., and Ji, X. (2013). RNase III: Genetics and function; structure and mechanism. *Annual Review of Genetics* 47, 405–31.

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

Cunningham, L. and Guest, J. R. (1998). Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. *Microbiology* 144, 2113–2123.

Daley, S. M. E., Kappell, A. D., Carrick, M. J., and Burnap, R. L. (2012). Regulation of the cyanobacterial CO_2-concentrating mechanism involves internal sensing of NADP+ and α-ketogutarate levels by transcription factor CcmR. *PLoS ONE* 7, 1–10.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25, 1974–1975.

Davies, F. K., Work, V. H., Beliaev, A. S., and Posewitz, M. C. (2014). Engineering limonene and bisabolene production in wild type and a glycogen-deficient mutant of *Synechoccus* sp. PCC 7002. *Frontiers in Bioengineering and Biotechnology* 2, 21.

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., and Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Molecular bioSystems* 5, 1512–1526.

Deana, A., Celesnik, H., and Belasco, J. G. (2008). The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* 451, 355–358.

del Solar, G., Acebo, P., and Espinosa, M. (1995). Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II. *Molecular Microbiology* 18, 913–24.

del Solar, G. and Espinosa, M. (1992). The copy number of plasmid pLS1 is regulated by two trans-acting plasmid products: the antisense RNA II and the repressor protein, RepA. *Molecular Microbiology* 6, 83–94.

del Solar, G., Diaz, R., and Espinosa, M. (1987). Replication of the *streptococcal* plasmid pMV158 and derivatives in cell-free extracts of *Escherichia coli*. *Molecular & General Genetics* 206, 428–435.

Delaye, L. and Moya, A. (2011). Abundance and distribution of the highly iterated palindrome 1 (HIP1) among prokaryotes. *Mobile Genetic Elements* 1, 159–168.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 471, 602–607.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 36, 465–469.

Desai, S. H. and Atsumi, S. (2013). Photosynthetic approaches to chemical biotechnology. *Current Opinion in Biotechnology* 24, 1031–1036.

Devroe, E., Kosuri, S., Berry, D., Afeyan, N., Skraly, F., Robertson, D., Green, B., and Ridley, C. (2009). "Hyperphotosynthetic organisms". United States Patent. WO2009062190 A3.

Dip, P. V., Kamariah, N., Subramanian Manimekalai, M. S., Nartey, W., Balakrishna, A. M., Eisenhaber, F., Eisenhaber, B., and Grüber, G. (2014). Structure, mechanism and ensemble formation of the alkylhydroperoxide reductase subunits AhpC and AhpF from *Escherichia coli*. *Acta Crystallographica Section D: Biological Crystallography* 70, 2848–2862.

Domain, F., Houot, L., Chauvat, F., and Cassier-Chauvat, C. (2004). Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation. *Molecular Microbiology* 53, 65–80.

Dombrowski, S., Brennicke, A., and Binder, S. (1997). 3'-inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *EMBO Journal* 16, 5069–5076.

Doolittle, W. F. (1973). Postmaturational cleavage of 23S ribosomal ribonucleic-acid and its metabolic control in blue-green alga *Anacystis Nidulans*. *Journal of Bacteriology* 113, 1256–1263.

Downing, W. L., Sullivan, S. L., Gottesman, M. E., and Dennis, P. P. (1990). Sequence and transcriptional pattern of the essential *Escherichia coli secE-nusG* operon. *Journal of Bacteriology* 172, 1621–1627.

Drider, D. and Condon, C. (2005). The continuing story of endoribonuclease III. *Journal of Molecular Microbiology and Biotechnology* 8, 195–200.

Ducat, D. C., Avelar-Rivas, J. A., Way, J. C., and Silvera, P. A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. *Applied and Environmental Microbiology* 78, 2660–2668.

Dühring, U., Axmann, I. M., Hess, W. R., and Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proceedings of the National Academy of Sciences of the United States of America* 103, 7054–7058.

Dunn, J. J. and Studier, F. W. (1973). T7 early RNAs and *Escherichia coli* ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease 3. *Proceedings of the National Academy of Sciences of the United States of America* 70, 3296–3300.

Durand, S., Gilet, L., and Condon, C. (2012). The essential function of *B. subtilis* RNase III is to silence foreign toxin genes. *PLoS Genet* 8, e1003181.

Eisenhut, M., Georg, J., Klähn Stephanand Sakurai, I., Mustila, H., Zhang, P., Hess, W. R., and Aro, E. M. (2012). The antisense RNA As1-flv4 in the cyanobacterium *Synechocystis* sp. PCC 6803 prevents premature expression of the *flv4-2* operon upon shift in inorganic carbon supply. *Journal of Biological Chemistry* 287, 33153–33162.

Elela, S. A., Igel, H., and Ares, M. (1996). RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* 85, 115–124.

Elhai, J. (1993). Strong and regulated promoters in the cyanobacterium *Anabaena* PCC 7120. *FEMS Microbiology Letters* 114, 179–184.

Emory, S. A., Bouvet, P., and Belasco, J. G. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes and Development* 6, 135–148.

Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research* 42, 2646–2659.

Espinosa, J., Forchhammer, K., Burillo, S., and Contreras, A. (2006). Interaction network in cyanobacterial nitrogen regulation: PipX, a protein that interacts in a 2-oxoglutarate dependent manner with PII and NtcA. *Molecular Microbiology* 61, 457–469.

Espinosa, J., Rodríguez-Mateos, F., Salinas, P., Lanza, V. F., Dixon, R., de la Cruz, F., and Contreras, A. (2014). PipX, the coactivator of NtcA, is a global regulator in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 111, E2423–E2430.

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Bréchemmier-Baey, D., and Putzer, H. (2005). Ribonucleases J1 and J2: Two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. *Nucleic Acids Research* 33, 2141–2152.

Evguenieva-Hackenberg, E. (2005). Bacterial ribosomal RNA in pieces. *Molecular Microbiology* 57, 318–325.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., and Salis, H. M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular Systems Biology* 10, 731.

Faubladier, M., Cam, K., and Bouché, J. P. (1990). *Escherichia coli* cell division inhibitor DicF-RNA of the *dicB* operon. Evidence for its generation in vivo by transcription termination and by RNase III and RNase E-dependent processing. *Journal of Molecular Biology* 212, 461–471.

Ferreira, R., Borges, V., Borrego, M. J., and Gomes, J. P. (2017). Global survey of mRNA levels and decay rates of *Chlamydia trachomatis* trachoma and lymphogranuloma venereum biovars. *Heliyon* 3, e00364.

Fickett, J. W. (1995). ORFs and Genes: How Strong a Connection? *Journal of Computational Biology* 2, 117–123.

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D., Li, W., Lomas, M., Veneziano, D., Vera, C., Vrugt, J. A., and Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlrococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences* 110, 9824–9829.

Flores, E., Frías, J. E., Rubio, L. M., and Herrero, A. (2005). Photosynthetic nitrate assimilation in cyanobacteria. *Photosynthesis Research* 83, 117–133.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532, 517–521.

Freed, E. F., Winkler, J. D., Weiss, S. J., Garst, A. D., Mutalik, V. K., Arkin, A. P., Knight, R., and Gill, R. T. (2015). Genome-wide tuning of protein expression levels to rapidly engineer microbial traits. *ACS Synthetic Biology* 4, 1244–1253.

Frey, J. and Bagdasarian, M. (1989). The Molecular Biology of IncQ Plasmids. *Promiscuous Plasmids of Gram-Negative Bacteria*, pp. 79–91.

Frias, J. E., Flores, E., and Herrero, A. (2000). Activation of the *Anabaena nir* operon promoter requires both NtcA (CAP family) and NtcB (LysR family) transcription factors. *Molecular Microbiology* 38, 613–625.

Gao, L., Wang, J., Ge, H., Fang, L., Zhang, Y., Huang, X., and Wang, Y. (2015). Toward the complete proteome of *Synechocystis* sp. PCC 6803. *Photosynthesis Research* 126, 203–219.

Gao, Y., Gong, Y., and Xu, X. (2013). RNase III-dependent down-regulation of *ftsH* by an artificial internal sense RNA in *Anabaena* sp. PCC 7120. *FEMS Microbiology Letters* 344, 130–137.

Garza-Sánchez, F., Schaub, R. E., Janssen, B. D., and Hayes, C. S. (2011). tmRNA regulates synthesis of the ArfA ribosome rescue factor. *Molecular Microbiology* 80, 1204–1219.

Gatewood, M. L., Bralley, P., and Jones, G. H. (2011). RNase III-dependent expression of the *rpsO-pnp* operon of *Streptomyces coelicolor*. *Journal of Bacteriology* 193, 4371–4379.

Gatti-Lafranconi, P., Dijkman, W. P., Devenish, S. R., and Hollfelder, F. (2013). A single mutation in the core domain of the *lac* repressor reduces leakiness. *Microbial Cell Factories* 12, 67.

Gautheret, D. and Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology* 313, 1003–1011.

Geerts, D., Bovy, A., Devrieze, G., Borrias, M., and Weisbeek, P. (1995). Inducible expression of heterologous genes targeted to a chromosomal platform in the cyanobacterium *Synechoccus* sp. PCC 7942. *Microbiology* 141, 831–841.

Ghassemian, M. and Straus, N. A. (2017). Fur regulates the expression of iron-stress genes. *Microbiology* 142, 1469–1476.

Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6, 343–345.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451.

Golden, S. S. and Sherman, L. A. (1984). Optimal conditions for genetic transformation of the cyanobacterium *Anacystis nidulans* R2. *Journal of Bacteriology* 158, 36–42.

González, A., Angarica, V. E., Sancho, J., and Fillat, M. F. (2014). The FurA regulon in *Anabaena* sp. PCC 7120: In silico prediction and experimental validation of novel target genes. *Nucleic Acids Research* 42, 4833–4846.

Griese, M., Lange, C., and Soppa, J. (2011). Ploidy in cyanobacteria. *FEMS Microbiology Letters* 323, 124–131.

Grossman, A. R., Schaefer, M. R., Chiang, G. G., and Collier, J. L. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiological Reviews* 57, 725–749.

Gründel, M., Scheunemann, R., Lockau, W., and Zilliges, Y. (2012). Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 158, 3032–3043.

Guerrero, F., Carbonell, V., Cossu, M., Correddu, D., and Jones, P. R. (2012). Ethylene synthesis and regulated expression of recombinant protein in *Synechocystis* sp. PCC 6803. *PLoS ONE* 7, e50470.

Gupta, A., Morby, A. P., Turner, J. S., Whitton, B. A., and Robinson, N. J. (1993). Deletion within the metallothionein locus of cadmium-tolerant *Synechococcus* PCC 6301 involving a highly iterated palindrome (HIP1). *Molecular Microbiology* 7, 189–195.

Gusarov Ivana nd Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Molecular Cell* 3, 495–504.

Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W., and Livny, J. (2012). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* 13, 734.

Haddad, N., Saramago, M., Matos, R. G., Prévost, H., and Arraiano, C. M. (2013). Characterization of the biochemical properties of *Campylobacter jejuni* RNase III. *Bioscience reports*, 889–901.

Hambraeus, G., Persson, M., and Rutberg, B. (2000). The *aprE* leader is a determinant of extreme mRNA stability in *Bacillus subtilis*. *Microbiology* 146, 3051–3059.

Hambraeus, G., von Wachenfeldt, C., and Hederstedt, L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Molecular Genetics and Genomics* 269, 706–714.

Hansen, K. D., Brenner, S. E., and Dudoit, S. (2010). Biases in Illumina tranascriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* 38, e131.

Hara, K., Araki, M., Okai, N., Wakai, S., Hasunuma, T., and Kondo, A. (2014). Development of bio-based fine chemical production through synthetic bioengineering. *Microbial Cell Factories* 13, 173.

Hedger, J., Holmquist, P. C., Leigh, K. A., Saraff, K., Pomykal, C., and Summers, M. L. (2009). Illumination stimulates cAMP receptor protein-dependent transcriptional activation from regulatory regions containing class I and class II promoter elements in *Synechocystis* sp. PCC 6803. *Microbiology* 155, 2994–3004.

Herrero, A., Muro-Pastor, A. M., and Flores, E. (2001). Nitrogen control in cyanobacteria. *Journal of Bacteriology* 183, 411–425.

Hickman, J. W., Kotovic, K. M., Miller, C., Warrener, P., Kaiser, B., Jurista, T., Budde, M., Cross, F., Roberts, J. M., and Carleton, M. (2013). Glycogen synthesis is a required component of the nitrogen stress response in *Synechococcus elongatus* PCC 7942. *Algal Research* 2, 98–106.

Higo, A., Isu, A., Fukaya, Y., and Hisabori, T. (2016). Efficient gene induction and endogenous gene repression systems for the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Plant and Cell Physiology* 57, 387–396.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77, 51–59.

Hobbs, E. C., Fontaine, F., Yin, X., and Storz, G. (2011). An expanding universe of small proteins. *Current Opinion in Microbiology* 14, 167–173.

Hood, R. D., Higgins, S. A., Flamholz, A., Nichols, R. J., and Savage, D. F. (2016). The stringent response regulates adaptation to darkness in the cyanobacterium *Synechococcus elongatus*. *Proceedings of the National Academy of Sciences*, 201524915.

Hook-Barnard, I. G. and Hinton, D. M. (2007). Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regulation and Systems Biology* 1, 275–293.

Horiuchi, M., Nakamura, K., Kojima, K., Nishiyama, Y., Hatakeyama, W., Hisabori, T., and Hihara, Y. (2010). The PedR transcriptional regulator interacts with thioredoxin to connect photosynthesis with gene expression in cyanobacteria. *The Biochemical Journal* 431, 135–140.

Hotto, A. M., Castandet, B., Gilet, L., Higdon, A., Condon, C., and Stern, D. B. (2015). *Arabidopsis* chloroplast mini-ribonuclease III participates in rRNA maturation and intron recycling. *Plant Cell* 27, 724–740.

Huang, C., Shen, C. R., Li, H., Sung, L., Wu, M., and Hu, Y. (2016). CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium *S. elongatus* PCC 7942. *Microbial Cell Factories* 15, 196.

Huang, H. H., Camsund, D., Lindblad, P., and Heidorn, T. (2010). Design and characterization of molecular tools for a synthetic biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Research* 38, 2577–2593.

Huang, H. and Lindblad, P. (2013). Wide-dynamic-range promoters engineered for cyanobacteria. *Journal of Biological Engineering* 7, 10.

Hui, M. P., Foley, P. L., and Belasco, J. G. (2014). Messenger RNA degradation in bacterial cells. *Annual Review of Genetics* 48, 537–559.

Ikeuchi, M. and Tabata, S. (2001). *Synechocystis* sp. PCC 6803 - a useful tool in the study of the genetics of cyanobacteria. *Photosynthesis Research* 70, 73–83.

Imamura, S. and Asayama, M. (2009). Sigma factors for cyanobacterial transcription. *Gene Regulation and Systems Biology* 3, 65–87.

Imashimizu, M., Tanaka, K., and Shimamoto, N. (2011). Comparative study of cyanobacterial and *E. coli* RNA polymerases: Misincorporation, abortive transcription, and dependence on divalent cations. *Genetics Research International* 2011, 1–11.

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., and Frishman, D. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 9, 102.

Ishii, A. and Hihara, Y. (2008). An AbrB-Like transcriptional regulator, Sll0822, is essential for the activation of nitrogen-regulated genes in *Synechocystis* sp. PCC 6803. *Plant Physiology* 148, 660–670.

Jacobsen, J. H. and Frigaard, N. U. (2014). Engineering of photosynthetic mannitol biosynthesis from CO_2 in a cyanobacterium. *Metabolic Engineering* 21, 60–70.

Jarrige, A., Mathy, N., and Portier, C. (2001). PNPase autocontrols its expression by degrading a double-stranded structure in the *pnp* mRNA leader. *The EMBO Journal* 20, 6845–6855.

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., Gingeras, T. R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research* 9, 1543–1551.

Jones, S. E., Leong, V., Ortega, J., and Elliot, M. A. (2014). Development, antibiotic production, and ribosome assembly in *Streptomyces venezuelae* are impacted by RNase J and RNase III deletion. *Journal of Bacteriology* 196, 4253–4267.

Kaberdin, V. R., Miczak, A., Jakobsen, J. S., Lin-Chao, S., McDowall, K. J., and von Gabain, A. (1998). The endoribonucleolytic N-terminal half of *Escherichia coli* RNase E is evolutionarily conserved in *Synechocystis* sp. and other bacteria but not the C-terminal half, which is sufficient for degradosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* 95, 11637–11642.

Kaga, N., Umitsuki, G., Nagai, K., and Wachi, M. (2002). RNase G-dependent degradation of the *eno* mRNA encoding a glycolysis enzyme enolase in *Escherichia coli*. *Bioscience, Biotechnology, and Biochemistry* 66, 2216–2220.

Kashino, Y., Lauber, W. M., Carroll, J. A., Wang, Q., Whitmarsh, J., Satoh, K., and Pakrasi, H. B. (2002). Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* 41, 8004–8012.

Kavalchuk, K., Madhusudan, S., and Schnetz, K. (2012). RNase III initiates rapid degradation of *proU* mRNA upon hypo-osmotic stress in *Escherichia coli*. *RNA Biology* 9, 96–107.

Keasling, J. D. (1999). Gene-expression tools for the metabolic engineering of bacteria. *Trends in Biotechnology* 17, 452–460.

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering* 14, 189–195.

Kelly, L., Mezulis, S., Yates, C., Wass, M., and Sternberg, M. (2015). The Phyre2 web portal for protein modelling, prediction, and analysis. *Nature Protocols* 10, 845–858.

Keren, N., Kidd, M. J., Penner-Hahn, J. E., and Pakrasi, H. B. (2002). A light-dependent mechanism for massive accumulation of manganese in the photosynthetic bacterium *Synechocystis* sp. PCC 6803. *Biochemistry* 41, 15085–15092.

Khodurksy, A., Bernstein, J., Peter, B., Rhodius, V., Wendisch, V., and Zimmer, D. (2003). *Escherichia coli* spotted double-strand DNA microarrays. *Methods in Molecular Biology* 224, 61–78.

Kido, M., Yamanaka, K., Mitani, T., Niki, H., Ogura, T., and Hiraga, S. (1996). RNase E polypeptides lacking a carboxyl-terminal half suppress a *mukB* mutation in *Escherichia coli*. *Journal of Bacteriology* 178, 3917–3925.

Kim, K. S., Manasherob, R., and Cohen, S. N. (2008). YmdB: A stress-responsive ribonuclease-binding regulator of *E. coli* RNase III activity. *Genes and Development* 22, 3497–3508.

King, T. C., Sirdeshmukh, R., and Schlessinger, D. (1984). RNase III cleavage is obligate for maturation but not for function of *Escherichia coli* pre-23S rRNA. *Proceedings of the National Academy of Sciences of the United States of America* 81, 185–188.

Kirst, H., Formighieri, C., and Melis, A. (2014). Maximizing photosynthetic efficiency and culture productivity in cyanobacteria upon minimizing the phycobilisome light-harvesting antenna size. *Biochimica et Biophysica Acta - Bioenergetics* 1837, 1653–1664.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J. J., Aryee, M. J., and Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–5.

Knoot, C. J., Ungerer, J., Wangikar, P. P., and Pakrasei, H. B. (2017). Cyanobacteria: promising biocatalysts for sustain chemical production. *Journal of Biological Chemistry*.

Kopf, M. and Hess, W. R. (2015). Regulatory RNAs in photosynthetic cyanobacteria. *FEMS Microbiology Reviews* 39, 301–315.

Krasikov, V., Aguirre von Wobeser, E., Dekker, H. L., Huisman, J., and Matthijs, H. C. P. (2012). Time-series resolution of gradual nitrogen starvation and its impact on photosynthesis in the cyanobacterium *Synechocystis* PCC 6803. *Physiologia Plantarum* 145, 426–439.

Kristoffersen, S. M., Haase, C., Weil, M. R., Passalacqua, K. D., Niazi, F., Hutchison, S. K., Desany, B., Kolstø, A., Tourasse, N. J., Read, T. D., and Økstad, O. A. (2012). Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium. *Genome Biology* 13, R30.

Kumar, K., Mella-Herrera, R. A., and Golden, J. W. (2010). Cyanobacterial heterocysts. *Cold Spring Harbor Perspectives in Biology* 2, 1–19.

Kumar, M., Sundaram, S., Gnansounou, E., Larroche, C., and Thakur, I. S. (2018). Carbon dioxide capture, storage and production of biofuel and biomaterials by bacteria: A review. *Bioresource Technology* 247, 1059–1068.

Kusakabe, T., Tatsuke, T., Tsuruno, K., Hirokawa, Y., Atsumi, S., Liao, J. C., and Hanai, T. (2013). Engineering a synthetic pathway in cyanobacteria for isopropanol production directly from carbon dioxide and light. *Metabolic Engineering* 20, 101–108.

Labarre, J., Chauvat, F., and Thuriaux, P. (1989). Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. *Journal of Bacteriology* 171, 3449–3457.

Lai, M. C. and Lan, E. I. (2015). Advances in metabolic engineering of cyanobacteria for photosynthetic biochemical production. *Metabolites* 5, 636–658.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* 8, 2180–96.

Lea-Smith, D. J., Bombelli, P., Dennis, J. S., Scott, S. A., Smith, A. G., and Howe, C. J. (2014). Phycobilisome-deficient strains of *Synechocystis* sp. PCC 6803 have reduced size and require carbon-limiting conditions to exhibit enhanced productivity. *Plant Physiology* 165, 705–714.

Lee, K., Bernstein, J. A., and Cohen, S. N. (2002). RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. *Molecular Microbiology* 43, 1445–1456.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419.

Lemke, J., Sanchez-Vazquez, P., Burgos, H., Hedberg, G., Ross, W., and Gourse, R. (2011). Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proceedings of the National Academy of Sciences* 108, 5712.

Leroy, A., Vanzo, N. F., Sousa, S., Dreyfus, M., and Carpousis, A. J. (2002). Function in *Escherichia coli* of the non-catalytic part of RNase E: role in the degradationof ribosome-free mRNA. *Molecular Microbiology* 45, 1231–43.

Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A., and Ecker, D. J. (2001). Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Research* 29, 3583–94.

Li, H., Shen, C. R., Huang, C. H., Sung, L. Y., Wu, M. Y., and Hu, Y. C. (2016). CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metabolic Engineering* 38, 293–302.

Li, L., Miles, S., Melville, Z., Prasad, A., Bradley, G., and Breeden, L. L. (2013). Key events during the transition from rapid growth to quiescence in budding yeast require posttranscriptional regulators. *Molecular Biology of the Cell* 24, 3697–3709.

Liao, J. C., Mi, L., Pontrelli, S., and Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology* 14, 288–304.

Lim, B., Ahn, S., Sim, M., and Lee, K. (2014). RNase III controls *mltD* mRNA degradation in *Escherichia coli*. *Current Microbiology* 68, 518–523.

Lim, B., Sim, M., Lee, H., Hyun, S., Lee, Y., Hahn, Y., Shin, E., and Lee, K. (2015). Regulation of *Escherichia coli* RNase III activity. *Journal of Microbiology* 53, 487–494.

Lim, B., Sim, S. H., Sim, M., Kim, K., Jeon, C. O., Lee, Y., Ha, N. C., and Lee, K. (2012). RNase III controls the degradation of *corA* mRNA in *Escherichia coli*. *Journal of Bacteriology* 194, 2214–2220.

Link, A. J., Robison, K., and Church, G. (1997). Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* 18, 1259–313.

Lioliou, E., Sharma, C. M., Caldelari, I., Helfer, A. C., Fechter, P., Vandenesch, F., Vogel, J., and Romby, P. (2012). Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genetics* 8, e1002782.

Liu, H., Zhang, H., Niedzwiedzki, D. M., Prado, M., He, G., Gross, M. L., and Blankenship, R. E. (2013). Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* 342, 1104–1107.

López-Aguilar, C., Ruiz-Masó, J. A., Rubio-Lepe, T. S., Sanz, M., and del Solar, G. (2013). Translation initiation of the replication initiator *repB* gene of promiscuous plasmid pMV158 is led by an extended non-SD sequence. *Plasmid* 70, 69–77.

Luciano, D. and Belasco, J. (2015). NAD in RNA: Unconventional headgear. *Trends in Biochemical Sciences* 40, 245–247.

Luciano, D. J., Vasilyev, N., Richards, J., Serganov, A., and Belasco, J. G. (2017). A novel RNA phosphorylation state enables 5' end-dependent degradation in *Escherichia coli*. *Molecular Cell* 67, 44–54.

Ludwig, M. and Bryant, D. A. (2011). Transcription profiling of the model cyanobacterium *Synechoccus* sp. strain PCC 7002 by Next-Gen (SOLiD™) sequencing of cDNA. *Frontiers in Microbiology* 2, 41.

Ludwig, M. and Bryant, D. a. (2012). *Synechoccus* sp. strain PCC 7002 transcriptome: Acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Frontiers in Microbiology* 3, 354.

Ludwig, M., Sultemeyer, D., and Price, G. D. (2000). Isolation of *ccmKLMN* genes from the marine cyanobacterium, *Synechoccus* sp. PCC 7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly. *Journal of Phycology* 36, 1109–1119.

Luque, I., Flores, E., and Herrero, A. (1994). Molecular mechanism for the operation of nitrogen control in cyanobacteria. *The EMBO Journal* 13, 2862–2869.

Lynch, S. and Gill, R. (2012). Synthetic biology: New strategies for directing designing. *Metabolic Engineering* 14, 205–211.

Łyzeń, R., Kochanowska, M., Wegrzyn, G., and Szalewska-Pałasz, A. (2009). Transcription from bacteriophage λ pR promoter is regulated independently and antagonistically by DksA and ppGpp. *Nucleic Acids Research* 37, 6655–6664.

Ma, A. T., Schmidt, C. M., and Golden, J. W. (2014). Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. *Applied and Environmental Microbiology* 80, 6704–6713.

Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research* 29, 4724–4735.

Mackie, G. A. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* 395, 720–723.

Mackie, G. A. (2013). RNase E: at the interface of bacterial RNA processing and decay. *Nature Reviews Microbiology* 11, 45–57.

Majumder, H. K., Bishayee, S., Chakraborty, P. R., and Maitra, U. (1977). Ribonuclease III cleavage of bacteriophage T3RNA polymerase transcripts to late T3 mRNAs. *Proceedings of the National Academy of Sciences of the United States of America* 74, 4891–4894.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R., and Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology* 13, 722–736.

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C., and Pfleger, B. F. (2015). Synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechoccus* sp. strain PCC 7002. *ACS Synthetic Biology* 4, 595–603.

Mathy, N., Bénard, L., Pellegrini, O., Daou, R., Wen, T., and Condon, C. (2007). 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 129, 681–692.

Matos, R. G., Fialho, A. M., Giloh, M., Schuster, G., and Arraiano, C. M. (2012). The *rnb* gene of *Synechocystis* PCC6803 encodes a RNA hydrolase displaying RNase II and not RNase R enzymatic properties. *PLoS ONE* 7, 1–12.

Matsunaga, J., Dyer, M., Simons, E. L., and Simons, R. W. (1996). Expression and regulation of the *rnc* and *pdxJ* operons of *Escherichia coli*. *Molecular Microbiology* 22, 977–989.

Matsunaga, J., Simons, E. L., and Simons, R. W. (1997). *Escherichia coli* RNase III (*rnc*) autoregulation occurs independently of *rnc* gene translation. *Molecular Microbiology* 26, 1125–1135.

Matsuoka, M., Takahama, K., and Ogawa, T. (2001). Gene replacement in cyanobacteria mediated by a dominant streptomycin-sensitive *rps12* gene that allows selection of mutants free from drug resistance markers. *Culture* 147, 2077–2087.

Mayer, J. E. and Schweiger, M. (1983). RNase III is positively regulated by T7 protein kinase. *Journal of Biological Chemistry* 258, 5340–5343.

McClure, R. S., Overall, C. C., Mcdermott, J. E., Hill, E. A., Markillie, L. M., Mccue, L. A., Taylor, R. C., Ludwig, M., Bryant, D. A., and Beliaev, A. S. (2016). Network analysis of

transcriptomics expands regulatory landscapes in *Synechoccus* sp. PCC 7002. *Nucleic Acids Research* 44, 8810–8825.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* 41, 597–600.

Miczak, A., Kaberdin, V. R., Wei, C., and Lin-Chao, S. (1996). Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proceedings of the National Academy of Sciences of the United States of America* 93, 3865–3869.

Minton, K. (1994). DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. *Molecular Microbiology* 13, 9–15.

Misra, T. K. and Apirion, D. (1979). RNase E , an RNA Processing Enzyme from *Escherichia coli*. *The Journal of Biological Chemistry* 254, 11154–11159.

Mitra, A., Kesarwani, A. K., Pal, D., and Nagaraja, V. (2011). WebGeSTer DB-A transcription terminator database. *Nucleic Acids Research* 39, 129–135.

Miyake, K., Abe, K., Ferri, S., Nakajima, M., Nakamura, M., Yoshida, W., Kojima, K., Ikebukuro, K., and Sode, K. (2014). A green-light inducible lytic system for cyanobacterial cells. *Biotechnology for Biofuels* 7, 56.

Mohanty, B. K., Giladi, H., Maples, V. F., and Kushner, S. R. (2008). Analysis of RNA decay, processing, and polyadenylation in *Escherichia coli* and other prokaryotes. *Methods in Enzymology* 447, 3–29.

Mori, T., Binder, B., and Johnson, C. H. (1996). Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proceedings of the National Academy of Sciences of the United States of America* 93, 10183–10188.

Mulo, P., Sakurai, I., and Aro, E. M. (2012). Strategies for *psbA* gene expression in cyanobacteria, green algae and higher plants: From transcription to PSII repair. *Biochimica et Biophysica Acta - Bioenergetics* 1817, 247–257.

Muro-Pastor, M. and Florencio, F. (2003). Regulation of ammonium assimilation in cyanobacteria. *Plant Physiology and Biochemistry* 41, 595–603.

Murphy, R. C., Gasparich, G. E., Bryant, D. A., and Porter, R. D. (1990). Nucleotide sequence and further characterization of the *Synechococcus* sp. strain PCC 7002 *recA* gene : complementation of a cyanobacterial *recA* mutation by the *Escherichia coli recA* gene. *Journal of Bacteriology* 172, 967–976.

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

Mutsuda, M., Michel, K. P., Zhang, X., Montgomery, B. L., and Golden, S. S. (2003). Biochemical properties of CikA, an unusual phytochrome-like histidine protein kinase that resets the circadian clock in *Synechococcus elongatus* PCC 7942. *Journal of Biological Chemistry* 278, 19102–19110.

Na, D., Lee, S., and Lee, D. (2010). Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Systems Biology* 4, 71.

Nair, U., Thomas, C., and Golden, S. S. (2001). Functional elements of the strong *psbAI* promoter of *Synechococcus elongatus* PCC 7942. *Journal of Bacteriology* 183, 1740–1747.

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., and Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant and Cell Physiology* 54, 1724–1735.

Ng, W. O., Zentella, R., Wang, Y., Taylor, J. S., and Pakrasi, H. B. (2000). PhrA, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase. *Archives of Microbiology* 173, 412–417.

Nicholson, A. W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiology Reviews* 23, 371–390.

Nishimura, T., Takahashi, Y., Yamaguchi, O., Suzuki, H., Maeda, S. I., and Omata, T. (2008). Mechanism of low CO_2-induced activation of the *cmp* bicarbonate transporter operon by a LysR family protein in the cyanobacterium *Synechococcus elongatus* strain PCC 7942. *Molecular Microbiology* 68, 98–109.

Nomura, M., Yates, J. L., Dean, D., and Post, L. E. (1980). Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proceedings of the National Academy of Sciences* 77, 7084–7088.

Nozzi, N. E. and Atsumi, S. (2015). Genome engineering of the 2,3-butanediol biosynthetic pathway for tight regulation in cyanobacteria. *ACS Synthetic Biology* 4, 1197–1204.

Nozzi, N. E., Oliver, J. W. K., and Atsumi, S. (2013). Cyanobacteria as a platform for biofuel production. *Frontiers in Bioengineering and Biotechnology* 1, 1–6.

Nudler, E. and Gottesman, M. E. (2002). Transcription termination and anti-termination in *E. coli*. *Genes to Cells* 7, 755–768.

Nurmohamed, S., Vincent, H. A., Titman, C. M., Chandran, V., Pears, M. R., Du, D., Griffin, J. L., Callaghan, A. J., and Luisi, B. F. (2011). Polynucleotide phosphorylase activity may be modulated by metabolites in *Escherichia coli*. *Journal of Biological Chemistry* 286, 14315–14323.

Oesterle, S., Gerngross, D., Schmitt, S., Roberts, T. M., and Panke, S. (2017). Efficient engineering of chromosomal ribosome binding site libraries in mismatch repair proficient *Escherichia coli*. *Scientific Reports* 7, 12327.

Oliver, J. W. K. and Atsumi, S. (2014). Metabolic design for cyanobacterial chemical synthesis. *Photosynthesis Research* 120, 249–261.

Oliver, J. W. K., Machado, I. M. P., Yoneda, H., and Atsumi, S. (2013). Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proceedings of the National Academy of Sciences* 110, 1249–1254.

Olvera, L., Mendoza-Vargas, A., Flores, N., Olvera, M., Sigala, J. C., Gosset, G., Morett, E., and Bolívar, F. (2009). Transcription analysis of central metabolism genes in *Escherichia coli*. Possible roles of σ38 in their expression, as a response to carbon limitation. *PLOS ONE* 4, e7466.

Omata, T., Gohta, S., Takahashi, Y., Harano, Y., and Maeda, S. I. (2001). Involvement of a CbbR homolog in low CO_2-induced activation of the bicarbonate transporter operon in cyanobacteria. *Journal of Bacteriology* 183, 1891–1898.

Orf, I., Schwarz, D., Kaplan, A., Kopka, J., Hess, W. R., Hagemann, M., and Klähn, S. (2016). CyAbrB2 contributes to the transcriptional regulation of low CO_2 acclimation in *Synechocystis* sp. PCC 6803. *Plant and Cell Physiology* 57, 2232–2243.

Osanai, T., Imamura, S., Asayama, M., Shirai, M., Suzuki, I., Murata, N., and Tanaka, K. (2006). Nitrogen induction of sugar catabolic gene expression in *Synechocystis sp.* PCC 6803. *DNA Research* 13, 185–195.

Ow, M. C., Perwez, T., and Kushner, S. R. (2003). RNase G of *Escherichia coli* exhibits only limited functional overlap with its essential homologue, RNase E. *Molecular Microbiology* 49, 607–622.

Pato, M. and Meyenburg, K. (1970). Residual RNA synthesis in *Escherichia coli* after inhibition of initiation of transcription by rifampicin. *Cold Spring Harbor Symposia on Quantitative Biology* 35, 497–504.

Peca, L., Kós, P. B., and Vass, I. (2007). Characterization of the activity of heavy metal-responsive promoters in the cyanobacterium *Synechocystis* PCC 6803. *Acta Biologica Hungarica* 58, 11–22.

Pei, X. Y., Bralley, P., Jones, G. H., and Luisi, B. F. (2015). Linkage of catalysis and 5' end recognition in ribonuclease RNase J. *Nucleic Acids Research* 43, 8066–8076.

Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328.

Perez, A. A., Gajewsky, J. P., Ferlez, B. H., Ludwig, M., Baker, C. S., Golbeck, J. H., and Bryant, D. A. (2017). Zn^{2+}-inducible expression platform for *Synechoccus* sp. strain PCC 7002 based on the *smtA* promoter/operator and *smtB* repressor. *Applied and Environmental Microbiology* 83, 1–14.

Peters, J. M., Vangeloff, A. D., and Landick, R. (2011). Bacterial transcription terminators: The RNA 3'-end chronicles. *Journal of Molecular Biology* 412, 793–813.

Peterson, E. S., McCue, L. A., Schrimpe-Rutledge, A. C., Jensen, J. L., Walker, H., Kobold, M. A., Webb, S. R., Payne, S. H., Ansong, C. K., Adkins, J. N., Cannon, W. R., and Webb-Robertson, B. M. (2012). VESPA: software to facilitate genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data. *BMC Genomics* 13, 131.

Petrov, A. S., Bernier, C. R., Hershkovits, E., Xue, Y., Waterbury, C. C., Hsiao, C., Stepanov, V. G., Gaucher, E. A., Grover, M. A., Harvey, S. C., Hud, N. V., Wartell, R. M., Fox, G. E., and Williams, L. D. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Research* 41, 7522–7535.

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

Pickens, L., Tang, Y., and Chooi, Y. (2014). Metabolic engineering for the production of natural products. *Annual Review of Chemical and Biomolecular Engineering* 2, 211–236.

Pinto, F. L., Thapper, A., Sontheim, W., and Lindblad, P. (2009). Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria. *BMC Molecular Biology* 10, 79.

Pitera, D. J., Paddon, C. J., Newman, J. D., and Keasling, J. D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metabolic Engineering* 9, 193–207.

Plunkett, G. and Echols, H. (1989). Retroregulation of the bacteriophage lambda *int* gene: limited secondary degradation of the RNase III-processed transcript. *Journal of Bacteriology* 171, 588–592.

Politi, N., Pasotti, L., Zucca, S., Casanova, M., Micoli, G., Cusella De Angelis, M., and Magni, P. (2014). Half-life measurements of chemical inducers for recombinant gene expression. *Journal of Biological Engineering* 8, 5.

Politz, M. C., Copeland, M. F., and Pfleger, B. F. (2013). Artificial repressors for controlling gene expresison in bacteria. *Chemical Communications* 49, 4325.

Portier, C., Dondon, L., Grunberg-Manago, M., and Regnier, P. (1987). The first step in the functional inactivation of the *Escherichia coli* polynucleotide phosphorylase is a ribonuclease III processing at the 5' end. *EMBO Journal* 6, 2165–2170.

Puigbò, P., Guzmán, E., Romeu, A., and Garcia-Vallvé, S. (2007). OPTIMIZER: A web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research* 35, 126–131.

Punta, M., Coggill, P., Eberhardt, R., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E., Eddy, S., Bateman, A., and Finn, R. (2012). The Pfam protein families databases. *Nucleic Acids Research* 30, 1–12.

Py, B., Causton, H., Mudd, E. A., and Higgins, C. F. (1994). A protein complex mediating mRNA degradation in *Escherichia coli*. *Molecular Microbiology* 14, 717–729.

Py, B., Higgins, C. F., Krisch, H. M., and Carpousis, A. J. (1996). A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* 381, 169–172.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183.

Ramey, C. J., Baron-Sola, A., Aucoin, H. R., and Boyle, N. R. (2015). Genome engineering in cyanobacteria: where we are and where we need to go. *ACS Synthetic Biology* 4, 1186–1196.

Ravindran, C. R. M., Suguna, S., and Shanmugasundaram, S. (2006). Electroporation as a tool to transfer the plasmid pRL489 in *Oscillatoria* MKU 277. *Journal of Microbiological Methods* 66, 174–176.

Redko, Y., Bechhofer, D. H., and Condon, C. (2008). Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. *Molecular Microbiology* 68, 1096–1106.

Redon, E., Loubière, P., and Cocaign-Bousquet, M. (2005). Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *Journal of Biological Chemistry* 280, 36380–36385.

Reeve, B., Hargest, T., Gilbert, C., and Ellis, T. (2014). Predicting translation initiation rates for designing synthetic biology. *Frontiers in Bioengineering and Biotechnology* 2, 1–6.

Régnier, P. and Grunberg-Manago, M. (1989). Cleavage by RNase III in the transcripts of the *metY-nusA-infB* operon of *Escherichia coli* releases the tRNA and initiates the decay of the downstream mRNA. *Journal of Molecular Biology* 210, 293–302.

Régnier, P. and Portier, C. (1986). Initiation, attenuation and RNase III processing of transcripts from the *Escherichia coli* operon encoding ribosomal protein S15 and polynucleotide phosphorylase. *Journal of Molecular Biology* 187, 23–32.

Reppas, N. and Ridley, C. (2010). "Methods and compositions for the recombinant biosynthesis of n-alkanes". United States Patent. 8183027.

Reppas, N. B., Wade, J. T., Church, G. M., and Struhl, K. (2006). The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Molecular Cell* 24, 747–757.

Reyrat, J., Pelicic, V., and Gicquel, B. (1998). Counterselectable markers : untapped tools for bacterial genetics and pathogenesis. *Infection and Immunity* 66, 4011–4017.

Richter, N., Zienert, A., and Hummel, W. (2011). A single-point mutation enables lactate dehydrogenase from *Bacillus subtilis* to utilize NAD+ and NADP+ as cofactor. *Engineering in Life Sciences* 11, 26–36.

Rische, T. and Klug, G. (2012). The ordered processing of intervening sequences in 23S rRNA of *Rhodobacter sphaeroides* requires RNase J. *RNA Biology* 9, 343–350.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature Biotechnology* 29, 24–26.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Robinson, P. J., Cranenburgh, R. M., Head, I. M., and Robinson, N. J. (1997). HIP1 propagates in cyanobacterial DNA via nucleotide substitutions but promotes excision at similar frequencies in *Escherichia coli* and *Synechococcus* PCC 7942. *Molecular Microbiology* 24, 181–189.

Rott, R., Zipor, G., Portnoy, V., Liveanu, V., and Schuster, G. (2003). RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from *Escherichia coli*. *Journal of Biological Chemistry* 278, 15771–15777.

Ruffing, A. M., Jensen, T. J., and Strickland, L. M. (2016). Genetic tools for advancement of *Synechoccus* sp. PCC 7002 as a cyanobacterial chassis. *Microbial Cell Factories* 15, 190.

Rustad, T. R., Minch, K. J., Brabant, W., Winkler, J. K., Reiss, D. J., Baliga, N. S., and Sherman, D. R. (2013). Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic Acids Research* 41, 509–517.

Sakamoto, T. and Bryant, D. A. (1997). Temperature-regulated mRNA accumulation and stabilization for fatty acid desaturase genes in the cyanobacterium *Synechoccus* sp. strain PCC 7002. *Molecular Microbiology* 23, 1281–1292.

Sakurai, I., Stazic, D., Eisenhut, M., Vuorio, E., Steglich, C., Hess, W. R., and Aro, E. (2012). Positive regulation of *psbA* gene expression by cis-encoded antisense RNAs in *Synechocystis* sp. PCC 6803. *Plant Physiology* 160, 1000–1010.

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

Sauter, M. and Sawers, R. G. (1990). Transcriptional analysis of the gene encoding pyruvate formate-lyase-activating enzyme of *Escherichia coli*. *Molecular Microbiology* 4, 355–364.

Sawers, G. and Bock, A. (1989). Novel transcriptional control of the pyruvate-lyase gene: upstream regulatory sequences and multiple promoters regulate anaerobic expression. *Journal of Bacteriology* 171, 2485–2498.

Schlessinger, D., Ono, M., Nikolaev, N., and Silengo, L. (1974). Accumulation of 30S preribosomal ribonucleic acid in an *Escherichia coli* mutant treated with chloramphenicol. *Biochemistry* 13, 4268–4271.

Schuergers, N. and Wilde, A. (2015). Appendages of the Cyanobacterial Cell. *Life* 5, 700–715.

Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M., and Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Research* 13, 1–8.

Seo, S. W., Yang, J. S., Kim, I., Yang, J., Min, B. E., Kim, S., and Jung, G. Y. (2013). Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metabolic Engineering* 15, 67–74.

Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *The EMBO Journal* 28, 3523–3533.

Shan, J. and Clokie, M. (2009). Preparation of RNA from bacteria infected with bacteriophages: a case study from the marine unicellular *Synechococcus* sp. WH7803 infected by phage S-PM2. *Methods in Molecular Biology* 502, 171–176.

Shi, C., Wang, S., Xia, E., Jiang, J., Zeng, F., and Gao, L. (2016). Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Scientific Reports* 6, 30135.

Shi, M., Zhang, X., Pei, G., Chen, L., and Zhang, W. (2017). Functional diversity of transcriptional regulators in the cyanobacterium *Synechocystis* sp. PCC 6803. *Frontiers in Microbiology* 8, 1–10.

Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K. W., Han, C. S., Rubin, E. M., Eisen, J. A., Woyke, T., Gugger, M., and Kerfeld, C. A. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences* 110, 1053–1058.

Sim, M., Lim, B., Sim, S. H., Kim, D., Jung, E., Lee, Y., and Lee, K. (2014). Two tandem RNase III cleavage sites determine *betT* mRNA stability in response to osmotic stress in *Escherichia coli*. *PLoS ONE* 9, e100520.

Sim, S. H., Yeom, J. H., Shin, C., Song, W. S., Shin, E., Kim, H. M., Cha, C. J., Han, S. H., Ha, N. C., Kim, S. W., Hahn, Y., Bae, J., and Lee, K. (2010). *Escherichia coli* ribonuclease III activity is downregulated by osmotic stress: consequences for the degradation of *bdm* mRNA in biofilm formation. *Molecular Microbiology* 75, 413–425.

Simonte, F. M., Dotsch, A., Galego, L., Arraiano, C., and Gescher, J. (2016). Investigation on the anaerobic propionate degradation by *Escherichia coli* K12. *Molecular Microbiology* 103, 55–66.

Smanski, M. J., Zhou, H., Claesen, J., Shen, B., Fischbach, M. A., and Voigt, C. A. (2016). Synthetic biology to access and expand nature's chemical diversity. *Nature Reviews Microbiology* 14, 135–149.

Smolke, C. D., Carrier, T. A., and Keasling, J. D. (2000). Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Applied and Environmental Microbiology* 66, 5399–5405.

Song, W., Kim, Y. H., Sim, S. H., Hwang, S., Lee, J. H., Lee, Y., Bae, J., Hwang, J., and Lee, K. (2014). Antibiotic stress-induced modulation of the endoribonucleolytic activity of RNase III

and RNase G confers resistance to aminoglycoside antibiotics in *Escherichia coli*. *Nucleic Acids Research* 42, 4669–4681.

Stanton, B. C., Nielsen, A. A. K., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology* 10, 99–105.

Stead, M. B., Marshburn, S., Mohanty, B. K., Mitra, J., Castillo, L. P., Ray, D., Van Bakel, H., Hughes, T. R., and Kushner, S. R. (2011). Analysis of *Escherichia coli* RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Research* 39, 3188–3203.

Steege, D. A., Cone, K. C., Queen, C., and Rosenberg, M. (1987). Bacteriophage lambda N gene leader RNA. RNA processing and translational initiation signals. *Journal of Biological Chemistry* 262, 17651–17658.

Steglich, C., Lindell, D., Futschik, M., Rector, T., Steen, R., and Chisholm, S. W. (2010). Short RNA half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biology* 11, R54.

Stern, D. B. and Gruissem, W. (1987). Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* 51, 1145–1157.

Stevens, S. E. and Porter, R. D. (1980). Transformation in *Agmenellum quadruplicatum*. *Proceedings of the National Academy of Sciences* 77, 6052–6056.

Stevens, S., Patterson, C., and Myers, J. (1973). The production of hydrogen peroxide by bluee-green algae: a survey. *Journal of Phycology*, 427–430.

Stucken, K., Ilhan, J., Roettger, M., Dagan, T., and Martin, W. F. (2012). Transformation and conjugal transfer of foreign genes into the filamentous multicellular cyanobacteria (subsection V) *Fischerella* and *Chlorogloeopsis*. *Current Microbiology* 65, 552–560.

Su, M., Ling, Y., Yu, J., Wu, J., and Xiao, J. (2013). Small proteins: Untapped area of potential biological importance. *Frontiers in Genetics* 4, 1–9.

Takahama, K., Matsuoka, M., Nagahama, K., and Ogawa, T. (2003). Construction and analysis of a recombinant cyanobacterium expressing a chromosomally inserted gene for an ethylene-forming enzyme at the *psbAI* locus. *Journal of Bioscience and Bioengineering* 95, 302–305.

Takahashi, T., Nakai, N., Muramatsu, M., and Hihara, Y. (2010). Role of multiple HLR1 sequences in the regulation of the dual promoters of the *psaAB* genes in *Synechocystis* sp. PCC 6803. *Journal of Bacteriology* 192, 4031–4036.

Takata, R., Izuhara, M., and Hori, K. (1989). Differential degradation of the *Escherichia coli* polynucleotide phosphorylase mRNA. *Nucleic Acids Research* 17, 7441–7451.

Tamura, M., Moore, C. J., and Cohen, S. N. (2013). Nutrient dependence of RNase E essentiality in *Escherichia coli*. *Journal of Bacteriology* 195, 1133–1141.

Taton, A., Lis, E., Adin, D. M., Dong, G., Cookson, S., Kay, S. A., Golden, S. S., and Golden, J. W. (2012). Gene transfer in *Leptolyngbya* sp. strain BL0902, a cyanobacterium suitable for production of biomass and bioproducts. *PLoS ONE* 7, e30901.

Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., Palenik, B., Peterson, T. C., Haerizadeh, F., Golden, S. S., and Golden, J. W. (2014). Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Research* 42, 1–16.

Taverniti, V., Forti, F., Ghisotti, D., and Putzer, H. (2011). *Mycobacterium smegmatis* RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Molecular Microbiology* 82, 1260–1276.

Topp, S., Reynoso, C. M., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., Shen, A., Puri, A. W., Komeili, A., Bertozzi, C. R., Scott, J. R., and Gallivan, J. P. (2010). Synthetic riboswitches that induce gene expression in diverse bacterial species. *Applied and Environmental Microbiology* 76, 7881–7884.

Ungerer, J. and Pakrasi, H. B. (2016). Cpf1 is a versatile tool for CRISPR genome editing across diverse species of cyanobacteria. *Scientific Reports* 6, 39681.

Unniraman, S., Prakash, R., and Nagaraja, V. (2002). Conserved economics of transcription termination in eubacteria. *Nucleic Acids Research* 30, 675–68.

Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., Krisch, H. M., and Carpousis, A. J. (1998). Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. *Genes and Development* 12, 2770–2781.

Varman, A. M., Xiao, Y., Pakrasi, H. B., and Tang, Y. J. (2013). Metabolic engineering of *Synechocystis* sp. strain PCC 6803 for isobutanol production. *Applied and Environmental Microbiology* 79, 908–914.

Vijayan, V., Jain, I. H., and O'Shea, E. K. (2011). A high resolution map of a cyanobacterial transcriptome. *Genome Biology* 12, R47.

Viola, S., Rühle, T., and Leister, D. (2014). A single vector-based strategy for marker-less gene replacement in *Synechocystis* sp. PCC 6803. *Microbial Cell Factories* 13, 4.

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

Wang, B., Wang, J., and Meldrum, D. R. (2012). Application of synthetic biology in cyanobacteria and algae. *Frontiers in Microbiology* 3, 1–15.

Wang, H. L., Postier, B. L., and Burnap, R. L. (2004). Alterations in global patterns of gene expression in *Synechocystis* sp. PCC 6803 in response to inorganic carbon limitation and the inactivation of *ndhR*, a LysR family regulator. *Journal of Biological Chemistry* 279, 5739–5751.

Wang, T., Shen, G., Balasubramanian, R., Mcintosh, L., Bryant, D. A., and Golbeck, J. H. (2004). The *sufR* Gene (sll0088 in *Synechocystis* sp. strain PCC 6803) functions as a repressor of the *sufBCDS* operon in iron-sulfur cluster biogenesis in cyanobacteria. *Journal of Bacteriology* 186, 956–967.

Wang, Y., Chen, L., and Zhang, W. (2016). Proteomic and metabolomic analyses reveal metabolic responses to 3-hydroxypropionic acid synthesized internally in cyanobacterium *Synechocystis* sp. PCC 6803. *Biotechnology for Biofuels* 9, 209.

Washio, T., Sasayama, J., and Tomita, M. (1998). Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Research* 26, 5456–63.

Wendt, K. E., Ungerer, J., Cobb, R. E., Zhao, H., and Pakrasi, H. B. (2016). CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Microbial Cell Factories* 15, 115.

Whitton, B. and Potts, M. (2000). Introduction to the Cyanobacteria. *The Ecology of Cyanobacteria*. Kluwer Academic Publishers, pp. 1–11.

Wolk, C. P., Vonshak, A., Kehoe, P., and Elhai, J. (1984). Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. *Microbiology* 81, 1561–1565.

Woodger, F. J., Bryant, D. A., and Price, G. D. (2007). Transcriptional regulation of the CO_2-concentrating mechanism in a euryhaline, coastal marine cyanobacterium, *Synechoccus* sp. strain PCC 7002: role of NdhR/CcmR. *Journal of Bacteriology* 189, 3335–3347.

Wu, J., Du, G., Chen, J., and Zhou, J. (2015). Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. *Scientific Reports* 5, 13477.

Xu, K., Huang, X., Wu, M., Wang, Y., Chang, Y., Liu, K., Zhang, J., Zhang, Y., Zhang, F., Yi, L., Li, T., Wang, R., Tan, G., and Li, C. (2014). A rapid, highly efficient and economical method of *Agrobacterium*-mediated in planta transient transformation in living onion epidermis. *PLoS ONE* 9, 1–7.

Xu, M. (2015). "Comparative Genomics Reveals Forces Driving the Evolution of Highly Iterated Palindrome-1 (HIP1) in Cyanobacteria ." PhD thesis. Carnegie Mellon University.

Xu, T., Li, Y. Y., Shi, Z., Hemme, C. L., Li, Y. Y., Zhu, Y., Van Nostrand, J. D., He, Z., and Zhou, J. (2015). Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Applied and Environmental Microbiology* 81, 4423–4431.

Xu, W., Huang, J., and Cohen, S. N. (2008). Autoregulation of *absB* (RNase III) expression in *Streptomyces coelicolor* by endoribonucleolytic cleavage of *absB* operon transcripts. *Journal of Bacteriology* 190, 5526–5530.

Yao, L., Cengic, I., Anfelt, J., and Hudson, E. P. (2016). Multiple gene repression in cyanobacteria using CRISPRi. *ACS Synthetic Biology* 5, 207–212.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C., and Shively, J. M. (2008). Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nature Reviews Microbiology* 6, 681–691.

Yoshimura, H., Hisabori, T., Yanagisawa, S., and Ohmori, M. (2000). Identification and characterization of a novel cAMP receptor protein in the cyanobacterium *Synechocystis* sp. PCC 6803. *Journal of Biological Chemistry* 275, 6241–6245.

Young, R. A. and Steitz, J. A. (1978). Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proceedings of the National Academy of Sciences of the United States of America* 75, 3593–3597.

Zahn, K., Inui, M., and Yukawa, H. (2000). Divergent mechanisms of 5' 23S rRNA IVS processing in the alpha-proteobacteria. *Nucleic Acids Research* 28, 4623–4633.

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

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163, 759–771.

Zhang, F., Ouellet, M., Batth, T. S., Adams, P. D., Petzold, C. J., Mukhopadhyay, A., and Keasling, J. D. (2012). Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. *Metabolic Engineering* 14, 653–660.

Zhang, J., Deng, X., Li, F., Wang, L. I., Huang, Q., and Zhang, C. (2014). RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterial-specific nonapeptide in the noncatalytic region RNase E. *RNA* 20, 568–579.

Zhang, S. R., Lin, G. M., Chen, W. L., Wang, L., and Zhang, C. C. (2013). ppGpp metabolism is involved in heterocyst development in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Journal of Bacteriology* 195, 4536–4544.

Zou, J., Gasparich, G. E., Stirewalt, V. L., Lorimier, R. de, and Bryant, D. A. (1999). The *cpcE* and *cpcF* Genes of *Synechoccus* sp. PCC 7002. Construction and phenotypic characterization of interposon mutants. *The Journal of Biological Chemistry* 267, 16138–16145.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31, 3406–3415.

Appendices

Appendix I: RNA spike-in protocol

Introduction

RNA spike-ins are an important and often overlooked element that can be critical in genome-wide experiments and analyses. It is absolutely essential that there is a linear relationship between inputs and outputs. Many scientists make the assumption that the concentration and yields of RNA and DNA are the same across conditions, which is often untrue leading to inaccurate conclusions. Traditional analysis pipelines normalize the total number of signals (reads) for each sample which can greatly skew results when there was differing amounts of input material for each sample. RNA spike-ins This issue is not widely discussed and most genome-wide analyses do not use RNA spike-ins. A few publications highlight this issue and some have even retrospectively normalized datasets that drew inaccurate conclusions due to the lack of RNA spike-in controls (Chen, Hu, et al. 2016).

The need for external RNA controls was recognized by industry scientists, who approached the National Institute of Standards in Technology (NIST) in 2003. The External RNA Controls Consortium (ERCC), a group of private, public, and academic scientists, was formed to address this problem. They produced a mixture of 96 DNA templates designed to mimic but not interfere with mammalian RNA, and wrote the Standard Reference Manual (SRM) 2374 that detailed their use. A R package (erccdashboard) was also developed as a spike-in control analysis tool. An ERCC RNA Spike-In mix is commercially available from ThermoFisher (catalog number 4456740). These RNA transcripts are a range of different sizes between 250 to 2,000 bases and mimic eukaryotic mRNA. The RNAs are synthesized with in vitro transcription, and the DNA templates are derived from *Bacillus subtilis* and *Methanocaldoccus jannaschii* that have been poly-A+ tailed. These sequences have some properties of endogenous transcripts including a spectrum of GC-contents and lengths (Jiang et al. 2011).

The use of these RNA spike-ins has been shown to greatly increase the quality of comparisons

between different biological samples, by providing normalization for several pervasive biases introduced in RNA-sequencing. There have been many examples of inconsistent read coverage along a transcript especially drops in coverage at the 3' end, presumably caused by the reduced number of reverse-transcriptase priming sites (Jiang et al. 2011). Short transcripts have also been shown to be underrepresented presumably for the same reasoning. Additionally, the sequence of the reverse-transcriptase priming sites has been shown to influence the amount of coverage (Hansen et al. 2010). The use of RNA spike-ins can remove both the biases involved in both sequencing and sample preparation.

These ERCC spike-ins are useful for the study of eukaryotic DNA and RNA, but unfortunately there is no standard set of spike-ins for use in prokaryotes. There are very few examples of RNA spike-ins used in the study of bacterial gene expression. RNA spike-ins were used as controls for studying transcript degradation in *E. coli* (Chen, Shiroguchi, et al. 2015), which were necessary because the pool of RNA was being degraded throughout time after the addition of the antibiotic rifampicin which stops nascent transcription. They used a combination of 6 RNA spike-ins derived from the ϕX174 (NC_001422) genome introduced at ratios between 0.1-100 copies per cell.

As we designed our experiment to examine transcript degradation in cyanobacteria (Chapter 5), we realized the need for normalization across samples because we were degrading large portions of mRNA throughout the timecourse. We adapted the protocol described in (Chen, Shiroguchi, et al. 2015) and have outlined this procedure below.

RNA spike-in preparation

The same six RNA spike-ins from ϕX174 as used in (Chen, Shiroguchi, et al. 2015) were used: gene D, gene F, gene G, gene H, fragment 190, and fragment 290. DNA templates for in vitro transcription were created by adding the T7 promoter sequence (5'-CGAAATTAATACGACTCACTATAGG-3') to the 5' end of the forward primer. The primers used to generate DNA templates are listed in Table A1.1

Table A1.1: Primers used to generate RNA spike-in DNA templates

Primer Name	Sequence	Description	Target
JCC423	cgaaatTAATACGACTCACTATAGGGTAAGAAA TCATGAGTCAAGTTACTGAAC	ϕX174 gene D-F	ϕX174 D
JCC424	TTACATCACTCCTTCCGCACGTAATTTT	ϕX174 gene D-R	ϕX174 D
JCC425	cgaaatTAATACGACTCACTATAGGGAGGATAA ATTATGTCTAATATTCAAACTG	ϕX174 gene F-F	ϕX174 F
JCC426	TCACGAAGTCATGATTGAATCG	ϕX174 gene F-R	ϕX174 F
JCC427	cgaaatTAATACGACTCACTATAGGGAGTTTAA TCATGTTTCAGACTTTTATTTC	ϕX174 gene G-F	ϕX174 G
JCC428	TCACTTAAGTGGCTGGAGACAAATAATC	ϕX174 gene G-R	ϕX174 G
JCC429	cgaaatTAATACGACTCACTATAGGGTGAGGTG ATTTATGTTTGGTGCTATTG	ϕX174 gene H-F	ϕX174 H
JCC430	TTATTTCCTAGACAAATTAGAGCCAATACC	ϕX174 gene H-R	ϕX174 H
JCC431	cgaaatTAATACGACTCACTATAGGGAATCGCG TAGAGGCTTTGCTA	ϕX174 frag190-F	ϕX174 190
JCC432	CAGCAAGAACCATACGACCAATATCAC	ϕX174 frag190-R	ϕX174 190
JCC433	cgaaatTAATACGACTCACTATAGGGTGAGCGT ATCGAGGCTCTTAAACCT	ϕX174 frag290-F	ϕX174 290
JCC434	TAGTGTTATTAATATCAAGTTGGGGGAGCAC	ϕX174 frag290-R	ϕX174 290

DNA templates were created in 50 µl PCR reactions containing 67.5 µl water, 20 µl 5X Phusion Buffer (New England Biolabs, M0530L), 5 µl each forward and reverse primer (10 µM), 1 µl dNTPs, 0.5 µl of 1 mg/µl ϕX174 DNA template (New England Biolabs, N3023S), and 1 µl Phusion enzyme. The reaction was run on a ProFlex PCR system with the following cycling conditions: 95°C for 3 minutes, 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, repeat to step 2 for a total of 35 times, 72°C for 10 minutes, and storage at 4°C. DNA was purified with ethanol precipitation (1/10 volume 3M sodium acetate and 2-3 volumes ethanol, stored at -80°C for $\geq$ 12 hours before centrifuging at 4°C for 30 minutes, washing with 75% ethanol, drying, and resuspending in water). RNA was quantified with a Nanodrop and used as a template for in vitro transcription using a T7 RiboMAX Express Large Scale RNA Production System (Promega, P1320) according to manufacturer instructions. In vitro transcription reactions were DNase-treated and RNA was purified by phenol/chloroform extraction followed by ethanol precipitation. RNA was quantified by Nanodrop.

Table A1.2: Copy number of spike-ins

RNA Spike-In	Copy Number added in (Chen, Shiroguchi, et al. 2015)	Copy number added in cyanobacterial dataset
gene D	0.1 RNA/cell	1 RNA/cell
gene F	1 RNA/cell	10 RNA/cell
gene G	1 RNA/cell	10 RNA/cell
gene H	0.01 RNA/cell	0.1 RNA/cell
fragment 190	100 RNA/cell	1000 RNA/cell
fragment 290	10 RNA/cell	100 RNA/cell

Determination of copy number of RNA spike-in to introduce

We were unsure if the range of copy numbers of RNA spike-ins in (Chen, Shiroguchi, et al. 2015) were appropriate for our cyanobacterial samples, so we did initial tests to see if we could detect expression of the spike-ins using reverse-transcription PCR. We designed primers to detect all six spike-ins along with three cyanobacterial targets with varying expression levels (A2512: medium; G0078: medium-low, and A2864: low). We collected samples from WT PCC 7002 and added the number of copies of RNA spike-ins as used previously in (Chen, Shiroguchi, et al. 2015) (**Table A1.2**) before extracting RNA with the hot phenol extraction protocol outlined in Chapter 4 and 5. We created cDNA using the iScript cDNA synthesis kit (BioRad # 1708890) using random hexamers according to manufacturer instructions.

We used the primers listed in **Table A1.3** to see if we could detect products from the reverse-transcribed cyanobacterial genes and RNA spike-ins. We ran 10 µl reactions with 5 µl 2X GoTaq Green Master Mix (Promega, M7123), 1 µl 10 µM mix of primers, 3 µl water, and 1 µl cDNA template or control with no reverse transcriptase. Reactions were amplified with the following cycling conditions: 95°C 3 minutes, 30 cycles of 95°C 30 seconds, 55°C 30 seconds, 72°C 30 seconds, and finally 72°C for 5 minutes. Products were run on a 1% agarose gel and visualized with ethidium bromide.

We could not detect product from the RNA spike-ins with the lowest copy numbers (gene D and gene H) even though we could detect expression of native low-expression cyanobacterial genes. We

Table A1.3: qPCR primers used for RNA spike-in detection.

Primer Name	Sequence	Primer Description	Target
JCC435	AAAGCCGCTGAATTGTTCGC	ϕX174 qPCR D-F	ϕX174 D
JCC436	TGACGCACGTTTTCTTCTGC	ϕX174 qPCR D-R	ϕX174 D
JCC437	TTTTAAAGCGCCGTGGATGC	ϕX174 qPCR F-F	ϕX174 F
JCC438	TTTTGAGATGGCAGCAACGG	ϕX174 qPCR F-R	ϕX174 F
JCC439	TGCCTGTTTGGTTCGCTTTG	ϕX174 qPCR G-F	ϕX174 G
JCC440	TCATGGCGACCATTCAAAGG	ϕX174 qPCR G-R	ϕX174 G
JCC441	ATTGCTGGCGGTATTGCTTC	ϕX174 qPCR H-F	ϕX174 H
JCC442	AGCACATCACCTTGAATGCC	ϕX174 qPCR H-R	ϕX174 H
JCC443	GCGTAGAGGCTTTGCTATTCAG	ϕX174 qPCR Frag190-F	ϕX174 190
JCC444	TCTAATCGGTCGTCAGCCAAC	ϕX174 qPCR Frag190-R	ϕX174 190
JCC445	TTTTCGTATGCAGGGCGTTG	ϕX174 qPCR Frag290-F	ϕX174 290
JCC446	TCATCACGAACGTCAGAAGC	ϕX174 qPCR Frag290-R	ϕX174 290
GG233	CCACAACTCTTTTGTGCCAAT	A2512 qPCR F	A2512
GG234	TTTGGTCTGTGGAGTCAATGTC	A2512 qPCR R	A2512
GG235	TGTTTAACAATGTTGGCACTGG	G0078 qPCR F	G0078
GG236	GCATACAGGAGCCCAGGAC	G0078 qPCR R	G0078
GG237	CGGTACGGATGACTCCAGTT	A2864 qPCR F	A2864
GG238	TGACGACCCGATGAGTTGTA	A2864 qPCR R	A2864

decided to increase the copy number of the spike-ins by 10-fold for all six spike-ins based on these results (**Table A1.3**). Ideally, these preliminary experiments would have been performed using Next-Gen RNA-sequencing and not reverse transcription PCR, but at this time of this experiment it was infeasible due to cost.

Introducing RNA spike-ins

For normalization with RNA spike-ins, they must be added proportionally according to the number of cells in each sample. Samples were taken and frozen as outlined in Chapter 5, and cells were taken from each sample for precise cells counts using a hemocytometer. A master mix of all six RNA spike-ins was created in the appropriate ratios. To control for the various processing steps in RNA-sequencing library preparation it is best to add the RNA spike-ins as early as possible, so the appropriate amount of master mix was added to each cell pellet according to the cell count. RNA was extracted, quantified, and prepared for Next-Generation Sequencing as outlined in Chapter 5.

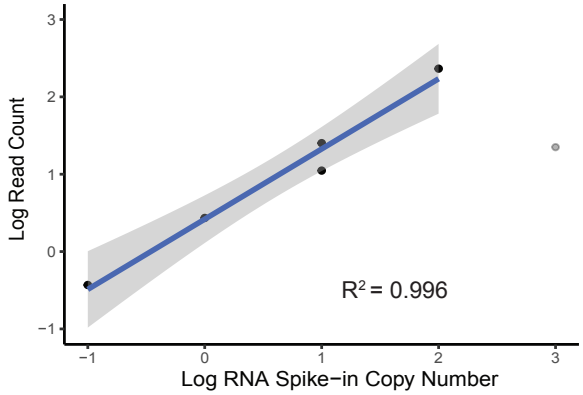


Figure A1.1: Relationship between read count and spike-in addition. A linear relationship between the log read count of the spike-ins (normalized for length) and the log spike-in copy number for spike-ins added between 0.1 and 100 copies/cell. The spike-in added at 1,000 copies/cell was retrospectively too high, and not used in the normalization calculation. The linear fit for the remaining 5 spike-ins is shown along with the shaded 95% confidence interval, $R^2 = 0.996$.

RNA spike-in quantification and normalization

RNA-sequencing files were aligned and processed according to standard protocols (outlined in Chapter 5). In addition to aligning reads to the cyanobacterial genome, reads were also aligned to the ϕX174 genome (NC_001422). The HTseq count function was used to determine the number of reads that aligned to the each spike-in. There should be a linear relationship between the log copy number added and the log read count output (normalized for length of the gene). We saw a linear relationship between the log copy number added and the log read count for five of the six spike-ins (**Figure A1.1**). There was significant deviation from this relationship for fragment 190 which was added at a concentration of 1000 RNA molecules/cell. The read counts from this spike-in appears to be saturated. All read counts from cyanobacterial genes fell in the copy number range of the lower 5 spike-ins (0.1-100 RNA molecules per cell).

We excluded all read counts and copy number information from fragment 190 for all subsequent analysis due to this saturation of read counts. All samples had a good correlation between log RNA spike-in copy number and read count with all R^2 values were between 0.971 and 0.999. Our experimental setup included tracking RNA degradation form three biological replicates of WT PCC

7002 following the addition of rifampicin to stop transcription. We were primarily concerned with the potential affects of significant RNA degradation in subsequent timepoints where the remaining mRNA would be sequenced with a high numbers of reads. The RNA spike-ins were added in the same ratios per cell across all samples, so should remain constant through all three timecourses. We calculated the sum of the reads that aligned to all five spike-ins and created a normalization factor for each sample to ensure constant reads through the timecourse. This normalization factor was also applied to all genes to ensure that read counts correspond to original RNA copy numbers.

This normalization method, using the sum of the reads that aligned to all of the RNA spike-ins was useful in determining original RNA copy number in the sample. This method however, does not correct for other RNA-sequencing biases such as variations in read coverage along a transcript. Future work to both validate a set of RNAs to be used as RNA spike-ins as well as bioinformatic tools to aid in normalization would be greatly appreciated in the field of genome-wide gene expression analysis in bacteria.

For future experiments we recommend using the copy numbers as outlined in (Chen, Shiroguchi, et al. 2015). We hope that future prokaryotic gene expression experiments include the use of RNA spike-ins and that a standard set of spike-ins can be developed to help with consistency and accuracy of genome-wide analyses in bacteria.

Appendix II: RNA-sequencing supplementary files

This thesis contains several RNA sequencing datasets from both *E. coli* and cyanobacteria, and all of the raw sequencing files and processed data is publicly available online. This appendix summarizes how to access these datasets as well as descriptions of the samples taken. Tables of a subset of the most pertinent information have been created for easy access.

Raw sequencing files (fastq) are the large files containing the sequences and quality information from millions of reads from the sequencer. These raw sequencing files have been deposited in the Sequence Read Archive (SRA) and can be accessed with the accession number found in **Table A2.4**.

These raw sequencing files were aligned to the genomes of the respective organisms and the number of reads that aligned to either each gene or each position in the genome was counted. These counts as well as other processed data files such as differentially expressed genes and calculated half-lives have been in the Gene Expression Omnibus (GEO) database. These files can be accessed with the accession number listed in **Table A2.4**.

In Chapter 2, potential RNase III cleavage sites were identified using a set of criteria to compared sequencing coverage from a WT *E. coli* strain and a mutant lacking RNase III (*rnc*-). The entire list of potential novel cleavage sites is listed in both the published paper as well as through the GEO, but a subset of the potential targets organized in decreasing levels of confidence can be seen below in **Table A2.5**. Only 19 or these putative sites were tested to see if RNase III could cleave them in vitro, and further exploration of these putative sites could reveal other gene expression control via RNase III.

Table A2.4: Summary of access to RNA-sequencing files

Name	Thesis Chapter	Description	SRA Accession Number	GEO Accession Number
E. coli RNase III	Chapter 2	Samples were taken from WT (MG1693) and a mutant that lacks RNase III (SK4455). Initial steady steady samples were taken as well as after the addition of rifampicin. Timepoints were taken at 0, 2.5, 5, 7.5, 10, and 20 minutes after rifampicin addition. A partial biological of WT was collected (0, 2.5, and 7 minutes).	SRP100654	GSE95318
Cyanobacteria RNase IIIs	Chapter 4	Steady-state samples from biological triplicates were taken from WT PCC 7002, all 3 single double RNase III mutants, all 3 double RNase III mutants, and the triple RNase III mutant.	SRP107964	GSE99279
Cyanobacteria Half-Lives	Chapter 5	Samples were taken from three biological replicates of WT PCC 7002 following the addition of rifampicin. Timepoints were taken at 0, 0.5, 1, 2.5, 5, 7.5, and 10 minutes. RNA spike-ins were added to cell pellets before lysis for normalization	SRP130967	GSE109174

Table A2.5: Top 100 putative RNase III sites in *E. coli*

Location	Annotation	Ratio 1	Ratio 2	WT A Reads	WT B Reads	*rnc*- Reads	Known?
4,166,537	intergenic *(murI-rrsB)*	415	587	41	29	17025	Y, rRNA
4,035,410	intergenic *(hemG-rrsA)*	314	925	53	18	16647	Y, rRNA
122,961	intergenic *(pdhR-aceE)*	6977	1550	2	9	13954	*
3,311,286	*psrO*	3782	3782	3	3	11347	Y
2,731,278	intergenic *(rrsG-clpB)*	487	534	23	21	11206	Y, rRNA
223,653	intergenic *(gmhB-rrsH)*	271	491	29	16	7853	Y, rRNA
3,428,875	intergenic *(rrsD-yrdA)*	340	234	20	29	6791	Y, rRNA
3,943,398	intergenic *(rrsC-gltU)*	1309	2619	4	2	5237	Y, rRNA
4,209,733	intergenic *(rrsE-gltV)*	4959	2480	1	2	4959	Y, rRNA
639,695	intergenic *(ahpC-ahpF)*	900	900	4	4	3598	*
1,799,278	intergenic *(pheM-rplT)*	639	799	5	4	3196	*
4,168,221	intergenic *(rrsB-gltT)*	123700	123700	0.01	0.01	1237	Y, rRNA
4,330,944	*proP*	66	83	15	12	992	
3,889,207	*tnaA*	301	129	3	7	903	
4,168,637	intergenic *(gltT-rrlB)*	120	419	7	2	837	Y, rRNA
3,427,176	intergenic *(ileU-rrsD)*	77500	775	0.01	1	775	Y, rRNA
3,397,879	*yhdE (rng)*	114	98	6	7	684	Y
3,889,742	*tnaA*	164	657	4	1	657	
225,355	intergenic *(rrsH-ileV)*	64300	643	0.01	1	643	Y, rRNA
4,037,115	intergenic *(rrsA-ileT)*	62600	62600	0.01	0.01	626	Y, rRNA
3,318,176	intergenic *(rimP-metY)*	102	609	6	1	609	Y

755,579	*sdhD*	545	42	1	13	545	
3,683,389	*dctA*	74	74	7	7	518	
4,331,699	*proP*	157	94	3	5	470	
4,037,523	*rrlA*	149	223	3	2	446	Y, rRNA
2,171,807	intergenic *(gatR-gatD)*	213	47	2	9	426	
758,515	*sucA*	30500	305	0.01	1	305	Y
4,207,874	intergenic *(purH-rrsE)*	29000	290	0.01	1	290	Y, rRNA
4,330,344	intergenic *(yjcZ-proP)*	28600	143	0.01	2	286	
3,609,652	*yhhQ*	58	77	4	3	231	
3,946,623	intergenic *(rrlC-rrfC)*	22700	227	0.01	1	227	Y, rRNA
2,729,186	intergenic *(rrlG-gltW)*	55	221	4	1	221	Y, rRNA
3,890,150	intergenic *(tnaA-tnaB)*	21500	21500	0.01	0.01	215	
4,040,443	intergenic *(rrlA-rrfA)*	20300	203	0.01	1	203	Y, rRNA
4,212,954	intergenic *(rrlE-rrfE)*	19500	65	0.01	3	195	Y, rRNA
3,941,526	intergenic *(yieP-rrsC)*	16200	41	0.01	4	162	Y, rRNA
951,147	intergenic *(pflA-pflB)*	15800	15800	0.01	0.01	158	*
3,890,912	*tnaB*	15700	157	0.01	1	157	
754,974	intergenic *(gltA-sdhC)*	15600	78	0.01	2	156	
2,351,931	*glpT*	14300	143	0.01	1	143	
2,237,178	*mglC*	133	13300	1	0.01	133	
4,307,817	*alsE*	126	63	1	2	126	
3,888,434	intergenic *(mnmE-tnaC)*	12100	12100	0.01	0.01	121	
4,171,559	intergenic *(rrlB-rrfB)*	12100	12100	0.01	0.01	121	Y, rRNA
228,677	intergenic *(rrlH-rrfH)*	11900	11900	0.01	0.01	119	Y, rRNA
2,090,031	*hisL*	11900	60	0.01	2	119	*
3,423,873	intergenic *(rrfD-rrlD)*	11700	117	0.01	1	117	Y, rRNA
2,726,273	intergenic *(rrfG-rrlG)*	11500	58	0.01	2	115	Y, rRNA
1,166,043	intergenic *(id176-ndh)*	10800	10800	0.01	0.01	108	*
4,219,637	*aceK*	35	10600	3	0.01	106	
159,578	*gluQ*	10400	52	0.01	2	104	
444,573	intergenic *(panE-yajQ)*	48	32	2	3	96	
1,622,816	intergenic *(yneM-mgrR)*	9100	91	0.01	1	91	
4,309,885	*alsA*	9100	46	0.01	2	91	
4,309,366	*alsC*	9100	30	0.01	3	91	
4,308,964	*alsC*	43	8500	2	0.01	85	
4,124,368	*cytR*	82	82	1	1	82	
160,018	*gluQ*	40	8000	2	0.01	80	
3,808,454	intergenic *(waaQ-waaA)*	7900	7900	0.01	0.01	79	
4,312,076	intergenic *(alsB-rpiR)*	7900	79	0.01	1	79	
3,625,154	*yhhI*	7800	7800	0.01	0.01	78	
3,902,404	*bglB*	7800	7800	0.01	0.01	78	
3,837,017	*setC*	7800	39	0.01	2	78	
2,237,626	*mglC*	77	7700	1	0.01	77	

1,087,223	*pgaC*	7600	7600	0.01	0.01	76	
2,236,473	*preA*	7500	75	0.01	1	75	
3,726,629	*wecH*	7400	7400	0.01	0.01	74	
329,508	*betT*	7300	7300	0.01	0.01	73	
2,238,048	*mglA*	7300	7300	0.01	0.01	73	
3,265,639	*tdcB*	7300	73	0.01	1	73	
4,258,391	*dinF*	37	37	2	2	73	
3,774,359	intergenic *(id496-mtlD)*	7200	72	0.01	1	72	
1,086,472	*pgaD*	7200	36	0.01	2	72	
3,219,193	intergenic *(aer-patA)*	7200	36	0.01	2	72	
3,371,184	*nanT*	7200	36	0.01	2	72	
3,706,530	*dppA*	71	71	1	1	71	
3,653,499	*yhiS*	7000	7000	0.01	0.01	70	
3,864,164	*yidP*	7000	7000	0.01	0.01	70	
4,029,925	*fadB*	7000	7000	0.01	0.01	70	
3,032,114	*uacT*	6900	69	0.01	1	69	
4,547,257	*fimD*	6900	69	0.01	1	69	
3,820,001	*ligB*	6800	6800	0.01	0.01	68	
4,454,370	*yjfF*	68	68	1	1	68	
3,218,246	*aer*	6700	6700	0.01	0.01	67	
3,362,576	*yhcA*	6700	6700	0.01	0.01	67	
3,712,153	intergenic *(yhjX-yhjY)*	6700	6700	0.01	0.01	67	
3,903,352	*bglB*	6700	34	0.01	2	67	
2,683,478	*yphH*	6600	6600	0.01	0.01	66	
3,948,917	*yifB*	6600	6600	0.01	0.01	66	
2,979,351	*lysR*	6600	66	0.01	1	66	
3,590,994	*id472*	6600	66	0.01	1	66	
3,842,636	*yicO*	6600	66	0.01	1	66	
4,284,124	*actP*	6600	66	0.01	1	66	
317,696	intergenic *(ykgQ-ykgB)*	6600	33	0.01	2	66	
618,903	*fepE*	6600	33	0.01	2	66	
1,372,502	*ycjN*	6500	6500	0.01	0.01	65	
2,779,152	*ypjA*	6500	6500	0.01	0.01	65	
3,274,632	*garP*	6500	6500	0.01	0.01	65	
4,203,506	*zraR*	6500	6500	0.01	0.01	65	
3,820,708	*ligB*	6500	65	0.01	1	65	

In Chapter 4 we highlight the differential expressed genes in the various single, double, and triple RNase III mutants compared to WT. These genes could either be directly affected by the absence of RNase III or be indirectly affected through other processes influence by RNase III. The genes are ranked by their False Discovery Rate (FDR), and those with the smallest FDR are the

most likely to be differentially expressed. The $\log_2$ fold change is also listed as well as the average reads from the WT and the mutant.

The most startling observation in this dataset is the large number of overlapping genes that are differentially expressed in the single RNase III mutants, but not the double or triple RNase III mutants. This group consists of all of the known members of the CcmR regulon as well as others that pertain to similar functions including bicarbonate transport and CO_2 transport. We also saw genes transhydrogenase genes as well as genes involved in glycolysis decreasing in expression in the single mutants only. These intriguing movements of genes involved in important metabolic processes may suggest that the single mutants can compensate for defects caused by the removal of a single RNase III by differentially regulating these genes, but upon the removal of two or all three RNase III genes, the cells can no longer compensate. The direct processing sites of RNase III remain elusive and highlight the complexity of their action and function within cyanobacteria.

In Chapter 5 we examined the global turnover of mRNA in PCC 7002 and calculated half-lives of all genes. We highlight here the top 80 most stable PCC 7002 genes (**Table A2.13**) and the top 80 least stable PCC 7002 genes (**Table A2.14**). Continued probing of this dataset may reveal sequences or structures that confer stability or decay and could be potentially used to influence the stability of a heterologous transcript.

Table A2.6: Top 40 differentially expressed genes in ΔA0061

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA0061
A0061	2.71E-148	-8.59	33.2	0.1
A1321	3.74E-63	3.14	364.6	3568.4
A2375	5.62E-63	3.38	9.8	112.7
A1503	5.77E-58	-1.9	439.2	131.2
A2670	1.17E-57	-2	297.3	83.1
G0011	6.48E-57	6.38	24.1	2217.7
A1514	1.72E-56	-2.29	204.0	46.4
A0473	2.00E-54	5.87	0.8	54.5
A0171	4.99E-54	6.48	17.9	1763.7
A2671	2.94E-52	-1.97	594.9	169.0
A2287	8.78E-52	5.36	2.5	111.6
A2371	1.40E-51	3.78	44.2	674.1
A0472	1.58E-51	6.81	9.1	1134.7
A0470	4.71E-49	7.22	34.2	5618.0
A0172	6.20E-49	6.25	23.7	2000.3
A0173	9.22E-49	6.18	17.5	1402.8
A2673	2.88E-48	-2.07	287.3	76.3
A1732	5.75E-48	4.2	34.4	703.1
A2372	1.84E-47	3.71	31.1	451.5
A2374	3.75E-47	3.19	6.3	63.9
A2492	1.31E-46	5.73	7.8	462.0
A0175	4.15E-46	5.64	9.2	510.2
A2254	3.77E-45	1.96	192.4	829.9
A2672	3.85E-45	-2.09	311.8	81.8
A1513	1.64E-44	-2.2	120.4	29.1
A2379	4.22E-43	2.7	6.4	46.4
A2674	8.10E-43	-1.83	570.6	179.0
A1520	1.19E-42	-1.57	117.5	43.8
A2676	5.06E-42	-1.93	872.1	255.7
A0174	2.47E-41	5.99	17.5	1234.9
A2376	1.74E-40	2.92	3.1	25.7
A2493	6.90E-39	5.81	4.3	264.6
A1509	3.20E-38	-2.19	407.0	99.0
A2675	3.52E-38	-1.78	436.4	141.7
A2380	1.39E-37	2.32	6.6	36.5
A1508	3.72E-36	-2.19	478.6	116.6
G0010	6.21E-36	3.65	47.4	659.6
A1505	1.26E-35	-1.8	593.1	189.4
A1531	1.49E-35	-1.7	98.9	33.9
A1506	2.12E-35	-2.11	425.8	109.8

Table A2.7: Top 40 differentially expressed genes in ΔA2542

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA2542
A2542	1.80E-140	-4.89	33.8	1.3
C0011	2.00E-139	3.43	49.9	517.5
C0010	3.00E-66	2.03	96.2	375.2
C0013	2.20E-58	1.93	87.4	320.8
C0009	1.30E-56	1.69	39.7	123.7
A2375	6.90E-56	3.17	9.8	84.8
C0012	9.80E-50	1.85	114.1	396.5
C0008	1.80E-48	2.09	294.7	1197.6
A1321	1.70E-46	2.63	364.6	2160.3
A0472	5.70E-45	6.21	9.1	648.1
A0473	2.90E-43	5.23	0.8	30.2
A0470	6.80E-43	6.57	34.2	3076.2
A0173	4.30E-42	5.6	17.5	811.3
A0171	6.50E-42	5.44	17.9	747.8
A2376	4.10E-41	2.95	3.1	23.1
G0011	7.76E-41	5.08	24.1	781.8
A1527	1.01E-40	-2.33	35.4	6.8
A0175	2.23E-40	5.18	9.2	318.8
A1531	2.23E-40	-1.85	98.9	26.7
C0007	2.23E-40	1.77	154.8	499.4
A0172	3.72E-39	5.37	23.7	940.5
A2379	4.02E-39	2.59	6.4	37.4
A2371	6.22E-39	3.21	44.2	394.0
A2670	1.01E-38	-1.61	297.3	93.9
A2372	4.54E-37	3.21	31.1	281.6
A1514	2.11E-36	-1.81	204.0	56.1
C0014	5.84E-36	1.66	51.4	157.0
A0174	1.23E-35	5.43	17.5	723.8
A2213	1.34E-35	-1.29	111.5	43.8
A2671	2.97E-35	-1.6	594.9	189.9
A2287	1.17E-34	4.25	2.5	44.9
A2106	2.29E-34	-1.05	108.4	50.1
C0006	5.49E-32	1.62	186.4	544.4
A2818	6.13E-32	1.15	52.4	112.5
A2374	8.76E-32	2.62	6.3	37.4
A2380	1.57E-31	2.15	6.6	28.2
A2672	1.18E-30	-1.69	311.8	93.6
A2673	1.37E-29	-1.58	287.3	92.9
A1732	2.36E-29	3.14	34.4	293.6
A2492	1.11E-28	4.21	7.8	138.1

Table A2.8: Top 40 differentially expressed genes in ΔA0384

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA0384
A0384	4.84E-135	-3.47	73.7	7.0
A0385	1.48E-40	1.22	180.5	438.2
A1060	2.59E-35	-2.03	440.0	111.8
A1059	2.20E-33	-2.01	934.1	241.4
A1065	1.43E-32	-2.01	1742.8	450.1
A2375	4.86E-31	2.38	9.8	52.8
A1064	1.13E-30	-2.06	1124.9	280.2
A1062	5.08E-30	-1.84	1332.2	385.9
A1531	5.08E-30	-1.61	98.9	33.8
A0472	6.76E-30	4.78	9.1	261.2
A1061	6.83E-29	-2.03	430.8	109.7
A0470	1.86E-28	5.01	34.2	1144.5
A1058	3.12E-28	-1.83	447.0	130.5
A1527	8.91E-28	-1.93	35.4	9.6
A0173	4.33E-27	4.24	17.5	343.9
A0175	4.86E-27	4.06	9.2	160.9
A1050	3.15E-26	-1.75	364.8	112.8
A2211	2.27E-25	-1.29	5600.0	2384.7
A0171	4.46E-25	3.95	17.9	287.8
A1057	9.08E-25	-1.76	221.1	68.1
A0174	1.96E-24	4.29	17.5	355.9
A0172	4.98E-24	3.96	23.7	383.8
A2213	5.60E-23	-1.06	111.5	55.5
A1879	1.30E-22	-2.03	37.9	9.5
A1520	1.45E-22	-1.15	117.5	54.9
A2819	2.77E-22	1.27	279.2	698.3
A2818	2.93E-22	0.99	52.4	108.4
A2379	5.01E-22	2	6.4	26.7
A1514	5.56E-22	-1.39	204.0	80.9
A1526	6.38E-22	-0.98	106.2	55.9
A1028	3.06E-21	-1.07	1061.6	528.1
A1063	3.85E-21	-1.85	403.2	116.1
A2372	1.15E-20	2.33	31.1	163.5
A2380	1.16E-20	1.81	6.6	24.0
A2063	3.01E-20	-1.14	475.1	223.7
A0473	5.09E-20	3.69	0.8	11.4
A1053	1.11E-19	-1.87	664.5	189.6
A2376	1.44E-19	2.18	3.1	14.4
A1246	3.97E-19	-0.88	230.5	129.8
A1048	4.95E-19	-1.4	611.1	240.2

Table A2.9: Top 40 differentially expressed genes in ΔA0061ΔA2542

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA0061ΔA2542
C0011	1.36E-180	3.99	49.9	932.1
A0061	1.02E-156	-9.46	33.2	0.0
A2542	1.60E-156	-4.95	33.8	1.3
A1060	3.15E-59	-2.66	440.0	81.5
A1059	3.88E-59	-2.72	934.1	166.5
A1050	3.00E-56	-2.63	364.8	69.0
A1058	1.69E-51	-2.52	447.0	91.5
A1064	5.54E-49	-2.65	1124.9	211.5
A1065	2.42E-48	-2.47	1742.8	369.9
A1531	5.29E-48	-2.02	98.9	28.7
C0009	7.14E-48	1.55	39.7	136.6
A1062	9.86E-48	-2.35	1332.2	307.3
C0010	1.96E-47	1.7	96.2	368.4
A1061	5.02E-46	-2.6	430.8	83.7
A1057	2.76E-45	-2.4	221.1	48.9
A2211	2.68E-43	-1.7	5600.0	2032.1
C0013	1.52E-42	1.63	87.4	318.6
A1051	2.75E-42	-2.66	653.6	121.9
A1520	3.18E-42	-1.57	117.5	46.8
A1049	2.03E-40	-2.15	634.8	168.9
A1527	6.99E-40	-2.24	35.4	8.7
A1879	6.99E-40	-2.76	37.9	6.5
A1063	2.73E-39	-2.58	403.2	79.0
A1053	3.35E-39	-2.71	664.5	119.3
C0008	1.19E-38	1.85	294.7	1249.4
A2818	1.97E-38	1.26	52.4	147.5
A1048	4.13E-38	-2.02	611.1	177.4
C0012	2.46E-37	1.59	114.1	404.3
A1052	3.26E-35	-2.76	477.9	83.2
C0014	8.07E-35	1.63	51.4	186.6
A1026	6.01E-34	-2.54	1329.9	269.4
A1055	8.11E-34	-2.41	450.8	99.7
A1056	1.34E-33	-2.52	372.3	76.2
A1300	2.97E-33	-1.32	79.2	37.5
A1036	1.16E-32	-1.22	129.3	65.1
A1027	2.57E-32	-2.47	1569.3	333.7
A1054	8.92E-30	-2.6	456.3	88.7
A1753	4.39E-28	-1.32	840.5	398.1
A1028	1.48E-27	-1.21	1061.6	540.2
A2063	7.98E-27	-1.31	475.1	225.3

Table A2.10: Top 40 differentially expressed genes in ΔA0061ΔA0384

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA0061ΔA0384
A0061	6.06E-124	-10.98	33.2	0.0
A0384	9.78E-105	-2.85	73.7	12.0
A2211	9.15E-63	-2.09	5600.0	1535.4
A1060	9.32E-54	-2.54	440.0	87.4
A1050	2.72E-51	-2.52	364.8	73.5
A1059	7.41E-50	-2.48	934.1	193.6
A1058	5.96E-44	-2.32	447.0	103.2
A0385	2.70E-43	1.26	180.5	499.1
A0750	1.17E-42	-1.59	354.6	136.9
A0734	2.36E-39	-1.52	1256.4	508.0
A1048	1.24E-38	-2.05	611.1	170.9
A1831	1.24E-38	1.68	64.0	238.5
A1051	2.09E-37	-2.5	653.6	134.4
A1062	2.38E-37	-2.06	1332.2	369.6
A1879	1.40E-36	-2.8	37.9	6.4
A1049	1.77E-36	-2.04	634.8	178.8
A1061	4.56E-36	-2.28	430.8	102.8
A0654	5.06E-35	1.5	87.4	287.5
A1246	6.45E-35	-1.21	230.5	115.2
A1057	1.11E-34	-2.1	221.1	59.7
A1053	9.29E-34	-2.51	664.5	135.6
A1064	8.49E-33	-2.12	1124.9	300.1
A0749	4.35E-32	-1.31	1975.1	926.1
A1065	9.54E-30	-1.9	1742.8	543.0
A2822	1.01E-29	1.1	33.7	84.0
A1052	2.56E-29	-2.5	477.9	97.9
A1276	9.49E-29	-1.07	707.4	391.8
A1036	5.97E-28	-1.16	129.3	67.0
A1055	1.48E-27	-2.17	450.8	116.4
A2244	2.78E-27	1.26	57.6	159.1
A0269	1.70E-25	-1.18	2016.8	1030.4
A1026	1.94E-25	-2.17	1329.9	342.3
A1063	1.60E-24	-2.01	403.2	116.5
A2818	1.70E-24	1.03	52.4	124.6
A1047	2.47E-24	-1.3	1209.8	571.2
A2212	3.43E-24	-1.52	1208.7	489.0
A1056	5.13E-24	-2.11	372.3	100.1
A1910	1.04E-23	-1.22	767.3	382.2
A2327	3.26E-23	-1.38	74.4	33.0
A1909	4.70E-23	-1.25	407.0	199.2

Table A2.11: Top 40 differentially expressed genes in ΔA2542ΔA0384

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA2542ΔA0384
A2542	1.10E-150	-5.51	33.8	0.8
A0384	2.61E-126	-3.03	73.7	8.7
C0011	1.81E-105	2.94	49.9	376.1
C0010	2.00E-66	2.03	96.2	385.6
C0009	6.92E-51	1.61	39.7	118.2
C0013	9.68E-47	1.72	87.4	282.4
C0012	1.23E-42	1.71	114.1	368.3
C0007	1.11E-35	1.67	154.8	481.6
C0008	2.63E-34	1.75	294.7	973.6
A0385	4.46E-34	1.11	180.5	381.8
C0014	1.62E-29	1.52	51.4	143.7
A1879	3.44E-29	-2.32	37.9	7.4
A1527	2.39E-25	-1.76	35.4	10.2
A2818	2.43E-25	1.04	52.4	106.0
C0003	2.43E-25	1.17	73.2	161.2
C0006	1.18E-21	1.34	186.4	462.9
A2822	7.91E-19	0.87	33.7	60.6
A1531	6.25E-17	-1.16	98.9	43.5
C0005	6.25E-17	1.36	168.8	424.7
A1520	1.71E-16	-0.97	117.5	58.7
A1880	4.14E-16	-3.46	4.6	0.4
A1537	4.53E-15	-0.75	105.2	61.6
A0576	5.51E-15	-1.31	23.5	9.0
C0004	5.66E-15	1.3	177.3	428.5
A1503	1.14E-14	-0.93	439.2	225.3
A1881	1.42E-14	-2.47	8.4	1.4
A0286	2.37E-14	-1.21	1347.6	571.9
A1330	2.97E-14	-0.97	228.8	114.0
C0015	1.27E-13	1.66	4.4	13.6
A0618	8.92E-12	-1	192.3	94.4
A0468	1.28E-11	0.9	143.0	262.0
A1878	2.25E-11	-2.05	6.0	1.3
A2819	2.60E-11	0.9	279.2	511.8
A1526	6.90E-11	-0.68	106.2	65.2
A2543	7.29E-11	-0.89	166.0	87.8
D0027	7.29E-11	0.88	225.4	405.5
A0360	1.14E-10	-0.77	48.2	27.4
A1285	2.51E-10	0.78	1300.6	2182.0
A1050	3.05E-10	-1.08	364.8	168.6
D0026	4.69E-10	0.79	88.4	149.9

Table A2.12: Top 40 differentially expressed genes in ΔA0061ΔA2542ΔA0384

Gene	False Discovery Rate	Log Fold Change	Average Reads WT	Average Reads ΔA0061ΔA2542ΔA0384
C0011	1.85E-212	4.39	49.9	1276.3
A0384	5.90E-134	-3.45	73.7	8.1
C0009	2.38E-132	2.63	39.7	300.5
C0010	2.80E-125	2.84	96.2	839.3
A0061	1.77E-122	-10.97	33.2	0.0
A2542	6.35E-118	-4.7	33.8	1.6
C0003	2.03E-75	2.02	73.2	362.2
A0654	1.98E-73	2.19	87.4	488.6
A1831	1.29E-69	2.27	64.0	377.3
A0602	7.10E-69	3.17	16.3	179.6
A1237	2.58E-68	3.41	14.5	188.9
C0013	7.50E-61	1.97	87.4	416.8
C0007	1.64E-53	2.05	154.8	783.4
A1671	6.97E-51	1.9	78.1	356.1
C0014	3.62E-49	1.95	51.4	242.7
C0012	6.42E-48	1.81	114.1	490.4
C0008	6.84E-46	2.03	294.7	1468.1
A1476	3.87E-45	3.84	5.7	100.4
G0140	1.19E-44	2.33	7.6	46.5
A2520	1.04E-43	2.84	5.8	50.7
C0002	9.09E-41	1.6	18.0	66.7
C0006	1.56E-40	1.84	186.4	815.0
C0001	1.39E-39	1.46	25.9	87.0
A1477	4.33E-37	3.54	9.4	134.0
C0005	3.27E-33	1.92	168.8	777.8
A0186	2.16E-29	2.15	30.5	165.8
C0004	2.41E-29	1.83	177.3	773.0
A1246	1.08E-28	-1.1	230.5	131.8
A1625	7.44E-28	1.47	14.1	47.7
A0486	1.26E-27	1.83	8.6	37.6
A0385	1.63E-27	1	180.5	439.5
A0397	2.27E-27	-1.23	1158.1	600.6
F0041	5.67E-27	1.56	18.2	65.8
A1879	1.71E-26	-2.25	37.9	9.7
A2211	1.71E-26	-1.32	5600.0	2757.5
C0015	1.72E-26	2.26	4.4	25.6
A1624	7.06E-25	1.41	61.9	200.7
A2563	1.09E-24	1.67	35.5	137.9
A0734	3.96E-24	-1.17	1256.4	680.7
A0140	1.66E-23	1.73	51.1	207.9

Table A2.13: Top 80 most stable PCC 7002 genes

Locus Tag	Gene	Half-Life (minutes)	Class
A0083	*prp2*	-	stable
A0139	*manR*	-	stable
A0140	*gifB*	-	stable
A0151		-	stable
A0159	*desB*	-	stable
A0232	*psbL*	-	stable
A0233	*psbJ*	-	stable
A0242	*sodB*	-	stable
A0282	*grxC*	-	stable
A0305		-	stable
A0314		-	stable
A0340	*ureJ*	-	stable
A0344	*rlpA*	-	stable
A0408		-	stable
A2563	*smtA*	5.29	decay
A1477		5.05	decay
A2462	*mscL*	4.99	decay
A2261	*petP*	4.95	decay
A1965	*hupB*	4.91	decay
A1311	*petM*	4.79	decay
A1589	*psaC*	4.61	decay
A1239		4.39	decay
A0231	*psbF*	4.31	decay
F0042		4.28	decay
A1387		4.28	decay
A1605	*psbQ*	4.24	decay
A0167	*petJ*	4.17	decay
A0230	*psbE*	4.11	decay
A1418	*psbA3*	4.10	decay
A2208	*amt*	4.10	decay
A0215	*msrA*	4.06	decay
A0811	*cpcG*	4.06	decay
F0041		4.01	decay
A2857	*prk*	3.99	decay
A1180		3.99	decay
A1301	*glpX*	3.96	decay
A1285	*glnB*	3.95	decay
E0029		3.94	decay
A0569	*ndhM*	3.93	decay
D0023	*csc2*	3.91	decay

Locus Tag	Gene	Half-Life (minutes)	Class
A1347	*psbY*	3.90	decay
A0320		3.89	decay
A2209	*cpcB*	3.87	decay
A0863	*cynS*	3.86	decay
A0808	*psbH*	3.86	decay
A2044	*rpoC1*	3.82	decay
A1393	*psaE*	3.81	decay
A2061	*tufA*	3.80	decay
A0186	*hliC*	3.79	decay
A2687		3.78	decay
A2286	*pstA*	3.77	decay
A0103		3.75	decay
A2305		3.75	decay
F0044	*cmr6*	3.75	decay
A2435		3.75	decay
A1560	*psbD1*	3.75	decay
A2521	*ycf23*	3.74	decay
A0374	*petG*	3.74	decay
A1490		3.74	decay
A0482		3.73	decay
A0176	*dsbG*	3.70	decay
A2448		3.70	decay
G0017		3.69	decay
A0682	*psaD*	3.68	decay
A2606		3.67	decay
A1910	*petA*	3.64	decay
A2380	*mrpB*	3.60	decay
A0147	*groEL2*	3.60	decay
A1772		3.59	decay
A2210	*cpcA*	3.59	decay
A2594		3.55	decay
A0735	*atpH*	3.54	decay
G0070		3.51	decay
A1909	*petC*	3.48	decay
F0086		3.48	decay
A1278	*ilvC*	3.48	decay
A2254		3.46	decay
A2696	*ybbK*	3.45	decay
A2379		3.44	decay
A0285		3.44	decay

Table A2.14: Top 80 least stable PCC 7002 genes

Locus Tag	Gene	Half-Life (minutes)	Class
A2112	*tRNA-Asn*	0.18	decay
A2138		0.20	decay
F0032		0.25	decay
A1156	*tRNA-His*	0.27	decay
A0304		0.28	decay
G0109		0.29	decay
E0003		0.31	decay
A2860		0.31	decay
A1669		0.32	decay
G0037		0.32	decay
A2137	*tRNA-Ser*	0.33	decay
A2086		0.33	decay
A1695		0.34	decay
A1668		0.34	decay
A2237		0.34	decay
A2741	*mutY*	0.35	decay
A1696		0.36	decay
A0018		0.36	decay
A2245		0.37	decay
A0089		0.37	decay
A2593		0.38	decay
A0061	*rnc*	0.38	decay
A0273		0.39	decay
A2753		0.39	decay
F0068		0.40	decay
A2669		0.40	decay
A2453		0.40	decay
A0937		0.40	decay
E0004		0.40	decay
A0008	*manA*	0.40	decay
A1786	*comFB*	0.41	decay
A2085		0.41	decay
A1429		0.41	decay
A2572		0.41	decay
A0859		0.41	decay
A1229		0.41	decay
A2503		0.41	decay
A2844		0.42	decay
A1502		0.42	decay
A0777		0.42	decay

Locus Tag	Gene	Half-Life (minutes)	Class
A1079		0.42	decay
A0920	*kaiB*	0.43	decay
A2752		0.43	decay
A2552		0.43	decay
A0909		0.43	decay
A1717		0.43	decay
A0429		0.43	decay
A1375	*pilG*	0.43	decay
A1680		0.43	decay
A1993	*chlR*	0.43	decay
A0036		0.43	decay
A1738		0.43	decay
A2349		0.43	decay
A2207		0.43	decay
A2068	*grxB*	0.43	decay
A0878	*nadD*	0.43	decay
A1128		0.43	decay
A1640		0.43	decay
A0906		0.43	decay
A1521	*obgE*	0.43	decay
A2676		0.43	decay
A1378		0.43	decay
A1810		0.43	decay
A2002		0.43	decay
A2246	*cruH*	0.43	decay
A1457	*devH*	0.43	decay
A2747		0.44	decay
A1855		0.44	decay
A2075	*clpS*	0.44	decay
A0359	*tolB*	0.44	decay
A1067		0.42	decay
A2494	*prp1*	0.42	decay
A1723		0.42	decay
A0090		0.42	decay
A2591	*pdxH*	0.42	decay
A0873	*rfrK*	0.42	decay
A1100	*menG*	0.42	decay
A0254		0.42	decay
A0211		0.42	decay
A2688		0.42	decay